OXFORD

Briefings in Bioinformatics, 2022, **23(6)**, 1–11
https://doi.org/10.1093/bib/bbac457
Advance access publication date 27 October 2022
Problem Solving Protocol

A new framework for drug–disease association prediction combing light-gated message passing neural network and gated fusion mechanism

Bao-Min Liu, Ying-Lian Gao, Dai-Jun Zhang, Feng Zhou, Juan Wang, Chun-Hou Zheng and Jin-Xing Liu

Corresponding authors. Ying-Lian Gao, Qufu Normal University Library, Qufu Normal University, Rizhao 276826, China. Tel.: 086-633-3981241; E-mail: yinliangao@126.com; Jin-Xing Liu, School of Computer Science, Qufu Normal University, Rizhao 276826, China. Tel.: 086-633-3981241; E-mail: sdcavell@126.com

Abstract

With the development of research on the complex aetiology of many diseases, computational drug repositioning methodology has proven to be a shortcut to costly and inefficient traditional methods. Therefore, developing more promising computational methods is indispensable for finding new candidate diseases to treat with existing drugs. In this paper, a model integrating a new variant of message passing neural network and a novel-gated fusion mechanism called GLGMPNN is proposed for drug–disease association prediction. First, a light-gated message passing neural network (LGMPNN), including message passing, aggregation and updating, is proposed to separately extract multiple pieces of information from the similarity networks and the association network. Then, a gated fusion mechanism consisting of a forget gate and an output gate is applied to integrate the multiple pieces of information to extent. The forget gate calculated by the multiple embeddings is built to integrate the association information into the similarity information. Furthermore, the final node representations are controlled by the output gate, which fuses the topology information of the networks and the initial similarity information. Finally, a bilinear decoder is adopted to reconstruct an adjacency matrix for drug–disease associations. Evaluated by 10-fold cross-validations, GLGMPNN achieves excellent performance compared with the current models. The following studies show that our model can effectively discover novel drug–disease associations.

Keywords: drug–disease association prediction, neural network, message passing, forget gate, output gate

Introduction

Over the past decades, there have been many significant advances occurring in life sciences, genomics and computing technologies [1, 2]. However, drug discovery is not developing as fast as it should. According to the data [3], a large number of drug candidates failed in phase 1 clinical trials. Meanwhile, it takes approximately 10–15 years and hundreds of millions of dollars for a new drug to successfully enter the market [3]. Hence, drug discovery is still a very time-consuming and expensive process. Predicting drug–disease associations is an important part of drug discovery [4]. Currently, driven by the accumulation of high-throughput data and the improvement of algorithms, the superiority of computational methods in finding new drug–disease candidate associations has been demonstrated [5, 6]. The existing computational methods can be roughly divided into the following three types: network diffusion-based methods, machine learning-based methods and deep learning-based methods [7, 8].

Due to the good interpretability of network diffusion-based methods, this type of methods has been increasingly applied to predict potential drug–disease associations. For example, Luo *et al.* [9] developed a network integration model combing random walk with restart with diffusion component analysis. The final

Bao-Min Liu received the BS degree in computer science and technology from Qingdao University, Qingdao, China, in 2021, where she is currently pursuing the master's degree in computer science and technology. Her research interests include data mining, pattern recognition and bioinformatics.
Ying-Lian Gao received the BS and MS degrees from Qufu Normal University, Rizhao, China, in 1997 and 2000, respectively. She is currently an associate professor with Qufu Normal University Library, Qufu Normal University. Her current interests include data mining and pattern recognition.
Dai-Jun Zhang received the BS degree in 2019; the Master degree candidate in Electronic and Information Engineering from Qufu Normal University, China. Her research interests include pattern recognition and bioinformatics.
Feng Zhou received the BS degree from the School of Information Science and Engineering, Qufu Normal University, Rizhao, China, in 2019, and the MS degree from the School of Computer Science, Qufu Normal University, Rizhao, China, in 2022. He is currently pursuing the PhD degree with the School of Medicine, Xiamen University. His research interests include feature selection, deep learning, pattern recognition and bioinformatics.
Juan Wang received the BS degree in applied electronic technology from QuFu Normal University, Rizhao, China, in 2000, and the MS degree in circuits and systems from Shandong University, Jinan, China, in 2003. She is an associate professor with the School of Computer Science, Qufu Normal University, Rizhao, China. Her research interests include pattern recognition and bioinformatics.
Chun-Hou Zheng received the BS degree in physics education and the MS degree in control theory and control engineering from Qufu Normal University, Rizhao, China, in 1995 and 2001, respectively, and the PhD degree in pattern recognition and intelligent system from the University of Science and Technology of China, Anhui, China, in 2006. He is currently a professor with the School of Computer Science and Technology, Hefei, Anhui, China, and the School of Computer Science, Qufu Normal University, Rizhao, China. His research interests include pattern recognition and bioinformatics.
Jin-Xing Liu received the BS degree in electronic information and electrical engineering from Shandong University, Jinan, China, in 1997, the MS degree in control theory and control engineering from Qufu Normal University, Jining, China, in 2003, and the PhD degree in computer simulation and control from the South China University of Technology, Guangzhou, China, in 2008. He is a professor with the School of Computer Science, Qufu Normal University, Rizhao, China. His research interests include pattern recognition, machine learning and bioinformatics.
Received: June 26, 2022. **Revised:** September 7, 2022. **Accepted:** September 23, 2022

representations of nodes are low-dimensional but contain informative topological properties. Xie *et al.* [10] proposed a model based on a two-step bipartite graph diffusion algorithm integrating linear neighborhood similarity and Gaussian interaction profile kernel similarity, called BGMSDDA. Compared with Luo's model, the model only uses the drug and disease information leading to the incompleteness of biology, but it effectively alleviates the impact of the low sparsity of association data, in which the weighted K nearest known neighbor (WKNKN) method [11] is deployed to reconstruct the drug–disease association matrix in preprocessing.

In recent years, machine learning-based methods, which make better use of the global patterns of associations, have also been widely used in drug–disease association prediction. For instance, Zhang *et al.* [12] developed robust SCMFDD, which adopted drug feature-based similarities and disease semantic similarity as constraints in matrix factorization. Luo *et al.* [13] adopted a fast singular value thresholding (SVT) algorithm to complete the drug–disease adjacency matrix with predicted scores for unknown drug–disease pairs. Furthermore, more drug and disease similarities are introduced in DRIMC [14]. The similarity network fusion method [15] is utilized to reconstruct similarity matrices. The probability of a drug–disease association is modelled by inductive matrix completion method. Moreover, the ensemble strategy is utilized in [16], named DTi2vec. First, node2vec [17] is applied to learn edge representations based on the heterogenous network. Then, the extracted features are fed to ensemble classifiers including adaptive boosting (AdaBoost) and extreme gradient boosting (XGBoost) [18]. In addition, Wang *et al.* [19] developed a CMAF model with integrating matrix factorization, label propagation and network consistency projection.

Although the above two methods have been generally put into practice in predicting drug–disease associations, there are still some challenges to overcome. For example, network diffusion-based methods have not improved enough [7], and machine learning-based methods need more diverse features that are not available to all drugs and diseases [20]. The efficiency of using deep learning-based methods to predict multiple biology associations has been affirmed, such as microRNA–disease associations [21–23] and microbe–disease associations [24]. Deep learning-based methods have also been utilized to predict drug–disease associations. For instance, Yu *et al.* [25] proposed an encoder–decoder architecture, named LAGCN. First, GCN is applied to capture information based on the heterogenous network. Then, the attention mechanism [26] is introduced to integrate the embeddings obtained from different layers. Meng *et al.* [27] proposed a new prediction model based on a novel weighted bilinear graph aggregator, called DRWBNCF. Different from LAGCN, DRWBNCF augments the conventional GCN by encoding the local nearest neighbors and their interactions.

In this paper, a deep-learning-based model is proposed for predicting drug–disease associations, which integrates a message passing neural network (MPNN) and a gated fusion mechanism composed of a forget gate between networks and an output gate, called GLGMPNN. Different from the existing models based on the heterogeneous network, a light-gated message passing neural network (LGMPNN) is proposed to encode information from the similarity networks and the association network. Then, a forget gate is developed to control the contribution of the association network instead of being fully used. The forget gate makes embeddings more comprehensive by fusing multiple pieces of information from different networks and alleviates the influence of the association network by weighting the node features obtained from the association network. Finally, to make further use of the similarity information, the output gate is designed to integrate the initial similarity features and the structure features captured from the networks by LGMPNN. The experiment under 10-fold cross-validation demonstrates that our model outperforms the compared models with higher robustness on two datasets. Additionally, it can effectively determine the potential candidate drugs and associations supported by authoritative databases.

The main contributions are summarized as follows:

1. A model combining a light-gated message passing neural network and a gated fusion mechanism is proposed for predicting drug–disease associations. To capture the topology information of the similarity networks and the association network, a light-gated message passing neural network (LGMPNN) is proposed to pass, aggregate and update messages from neighbour nodes, in which the trainable feature transformation matrix is removed to accelerate the training process and improve the performance.
2. A forget gate is designed to extract the useful association information, making the node features encoded from the similarity networks more comprehensive. Then, the output gate is proposed to fuse the topology information mined by LGMPNN and the initial similarity information.

The remaining sections of this paper are as follows. The Materials section introduces the details of the datasets. The Method section summarizes the procedures of the GLGMPNN model. The Results and Discussions section shows the favourable performance of the GLGMPNN model. The Conclusions section is a summary of this paper.

Table 1. Details of datasets

Dataset	Drugs	Diseases	Associations	Sparsity
Fdataset	593	313	1933	0.0104
Cdataset	663	409	2352	0.0087

Materials

Datasets

In this paper, Fdataset [28] and Cdataset [29] are used to verify the performance of the GLGMPNN model. The details of the datasets are summarized in Table 1. Fdataset includes 593 drugs collected from DrugBank (DB) [30], 312 diseases extracted from the Online Mendelian Inheritance in Man (OMIM) [31], and 1933 known drug–disease associations proceeded from [28]. Cdataset consists of 663 drugs, 409 diseases and 2352 associations derived from [29]. Likewise, the drug and disease data in Cdataset are from DB and OMIM. DB has provided detailed, confirmed and up-to-date information about drugs, targets and related factors for conducting quantitative academic research from birth. OMIM focuses on human genetic variation and phenotypic traits.

Construction of networks

Each dataset consists of three matrices: a drug–drug similarity matrix $\mathbf{S}_{\mathrm{dr}} \in R^{n\times n}$, a disease–disease similarity matrix $\mathbf{S}_{\mathrm{di}} \in R^{m\times m}$ and a drug–disease association matrix $\mathbf{A} \in R^{n\times m}$. For a drug–disease pair $(dr_{\mathrm{i}}, di_{\mathrm{j}})$, $\mathbf{A}(\mathrm{i},\mathrm{j}) = 1$ if the association between drug dr_{i} and disease di_{j} exists. Otherwise, $\mathbf{A}(\mathrm{i},\mathrm{j}) = 0$. The similarity score between drugs is based on their fingerprints, which are

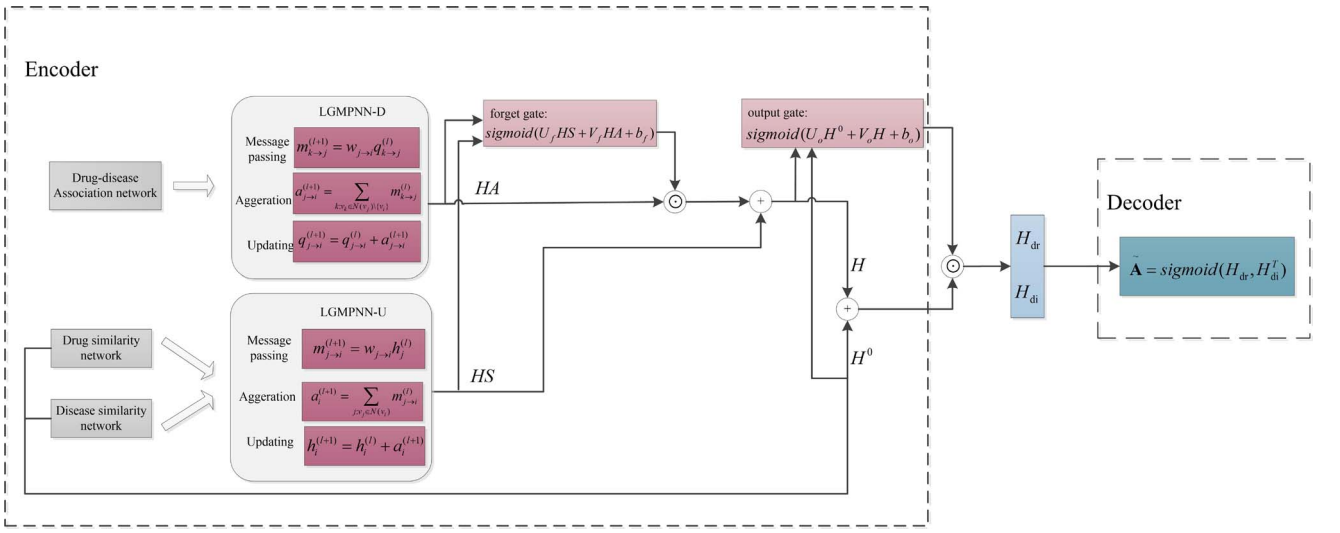

Figure 1. Flowchart of GLGMPNN.

obtained by the Chemical Development Kit processing the chemical structure of SMILES (simplified molecular input line entry system) [32], according to the 2D Tanimoto score. The disease similarity is calculated by using MimMiner [33], a web interface consisting of the disease phenotype similarity data. The disease–disease similarity is based on the MeSH terms in the medical descriptions of diseases.

On the similarity networks, the weight between edges is the similarity. Then, the top k nodes in the similarity are screened out for each drug (disease) node. On the association network, if an edge between a drug node and a disease node exists, the weight of the edge is 1; if not, the weight is 0.

Method

Overview

First, the process of MPNN on undirected graphs is introduced. Then, the GLGMPNN model is established in the Encoder section. LGMPNN is proposed to extract the feature representations of drug and disease nodes from the similarity networks and the association network. The forget gate is developed to integrate the two types of representations, making the representations more informative. The output gate is used to fuse the similarity information and the output of LGMPNN containing the topology information of the networks. The flowchart is shown in Figure 1.

Message Passing Neural Networks

MPNN [34] is a general framework abstracted from the existing graph neural learning models, which consists of three steps [35]: message passing, aggregation and updating. Given an undirected graph G with node features h and edge features q, the three steps are as follows:

$$m_{j\to i}^{(l+1)} = M(h_j^{(l)}, h_i^{(l)}, q_{j\to i}), \forall j : v_j \in N(v_i), \tag{1}$$

$$a_i^{(l+1)} = A(\{m_{j\to i}^{(l+1)}\}_{:v_j\in N(v_i)}), \tag{2}$$

$$h_i^{(l+1)} = U(h_i^{(l)}, a_i^{(l+1)}). \tag{3}$$

At the lth iteration, the message m is passed from the neighbor nodes ($\forall j : v_j \in N(v_i)$) according to Eq. (1). M denotes a message function. Obviously, the message is influenced by the node features $h^{(l)}$ obtained from the last iteration and the edge features $q_{j\to i}$. It should be noted that edge features are not necessary. Then, the calculated message from the neighbour nodes of v_i is aggregated in Eq. (2). A denotes an aggregation function. Finally, node features are updated by using an update function U.

GLGMPNN is based on the above standard MPNN on an undirected graph in which node features are updated during the process and MPNN on a directed graph in which edge features are updated during the process.

Encoder

Light-Gated Message Passing Neural Network

Most existing models [16, 36, 37] for predicting drug–disease associations are based on the heterogeneous network, importing heterogeneous information to improve the performance. However, the construction of a heterogeneous network may enlarge the network scale and introduce some noise, leading to a decline in predicting capability. Therefore, the similarity networks and the association network are represented as the undirected networks and the directed network [38], respectively. Then, LGMPNN based on the MPNN variants [35, 39] is applied on the undirected networks (LGMPNN-U) and on the directed network (LGMPNN-D) to represent nodes.

The drug and disease node representations are initialized as follows:

$$H^0 = \begin{bmatrix} \mathbf{S}_{dr} & 0 \\ 0 & \mathbf{S}_{di} \end{bmatrix}. \tag{4}$$

The message passing, aggregation and updating process of LGMPNN-U on the similarity networks are as follows:

$$m_{j\to i}^{(l+1)} = w_{j\to i} h_j^{(l)}, \forall j : v_j \in N(v_i), \tag{5}$$

$$a_i^{(l+1)} = \sum_{j:v_j\in N(v_i)} m_{j\to i}^{(l)}, \tag{6}$$

$$h_i^{(l+1)} = h_i^{(l)} + a_i^{(l+1)}, \tag{7}$$

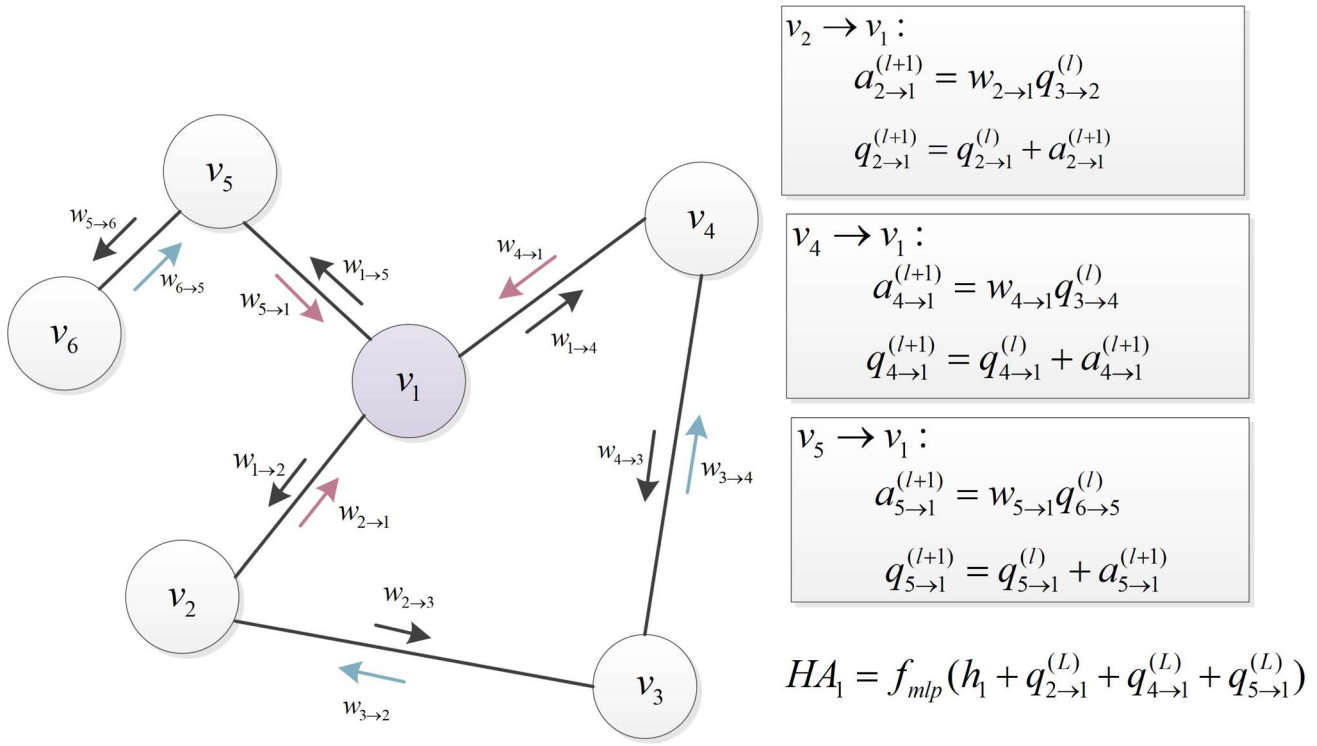

Figure 2. Workflow of LGMPNN-D on the association network.

where $w_{j\rightarrow i}$ denotes the weight that controls the information flowing from the adjacent nodes of v_i in Eq. (5), a_i denotes the aggregated message and h denotes node features. In the MPNN variant [35], the weight is a learnable matrix which can be regarded as feature transformation. However, inspired by LightGCN [40], the similarity scores are employed as weights in LGMPNN-U. The feature transformation and nonlinear activation are all removed in LightGCN model. The model achieves better performance, demonstrating that these operations are not necessary. Therefore, the simple constant weight is adopted to reduce the parameters trained in neural networks and improve the performance. After the set last iteration L, HS_i, which denotes the embeddings of node v_i encoded on the similarity networks, is given by

$$HS_i = f_{mlp}(h_i + \sum h_i^{(L)}), \tag{8}$$

where f_{mlp} denotes a nonlinear function.

Different from the similarity networks, which only include one entity, drug and disease nodes are present in the association network at the same time. Therefore, LGMPNN on the association network is expected to aggregate messages from neighbour drug nodes and disease nodes at one iteration. Based on this requirement, LGMPNN-D is proposed for the association network.

The node embeddings are initialized as Eq. (4). First, the association network is transferred from an undirected graph to a directed graph. Then, the three steps of LGMPNN-D on the association network are formulated as follows:

$$m_{k\rightarrow j}^{(l+1)} = w_{j\rightarrow i} q_{k\rightarrow j}^{(l)}, \forall k : v_k \in N(v_j) \backslash \{v_i\}, \tag{9}$$

$$a_{j\rightarrow i}^{(l+1)} = \sum_{k:v_k \in N(v_j)\backslash\{v_i\}} m_{k\rightarrow j}^{(l)}, \tag{10}$$

$$q_{j\rightarrow i}^{(l+1)} = q_{j\rightarrow i}^{(l)} + a_{j\rightarrow i}^{(l+1)}. \tag{11}$$

Compared with LGMPNN-U based on the standard MPNN, the edge features, not node features, are updated in LGMPNN-D, which can reduce the redundancy of nodes [35]. As shown in Eq. (9), the weight $w_{j\rightarrow i}$ can be learned from embeddings of head node v_j and tail node v_i [35]. Out of the same purpose in LGMPNN-U, edge weights in LGMPNN-D are constants. The weight $w_{j\rightarrow i}$ is 0 or 1, depending on the existence of the association and $m_{k\rightarrow j}$ denotes the message passed from the neighbouring edges $e_{k\rightarrow j}$ of the edge $e_{j\rightarrow i}$. Node v_j is the common vertex of these neighbouring edges. In Eq. (10), $a_{j\rightarrow i}$ denotes the aggregated message. In Eqs (9) and (10), q denotes the edge features. The initial feature representations $q_{j\rightarrow i}^{(0)}$ are calculated as follows:

$$q_{j\rightarrow i}^{(0)} = w_{j\rightarrow i} h_j. \tag{12}$$

Then, after the set last iteration L, HA_i, which denotes the embeddings of node v_i encoded on the association network, is given by:

$$HA_i = f_{mlp}\left(h_i + \sum_{j:v_j \in N(v_i)} q_{j\rightarrow i}^{(L)}\right), \tag{13}$$

where f_{mlp} denotes a nonlinear function,with the same architecture as f_{mlp} in LGMPNN-U.

As shown in Figure 2, v_1 is assumed to be a drug node. According to the definition of the association network, v_3 and v_6 are drug nodes, and v_2, v_4 and v_5 are disease nodes. From the learning process illustrated in Figure 2, all features of edges tailing the drug node v_1 are updated with its initial disease message and adjacent drug message. For example, $q_{2\rightarrow1}^{(1)}$ incorporates $q_{2\rightarrow1}^{(0)}$ obtained from

embeddings of disease node v_2, and $a^{(1)}_{2\to1}$ aggregates the message from the drug node v_3 after the first iteration.

Forget gate and output gate

The performance of most existing models is easily influenced by the sparsity of the association data [10]. To lessen the recline on the associations and make the node embeddings more informative, the forget gate inspired by the long short-term memory (LSTM) [41] is proposed to merge HS and HA as follows:

$$F = sigmoid(U_f HS + V_f HA + b_f), \tag{14}$$

$$H = HS + F \odot HA, \tag{15}$$

where U_f and V_f are automatically trainable matrices, b_f is the learnable bias vector and F is the forget gate controlling the contribution of the association network. $\odot$ denotes the Hadamard product, pointwise multiplication.

Finally, the output gate is combined to fuse the initial similarity information and the output of LGMPNN, which makes better use of the similarity information and the topology information. The final output is modified as follows:

$$O = sigmoid(U_o H^0 + V_o H + b_o), \tag{16}$$

$$\begin{bmatrix} H_{\mathrm{dr}} \\ H_{\mathrm{di}} \end{bmatrix} = O \odot tanh(H^0 + H), \tag{17}$$

where O denotes the output gate calculated by the learnable parameters U_o, V_o and b_o in Eq. (16). H_{dr} and H_{di} represent the final features of the drug and disease nodes, respectively. Considering the range value of F and O, the activation functions in the forget gate and the output gate are the same as those in LSTM.

Decoder

After getting the final embeddings, a bilinear decoder [42] is used to reconstruct the adjacency matrix. The decoder is formulated as

$$\widetilde{\mathbf{A}} = f(H_{\mathrm{dr}}, H_{\mathrm{di}}) = sigmoid(H_{\mathrm{dr}}, H_{\mathrm{di}}^T), \tag{18}$$

where $\widetilde{\mathbf{A}} \in R^{n\times m}$ is the predicted probability matrix and $\widetilde{\mathbf{A}}(\mathrm{i},\mathrm{j})$ represents the association probability between drug dr_{i} and disease di_{j}.

Optimization

The known associations are taken as positive samples, and the unknown or unobserved associations are taken as negative samples. The low sparsity of Fdataset and Cdataset indicates that the number of positive samples is much less than that of negative samples. Therefore, the weighted cross-entropy loss function and a balance parameter λ are applied to reduce the impact of the unbalance of samples. The formula is as follows:

$$loss = -\frac{1}{n\times m}\left(\lambda \times \sum_{(\mathrm{i},\mathrm{j})\in S^p} \log \widetilde{\mathbf{A}}_{\mathrm{ij}} + \sum_{(\mathrm{i},\mathrm{j})\in S^n} (1 - \log \widetilde{\mathbf{A}}_{\mathrm{ij}})\right), \tag{19}$$

where S^p and S^n represent the sets of positive and negative samples, respectively. The numbers of drugs and diseases are n and m. The unbalance parameter is denoted as $\lambda = |S^n| / |S^p|$, where $|S^p|$ and $|S^n|$ are the numbers of elements in S^p and S^n, respectively.

In our model, the Adam optimizer [43] is used in optimization. In addition, regular and edge dropouts are introduced to effectively generalize the unobserved data.

Results and Discussions

To show the superiority of GLGMPNN model, we compare it with DRRS [5], BNNR [44], SCPMF [45], LAGCN [25] and DRWBNCF [27] on Fdataset and Cdataset.

- DRRS: A fast SVT algorithm is applied to complete the association matrix of the constructed heterogenous network.
- BNNR: A bounded nuclear norm regularization is incorporated in noisy matrix completion for predicting associations.
- SCPMF: Similarity constraints are introduced as constraints into the probabilistic matrix factorization process.
- LAGCN: The GCN is utilized to obtain embedding from the heterogenous network, and layer attention is introduced to combine the embeddings of multiple layers. Then, a bilinear decoder is used for predicting associations.
- DRWBNCF: A new weighted bilinear graph convolution operation is proposed to integrate the known drug–disease association and the drug's and disease's neighbourhood and neighbourhood interactions. Then, the MLP is applied to predict associations.

The parameters of the compared models are chosen as the optimal values from their papers.

In our model, there are several parameters to set, such as the initial learning rate, max training epoch, dimension of nodes, iteration of LGMPNN, edge drop rate and regular drop rate. Considering the time cost and the performance of our model, we empirically adjust the parameters. Finally, the learning rate of 0.01, max training epoch of 350, dimension of nodes of 128, iteration of 1, edge drop rate of 0.4 and regular drop rate of 0.2 are adopted in the experiments.

Performance in the cross-validation

In this paper, 10-fold cross-validation (10-CV) is applied to evaluate the performance of models. The dataset is randomly divided into 10 subsample sets. In each fold, a subsample set is used to verify the model performance, and the remaining samples are used for training. Cross-validationn can avoid unfitting and overfitting, increasing the reliability of the obtained results. Meanwhile, the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPR) are used as metrics. The higher values of AUROC and AUPR represent the better classification capacity of models.

The curves are plotted by true positive rate (TPR), false positive rate (FPR), recall and precision.

$$TPR = recall = \frac{TP}{TP + FN}, \tag{20}$$

$$FPR = \frac{FP}{TN + FP}, \tag{21}$$

$$precision = \frac{TP}{TP + FP}. \tag{22}$$

Table 2. Performance of all models by 10-CV

	Fdataset		Cdataset	
Model	**AUROC**	**AUPR**	**AUROC**	**AUPR**
DRRS	0.928 ± 0.001	0.378±0.006	0.948±0.002	0.402±0.005
BNNR	0.934 ± 0.001	0.440±0.004	0.950±0.002	0.471±0.004
SCPMF	0.899 ± 0.002	0.357±0.006	0.912±0.002	0.421±0.005
LAGCN	0.921 ± 0.002	0.302±0.002	0.940±0.001	0.352±0.004
DRWBNCF	0.924 ± 0.001	0.491±0.006	0.940±0.002	0.566±0.007
GLGMPNN	**0.945 ± 0.001**	**0.513 ± 0.005**	**0.957±0.001**	**0.599±0.006**

Note: The best results are in bold.

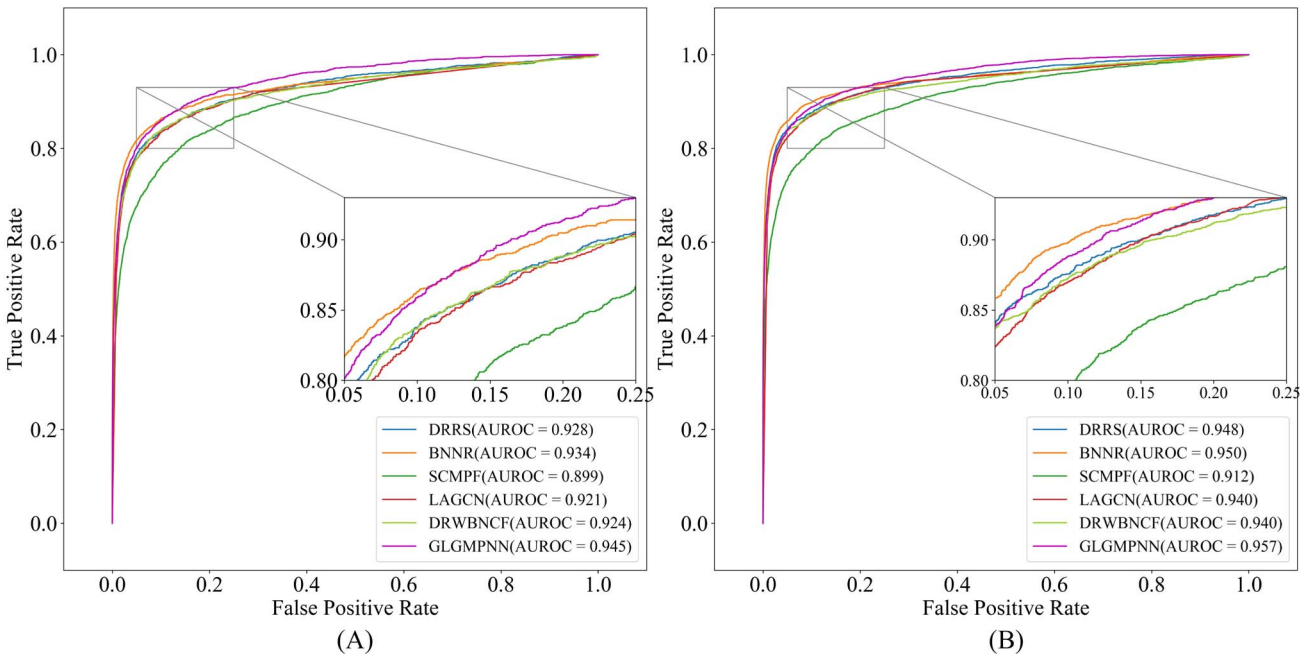

Figure 3. ROC curves obtained by all models on Fdataset (**A**) and Cdataset (**B**).

TP and FN represent the correctly and incorrectly predicted positive samples, respectively. Likewise, TN and FP represent the correctly and incorrectly predicted negative samples, respectively.

In Table 2 and Figure 3, the average AUROC and AUPR obtained by GLGMPNN on Fdataset are 0.945 and 0.513, respectively, both rank as first in all models. The second AUROC is obtained by BNNR (0.934), 1.1% lower than our model. GLGMPNN performs 2.2% higher than the second AUPR of 0.491 obtained by DRWBCNF. On Cdataset, GLGMPNN model also achieves the best scores in terms of AUROC and AUPR.

LGMPNN utilized to aggregate neighbourhood information in the GLGMPNN model can be regarded as a simplification of GCN. Compared with the other GCN-based models, LAGCN and DRWBCNF, it is obvious that the performance of GLGMPNN is greatly improved. The node embeddings learned by GLGMPNN are more comprehensive for analysing whether the associations between drugs and diseases are available. Furthermore, LAGCN and DRWBNCF perform worse than the matrix completion-based models (DRRS, BNNR) in terms of AUROC, where GLGMPNN performs better AUROC. The result indicates that GLGMPNN aggregates more significant neighbourhood information and captures more effective network topology information. As shown in Figure 4, the 3 lines in each half of the violin plots represent the third quartile, the median and the first quartile from top to bottom.

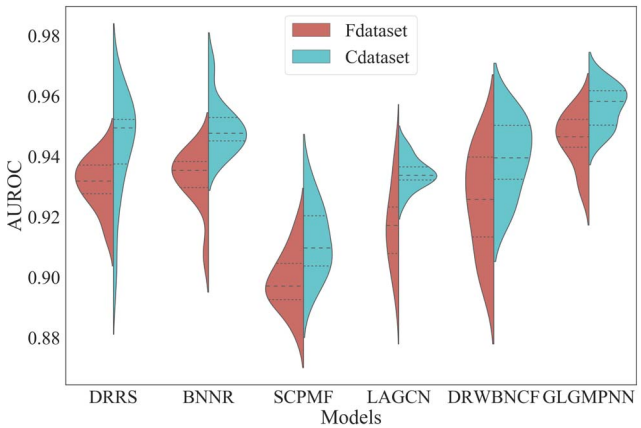

Figure 4. Violin plots of AUROC scores obtained by 10 times 10-CV on Fdataset and Cdataset.

Obviously, the AUROC scores achieved by GLGMPNN are closer both on Fdataset and Cdataset. The results demonstrate that GLGMPNN has good performance in predicting associations and is less influenced by the changes in data.

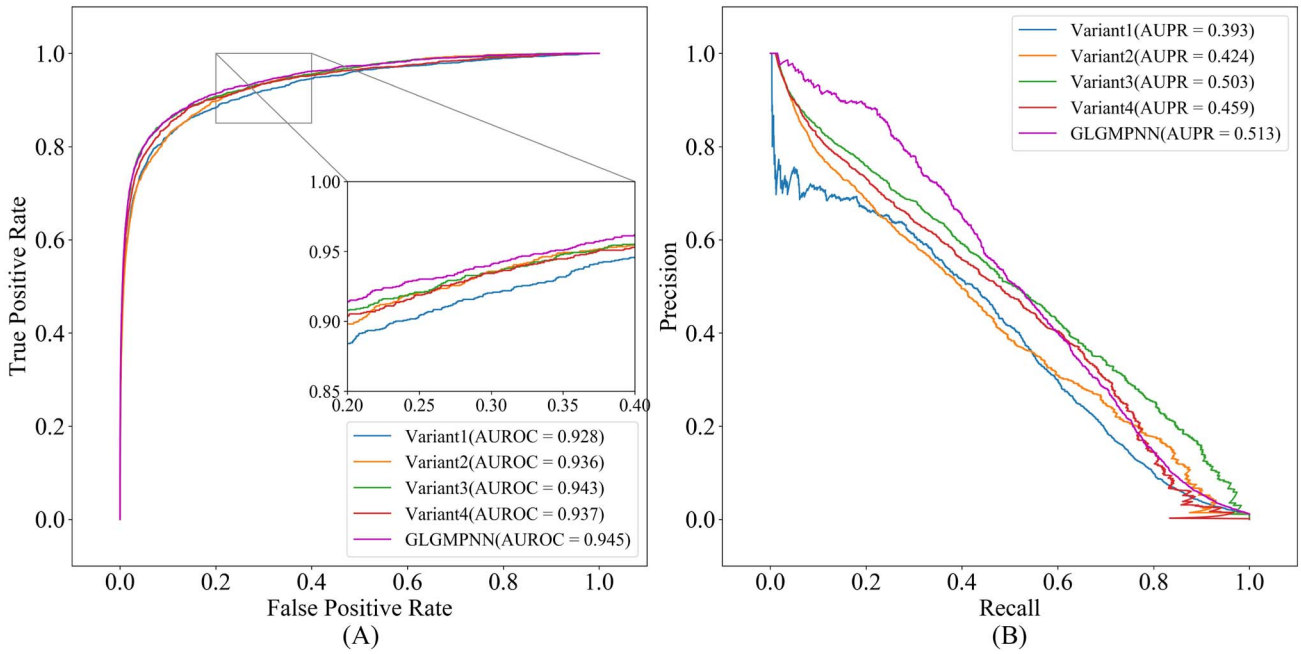

Figure 5. ROC curves (**A**) and PR curves (**B**) obtained by GLGMPNN and its variants on Fdataset.

Table 3. Performance of GLGMPNN and its variants on Fdataset

Model	AUROC	AUPR
Variant1	0.928±0.001	0.393±0.005
Variant2	0.936±0.002	0.424±0.006
Variant3	0.943±0.001	0.503±0.006
Variant4	0.937±0.001	0.459±0.005
GLGMPNN	**0.945±0.001**	**0.513±0.005**

Note: The best results are in bold.

Ablation analysis

In Table 3 and Figure 5, GLGMPNN and its variants are compared by 10-CV on Fdataset. The same decoder is applied among all models, which is used to calculate the association probability scores.

GLGMPNN mainly consists of three components: LGMPNN, the forget gate and the output gate. LGMPNN is applied to learn node embeddings from the association and the similarity networks. The introduction of the forget gate controls how much information from the association network needs to be forgotten to reduce dependence on known associations and increase the utilization of similarity information. Therefore, the forget gate must follow the use of LGMPNN-D and LGMPNN-U, which is calculated by the outputs of them. The output gate is used to further extract the initial similarity embeddings and the output of LGMPNN. The following variants are designed to present the capacity of different components. Variant1 and Variant2 test the capacity of LGMPNN-D and LGMPNN-U, respectively. Variant 3 tests the capacity of the forget gate. Variant 4 tests the capacity of the output gate based on LGMPNN and the forget gate. The variants are summarized as follows:

- Variant1 only consists of the similarity network. After LGMPNN applied in the similarity network, the output gate fuses the initial similarity embeddings and the output of LGMPNN-U.
- Likewise, Variant2 consists of the association network. Then, the embeddings are fed to the output gate.
- In Variant3, the forget gate is removed. The output of the LGMPNN-D is directly added to the output of the LGMPNN-U. Then, the fused embeddings and the initial similarity embeddings are fed to the output gate.
- In Variant4, the output of LGMPNN-D is integrated to the output of LGMPNN-U by the forget gate. Without processed by the output gate, the embeddings are fed to the decoder.

According to Table 3 and Figure 5, Variant1 and Variant2 perform well in predicting associations, demonstrating that LGMPNN can effectively capture the topology information of the association and similarity networks, respectively. Additionally, GLGMPNN which combine the similarity network and the association network significantly improves the AUROC and AUPR. This result shows that the similarity networks and the association network are all necessary and their combination by the forget gate can achieve better performance.

Furthermore, the AUPR score of GLGMPNN is much better than those of Variant3 and Variant4. It is a remarkable fact that AUPR is in favour of the identification of positive samples in sparse datasets [46], which is more informative than AUROC especially for biology data. This result demonstrates that the incorporation of the forget fate and the output gate provides more useful information for predicting associations than the individual use of the output gate and the forget gate.

To further demonstrate the capacity of LGMPNN and the gated fusion mechanism, four models based on existing variants of MPNN and fusion methods and GLGMPNN are compared in Figure 6. The four models consist of GCN, GAT (graph attention network), AVE (LGMPNN with average fusion) and ATT (LGMPNN with attention mechanism in LAGCN).

As shown in Figure 6, all models based on LGMPNN outperform GCN and GAT, even with the simple average fusion. The result demonstrates that LGMPNN has a better capacity for learning network information than GCN and GAT. Moreover, ATT has better results than AVE, which reveals that the attention mechanism

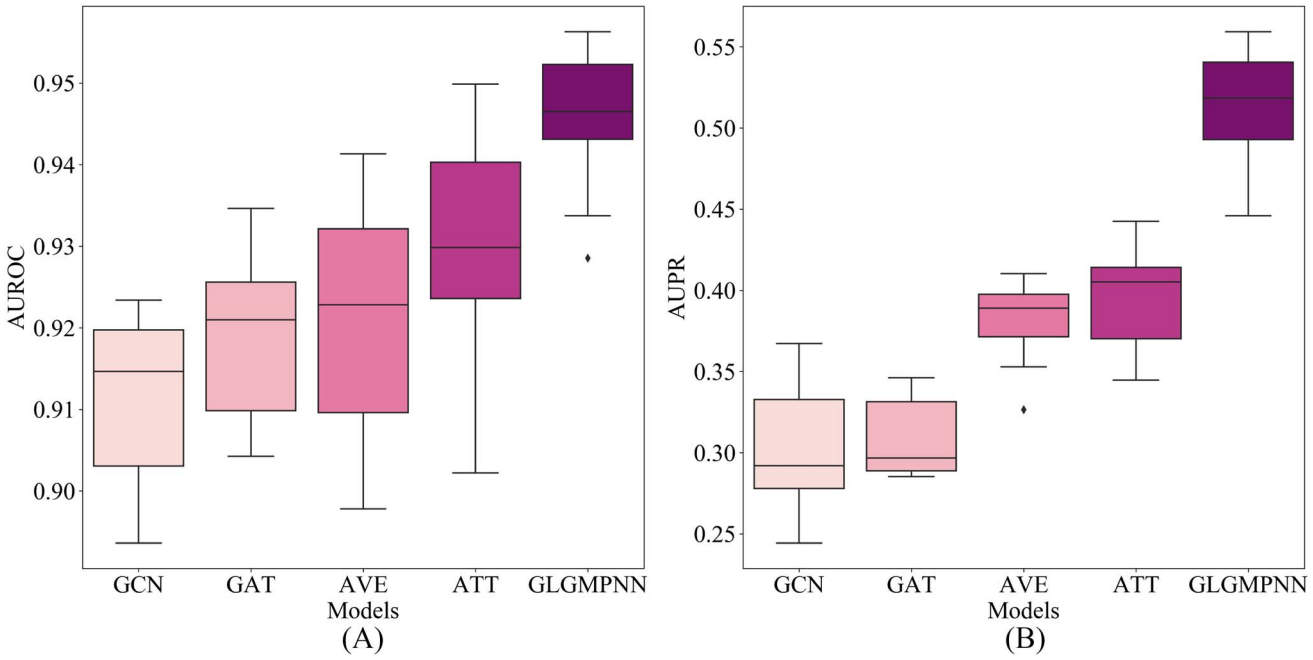

Figure 6. Box plots of AUROC (**A**) and AUPR (**B**) scores obtained by 10 times 10-CV on Fdataset.

can strengthen the performance of fusion more than a simple average operation. GLGMPNN significantly improves the performance compared with AVE and ATT, illustrating that the gated fusion mechanism composed of the forget gate and the output gate can better integrate multiple pieces of information captured by LGMPNN.

Discovering candidates for new diseases

To verify the performance of GLGMPNN in predicting new candidate diseases, we use leave-one-fold cross-validation (LOOCV) on Fdataset. First, the known associations about disease di_j are deleted as the testing set. The remaining associations are used as the training samples. Figure 7 shows the performance of all models. As shown, GLGMPNN achieves the highest score 0.815 in terms of AUROC, which is 1.7% better than BNNR based on matrix completion and 4% better than DRWBNCF based on GCN. The AUROCs of LAGCN (0.748) and DRWBNCF (0.775) are both lower than BNNR (0.798), which means these GCN-based models rely more on the associations, resulting the insufficiency of predicting candidates for new diseases. However, GLGMPNN performs better than BNNR. Despite the lack of the known associations, GLGMPNN can better extract the similarity information to discover novel drugs for unknown diseases than other models based on matrix completion and GCN.

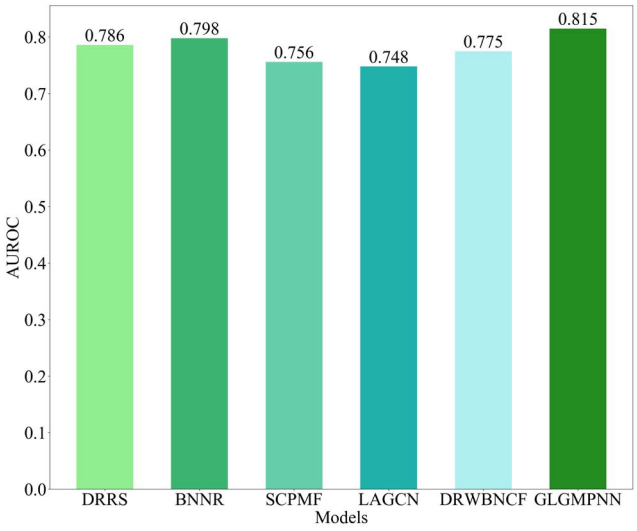

Figure 7. Performance of discovering candidates for new diseases on Fdataset.

Case study

Furthermore, we verify the ability of predicting novel drug–disease associations on Fdataset. All known drug–disease associations are used as a training set to predict new candidate associations. The top 10 highest scored candidate associations are listed in Table 4. All listed associations are unknown on Fdataset. Most of the predicted associations can be proved by the public literature and other available sources, such as DB, Comparative Toxicogenomics Database (CTD) [47] and Pubchem [48]. For example, prochlorperazine is a typical drug used in the treatment of schizophrenia that inhibits D2 dopamine receptors in the brain [49]. Generally, leuprolide has effects in treating advanced prostate cancer, central precocious puberty and endometriosis [50]. It can lower gonadotropin levels by binding to gonadotropin-releasing hormone (GnRH) receptor. In addition, the investigation in [51] demonstrates that leuprolide preforms better in identifying idiopathic hypogonadotropic hypogonadism. Moreover, baclofen can be used to treat dystonia of some patients with Parkinson's disease [52].

Then, the top 10 scored drugs for Alzheimer's disease (AD) are listed in Table 5. AD seriously affects the daily life of patients. To date, the precise aetiology of AD remains unknown [53]. Only a few approaches may temporarily relieve or improve symptoms of AD. Therefore, finding candidate drugs for AD has important clinical implications. As listed in Table 5, all the drug candidates predicted

Table 4. Top 10 scored associations on Fdataset

Drug	Disease	Evidence
Prochlorperazine	Schizophrenia	DB
Leuprolide	IHH[1]	[51]
Tretinoin	HCVAD[2]	CTD
Dexrazoxane	WPW Syndrome[3]	NA
Dexrazoxane	CMD2A[4]	CTD
Baclofen	Parkinson Disease, Juvenile, of Hunt	CTD
Cyproheptadine	DYT9[5]	NA
Cladribine	WPW Syndrome[3]	NA
Cyproheptadine	OCD[6]	CTD
Ergocalciferol	AVED[7]	NA

[1]Idiopathic Hypogonadotropic Hypogonadism [2]Hypercarotenemia and Vitamin A Deficiency [3]Wolff–Parkinson–White Syndrome [4]Cardiomyopathy, Dilated, 2a [5]Dystonia 9 [6]Obsessive–Compulsive Disorder [7]Ataxia with Vitamin E Deficiency

Table 5. Top 10 scored drugs for Alzheimer's disease on Fdataset

Rank	Drug	Evidence
1	Dantrolene	CTD
2	Vitamin E	DB/CTD
3	Haloperidol	DB/CTD/Pubchem
4	Mecamylamine	CTD
5	Cyproheptadine	CTD
6	Memantine	DB/CTD/Pubchem
7	Baclofen	CTD
8	Tetrabenazine	[54, 58]
9	Diltiazem	CTD
10	Azathioprine	CTD

by the GLGMPNN model have been confirmed to have associations with AD. For instance, [54] found that tetrabenazine has effects on neural function similar to the calcineurin inhibitors cyclosporine A (CsA), which truly reduces the incidence of AD [55]. Diltiazem also has been demonstrated to play a beneficial role in treating aluminum chloride-induced dementia, for which diltiazem can suppress amyloid beta production related to the development of AD [56]. In [57], the result showed that thiopurine, including azathioprine, can inhibit Rac1 (Ras-related C3 botulinum toxin substrate 1) activation, which may lead to the synaptic degeneration, a predictor of clinical AD symptoms. Hence, patients with longer thiopurine exposure time have a lower rate of AD.

Conclusions

In this paper, GLGMPNN is proposed to precisely discover candidate drugs for treating diseases. First, LGMPNN-U and LGMPNN-D are designed on the similarity networks and the association network, respectively. Then, the forget gate and the output gate are combined to fuse the embeddings obtained from multiple networks. The case study shows that GLGMPNN has good performance in predicting candidate associations.

Although GLGMPNN achieves good results, there are still some shortcomings. The similarity information only consists of the drug chemical structure and the disease phenotype. In the future, more biology information, such as drug–drug similarity based on side effects and disease–disease similarity based on genetic signatures, should be introduced to make the similarity networks more comprehensive. Additional methods should focus on mining the differences of multiple similarities and fusing them more effectively for predictions. Besides, LGMPNN-U and LGMPNN-D are applied on the different networks, respectively. In future work, LGMPNN is expected to be used in the heterogeneous network, mining more comprehensive information between different biology entities.

Key Points

- A deep learning-based model combining the proposed LGMPNN and the gated fusion mechanism is developed for predicting drug–disease associations, called GLGPMNN.
- LGPMNN-U and LGMPNN-D are applied to capture the topology information of the similarity networks and the association network, respectively.
- The trainable matrices in LGMPNN-U and LGMPNN-D are removed to accelerate the training process.
- The gated fusion mechanism composed of the forget gate and the output gate fuses the similarity information and the useful association information to improve the performance of the GLGPMNN model.

Funding

National Natural Science Foundation of China [62172254, 62172253].

Data availability

The implementation of GLGMPNN is available at: https://github.com/bdtree/GLGMPNN.

References

1. McGuire AL, Gabriel S, Tishkoff SA, *et al.* The road ahead in genetics and genomics. *Nat Rev Genet* 2020;**21**(10):581–96.
2. May M. Life science technologies: Big biological impacts from big data. *Science* 2014;**344**(6189):1298–300.
3. Avorn J. The $2.6 Billion Pill - Methodologic and Policy Considerations. *N Engl J Med* 2015;**372**(20):1877–9.
4. Zou J, Zheng MW, Li G, *et al.* Advanced systems biology methods in drug discovery and translational biomedicine. *Biomed Res Int* 2013;**2013**:1–8.
5. Cheng F, Liu C, Jiang J, *et al.* Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput Biol* 2012;**8**(5):e1002503.
6. Campillos M, Kuhn M, Gavin A-C, *et al.* Drug target identification using side-effect similarity. *Science* 2008;**321**(5886):263–6.
7. Luo H, Li M, Yang M, *et al.* Biomedical data and computational models for drug repositioning: a comprehensive review. *Brief Bioinform* 2021;**22**(2):1604–19.
8. Zhang ZC, Zhang XF, Wu M, *et al.* A graph regularized generalized matrix factorization model for predicting links in biomedical bipartite networks. *Bioinformatics* 2020;**36**(11):3474–81.
9. Luo Y, Zhao X, Zhou J, *et al.* A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun* 2017;**8**(1):573.
10. Xie G, Li J, Gu G, *et al.* BGMSDDA: a bipartite graph diffusion algorithm with multiple similarity integration for drug-disease association prediction. *Mol Omics* 2021;**17**(6):997–1011.

11. Ezzat A, Zhao P, Wu M, *et al.* Drug-target interaction prediction with graph regularized matrix factorization. *IEEE/ACM Trans Comput Biol Bioinform* 2017;**14**(3):646–56.
12. Zhang W, Yue X, Lin W, *et al.* Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinformatics* 2018;**19**:233.
13. Luo H, Li M, Wang S, *et al.* Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics* 2018;**34**(11):1904–12.
14. Zhang W, Xu H, Li X, *et al.* DRIMC: an improved drug repositioning approach using bayesian inductive matrix completion. *Bioinformatics* 2020;**36**(9):2839–47.
15. Wang B, Mezlini AM, Demir F, *et al.* Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014;**11**(3):333–7.
16. Thafar MA, Olayan RS, Albaradei S, *et al.* DTi2Vec: Drug-target interaction prediction using network embedding and ensemble learning. *J Chem* 2021;**13**(1):1–18.
17. Grover A, Leskovec J. node2vec: Scalable Feature Learning for Networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, United States: Association for Computing Machinery, 2016, 855–64.
18. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;**29**:1189–232.
19. Wang J, Wang W, Yan C, *et al.* Predicting Drug-Disease Association Based on Ensemble Strategy. *Front Genet* 2021;**12**:666575.
20. Zhang W, Yue X, Huang F, *et al.* Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network. *Methods* 2018;**145**:51–9.
21. Chen X, Li T-H, Zhao Y, *et al.* Deep-belief network for predicting potential miRNA-disease associations. *Brief Bioinform* 2020;**22**(3):bbaa186.
22. Wang C-C, Li T-H, Huang L, *et al.* Prediction of potential miRNA-disease associations based on stacked autoencoder. *Brief Bioinform* 2022;**23**(2):bbac021.
23. Zhou F, Yin MM, Jiao CN, *et al.* Predicting miRNA-Disease Associations Through Deep Autoencoder With Multiple Kernel Learning. *IEEE Trans Neural Netw Learn Syst* 2021;1–10. https://doi.org/10.1109/TNNLS.2021.3129772.
24. Long Y, Luo J, Zhang Y, *et al.* Predicting human microbe-disease associations via graph attention networks with inductive matrix completion. *Brief Bioinform* 2021;**22**(3):bbaa146.
25. Yu Z, Huang F, Zhao X, *et al.* Predicting drug-disease associations through layer attention graph convolutional network. *Brief Bioinform* 2021;**22**(4):bbaa243.
26. Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. New York, NY, United States: Curran Associates Inc., 2017, 6000–10.
27. Meng Y, Lu C, Jin M, *et al.* A weighted bilinear neural collaborative filtering approach for drug repositioning. *Brief Bioinform* 2022;**23**(2):bbab581.
28. Gottlieb A, Stein GY, Ruppin E, *et al.* PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol* 2011;**7**(1):496.
29. Luo H, Wang J, Li M, *et al.* Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics* 2016;**32**(17):2664–71.
30. Wishart DS, Feunang YD, Guo AC, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**(D1):D1074–82.
31. Hamosh A, Scott AF, Amberger JS, *et al.* Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2005;**33**(Database issue):D514–7.
32. Weininger DSMILES. a chemical language and information system. 1. introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 1988;**28**(1):31–6.
33. van Driel MA, Bruggeman J, Vriend G, *et al.* A text-mining analysis of the human phenome. *Eur J Hum Genet* 2006;**14**(5):535–42.
34. Gilmer J, Schoenholz SS, Riley PF. Neural message passing for Quantum chemistry. In: *Proceedings of the 34th International Conference on Machine Learning*. New York, NY, United States: PMLR, 2017, 1263–72.
35. Nyamabo AK, Yu H, Liu Z, *et al.* Drug-drug interaction prediction with learnable size-adaptive molecular substructures. *Brief Bioinform* 2022;**23**(1):bbab441.
36. Zhang H, Cui H, Zhang T, *et al.* Learning multi-scale heterogenous network topologies and various pairwise attributes for drug-disease association prediction. *Brief Bioinform* 2022;**23**(2):bbac009.
37. Cai L, Lu C, Xu J, *et al.* Drug repositioning based on the heterogeneous information fusion graph convolutional network. *Brief Bioinform* 2021;**22**(6):bbab319.
38. Bang S, Kim JH, Shin H. Causality modeling for directed disease network. *Bioinformatics* 2016;**32**(17):i437–44.
39. Yang K, Swanson K, Jin W, *et al.* Analyzing learned molecular representations for property prediction. *J Chem Inf Model* 2019;**59**(8):3370–88.
40. He X, Deng K, Wang X, *et al.* LightGCN: Simplifying and powering graph convolution network for recommendation. In: *International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, United States: Association for Computing Machinery, 2020, 639–48.
41. Greff K, Srivastava RK, Koutník J, *et al.* LSTM: a search space odyssey. *IEEE Trans Neural Netw Learn Syst* 2016;**28**(10):2222–32.
42. Huang Y-A, Hu P, Chan KCC, *et al.* Graph convolution for predicting associations between miRNA and drug resistance. *Bioinformatics* 2020;**36**(3):851–8.
43. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: *International Conference for Learning Representations*. New York, NY, United States: arXiv.org, 2015, 1–14.
44. Yang M, Luo H, Li Y, *et al.* Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* 2019;**35**(14):i455–63.
45. Meng Y, Jin M, Tang X, *et al.* Drug repositioning based on similarity constrained probabilistic matrix factorization: COVID-19 as a case study. *Appl Soft Comput* 2021;**103**:107135.
46. Pliakos K, Vens C. Network inference with ensembles of bi-clustering trees. *BMC Bioinformatics* 2019;**20**(1):525.
47. Davis AP, Grondin CJ, Johnson RJ, *et al.* Comparative toxicogenomics database (CTD): update 2021. *Nucleic Acids Res* 2021;**49**(D1):D1138–43.
48. Kim S, Chen J, Cheng T, *et al.* PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* 2019;**47**(D1):D1102–9.
49. Stępnicki P, Kondej M, Kaczor AA. Current concepts and treatments of schizophrenia. *Molecules* 2018;**23**(8):2087.
50. Wang L, Jiang Q, Wang M, *et al.* The effect of triptorelin and leuprolide on the level of sex hormones in girls with central precocious puberty and its clinical efficacy analysis. *Transl Pediatr* 2021;**10**(9):2307.
51. Wei C, Crowne EC. Recent advances in the understanding and management of delayed puberty. *Arch Dis Child* 2016;**101**(5):481–8.

52. Bellows S, Jankovic J. Treatment of dystonia and tics. *Clin Park Relat Disord* 2020;**2020**(2):12–9.
53. Reitz C, Brayne C, Mayeux R. Epidemiology of Alzheimer disease. *Nat Rev Neurol* 2011;**7**(3):137–52.
54. Tucker Edmister S, Del Rosario HT, Ibrahim R, *et al*. Novel use of FDA-approved drugs identified by cluster analysis of behavioral profiles. *Sci Rep* 2022;**12**(1):6120.
55. Taglialatela G, Rastellini C, Cicalese L. Reduced incidence of dementia in solid organ transplant patients treated with calcineurin inhibitors. *J Alzheimers Dis* 2015;**47**(2): 329–33.
56. Rani A, Sodhi RK, Kaur A. Protective effect of a calcium channel blocker "diltiazem" on aluminum chloride-induced dementia in mice. *Naunyn Schmiedebergs Arch Pharmacol* 2015;**388**(11): 1151–61.
57. Sutton SS, Magagnoli J, Cummings T, *et al*. Association between thiopurine medication exposure and Alzheimer's disease among a cohort of patients with inflammatory bowel disease. *Alzheimers Dement (N Y)* 2019;**5**:809–13.
58. Bukhari SNA. Dietary Polyphenols as Therapeutic Intervention for Alzheimer's Disease: A Mechanistic Insight. *Antioxidants* 2022;**11**(3):554.