Briefings in Bioinformatics, 00(00), 2020, 1–11

doi: 10.1093/bib/bbaa243
Problem solving protocol

Predicting drug–disease associations through layer attention graph convolutional network

Zhouxin Yu†, Feng Huang†, Xiaohan Zhao, Wenjie Xiao and Wen Zhang

Corresponding author: Wen Zhang, College of Informatics, Huazhong Agricultural University, Wuhan, Hubei 430070, China; Hubei Engineering Technology Research Center of Agricultural Big Data, Wuhan, Hubei 430070, China. E-mail: zhangwen@mail.hzau.edu.cn
†These authors contributed equally to this work.

Abstract

Background: Determining drug–disease associations is an integral part in the process of drug development. However, the identification of drug–disease associations through wet experiments is costly and inefficient. Hence, the development of efficient and high-accuracy computational methods for predicting drug–disease associations is of great significance.
Results: In this paper, we propose a novel computational method named as layer attention graph convolutional network (LAGCN) for the drug–disease association prediction. Specifically, LAGCN first integrates the known drug–disease associations, drug–drug similarities and disease–disease similarities into a heterogeneous network, and applies the graph convolution operation to the network to learn the embeddings of drugs and diseases. Second, LAGCN combines the embeddings from multiple graph convolution layers using an attention mechanism. Third, the unobserved drug–disease associations are scored based on the integrated embeddings. Evaluated by 5-fold cross-validations, LAGCN achieves an area under the precision–recall curve of 0.3168 and an area under the receiver–operating characteristic curve of 0.8750, which are better than the results of existing state-of-the-art prediction methods and baseline methods. The case study shows that LAGCN can discover novel associations that are not curated in our dataset.
Conclusion: LAGCN is a useful tool for predicting drug–disease associations. This study reveals that embeddings from different convolution layers can reflect the proximities of different orders, and combining the embeddings by the attention mechanism can improve the prediction performances.

Key words: drug; disease; drug–disease association prediction; layer attention; graph convolutional network

Introduction

Drug development is an extremely lengthy and costly process, which costs 2.6 billion dollars and takes 12 years on average for a new drug [1]. Determining the associated diseases of a new drug (such as off-label indications and side-effects) is an important part of drug development. Computationally, identifying drug–disease associations efficiently picks out candidate associations and guides wet experiments for further validation, and hence can accelerate drug development. The development of high-accuracy (ACC) computational methods is of far-reaching significance with still great challenges and has attracted continuous attention.

The previous computational methods for predicting drug–disease associations can be roughly divided into three categories

Zhouxin Yu is a student of the College of Informatics, Huazhong Agricultural University. His research interests include bioinformatics and machine learning.
Feng Huang is a PhD student of the College of Informatics, Huazhong Agricultural University. His research interests include bioinformatics and machine learning.
Xiaohan Zhao is a student of the College of Informatics, Huazhong Agricultural University. Her research interests include bioinformatics and machine learning.
Wenjie Xiao is a student of the Information School, University of Washington. Her research interests include bioinformatics and machine learning.
Wen Zhang obtained a bachelor's degree and a master's degree in computational mathematics from Wuhan University in 2003, 2006, and got a doctoral degree in computer science from Wuhan University in 2009. He is now a professor in the College of Informatics, Huazhong Agricultural University, People's Republic of China. His research interests include bioinformatics and machine learning.
Submitted: 29 June 2020; **Received (in revised form):** 16 August 2020

[2, 3], i.e. network diffusion-based methods, machine learning-based methods and deep learning-based methods.

Network diffusion-based methods generally link drugs with diseases through message propagation on the paths bridging different networks [4–7]. For example, Wang *et al.* [8] designed a triple-layer heterogeneous graph-based inference method (TL-HGBI) to infer potential links between drugs and diseases. Luo *et al.* applied random walkers respectively on the drug–disease bipartite network [9] and on a drug–target–disease heterogeneous network [10] to predict novel drug–disease associations. Although the network diffusion-based methods have advantages of good interpretability, their performances are not satisfactory [2].

Machine learning techniques have been widely adopted to develop more accurate models for drug–disease association prediction. For example, a number of feature-based classification methods [11–15] take drug–disease pairs as samples, encode the side information of drugs and diseases into feature vectors to characterize the samples and then train classifiers to distinguish whether the associations exist or not. However, feature-based classification methods heavily rely on the extraction of features and the selection of negative samples. Thus, a surge of more sophisticated techniques, such as sparse subspace learning [16], semi-supervised graph cut [17], label propagation [18], regularized least squares [19], matrix factorization [20–22] and matrix completion [23–26], have been applied to the drug–disease association prediction. In particular, matrix factorization and matrix completion techniques are of great popularity in community, due to their flexibility in integrating prior information, and have shown promising results in predicting drug–disease associations, but it is challenging to deploy them on large-scale data because of high-complexity matrix operations.

Deep learning methods have been demonstrated to be more effective in many tasks, including but not limited to face recognition, question answering system, computational biology [27, 28], and also have successful applications in the drug–disease association prediction [29–31]. For example, Zeng *et al.* [29] recently developed a network-based deep learning approach, termed deepDR. It firstly calculated positive pointwise mutual information (PPMI) matrices from 10 drug-related networks and use them as features, then fused PPMI matrices by multimodal deep autoencoder and finally exploited the fused features to infer new applications for existing drugs by collective variational autoencoder. The advantage of deepDR is taking full use of topological information of drug similarity networks. However, deepDR does not take the side information of diseases into account. DeepDR has two separate components instead of a full end-to-end framework, and this may have an impact on the performances of prediction models.

Graph convolutional network (GCN) [32], extending convolutional neural networks for processing graph data, is readily embedded in end-to-end architectures to perform specific tasks with graph inputs, captures structural information of graphs via message passing between the nodes of graphs and retains high interpretability. Recently, it demonstrates convincing performances on biomedical network analysis, such as microRNA (miRNA)–disease association prediction [33], polypharmacy side-effect prediction [34] and miRNA–drug resistance association prediction [35].

In this paper, we develop a novel end-to-end layer attention graph convolutional network (LAGCN) method for predicting drug–disease associations. We first construct heterogeneous networks by bridging the known drug–disease associations, drug–drug similarities and disease–disease similarities. Then, we use the graph convolution operation on the heterogeneous network to learn the embeddings of drugs and diseases. Given that the embeddings from multiple convolution layers reflect proximities of different orders [36] between nodes in the network, we resort to the attention mechanism [37] to integrate all useful structural information from multiple graph convolution layers. Finally, the predictive scores for unobserved drug–disease associations are given by a well-defined score function based on the integrated embeddings. According to the reliable *in silico* experiments, our proposed method LAGCN achieves the area under the precision–recall curve (AUPR) score of 0.3168 and area under the receiver–operating characteristic curve (AUC) score of 0.8750, and performs better than other state-of-the-art methods and baseline methods.

The main contributions of this work are summarized as follows:

- We propose a full end-to-end graph-based deep learning method, termed LAGCN, for effectively predicting drug–disease associations.
- LAGCN utilizes a GCN to capture structural information from the heterogeneous network composed of drug–disease associations, drug–drug similarities and disease–disease similarities.
- The attention mechanism is introduced to combine the embeddings from different convolution layers, which leads to a more informative representation of drugs and diseases.

Materials

Datasets

The data in our previous studies [20, 38] are assembled as the main dataset in this paper. The main dataset contains 18 416 drug–disease associations between 269 drugs and 598 diseases derived from Comparative Toxicogenomics Database (CTD) [39]. The comprehensive information about drugs, such as targets, enzymes, drug–drug interactions, pathways and substructures, is obtained from DrugBank database [40]. The diseases are normalized through standard terms from Medical Subject Headings (MeSH). Considering that therapeutic associations may have special significance for drug discovery, we also extract 6244 therapeutic associations annotated in CTD from the main dataset as the therapeutic dataset. Detailed information about two datasets is summarized in Table 1.

Construction of the heterogeneous network

Drug–drug similarities

Drugs usually have different features describing biological or chemical characteristics. One drug can be encoded as a binary feature vector where each element means the presence or absence of features descriptor. Since we have different types of features, we can convert drugs into multiple types of feature vectors and calculate various drug–drug similarities based on these features by using different similarity measures. To the best of our knowledge, Jaccard index [29, 41] and Cosine similarity [21] are two prevailing measures for the drug–drug similarities.

Jaccard index between two binary feature vectors x_i and x_j is calculated by

$$S_{ij}^{r} = \frac{|x_i \cap x_j|}{|x_i \cup x_j|} \quad (1)$$

Table 1. Summary of two datasets

Dataset	Drugs	Diseases	Known associations	Drug features				
				Target	Enzyme	Drug–drug interactions	Pathway	Substructure
Main dataset	269	598	18 416	623	247	2086	465	881
Therapeutic dataset	269	598	6244	623	247	2086	465	881

where $\left|x_i \cap x_j\right|$ denotes the number of cases where both elements in x_i and the corresponding ones of x_j are equal to 1, and $\left|x_i \cup x_j\right|$ denotes the number of the cases where either the elements of x_i or the corresponding ones of x_j are equal to 1.

Cosine similarity between two binary feature vectors x_i and x_j is calculated by

$$S_{ij}^{r} = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|} \tag{2}$$

where $\|x_i\|$ denotes the L2-norm of x_i.

In this work, we adopt the Jaccard index to calculate the drug–drug similarities for our prediction methods and also consider Cosine similarity. Jaccard index and Cosine similarity are compared in the section of 'Results and Discussion'. Since we have five types of drug features in our datasets, we also calculate drug–drug similarities based on different features and compare these similarities.

Disease–disease similarities

MeSH descriptors of diseases can be represented as hierarchical directed acyclic graphs (DAGs). As described in [42], disease–disease similarities can be calculated using the DAG structures. For a disease d, we represent its hierarchical relationship by $\mathrm{DAG}(d) = \left(\mathcal{N}(d), \mathcal{E}(d)\right)$, where $\mathcal{N}(d)$ is the set of nodes containing d and its ancestors, and $\mathcal{E}(d)$ denotes the set of direct links from parent nodes to their child nodes. Based on this DAG structure, the contribution of a node n in $\mathrm{DAG}(d)$ to the semantic value of disease d is given by

$$C_d(n) = \begin{cases} 1 & \text{if } n = d \\ \max\left\{\Delta * C_d(n') \mid n' \in \text{children of } n\right\} & \text{if } n \neq d \end{cases} \tag{3}$$

where Δ is a contribution factor ranging from 0 to 1, and here Δ is set to 0.5. The semantic value of disease d is defined as $\mathrm{DV}(d) = \sum_{n \in \mathcal{N}(d)} C_d(n)$. It is believed that diseases with more common ancestors in the DAG are prone to have higher semantic similarities. According to this hypothesis, we calculate the semantic similarity between the two diseases d_i and d_j by

$$S_{ij}^{d} = \frac{\sum_{n \in \mathcal{N}(d_i) \cap \mathcal{N}(d_j)} \left(C_{d_i}(n) + C_{d_j}(n)\right)}{\mathrm{DV}(d_i) + \mathrm{DV}(d_j)} \tag{4}$$

Heterogeneous network

The heterogeneous network is constructed based on drug–disease associations, disease–diseases similarities and disease–disease similarities.

We denote drug–disease associations as a binary matrix $A \in \{0,1\}^{N \times M}$, where M, N denote the number of diseases and drugs, respectively. A_{ij} is equal to 1, if a drug r_i has association with a disease d_j; otherwise $A_{ij} = 0$. The pairwise similarities between N drugs are denoted as a similarity matrix S^r with S_{ij}^r as its (i,j)th entry; the pairwise similarities between M diseases are denoted as a similarity matrix S^d with S_{ij}^d as its (i,j)th entry. We normalize the similarity matrices by $\sim S^r = {D_r}^{-\frac{1}{2}} S^r {D_r}^{-\frac{1}{2}}$ and $\sim S^d = {D_d}^{-\frac{1}{2}} S^d {D_d}^{-\frac{1}{2}}$, where $D_r = \mathrm{diag}(\sum_j S_{ij}^r)$ and $D_d = \mathrm{diag}(\sum_j S_{ij}^d)$. Finally, we construct the heterogeneous network defined by the adjacency matrix:

$$A_H = \begin{bmatrix} \sim S^r & A \\ A^T & \sim S^d \end{bmatrix} \in \mathbb{R}^{(N+M)\times(N+M)} \tag{5}$$

Layer attention graph convolutional network

In this section, we introduce the LAGCN for drug–disease association prediction. The workflow of LAGCN is briefly shown in Figure 1.

Method architecture

GCN [32] is a multilayer connected neural network architecture for learning low-dimensional representations of nodes from graph-structured data. Each layer of GCN aggregates neighbor's information to reconstruct the embeddings as the inputs of the next layer through the direct links of graphs.

Specifically, given a network with the corresponding adjacency matrix G, the layerwise propagation rule of GCN is formulated as

$$H^{(l+1)} = f\left(H^{(l)}, G\right) = \sigma\left(D^{-\frac{1}{2}}\, G\, D^{-\frac{1}{2}}\, H^{(l)}\, W^{(l)}\right) \tag{6}$$

where $H^{(l)}$ is the embeddings of nodes at the lth layer, $D = \mathrm{diag}(\sum_j G_{ij})$ is the degree matrix of G, $W^{(l)}$ is a layer-specific trainable weight matrix and $\sigma(\cdot)$is a non-linear activation function.

To build a GCN-based encoder for learning the low-dimensional representations of drugs and diseases, we consider combining node similarities and directly linked association information through deploying GCN on our constructed heterogeneous graph A_H. First, we introduce a penalty factor μ to control the contribution of similarities in the propagation process of GCN. To be specific, we set the input graph G as

$$G = \begin{bmatrix} \mu \sim S^r & A \\ A^T & \mu \sim S^d \end{bmatrix} \tag{7}$$

Then, we initialize embeddings as

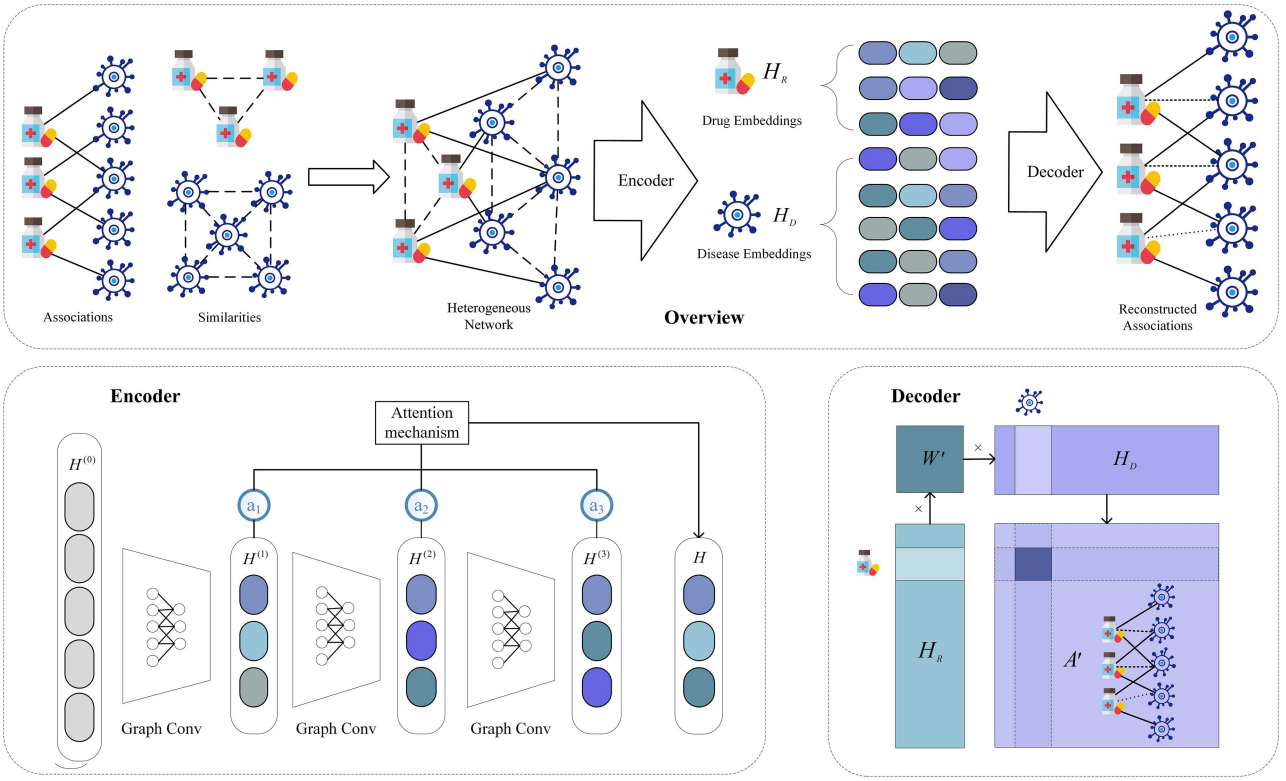

Figure 1. The workflow of layer attention graph convolutional network.

$$H^{(0)} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \tag{8}$$

With the above settings, the first layer of our GCN encoder is formulated as

$$H^{(1)} = \sigma\left(D^{-\frac{1}{2}} G D^{-\frac{1}{2}} H^{(0)} W^{(0)}\right) \tag{9}$$

where $W^{(0)} \in \mathbb{R}^{(N+M)\times k}$ is an input-to-hidden weight matrix, $H^{(1)} \in \mathbb{R}^{(N+M)\times k}$ is the first-layer embeddings of the nodes (drugs and diseases) of the heterogeneous network A_H, k is the dimensionality of the embeddings and G is defined in Equation (7). The subsequent layers of our GCN encoder follow the Equation (6) for $l = 1, 2, \ldots, L$ with $W^{(l)} \in \mathbb{R}^{k\times k}$ and G defined in Equation (7). After L iterations, we can obtain L k-dimensionality embeddings from different graph convolution layers. Exponential linear unit [43] is used as the non-linear activation function in all graph convolution layers, which not only accelerates learning procedure but also significantly enhances generalization performance.

The embeddings at different layers capture different structural information of the heterogeneous network. For example, the first layer harvests the direct link information, and higher layers capture multihop neighbor information (high-order proximity) by iteratively updating the embeddings [44, 45]. Considering the contributions of different embeddings at different layers are inconsistent, we introduce an attention mechanism to combine these embeddings and obtain final embeddings of drugs and diseases as $\begin{bmatrix} H_R \\ H_D \end{bmatrix} = \sum a_l H^l$, where $H_R \in \mathbb{R}^{N\times k}$ is the final embeddings of drugs, $H_D \in \mathbb{R}^{M\times k}$ is the final embeddings of diseases and a_l is auto-learned by neural networks and initialized as $1/(l+1), l = 1, 2, \ldots, L$.

To reconstruct the adjacency matrix for drug–disease associations, a bilinear decoder $A' = f(H_R, H_D)$ created by [33] is adopted:

$$A' = \text{sigmoid}\left(H_R W' H_D^T\right) \tag{10}$$

where $W' \in \mathbb{R}^{k\times k}$ is a trainable matrix. The predicted score for the association between drug r_i and disease d_j is given by the corresponding (i,j)th entry of A', denoted as a'_{ij}.

Optimization

From a dataset with N drugs and M diseases, we take drug–disease association pairs as positive instances and take other pairs as negative instances. Herein, the set of positive instances and the set of negative instances are respectively denoted as $\mathcal{Y}^+$ and $\mathcal{Y}^-$. Differentiating two types of drug–disease pairs is a binary classification problem. However, the number of associations is much less than that of the drug–disease pairs, which have no observed associations. Here, we adopt the weighted cross-entropy as loss function:

$$\text{Loss} = -\frac{1}{N\times M}\left(\lambda \times \sum_{(i,j)\in\mathcal{Y}^+} \log a'_{ij} + \sum_{(i,j)\in\mathcal{Y}^-} \log\left(1 - a'_{ij}\right)\right) \tag{11}$$

where (i,j) denotes the pair for drug r_i and disease d_j. $\lambda = \frac{|\mathcal{Y}^-|}{|\mathcal{Y}^+|}$, $|\mathcal{Y}^+|$ and $|\mathcal{Y}^-|$ are the number of instances in $\mathcal{Y}^+$ and $\mathcal{Y}^-$. The

Table 2. Performances of LAGCN with different drug–drug similarities and without similarities

Similarity measures	Drug feature	AUPR	AUC	RE	SP	ACC	F1
Jaccard	Target	0.3168	0.8750	0.3600	0.9760	0.9605	0.3150
	Enzyme	0.3166	0.8758	0.3567	0.9764	0.9608	0.3149
	Drug interaction	0.3163	0.8761	0.3533	0.9772	0.9615	0.3167
	Pathway	0.3149	0.8761	0.3603	0.9758	0.9603	0.3141
	Substructure	0.3115	0.8765	0.3475	0.9778	0.9619	0.3153
	Average similarity	0.3165	0.8763	0.3773	0.9739	0.9588	0.3159
	Concatenated feature-based similarity	0.3134	0.8761	0.3505	0.9776	0.9618	0.3162
Cosine	Target	0.3149	0.8737	0.3717	0.9739	0.9587	0.3121
	Enzyme	0.3137	0.8739	0.3701	0.9746	0.9593	0.3144
	Drug interaction	0.3136	0.8753	0.3464	0.9776	0.9617	0.3130
	Pathway	0.3149	0.8746	0.3591	0.9758	0.9603	0.3131
	Substructure	0.3110	0.8748	0.3609	0.9755	0.9600	0.3128
	Average similarity	0.3113	0.8743	0.3606	0.9756	0.9601	0.3132
	Concatenated features-based similarity	0.3131	0.8754	0.3596	0.9757	0.9602	0.3125
	LAGCN-NH	0.2952	0.8455	0.3577	0.9491	0.9342	0.2918

RE, recall; SP, specificity; F1, F1-measure.

weight factor λ imposes the importance of observed associations to reduce the influence of data imbalance.

All the trainable weight matrices ($W^{(l)}$ and W') are initialized by the Xaiver initialization method [46]. Then, we use Adam optimizer [47] to minimize the loss function. Adam optimizer can update the weights of neural network iteratively based on training data. To prevent over-fitting, we introduce node dropout [48] and regular dropout [49] to the graph convolution layers. This node dropout can be considered as training of different models on various small subnetworks, and unknown drug–disease pairs are predicted by integrating these small models [50]. Besides, the cyclic learning rate [51] is used during the optimization. A simple cyclic learning rate makes a change in the learning rate between the maximum learning rate and the minimum, helping us to balance the training speed and ACC.

Results and discussion

Experimental setting

In our experiments, we adopted 5-fold cross-validation (5-CV) to evaluate the performances of prediction methods. All known drug–disease associations are randomly split into five equal-sized subsets. The cross-validation process is repeated fives times, and every subset is used as the testing set in turn while the remaining four subsets are used as the training set. In each fold, a prediction model is constructed on known associations in the training set and is used to predict associations in the testing set. We adopt the AUPR and the AUC as primary metrics, for they can measure the performances of methods without any specific threshold. Besides, the threshold-based metrics are also calculated, i.e. recall (also known as sensitivity), specificity, ACC, precision and F1-measure (F1).

There are several hyperparameters in LAGCN such as the dimensionality of embeddings k, the number of layers L, the initial learning rate of optimizer lr, the total training epochs of LAGCN α, two dropout rates (node dropout and regular dropout) β, γ and the penalty factor μ in the heterogeneous network. We consider different combinations of these parameters from the ranges $\alpha \in \{500, 1000, 2000, 4000\}$, $\beta, \gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ and $\mu \in \{2, 4, 6, 8, 10\}$. By adjusting the parameters empirically, we set the parameter $k = 64, L = 3, lr = 0.008, \alpha = 4000, \beta = 0.6, \gamma = 0.4$ and $\mu = 6$ for LAGCN in the following experiments.

Results of LAGCN

Influence of different heterogeneous networks

LAGCN makes use of the drug–disease heterogeneous network to build the prediction model. The drug–disease heterogeneous network consists of known drug–disease associations, drug–drug similarities and disease–disease similarities. Since we consider five types of drug features and two similarity measures, we can train LAGCN on different heterogeneous networks based on different drug–drug similarities, and then discuss how these drug–drug similarities influence the performances of LAGCN.

LAGCN models based on heterogeneous networks with different drug–drug similarities are evaluated by 5-CV on the main dataset, and the corresponding results are displayed in Table 2. Jaccard index leads to slightly better results than Cosine similarity measure, and similarities based on different features produce similar performances. These results indicate that LAGCN is robust, regardless of similarity measures and drug features. Drug target-based similarities (by both Jaccard index and Cosine similarity) leads to the highest AUPR score.

Moreover, we only use the drug–disease associations to construct the network, and then build a reduced version of LAGCN based on this network, named as LAGCN-NH. According to Table 2, LAGCN-NH produces lower AUPR score and AUC score than all LAGCN models, which demonstrates that the drug–drug similarities and disease–disease similarities in the heterogeneous network contain useful information and lead to the improved performances of LAGCN.

We also integrate different drug feature-based similarities using two simple strategies, and build LAGCN models. The average similarity strategy calculates the average of drug–drug similarities based on different features to obtain the integrated similarities. The concatenated feature-based similarity strategy firstly concatenate different drug feature vectors and then calculate drug–drug similarities based on the concatenated

Table 3. Performance of LAGCN based on different embeddings

Models	AUPR	AUC	RE	SP	ACC	F1
LAGCN	0.3168	0.8750	0.3600	0.9760	0.9605	0.3150
LAGCN-AVE	0.2912	0.8675	0.3550	0.9732	0.9576	0.2971
LAGCN-CON	0.3006	0.8738	0.3467	0.9771	0.9612	0.3106
LAGCN-L1	0.2928	0.8765	0.3486	0.9777	0.9618	0.3155
LAGCN-L2	0.2921	0.8629	0.3489	0.9725	0.9568	0.2896
LAGCN-L3	0.2724	0.7319	0.4108	0.7788	0.7686	0.1909

RE, recall; SP, specificity; F1, F1-measure.

feature vectors. The results in Table 2 show that integrating different drug feature-based similarities do not necessarily lead to improved performances. The possible reason is that the known drug–disease associations make the major contribution to the prediction, and drug features bring supplementary information but have redundant information between them.

Based on the above discussion, we adopt the target-based drug–drug similarities calculated by the Jaccard index, MeSH-based disease–disease similarities and drug–disease associations to construct the heterogeneous network and then build the LAGCN models in the following study.

Effect of layer attention mechanism

Layer attention is one component of the network architecture of LAGCN and is in charge of managing and quantifying the interdependence of different convolution layers. Here, we discuss the effect of the layer attention mechanism.

We only use the embeddings at the lth layer of LAGCN with $l = 1, 2, 3$ to build LAGCN models, abbreviated as LAGCN-L1, LAGCN-L2 and LAGCN-L3. The results of all models evaluated by 5-CV on the main dataset are shown in Table 3. LAGCN-L1 and LAGCN-L2 produce better results than LAGCN-L3, indicating that the first-layer embeddings and second-layer embeddings contain more information than the third-layer embeddings. The results may be caused by the over-smoothing of GCN [52]. LAGCN that integrates embeddings at three layers produces better results than LAGCN-L1, LAGCN-L2 and LAGCN-L3.

It is believed that the lth convolution layer of GCN captures the lth-order proximity, and the attention weights denote the contributions of the embeddings at different convolution layers to the final embeddings. We implement 20 runs of 5-CV for LAGCN, and visualize the attention weights for three convolution layers in Figure 2. Three layers have different attention weights, and first layer > second layer > third layer, which meets our expectation that lower-order proximity has a greater contribution, and higher-order proximity has a lower contribution. The results also help to explain the performances of LAGCN-L1, LAGCN-L2 and LAGCN-L3 in Table 3. Therefore, paying different attention to convolution layers is necessary for building the high-ACC prediction models.

Furthermore, we consider other approaches to combining embeddings at different layers. LAGCN-AVE assigns uniform weights to different embeddings; LAGCN-CON directly concatenates different embeddings. As shown in Table 3, LAGCN produces better results than LAGCN-AVE and LAGCN-CON, showing the effectiveness of attention mechanism in LAGCN.

Comparison with other methods

In this section, we compare LAGCN with five state-of-the-art drug–disease prediction methods [8, 20, 24, 25, 29] and a baseline method [35]. We replicate them according to their publications

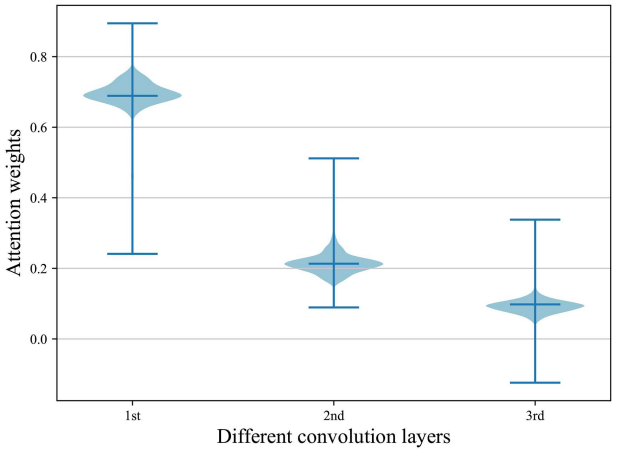

Figure 2. Attention weights for three convolution layers in LAGCN.

or using publicly available programs. The performances of all methods are shown in Table 4.

- TL-HGBI [8] constructed a three-layer heterogeneous network, and then an iterative updating algorithm was proposed to infer the probabilities of new drug–disease associations.
- SCMFDD [20] projected drug–disease association into two low-rank spaces uncovering latent features for drugs and diseases, and then introduced similarity constraints to smooth the features.
- DRRS [24] deployed a matrix completion-based recommendation system on a drug–disease heterogeneous network to predict drug–disease associations.
- BNNR [25] proposed a bounded nuclear norm regularization method to complete a drug–disease heterogeneous network.
- DeepDR [29] calculated PPMI matrices based on drug-related networks as drug features, and then proposed a multimodal deep autoencoder for fusing the features and a collective variational autoencoder for mining new associations. Here, we implement deepDR on our dataset by calculating six PPMI matrices from a drug–disease network and five drug–drug similarity networks.
- NIMCGCN [35] used GCNs to learn latent feature representations of miRNA and disease from the similarity networks, and then put the learned features into a neural inductive matrix completion model to obtain a reconstructed association matrix. NIMCGCN is a GCN-based method proposed for the miRNA–disease association prediction, and we adopt it as the baseline method.

According to Table 4, LAGCN outperforms all comparison methods in terms of most evaluation metrics. Compared

Table 4. Performance of comparison methods on two datasets

Dataset	Methods	AUPR	AUC	RE	SP	ACC	F1
Main dataset	TL-HGBI	0.0665	0.7029	0.2545	0.9284	0.9114	0.1266
	SCMFDD	0.2659	0.8727	0.3430	**0.9783**	**0.9623**	0.3143
	DRRS	0.1321	0.8429	0.3276	0.9468	0.9324	0.2178
	DeepDR	0.1353	0.8211	0.2959	0.9567	0.9400	0.1991
	BNNR	0.2262	0.8567	0.3403	0.9738	0.9578	0.2894
	NIMCGCN	0.2002	0.8533	0.3083	0.9739	0.9572	0.2661
	LAGCN	**0.3168**	**0.8750**	**0.3600**	0.9760	0.9605	**0.3150**
1pt							
Therapeutic dataset	TL-HGBI	0.0388	0.7401	0.1151	0.9830	0.9761	0.0720
	SCMFDD	0.1383	0.8754	0.2774	0.9871	0.9815	0.1934
	DRRS	0.1494	0.8873	0.2726	0.9907	0.9849	0.2249
	DeepDR	0.1011	0.8572	0.2327	0.9866	0.9806	0.1610
	BNNR	0.1832	0.8794	0.2859	**0.9918**	**0.9861**	**0.2477**
	NIMCGCN	0.0899	0.8075	0.2225	0.9859	0.9798	0.1525
	LAGCN	**0.3431**	**0.8902**	**0.5825**	0.9549	0.9520	0.1630

RE, recall; SP, specificity; F1, F1-measure.

Table 5. Performance of prediction models in deepDR's dataset

Methods	AUPR	AUC	RE	SP	ACC	F1
DeepDR	0.9201	0.9021	0.7789	0.8919	0.8354	0.8254
LAGCN	0.9487	0.9406	0.9044	0.8417	0.8731	0.8770
	0.9480	0.9391	0.9176	0.8294	0.8735	0.8789
	0.9447	0.9360	0.9079	0.8255	0.8667	0.8720
	0.9393	0.9314	0.8507	0.8787	0.8647	0.8627
	0.9404	0.9342	0.8675	0.874	0.8708	0.8703
	0.9378	0.9308	0.864	0.8652	0.8646	0.8645
	0.9416	0.935	0.872	0.8543	0.8631	0.8643
	0.9377	0.9302	0.861	0.8526	0.8568	0.8575
	0.9389	0.9311	0.8421	0.8865	0.8643	0.8612
	0.9424	0.9316	0.9277	0.8046	0.8661	0.8739

RE, recall; SP, specificity; F1, F1-measure.

with the matrix factorization and completion-based methods (SCMFDD, DRRS and BNNR), the network diffusion-based method TL-HGBI performs worse, whereas LAGCN achieves 52% improvement on average over them in terms of AUPR. GCN-based methods (NIMCGCN and LAGCN) perform better than deepDR, demonstrating that GCN may lead to better aggregation of network topological information.

DeepDR depends on their dataset that contains 10 networks, whereas LAGCN only uses drug–drug similarities, disease–disease similarities and drug–disease associations. We also compare LAGCN with deepDR based on the deepDR's dataset. Following deepDR's experimental setting, we randomly select 20% of the observed drug–disease associations and a matching number of randomly sampled unknown associations as the testing set, and the remaining 80% associations are used to train the models. Among 10 drug-related networks used in deepDR's dataset, six types of drug–drug similarities are respectively used as input of our proposed method LAGCN; other networks are also converted into drug–drug similarities by Jaccard index and then respectively used as input of LAGCN. We set $\tilde{S}^{d} = 0$ in LAGCN due to the absence of disease–disease similarities in deepDR's dataset. DeepDR is implemented by using the publicly available source code and the recommended parameter settings. As shown in Table 5, LAGCN models that use different similarities produce consistently higher AUC scores and AUPR scores than deepDR, which further shows the robustness of LAGCN. Although both LAGCN and deepDR are deep learning-based methods, the full end-to-end framework of LAGCN can help to improve the performances.

The known drug–disease associations are an essential resource for predicting potential drug–disease associations. The number of observed associations could greatly affect method performance. To test the robustness of LAGCN and comparison methods, SCMFDD, deepDR, BNNR and NIMCGCN, we randomly remove a fraction of known associations in the main dataset at a ratio $r \in \{80\%, 85\%, 90\%, 95\%, 100\%\}$ and implement 5-CV to evaluate methods. As displayed in Figure 3, the number of drug–disease associations is an important factor for the drug–disease association prediction, and more associations can result in better prediction models. However, LAGCN can produce the most robust and highest performances across different data richness among these methods.

Case study

In this section, we build an LAGCN model using all drug–disease associations and then predict novel associations. Because all known associations have been used to construct the prediction

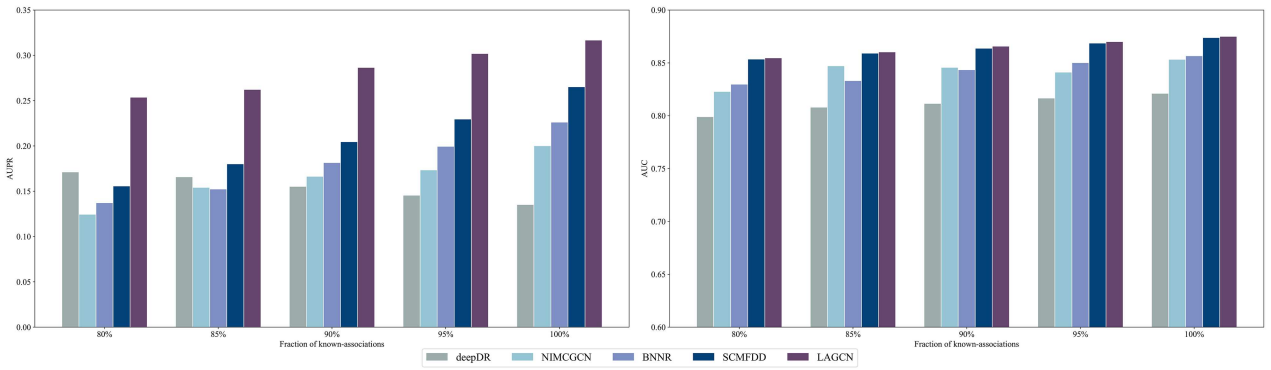

Figure 3. Performances of methods based on different fractions of known associations.

Table 6. Top 10 drug–disease associations predicted by LAGCN

Drug	Disease	Evidence
Cocaine	Migraine disorders	[57]
Tamoxifen	Depressive disorder	[53]
Clozapine	Hyperhomocysteinemia	NA
Cimetidine	Hypotension, orthostatic	NA
Phenytoin	Splenomegaly	[58]
Clozapine	Hyperhidrosis	[59]
Cocaine	Endomyocardial fibrosis	NA
Sirolimus	Sinoatrial block	NA
Dexamethasone	Multiple myeloma	[54]
Tamoxifen	Dementia	NA

Table 7. Top 10 associated diseases (associated drugs) for a given drug (disease) predicted by LAGCN

Drug	Disease	Evidence
Carbamazepine	Intraoperative complications	NA
	Movement disorders	[55]
	Bradycardia	[60]
	Tremor	[61]
	Carcinoma, squamous cell	NA
	Liver cirrhosis	NA
	Leukemia, myelogenous, chronic, BCR-ABL positive	NA
	Chest pain	[62]
	Cocaine-related disorders	[63]
	Lethargy	[64]
Disease	Drug	Evidence
Breast neoplasms	Doxorubicin	[65]
	Vincristine	[66]
	Dexamethasone	[67]
	Sirolimus	NA
	Tretinoin	[68]
	Indomethacin	[69]
	Sorafenib	NA
	Cytarabine	[70]
	Mitoxantrone	[71]
	Tamoxifen	[56]

model, the predicted associations require verification by public literature or other available sources. Top 10 drug–disease associations predicted by LAGCN are listed in Table 6, and we can find evidence to confirm five out of them. For example, tamoxifen is capable of prolonging the lives of premenopausal women with breast cancer and decreasing the probability of recurrence, but it led to acute depression symptoms in a 34-year-old breast cancer patient [53]. Dexamethasone is a type of corticosteroid medication. It was used in the treatment of many conditions, including rheumatic problems, several skin diseases, severe allergies, asthma, chronic obstructive lung disease, croup, brain swelling, ocular pain following ophthalmic surgery and antibiotics in tuberculosis. It was also used as a direct chemotherapeutic agent in the treatment of multiple myeloma, in which dexamethasone is given in combination with lenalidomide [54].

Furthermore, we check upon the top 10 candidate diseases for carbamazepine and the top 10 candidate drugs for breast neoplasms. Table 7 shows the results of our experiments, and some of the predictions can be confirmed. For example, carbamazepine is an anticonvulsant medication used primarily in the treatment of epilepsy and neuropathic pain, but De Sarro *et al.* [55] have proved that carbamazepine also leads to movement disorders by potentiating the anticonvulsant activity in the DBA/2 mice animal model. Breast cancer is the leading type of cancer in women, accounting for 25% of all cases according to Wikipedia, and countless researchers have been devoting themselves to finding treatment of it. According to Tsai *et al.* [56], tamoxifen and fulvestrant are widely used therapeutic agents and are considered to alter estrogen receptor (ER) signaling in ER-positive breast cancers.

Therefore, case studies demonstrate that LAGCN can help to identify novel associations as well as associated diseases (associated drugs) for a given drug (disease).

Conclusions

In this paper, we establish an LAGCN for identifying the latent drug–disease associations. In contrast to existing methods that utilize the bipartite graphs, LAGCN captures the topological information of a heterogeneous network constructed from drug–disease associations, drug–drug similarities and disease–disease similarities. LAGCN achieves good performances in predicting drug–disease associations by adaptively combining embeddings at different convolution layers with an attention mechanism,

and outperforms other drug–disease association prediction methods and the baseline method.

In the future, we will consider more biological entities involved in the drug–disease associations, such as genes, miRNAs and targets, and build a heterogeneous network with more types of entities and links for the drug–disease association prediction. Although GCN is a powerful method for analyzing the networks, it suffers from the problem of over-smoothing, and we will use data augmentation techniques to relieve over-smoothing in deep GCN.

Data availability

The datasets were derived from the following sources in the public domain: the drug–disease associations from http://ctdbase.org/, the drug features from https://www.drugbank.ca/ and the disease MeSH descriptors from https://meshb.nlm.nih.gov/. The implementation of LAGCN and the preprocessed data is available at https://github.com/storyandwine/LAGCN.

Key Points

- We propose a full end-to-end graph-based deep learning method, termed LAGCN, for predicting drug–disease associations.
- LAGCN utilizes a GCN to capture structural information from the heterogeneous network composed of associations and similarities.
- The attention mechanism is introduced to combine the embeddings at different convolution layers, which leads to more informative representations of drugs and diseases.

Funding

This work was supported by the National Natural Science Foundation of China [61772381, 62072206, 61572368]; National Key Research and Development Program [2018YFC0407904]; and Huazhong Agricultural University Scientific & Technological Self-innovation Foundation. The funders have no role in study design, data collection, data analysis, data interpretation or writing of the manuscript.

References

1. Chan HCS, Shan H, Dahoun T, *et al*. Advancing drug discovery via artificial intelligence. *Trends Pharmacol Sci* 2019; **40**:592–604.
2. Luo H, Li M, Yang M, *et al*. Biomedical data and computational models for drug repositioning: a comprehensive review. *Brief Bioinform* 2020. https://doi.org/10.1093/bib/bbz176.
3. Zhang Z-C, Zhang X-F, Wu M, *et al*. A graph regularized generalized matrix factorization model for predicting links in biomedical bipartite networks. *Bioinformatics* 2020; **36**:3474–81.
4. von Eichborn J, Murgueitio MS, Dunkel M, *et al*. PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Res* 2010;**39**:D1060–6.
5. Wiegers TC, Davis AP, Cohen KB, *et al*. Text mining and manual curation of chemical-gene-disease networks for the comparative toxicogenomics database (CTD). *BMC Bioinformatics* 2009;**10**:326.
6. Wang L, Wang Y, Hu Q, *et al*. Systematic analysis of new drug indications by drug-gene-disease coherent subnetworks. *CPT Pharmacometrics Syst Pharmacol* 2014;**3**:e146.
7. Zickenrott S, Angarica VE, Upadhyaya BB, *et al*. Prediction of disease-gene-drug relationships following a differential network analysis. *Cell Death Dis* 2016;**7**:e2040.
8. Wang W, Yang S, Zhang X, *et al*. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* 2014;**30**:2923–30.
9. Luo H, Wang J, Li M, *et al*. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics* 2016;**32**:2664–71.
10. Luo H, Wang J, Li M *et al*. Computational drug repositioning with random walk on a heterogeneous network. *IEEE/ACM Trans Comput Biol Bioinform* 2019;**6**(16):1890–1900.
11. 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**:496.
12. Yang L, Agarwal P. Systematic drug repositioning based on clinical side-effects. *PLoS One* 2011;**6**:e28025.
13. Wang K, Sun J, Zhou S, *et al*. Prediction of drug-target interactions for drug repositioning only based on genomic expression similarity. *PLoS Comput Biol* 2013;**9**:e1003315.
14. Oh M, Ahn J, Yoon Y. A network-based classification model for deriving novel drug-disease associations and assessing their molecular actions. *PLoS One* 2014;**9**:e111668.
15. Yang K, Zhao X, Waxman D, *et al*. Predicting drug-disease associations with heterogeneous network embedding. *Chaos* 2019;**29**:123109.
16. Liang X, Zhang P, Yan L, *et al*. LRSSL: predict and interpret drug–disease associations based on data integration using sparse subspace learning. *Bioinformatics* 2017;**33**: 1187–96.
17. Wu G, Liu J, Wang C. Semi-supervised graph cut algorithm for drug repositioning by integrating drug, disease and genomic associations. In: *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Los Alamitos, CA: IEEE Computer SOC, 2016.
18. Zhang W, Yue X, Chen Y *et al*. Predicting drug-disease associations based on the known association bipartite network. In: *IEEE International Conference on Bioinformatics and Biomedicine*. New York, NY: IEEE, 2017. pp. 503–9.
19. Lu L, Yu H. DR2DI: a powerful computational tool for predicting novel drug-disease associations. *J Comput Aided Mol Des* 2018;**32**:633–42.
20. Zhang W, Yue X, Lin W, *et al*. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinformatics* 2018;**19**:233.
21. Xuan P, Cao Y, Zhang T, *et al*. Drug repositioning through integration of prior knowledge and projections of drugs and diseases. *Bioinformatics* 2019;**35**:4108–19.
22. Zhang P, Wang F, Hu J. Towards drug repositioning: a unified computational framework for integrating multiple aspects of drug similarity and disease similarity. *AMIA Annu Symp Proc* 2014;**2014**:1258–67.
23. Yang M, Luo H, Li Y, *et al*. Overlap matrix completion for predicting drug-associated indications. *PLoS Comput Biol* 2019;**15**:e1007541.
24. Luo H, Li M, Wang S, *et al*. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics* 2018;**34**:1904–12.
25. Yang M, Luo H, Li Y, *et al*. Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* 2019; **35**:i455–63.

26. Zhang W, Xu H, Li X, *et al*. DRIMC: an improved drug repositioning approach using Bayesian inductive matrix completion. *Bioinformatics* 2020;**36**:2839–47.
27. Angermueller C, Pärnamaa T, Parts L, *et al*. Deep learning for computational biology. *Mol Syst Biol* 2016;**12**:878–8.
28. Yue X, Gutierrez BJ, Sun H. Clinical reading comprehension: a thorough analysis of the emrQA dataset. In: *ACL*. Stroudsburg, PA: Assoc Computational Linguistics-ACL, 2020.
29. Zeng X, Zhu S, Liu X, *et al*. deepDR: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics* 2019;**35**:5191–8.
30. Li Z, Huang Q, Chen X, *et al*. Identification of drug-disease associations using information of molecular structures and clinical symptoms via deep convolutional neural network. *Front Chem* 2020;**7**:924.
31. Xuan P, Ye Y, Zhang T, *et al*. Convolutional neural network and bidirectional long short-term memory-based method for predicting drug-disease associations. *Cells* 2019;**8**:705.
32. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: *International Conference on Learning Representations (ICLR)*, 2017. https://iclr.cc/archive/www/2017.html.
33. Y-A H, Hu P, KCC C, *et al*. Graph convolution for predicting associations between miRNA and drug resistance. *Bioinformatics* 2019;**36**:851–8.
34. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018;**34**:i457–66.
35. Li J, Zhang S, Liu T, *et al*. Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction. *Bioinformatics* 2020;**36**(8):2538–46.
36. Yang J-H, Chen C-M, Wang C-J *et al*. HOP-rec: high-order proximity for implicit recommendation. In: *Proceedings of the 12th ACM Conference on Recommender Systems*. New York, NY: Assoc Computing Machinery, 2018, pp. 140–4.
37. Vaswani A, Shazeer N, Parmar N *et al*. Attention is all you need. In: *Advances in Neural Information Processing Systems*. La Jolla, California: Neural Information Processing Systems (Nips), 2017, pp. 5998–6008.
38. Zhang W, Huang F, Yue X *et al*. Prediction of drug-disease associations and their effects by signed network-based non-negative matrix factorization. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. New York, NY: IEEE, 2018, pp. 798–802.
39. Davis AP, Grondin CJ, Johnson RJ, *et al*. The comparative toxicogenomics database: update 2017. *Nucleic Acids Res* 2017;**45**:D972–8.
40. Law V, Knox C, Djoumbou Y, *et al*. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 2014;**42**:D1091–7.
41. Deng Y, Xu X, Qiu Y, *et al*. A multimodal deep learning framework for predicting drug-drug interaction events. *Bioinformatics* 2020. https://doi.org/10.1093/bioinformatics/btaa501.
42. Wang D, Wang J, Lu M, *et al*. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* 2010;**26**:1644–50.
43. Clevert D-A, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). In: *International Conference on Learning Representations (ICLR)*, 2016.
44. Wang X, He X, Wang M *et al*. Neural graph collaborative filtering. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY: Assoc Computing Machinery, 2019, pp. 165–74.
45. He X, Deng K, Wang X *et al*. LightGCN: simplifying and powering graph convolution network for recommendation. In: *SIGIR*, 2020. https://sigir.org/sigir2020/accepted-papers/.
46. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–56. http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf.
47. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: *International Conference on Learning Representations (ICLR)*, 2015. https://iclr.cc/archive/www/2015.html.
48. van den Berg R, Kipf TN, Welling M. Graph convolutional matrix completion. In: *KDD*, 2018. https://www.kdd.org/kdd2018/deep-learning-day.
49. Srivastava N, Hinton G, Krizhevsky A, *et al*. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;**15**:1929–58.
50. Zhu L, Hong Z, Zheng H. Predicting gene-disease associations via graph embedding and graph convolutional networks. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019, pp. 382–9.
51. Smith LN. Cyclical learning rates for training neural networks. In: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. New York, NY: IEEE, 2017, pp. 464–72.
52. Li Q, Han Z, Wu X-M. Deeper insights into graph convolutional networks for semi-supervised learning. In: *Thirty-Second AAAI Conference on Artificial Intelligence*. Palo Alto, CA, 2018.
53. Bourque F, Karama S, Looper K, *et al*. Acute tamoxifen-induced depression and its prevention with venlafaxine. *Psychosomatics* 2009;**50**:162–5.
54. Jakobsen Falk I, Lund J, Gréen H, *et al*. Pharmacogenetic study of the impact of ABCB1 single-nucleotide polymorphisms on lenalidomide treatment outcomes in patients with multiple myeloma: results from a phase IV observational study and subsequent phase II clinical trial. *Cancer Chemother Pharmacol* 2018;**81**:183–93.
55. De Sarro G, Paola EDD, Gratteri S, *et al*. Fosinopril and zofenopril, two angiotensin-converting enzyme (ACE) inhibitors, potentiate the anticonvulsant activity of antiepileptic drugs against audiogenic seizures in DBA/2 mice. *Pharmacol Res* 2012;**65**:285–96.
56. Tsai C-F, Cheng Y-K, Lu D-Y, *et al*. Inhibition of estrogen receptor reduces connexin 43 expression in breast cancers. *Toxicol Appl Pharmacol* 2018;**338**:182–90.
57. Srikiatkhachorn A, Suwattanasophon C, Ruangpattanatawee U, *et al*. 2002 Wolff Award. 5-HT2A receptor activation and nitric oxide synthesis: a possible mechanism determining migraine attacks. *Headache* 2002;**42**:566–74.
58. Schlaifer D, Arlet P, De la Roque PM, *et al*. Antiepileptic drug-induced lymphoproliferative disorder associated with acquired C1 esterase inhibitor deficiency and angioedema. *Eur J Haematol* 1992;**48**:274–5.
59. Szymanski SJD, Leipzig R, Masiar S, *et al*. Anticholinergic delirium caused by retreatment with clozapine. *Am J Psychiatry* 1991;**148**:1752.
60. Kumada T, Hattori H, Doi H, *et al*. Postoperative complete atrioventricular block induced by carbamazepine in a patient with congenital heart disease. *No To Hattatsu* 2005;**37**:257–61.

61. Horvath J, Coeytaux A, Jallon P, *et al*. Carbamazepine encephalopathy masquerading as Creutzfeldt–Jakob disease. *Neurology* 2005;**65**:650.
62. Wittchen F, Spencker S, Zinke S, *et al*. Leistungsknick, Thoraxschmerz und Polyserositis bei einem 35-jährigen Patienten mit antikonvulsiver Therapie. *Internist* 2006;**47**: 69–75.
63. Brady KT, Sonne SC, Malcolm RJ, *et al*. Carbamazepine in the treatment of cocaine dependence: subtyping by affective disorder. *Exp Clin Psychopharmacol* 2002;**10**: 276.
64. Reynolds ER, Stauffer EA, Feeney L, *et al*. Treatment with the antiepileptic drugs phenytoin and gabapentin ameliorates seizure and paralysis of Drosophila bang-sensitive mutants. *J Neurobiol* 2004;**58**:503–13.
65. Wang X, Ji C, Zhang H, *et al*. Identification of a small-molecule compound that inhibits homodimerization of oncogenic NAC1 protein and sensitizes cancer cells to anticancer agents. *J Biol Chem* 2019;**294**:10006–17.
66. Cassidy J, Merrick MV, Smyth JF, *et al*. Cardiotoxicity of mitozantrone assessed by stress and resting nuclear ventriculography. *Eur J Cancer Clin Oncol* 1988;**24**:935–8.
67. Hofstra L, Van Der Graaf W, De Vries E, *et al*. Ataxia following docetaxel infusion. *Ann Oncol* 1997;**8**:812–3.
68. Alsafadi S, Even C, Falet C, *et al*. Retinoic acid receptor alpha amplifications and retinoic acid sensitivity in breast cancers. *Clin Breast Cancer* 2013;**13**:401–8.
69. Ackerstaff E, Gimi B, Artemov D, *et al*. Anti-inflammatory agent indomethacin reduces invasion and alters metabolism in a human breast cancer cell line. *Neoplasia* 2007;**9**:222–35.
70. Feldman LD, Buzdar AU, Blumenschein GR. High-dose 1-beta-D-arabinofuranosylcytosine in advanced breast cancer. *Oncology* 1985;**42**:273–4.
71. Di Costanzo F, Manzione L, Gasperoni S, *et al*. Paclitaxel and mitoxantrone in metastatic breast cancer: a phase II trial of the Italian Oncology Group for Cancer Research. *Cancer Invest* 2004;**22**:331–7.