OXFORD

Briefings in Bioinformatics, 2022, **23(3)**, 1–13
https://doi.org/10.1093/bib/bbac123
Problem Solving Protocol

Heterogeneous multi-scale neighbor topologies enhanced drug–disease association prediction

Ping Xuan, Xiangfeng Meng, Ling Gao, Tiangang Zhang and Toshiya Nakaguchi
Corresponding author: Tiangang Zhang. E-mail: zhang@hlju.edu.cn

Abstract

Motivation: Identifying new uses of approved drugs is an effective way to reduce the time and cost of drug development. Recent computational approaches for predicting drug–disease associations have integrated multi-sourced data on drugs and diseases. However, neighboring topologies of various scales in multiple heterogeneous drug–disease networks have yet to be exploited and fully integrated.
Results: We propose a novel method for drug–disease association prediction, called MGPred, used to encode and learn multi-scale neighboring topologies of drug and disease nodes and pairwise attributes from heterogeneous networks. First, we constructed three heterogeneous networks based on multiple kinds of drug similarities. Each network comprises drug and disease nodes and edges created based on node-wise similarities and associations that reflect specific topological structures. We also propose an embedding mechanism to formulate topologies that cover different ranges of neighbors. To encode the embeddings and derive multi-scale neighboring topology representations of drug and disease nodes, we propose a module based on graph convolutional autoencoders with shared parameters for each heterogeneous network. We also propose scale-level attention to obtain an adaptive fusion of informative topological representations at different scales. Finally, a learning module based on a convolutional neural network with various receptive fields is proposed to learn multi-view attribute representations of a pair of drug and disease nodes. Comprehensive experiment results demonstrate that MGPred outperforms other state-of-the-art methods in comparison to drug-related disease prediction, and the recall rates for the top-ranked candidates and case studies on five drugs further demonstrate the ability of MGPred to retrieve potential drug–disease associations.

Introduction

The development of a novel drug typically takes 10 to 15 years and 0.8 to 1.5 billion dollars to complete [1–5]. In addition, the likelihood of a new drug being successfully approved by a regulatory agency is usually less than 10% [6]. It is therefore important to discover new uses of approved drugs, namely drug repositioning [7, 8]. The tolerability and pharmacological characteristics of approved drugs have been clarified, and these drugs can be used more effectively and safely in clinics. More importantly, the time and economic cost of drug repositioning are much less than that of new drug R&D from scratch, which may shorten the drug R&D cycle to 6.5 years and reduce the expense to $300 million [9–11].

By predicting the probability of association between approved drugs and diseases, we screened reliable candidate diseases, which was conducive to further experimental verification. Previously, there were three classes of approaches used to study drug–disease associations. The first class only integrates similar and associative data between drugs and diseases for predicting the association. A support vector machine based classifier was employed to evaluate the associative probability by integrating data on drugs and diseases [12]. Methods of similarity constrained matrix factorization [13] and nonnegative matrix factorization [14] were utilized to integrate drug characteristics, the semantic similarities of diseases and known drug–disease associations. Gottlieb *et al.* [15]

Ping Xuan is a professor at the School of Computer Science and Technology, Heilongjiang University, Harbin, China. She is also a professor at the School of Computer Science, Shaanxi Normal University. Her current research interests include deep learning, complex network analysis and medical image analysis.
Xiangfeng Meng is studying for her master's degree in the School of Computer Science and Technology at Heilongjiang University, Harbin, China. Her research interests include complex network analysis and deep learning.
Ling Gao is studying for her master's degree in the School of Computer Science and Technology at Heilongjiang University, Harbin, China. Her research interests include complex network analysis and deep learning.
Tiangang Zhang is an associate professor of the School of Mathematical Science, Heilongjiang University, Harbin, China. His current research interests include complex network analysis and computational fluid dynamics.
Toshiya Nakaguchi is a professor at the Center for Frontier Medical Engineering, Chiba University, Chiba, Japan. His current research interests include medical image processing, machine learning, image-guided surgery and biomedical measurement.
Received: January 1, 2022. **Revised:** February 20, 2022. **Accepted:** March 15, 2022

presented a novel algorithm, called PREDICT, to infer novel drug indications. A few approaches have utilized random walk or bi-random walk on the established bi-layer network for predictive purposes [16–18]. However, the above methods only use the drug and disease similarities and associative information, ignoring the side effects and target information of the drug.

The second class of approaches took advantage of multi-source data for an associative prediction. A three-layer heterogeneous network model (TL-HGBI) [19] was proposed to conduct drug repositioning using the target information integrated by a heterogeneous network. Liang *et al.* [20] integrated information on multiple attributes of drugs and employed a sparse subspace learning approach to predict the associations. Non-negative matrix factorization [21], inference probability matrix factorization [22] and graph regularization matrix factorization [23] were used to discover novel effects for existing drugs. Several approaches had been employed to conduct associative predictions using a random walk on the constructed network [24–27]. Nevertheless, the above approaches are all based on shallow models, and it is difficult to capture the complex nonlinear relationship among multi-sourced data.

The third class of methods was based on a deep learning approach to deeply integrate various drug and disease data and improve the prediction performance. A few approaches have applied convolutional neural network (CNN) to drug repositioning [28–30]. A Gaussian kernel autoencoder had been used to screen the candidates [31]. A model based on the combination of a CNN and Bi-LSTM was established for associative predictions [32]. Several methods used GCN [33–35], GNN [36] and sparse autoencoder deep neural network [37] to learn the low-dimensional features of drugs and diseases to evaluate the association scores. However, these approaches neglect the multi-scale neighboring topologies of the nodes and the multi-view attribute information of multiple heterogeneous networks.

We presented a novel predictive model, MGPred, which integrated and learned pairwise attributes and multi-scale neighboring topologies of drugs and diseases nodes. The contribution points are summarized as follows.

- To associate various types of drug similarities, we establish three drug–disease bi-layer networks. Based on these three networks, the topological representations of drug and disease nodes are learned separately for extracting the specific topology of each network.
- We propose a learning module based on multiple graph convolutional autoencoders with parameter sharing to enhance the integration of various similarities and associations of drugs and diseases from multiple views. This module contains a new multi-scale embedding mechanism to generate a multi-scale neighboring topology embedding.
- Since neighboring topologies at different scales have different contributions to drug–disease association prediction, we design a scale-level attention mechanism to distinguish their contributions and obtain multi-scale neighboring topology encoding for each drug–disease network.
- A CNN encoding module with multiple receptive fields is proposed to encode the attribute representations of drug–disease node pairs. The attributes of node pairs in different ranges are adaptively integrated using learnable and adjustable sizes of the receptive field. The improvement in the MGPred performance is demonstrated through a comparison with six advanced predictive models, the recall rates in top-ranked candidates and case studies of five drugs.

Materials and Methods

We present MGPred for predicting candidate diseases associated with a given drug. Three drug–disease bi-layer networks are created by taking advantage of multi-sourced data on drugs and diseases (Figure 1). MGPred consists of two modules, as shown in Figure 2. In the first module, the multi-scale neighboring topological representations of the drug and disease nodes are learned using multiple graph convolutional autoencoders with parameter sharing. For this module, we also design a novel scale-level attention mechanism to integrate the multi-scale neighboring topological representations of drug and disease nodes. Afterward, a module based on CNN with different receptive fields is constructed which is used to encode the low-dimensional attribute representations of drug–disease node pairs. Finally, the multiscale neighboring topological representation and the attribute representation of the node pair are passed through the fully connected layer and combined to form the final drug–disease association score. We also demonstrate the schematic diagram and algorithm of MGPred in Supplementary Figure 1 and Figure 2, respectively.

Dataset

From previous research, information related to drug–disease association, the chemical substructures of drugs, the structural domains and gene ontology annotations of drug target proteins and the semantic similarities of diseases were acquired [20]. At the same time, we acquired 3051 known associations between 763 drugs and 608 diseases from the Unified Medical Language System(UMLS) [38, 39]. The chemical fingerprints were acquired from PubChem, and this information was employed to build drug-related chemical substructure profiles [40]. PubChem is a database of the National Institutes of Health (NIH). The domain and gene ontology annotation information of drug target proteins were acquired from the InterPro and UniProt databases,

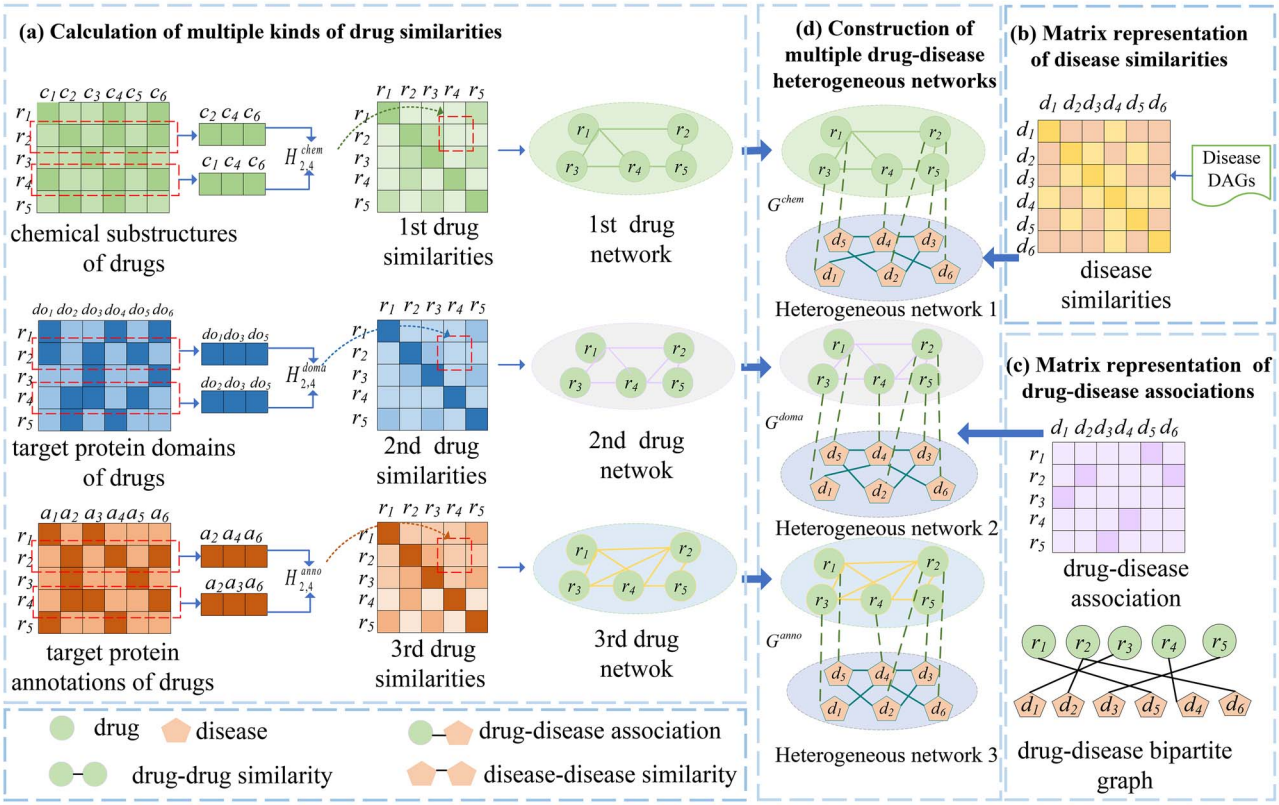

Figure 1. Construction of three drug–disease heterogeneous networks according to multiple kinds of drug similarities.

respectively [41, 42]. The semantic similarity of diseases was computed by constructing directed acyclic graphs (DAGs) of diseases, which was proposed by Wang *et al.* [43]. The disease terms for constructing the DAGs were taken from the U.S. National Library of Medicine(ULM).

Multi-source data representation

Representation of drug similarities

Based on multiple types of drug-related data, we acquire three kinds of drug similarity matrices (Figure 1A). This is expressed as follows:

$$H_r^{\upsilon} = \left\{ H_r^{chem}, \quad H_r^{doma}, \quad H_r^{anno} \right\}, \tag{1}$$

where $\upsilon = (chem, doma, anno)$, the chemical substructure similarity matrix of the drug, H_r^{chem}, is obtained based on the biological premise that the more chemical substructures that r_i and r_j have in common, the more similar their functions will be. H_r^{chem} is calculated by utilizing the chemical substructure information and cosine similarity measure. In addition, $H_r^{chem} \in R^{N_r \times N_r}$, where N_r denotes the number of drugs, i.e. 763. When r_i and r_j have more common target protein domains or more gene ontology annotations, the similarity between r_i and r_j will increase [44, 45]. Based on these premises, the domain similarity of the drug target proteins $H_r^{doma} \in R^{N_r \times N_r}$ and the functional annotation similarity of the targets $H_r^{anno} \in R^{N_r \times N_r}$ are also calculated using the cosine similarity measure. The values of H_r^{chem}, H_r^{doma} and H_r^{anno} are within [0, 1]. The larger the value is, the more similar r_i and r_j.

Representation of disease similarities

Diseases are usually expressed by DAGs containing their related terms. Utilizing their DAGs, the semantic similarity between two diseases is computed, the method for which is presented by Wang *et al* [46]. The similarity between N_d diseases from the matrix $H^d \in R^{N_d \times N_d}$ (Figure 1B) and N_d indicates the number of diseases, i.e. 681. $H_{i,j}^d \in [0, 1]$, and the higher the value, the higher the semantic similarity between d_i and d_j.

Representation of drug–disease associations

$B^{rd} \in R^{N_r \times N_d}$ is the drug–disease association matrix, which records the relationship between N_r drugs and N_d diseases. If r_i is associated with d_j, then $B_{i,j}^{rd}$ is 1; otherwise, it is 0.

Multiple drug–disease heterogeneous networks

Facing three different drug similarities, three drug–disease heterogeneous networks $G^{\upsilon}(\upsilon = chem, doma, anno)$ are constructed (Figure 1D). Each heterogeneous network contains two types of nodes $F = \{F^R \cup F^D\}$ and three types of edges $E = \{E_{\upsilon}^{RR} \cup E^{DD} \cup E^{RD}\}$. E_{υ}^{RR} is the edge of the drug–drug similarity, which is built based on the υth drug similarity. In addition, E^{DD} represents the edge between disease and disease, which is constructed using the semantic similarity of the disease. Taking advantage of the known association data, the drug–disease side is established, which is represented using E^{RD}. If a connection is observed between nodes $f_i \in F$ and $f_j \in F$, then $e_{ij} \in E$.

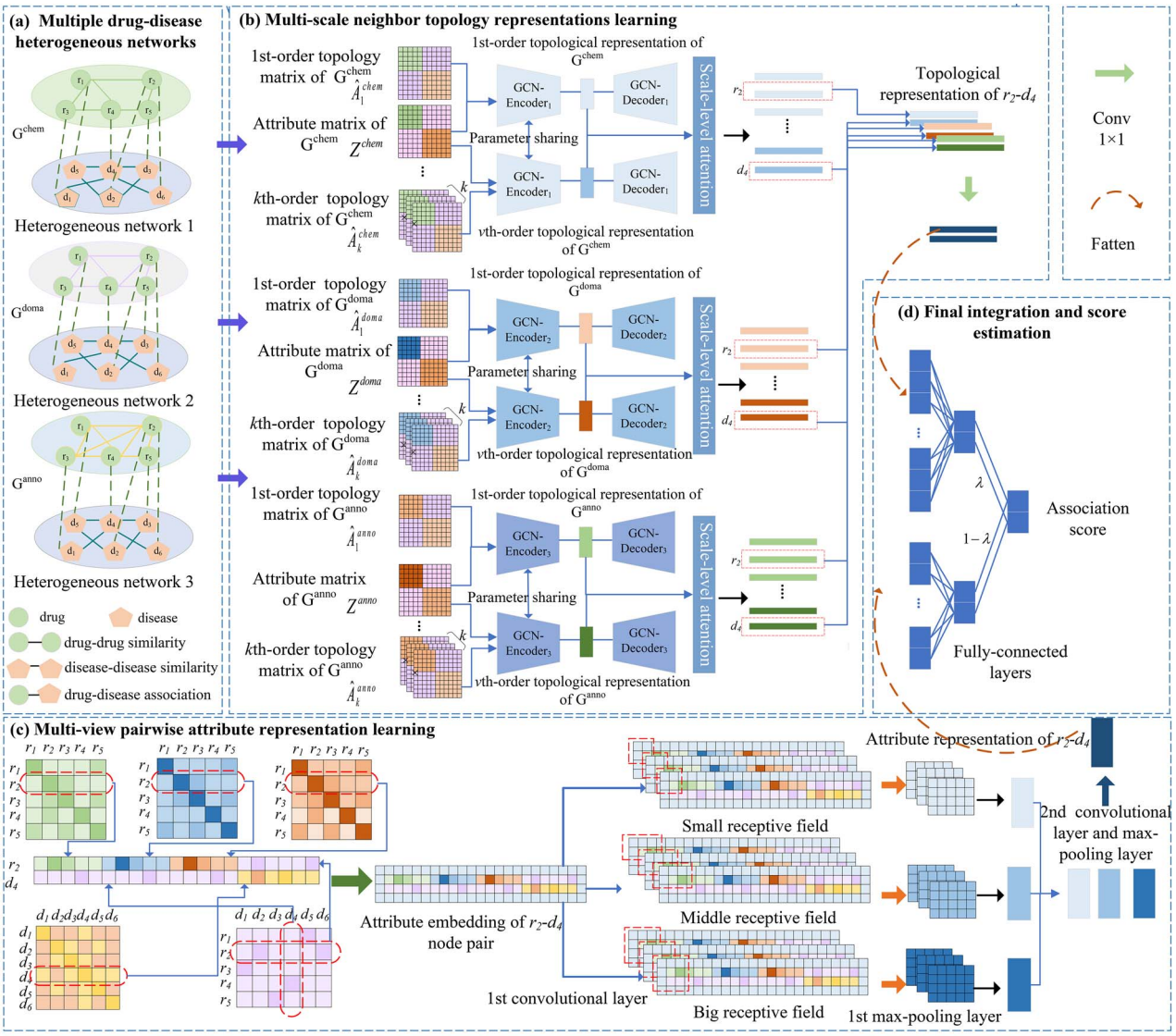

Figure 2. Framework of MGPred model. **(A)** construct three drug–disease heterogeneous networks based on multiple kinds of drug similarities **(B)** learn the multi-scale neighbor topologies of all drug and disease nodes **(C)** learn attribute representations of r_i-d_j node pair **(D)** deeply integrate multi-scale topological representations, attribute representations and estimate association score.

Adjacency matrices of multiple drug–disease bi-layer networks

Suppose that the adjacency matrix of the υth bi-layer network is expressed as

$$A^{\upsilon} = \begin{bmatrix} H_r^{\upsilon} & B^{rd} \\ B^{rd^T} & H^d \end{bmatrix}, \tag{2}$$

where $A^{\upsilon} \in R^{N_q \times N_q}(N_q = N_r + N_d)$ and B^{rd^T} is the transposition of B^{rd}. In addition, ($\upsilon = chem, doma, anno$) represented the drug similarity calculated based on different angles.

Multi-scale neighbor topologies and attributes extraction based on embedding strategy

Extraction of multi-scale neighboring topological embeddings

The direct neighbor node for each $r_i(d_j)$ node in the υth bi-layer network is its first-order neighboring topology, and its k-order neighboring topology is formed after k hops from that node. The first-order neighboring topology and k-order neighboring topology reflect the neighboring topology of the node at different scales, and neighboring topologies at different scales are extremely important for predicting drug–disease associations. Therefore, to study the low-dimensional topological representations of all nodes at different scales, we establish a framework based on multiple graph convolutional autoencoders with shared parameters. The first-order neighboring topology embedding of the υth bi-layer network is $\hat{A}_1^{\upsilon}$ (Figure 3),

$$\hat{A}_1^{\upsilon} = (D^{\upsilon})^{-\frac{1}{2}} A^{\upsilon} (D^{\upsilon})^{-\frac{1}{2}}, \tag{3}$$

where $(D^{\upsilon})_{ii} = \sum_j (A^{\upsilon})_{ij} \in R^{N_q \times N_q}$ and D^{υ} is the degree matrix of A^{υ} and is also a diagonal matrix. In addition,

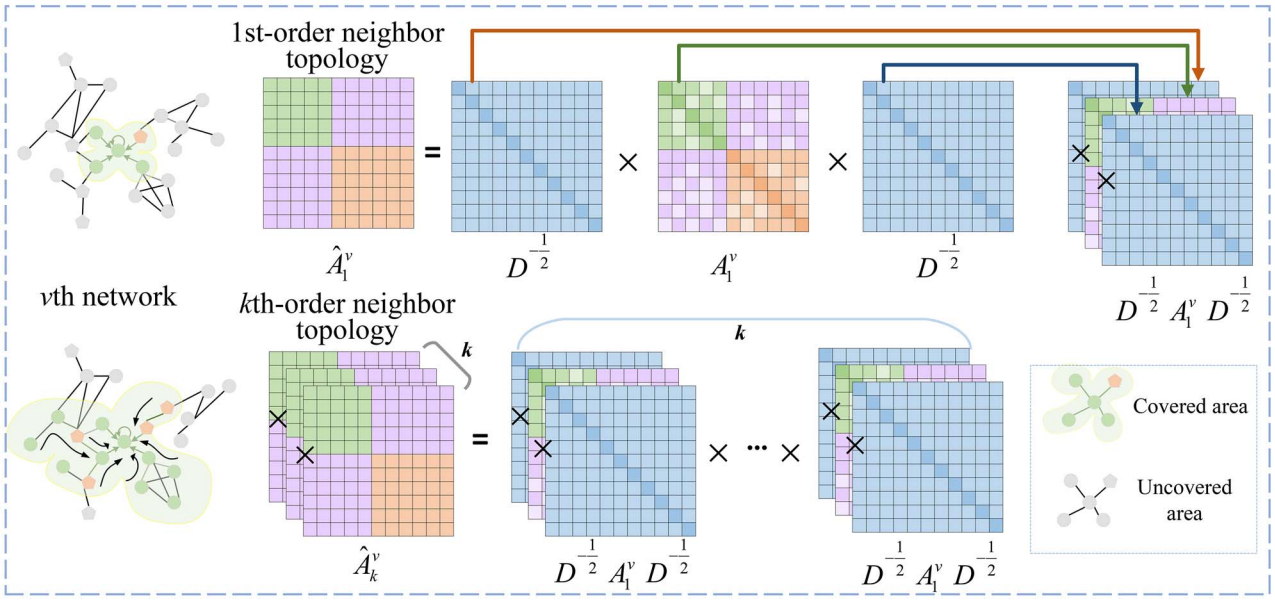

Figure 3. Calculation of kth-order neighbor topology of the vth network

$A^v = A^v + I$ is considered first-order neighboring topological embedding after joining the self-connections, and I denote the identity matrix. The kth-order neighboring topological embedding of the vth network is generated as follows:

$$\hat{A}_k^v = (D^v)^{-\frac{1}{2}} A^v [(D^v)^{-1} A^v]^{k-1} (D^v)^{-\frac{1}{2}}. \quad (4)$$

Similarly, we added a self-connection to A^v and $\hat{A}_k^v$ incorporates the topology information of k-hop neighboring nodes.

Extraction of node attribute embeddings

A_i^v records the similarity between r_i and N_r drugs, and the association between N_d diseases, which can be treated as the attribute vector of r_i. Likewise, A_{N_r+j} can be viewed as the attribute vector of d_j. Thus, A^v can be regarded as the attribute matrix of the vth network. We name the attribute matrix of the vth network as Z^v and $Z^v = A^v$. We build different graph convolutional autoencoders with shared parameters for the three networks G^{chem}, G^{doma} and G^{anno}, respectively.

Multi-scale neighbor topological encoding and decoding by graph convolutional autoencoder

Each network has neighboring topological embeddings of different scales $\hat{A}_1^v, \hat{A}_2^v \ldots \hat{A}_k^v \ldots \hat{A}_{N_c}^v$ and the corresponding attribute matrix Z^v. N_c is the number of all scales. We set up special graph convolution autoencoders with shared parameters to encode and learn the low-dimensional topological representations of the drugs (diseases) at the kth scale. Because the encoding and decoding processes of the three networks are similar, we only take the vth network as an example to introduce the encoding and decoding processes.

Encoding of the *k*th-scale neighboring topology. The encoder is composed of two layers of graph convolution. Given kth-scale neighboring topological embedding $\hat{A}_k^v$ and attribute matrix Z^v, the output feature maps of the first and ℓ-*th* layers of the graph convolutional encoder are denoted as

$$Z_{en}^{v(1)} = \phi(\hat{A}_k^v Z^v W_{en}^1) \quad (5)$$

$$Z_{en}^{v(\ell)} = \phi(\hat{A}_k^v Z_{en}^{v(\ell-1)} W_{en}^{\ell})$$
$$\ell = 2, \ldots\ldots, L_{EN}, \quad (6)$$

where L_{EN} denotes the total number of encoding layers and ϕ is a nonlinear activation function ReLU. We employ $W_{en}^{(1)}$ and $W_{en}^{(\ell)}$ to denote the weight matrices of the first and ℓ-*th* layers of the graph convolution encoder, respectively. In addition, $Z_{en}^{v(1)}$ and $Z_{en}^{v(\ell)}$ are the corresponding encoding outputs of the first layer and the ℓ-*th* layer. Each line denotes the feature vector of $r_i(d_j)$ node after a dimensionality reduction using the graph convolution encoder. The output of the last layer of the encoder is denoted as $Z_{en}^{v(L_{EN})}$, which contains the low-dimensional topological representations of all drug and disease nodes. Let us rename it as Z_k^v.

Decoding of the *k*th-scale neighboring topology. We utilize the decoder to upgrade the feature of the encoder back to the original dimension, and then compute the loss between the reconstructed matrices $\hat{Z}_k^v$ and the original matrices Z^v to obtain a better encoding for drug–disease association prediction. The decoder consists of two layers of graph convolution. Z_k^v and $\hat{A}_k^v$ are utilized as inputs to the graph convolutional decoder. The decoding $Z_{de}^{v(1)}$ and $Z_{de}^{v(\ell)}$ of the first and ℓ-*th* layers are obtained as

follows:

$$Z_{de}^{\upsilon(1)} = \phi(\hat{A}_k^{\upsilon} Z_k^{\upsilon} W_{de}^{(1)}) \tag{7}$$

$$Z_{de}^{\upsilon(\ell)} = \phi(\hat{A}_k^{\upsilon} Z_{de}^{\upsilon(\ell-1)} W_{de}^{(\ell)})$$
$$\ell = 2,\ldots\ldots,L_{DE}, \tag{8}$$

where L_{DE} is the total number of decoded layers. The weight matrices of the first and ℓ-*th* layers of the decoding layer are denoted by $W_{de}^{(1)}$ and $W_{de}^{(\ell)}$, respectively. $Z_{de}^{\upsilon(1)}$ and $Z_{de}^{\upsilon(\ell)}$ represent the outputs of the corresponding decoding layers. The activation function of the first layer of the graph convolution is ReLU, and the activation function of the second layer is Sigmoid. The output $Z_{de}^{\upsilon(L_{DE})}$ of the last decoding layer is acquired and renamed as $\hat{Z}_k^{\upsilon}$.

Optimization. We have three drug–disease networks, i.e. G^{chem}, G^{doma} and G^{anno}, and each network has N_c scale neighboring topologies. Because the process of calculating the loss of the three networks is similar, we calculate the loss of G^{chem} as an example to describe this process. The loss of G^{chem}, $Loss_{chem}$, is

$$Loss_{chem} = \sum_{k=1}^{N_C} \left\| Z^{chem} - \hat{Z}_k^{chem} \right\|_F^2, \tag{9}$$

where N_c denotes the total number of scales and F represents the Frobenius norm. Similarly, the losses of G^{doma} and G^{anno} are as follows:

$$Loss_{doma} = \sum_{k=1}^{N_C} \left\| Z^{doma} - \hat{Z}_k^{doma} \right\|_F^2, \tag{10}$$

$$Loss_{anno} = \sum_{k=1}^{N_C} \left\| Z^{anno} - \hat{Z}_k^{anno} \right\|_F^2. \tag{11}$$

The final loss of the graph convolutional encoding module is

$$Loss_c = Loss_{chem} + Loss_{doma} + Loss_{anno}. \tag{12}$$

Adam algorithm was utilized to optimize the loss function, and the graph convolution autoencoder was trained using the backpropagation algorithm.

Scale-level attention mechanism

We have learned the low-dimensional topological representations of $r_i(d_j)$ containing k-scale neighboring topologies in different heterogeneous networks and denote it as Z_k^{υ}. Let $Z_k^{\upsilon}(r_i)$ and $Z_k^{\upsilon}d_j)$ be the low-dimensional topological representations of r_i and d_j after a dimensionality reduction through graph convolution. Because the encoding of $r_i(d_j)$ of different scales has different contributions to the final prediction, a new scale-level attention mechanism is presented to obtain the topological representations of attention-enhanced $r_i(d_j)$. Because the scale attention mechanism of r_i and d_j is similar, we take r_i as an example to describe the mechanism. In the υth network, the attention score of the kth-scale level of r_i is s_k^{υ},

$$s_k^{\upsilon} = t^{sca} \tanh(W^{sca} Z_k^{\upsilon}(r_i) + b^{sca}), \tag{13}$$

where $k = 1, 2 \ldots N_c$ represents the size of the scale. In addition, W^{sca} is the weight matrix. The weight and bias vectors are represented by t^{sca} and b^{sca}, respectively. The normalized attention weight is $\beta_{r_i,k}^{\upsilon}$:

$$\beta_{r_i,k}^{\upsilon} = \frac{\exp(s_k^{\upsilon})}{\sum_{j \in N_C} \exp(s_j^{\upsilon})}, \tag{14}$$

where exp is an exponential function and the value of $\beta_{r_i,k}^{\upsilon}$ is within the range of [0,1]. The topology of r_i enhanced through the scale attention is expressed as $\tilde{Z}_k^{\upsilon}(r_i)$,

$$\tilde{Z}_k^{\upsilon}(r_i) = \sum\nolimits_{k \in N_C} (\beta_{r_i,k}^{\upsilon} Z_k^{\upsilon}(r_i) + Z_k^{\upsilon}(r_i)), \tag{15}$$

where $\tilde{Z}_k^{\upsilon}(r_i)$ represents the enhanced topological representation of r_i based on the kth-scale neighboring topology learning in the υth network. Based on the kth-scale neighboring topologies $\hat{A}_k^{\upsilon}$ and attribute matrices Z^{υ} of three different networks, three enhanced r_i topological representations, i.e. $\tilde{Z}_k^{chem}(r_i)$, $\tilde{Z}_k^{doma}(r_i)$ and $\tilde{Z}_k^{anno}(r_i)$ are learned, respectively. To take full advantage of $\tilde{Z}_k^{chem}(r_i)$, $\tilde{Z}_k^{doma}(r_i)$ and $\tilde{Z}_k^{anno}(r_i)$, we stack the attention-enhanced representations of r_i up and down and then use 1×1 convolution to fuse them. The final topological representation of r_i, $\tilde{Z}_{r_i,k}^{\upsilon}$, is defined as follows:

$$\tilde{Z}_{r_i,k}^{\upsilon} = conv(\tilde{Z}_k^{chem}(r_i) || \tilde{Z}_k^{doma}(r_i) || \tilde{Z}_k^{anno}(r_i)), \tag{16}$$

where *conv* represents 1×1 convolution operation and $||$ represents the up and down stacking operation. The final topological representation Z^{topo} is acquired.

Drug–disease pairwise attribute extraction based on embedding strategy

If r_i and d_j are similar or related to more common drugs and diseases, they are more likely to be related. Based on this biological premise, an embedding strategy was presented to integrate the similarity and association of r_i-d_j and acquire the attribute embedding matrix of the node pairs.

Given three kinds of drug similarity matrices H_r^{chem}, H_r^{doma} and H_r^{anno}; disease similarity matrix H^d; and drug–disease association matrix B^{rd}, the attribute embedding matrix T of $r_i - d_j$ was constructed (Figure 2C). First, facing drug similarity at the chemical substructure level, to embed the connections from the common drug, we stack the i-th row of H_r^{chem}, $(H_r^{chem})_{i,*}$, up and down with

the transpose of the j-th column of B^{rd}, $B^{r\,d\,T}_{*,j}$, which is expressed as follows:

$$t_1 = \left[(H_r^{chem})_{i,*}; B^{rd\,T}_{*,j}\right], \tag{17}$$

where $(H_r^{chem})_{i,*}$ contains similarities at the chemical substructure level between r_i and N_r drugs, and ; indicates the up-down stack operation. In addition, $B^{rdT}_{*,j}$ records the association between d_j and all drugs. Similarly, embeddings based on the domain level and annotation level of the drug target protein are obtained as

$$t_2 = \left[(H_r^{doma})_{i,*}; B^{rd\,T}_{*,j}\right]; \tag{18}$$

$$t_3 = \left[(H_r^{anno})_{i,*}; B^{rd\,T}_{*,j}\right]. \tag{19}$$

Among them, $(H_r^{doma})_{i,*}$ and $(H_r^{anno})_{i,*}$ cover the similarity between r_i and all drugs at the level of the domain and the annotation, respectively. Finally, the association of common diseases is captured by splicing up and down the i-th row of B^{rd}, $B^{rd}_{i,*}$ and the j-th row of H^d, $H^d_{*,j}$, as follows:

$$t_4 = \left[B^{rd}_{i,*}; H^d_{j,*}\right], \tag{20}$$

where $H^d_{*,j}$ covers the semantic similarity between d_j and all diseases. Finally, the attribute embedding matrix $T \in R^{2\times Ng}$ of r_i-d_j is constructed, and $N_g = 3N_r + N_d$.

$$\begin{aligned} T &= [t_1 t_2 t_3 t_4] \\ &= \begin{bmatrix} (H_r^{chem})_{i,*} & (H_r^{doma})_{i,*} & (H_r^{anno})_{i,*} & B^{rd}_{i,*} \\ B^{rd\,T}_{*,j} & B^{rd\,T}_{*,j} & B^{rd\,T}_{*,j} & H^d_{j,*} \end{bmatrix} \end{aligned} \tag{21}$$

Paired attribute encoding based on CNN

Given the attribute embedding matrix of r_i-d_j, CNN with multiple receptive fields are applied to learn the attribute representation of r_i-d_j. To capture the marginal information of the attribute embedding matrix T, we conduct a zero-padded operation on T. Set $F^{\varphi} \in R^{n^{\varphi}_w \times n^{\varphi}_h \times n_{conv}}$ as the filter, where $\varphi = (1, 2, 3)$ denotes filters of different scales, and n^{φ}_h and n^{φ}_w are the length and height of the filter, respectively. In addition, n_{conv} represents the number of filters used. When the m-th filter F^{φ}_m slides to the $T_{i,j}$ position,

$$T^{\varphi}_{m,i,j} = T(i : i + n^{\varphi}_w, j : j + n^{\varphi}_h), \tag{22}$$

where $i \in [1, 4-n^{\varphi}_w + 1]$, $j \in \left[1, 2 + n_g - n^{\varphi}_h + 1\right]$, $m \in [1, n_{conv}]$ and $n_g = 3N_r + N_d$. The filter $F^{\varphi}_{m,i,j}$ of the φth scale slides on $T^{\varphi}_{i,j}$ to obtain the element value $C^{\varphi}_m(i,j)$ at the m-th feature map C^{φ}_m as

$$C^{\varphi}_m(i,j) = \sigma(F^{\varphi}_{m,i,j} * T^{\varphi}_{m,i,j} + b(m)), \tag{23}$$

where b is the bias term and $C^{\varphi}_m(i,j)$ is the element value of the i-th row and j-th column in the m-th feature map. Moreover, σ is ReLU activation function. Max-pooling can be used to select the most important attributes in C^{φ}_m. We utilize n_a and n_b to denote the length and height of the filter in the pooling layer, respectively. The i-th row and j-th column in the m-th feature map are represented by $P_{m,i,j}$, which is calculated as follows:

$$P_{m,i,j} = \mathbf{max}(C^{\varphi}_m(i : i + n_b, j : j + n_a)). \tag{24}$$

Given that the filter sizes of the three scales are $n^1_w \times n^1_h = 2 \times 2$, $n^2_w \times n^2_h = 2 \times 3$ and $n^3_w \times n^3_h = 3 \times 3$, and that the corresponding outputs are $T_{(1)}$,$T_{(2)}$ and $T_{(3)}$, we further use max-pooling and acquire $\hat{T}_{(1)}$, $\hat{T}_{(2)}$ and $\hat{T}_{(3)}$. We then splice them at the left and right to form T_L. T_L is sent to the second convolutional layer and the pooling layer for fusion, and finally acquire deeply integrated attribute representation $\hat{T}$ of r_i-d_j.

Final integration and optimization

We flatten the given topological representation Z^{topo} into vector $\hat{z}^{topo}$ and send it to the fully connected layer and the softmax layer to learn the association score of r_i-d_j. The association score, q^{topo}, is defined as follows:

$$q^{topo} = soft\,\mathbf{max}(W^{topo}\hat{z}^{topo} + b^{topo}), \tag{25}$$

where W^{topo} is the weight matrix and b^{topo} is the bias vector. In addition, $q^{topo} = \left[q^{topo}_0, q^{topo}_1\right]$, where q^{topo}_1 and q^{topo}_0 record the probability that r_i-d_j is related and unrelated, respectively. The cross-entropy loss function between the drug and disease associative prediction score q^{topo} and the true label is defined as

$$Loss_{gcn} = -\sum_{i=1}^{N_t} [y_l \times \mathbf{log}\,(q^{topo})_1 + (1 - y_l) \times \mathbf{log}(1 - (q^{topo})_0)], \tag{26}$$

where N_t represents a set of training samples and y_l represents the true label. It is known that $y_l = 1$ when the r_i is associated with the d_j, and $y_l = 0$ when there is no such association.

Given that the attribute of r_i-d_j represents $\hat{T}$, it is flattened into a vector $\hat{t}$. Then $\hat{t}$ is fed into the full connection and softmax layers. The predicted correlation score q^{attr} and the cross-entropy loss of another module are

calculated as follows:

$$q^{attr} = soft\,\mathbf{max}(W^{attr}\hat{t} + b^{attr}); \tag{27}$$

$$\mathrm{Loss}_{cnn} = -\sum_{i=1}^{N_t}[y_l \times \log{(q^{attr})_1} + (1 - y_l) \times \log(1 - (q^{attr})_0)], \tag{28}$$

where $(q^{attr})_1$ and $(q^{attr})_0$ represent the r_i-d_j related probability and the non-related probability, respectively. We use the Adam algorithm and backpropagation algorithm to optimize and update Loss_{gcn} and Loss_{cnn}. Finally, q^{topo} and q^{attr} are weighted and integrated to obtain the final predicted score Y:

$$Y = \lambda \times q^{topo} + (1 - \lambda)q^{attr}, \tag{29}$$

where $\lambda \in [0, 1]$, and λ is the regulation parameter balancing the contributions of q^{topo} and q^{attr}.

Experiment results and discussions

Experiment setup and evaluation metrics

For the multi-scale neighboring topology encoding module, the layers of the graph convolution encoding and decoding and the scale of the neighboring topology are both set to 2. The output feature dimensions of the two graph convolutional encoding layers are 1000 and 700, and the output dimensions of the decoding layers are 1000 and 1444, respectively. The filter sizes of the first convolution layer of the attribute encoding module are 2×2, 2×3 and 3×3, respectively. The sizes of the second convolution layer and pooling layer are both 2×2. In the first and second convolutional layers, we use 16 and 32 filters, respectively.

MGPred and other approaches used to infer drug-related diseases are evaluated using 5-fold cross-validation. Known drug–disease associations and undiscovered associations are treated as positive and negative samples, respectively. First, we randomly divide positive examples into five parts. In each fold cross-validation, four positive samples and the same number of negative samples are randomly selected as the training set. The remaining positive and negative samples are used as the test sets. The area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC) and the top k recall rates are used as evaluation indicators. Because biologists are more willing to select the top-ranked candidates for further experimental verification, we also calculated the recall rate of top $k \in [30, 240]$. The higher the value, the more correct samples are identified.

Ablation experiment

We performed the ablation experiments to demonstrate the contribution of three kinds of drug similarities for the association prediction. These similarities include the drug chemical substructure similarity (DCS), the drug target domain similarity (DDS) and the drug target annotation similarity (DAS). As shown in Table 1, the whole model of MGPred got the best prediction performance when it utilized all these drug similarities. When MGPred was trained without DCS, its AUROC and AUPRC are reduced by 3.3% and 6.2%, respectively. When there is no DDS for training, AUROC and AUPRC decreased by 1.1% and 3.2%, respectively. When DAS was removed from the training process, AUROC and AUPRC dropped down by 2.2% and 4.1%, respectively. The experimental results indicated DCS has the most contribution to the increased prediction performance. It might be because there is more complete information of chemical structures of drugs, while there are less known targets for each drug.

Table 1. Results of the ablation studies.

DCS	DDS	DAS	Average AUROC	Average AUPRC
×	✓	✓	0.951	0.372
✓	×	✓	0.973	0.402
✓	✓	×	0.962	0.393
✓	✓	✓	**0.984**	**0.434**

Comparison with other methods

To evaluate the performance of MGPred, we compare it with six advanced approaches: GFPred [34], CBPred [32], SCMFDD [13], LRSSL [20], MBiRW [18] and HGBI [19]. To make the comparison results more convincing, MGPred and all comparison methods use the same dataset for the training and testing, and each comparison method uses the optimal parameter settings provided in the corresponding literature. We also briefly describe these six methods as follows.

HGBI[19]: HGBI constructed a triple-layer network composed of the drug, disease and target nodes, and the similarity or interaction connections among them. It evaluated the association score between a pair of drug and disease nodes based on information flowing in the network.

MBiRW[18]: MBiRW built a drug–disease heterogeneous network by utilizing the drug similarities, the disease similarities and the associations between drugs and diseases. It inferred the potential drug–disease associations by bi-directional random walk on the network.

LRSSL[20]: LRSSL integrated the chemical substructures of drugs, the domains of drug targets and the functional annotations of targets. It presented a method based on sparse subspace learning to predict the new indications for drugs.

SCMFDD[13]: SCMFDD proposed a matrix factorization approach with similarity constraint for improving the performance of predicting candidate drug–disease associations.

CBPred[32]: CBPred focused on the drug similarities based on their chemical structures, the disease

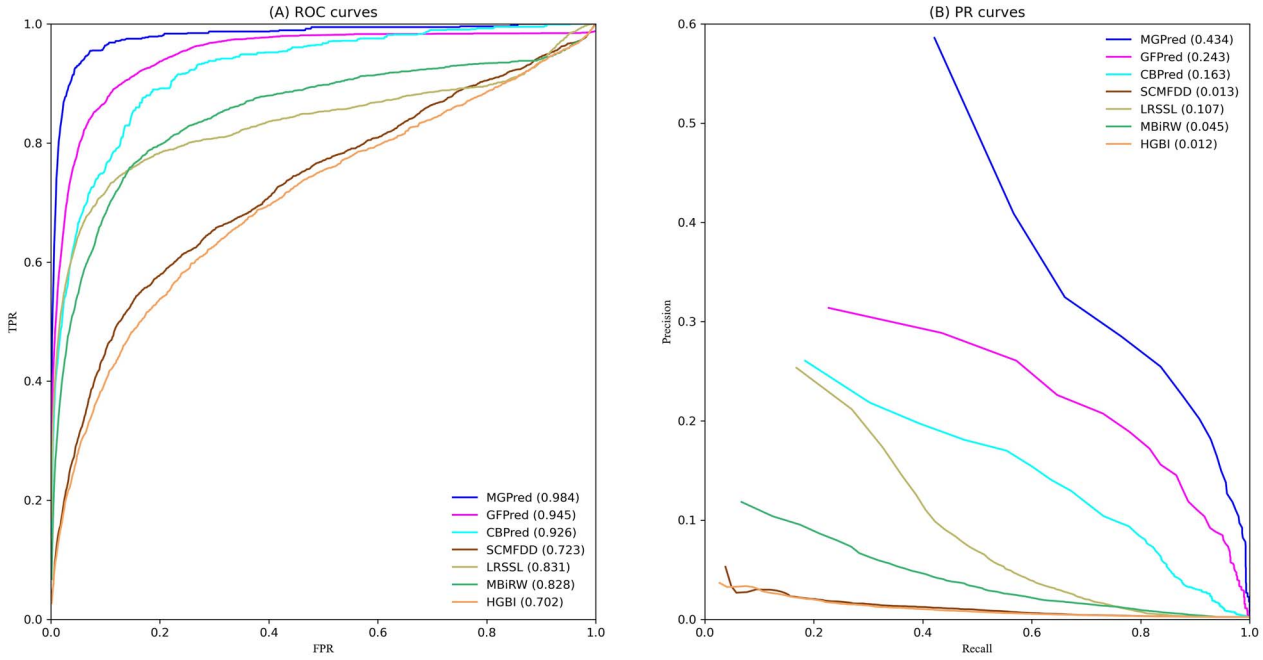

Figure 4. ROC and PR curves of different drug–disease association prediction methods.

Table 2. The statistical results of the paired Wilcoxon test on the AUROCs and AUPRCs over all the drugs by comparing MGPred and all other six methods.

	GFPred	CBPred	SCMFDD	LRSSL	MBiRW	HGBI
P-value of AUROC	6.6204e-05	2.8986e-13	1.7375e-53	4.9072e-30	1.4141e-38	2.2557e-58
P-value of AUPRC	2.3860e-38	3.6333e-42	7.8336e-76	8.4827e-56	4.7779e-63	2.0043e-82

similarity and the observed drug–disease associations. It proposed a method based on CNN and bi-directional long short-term memory network to estimate the drug–disease association scores.

GFPred[34]: GFPred utilized multiple kinds of drug similarities, and it constructed a prediction model based on graph convolutional auto-encoders.

We provide the average ROC and PR curves obtained for all drugs using various approaches (Figure 4). Our approach achieves the highest average AUROC (0.984), which is 3.9% higher than the second-ranked GFPred; 5.8% higher than CBpred, 15.3% and 15.6% better than LRSSL and MBiRW, respectively; 26.1% higher than SCMFDD; and 28.2% higher than the last ranked HGBI. As biologists are more willing to validate top-ranked candidate drugs, we show the partial ROC and partial AUC of FPR 0.1 (Supplementary Figure 3). The AUPRC value of MGPred is the best (0.434), and is 19.1%, 27.1%, 32.7%, 38.9%, 42.1% and 42.2% higher than those of GFPred, CBpred, LRSSL, MBiRW, SCMFDD and HGBI, respectively. The paired Wilcoxon test is also employed as an indicator to verify the performance of MGPred. It can be seen from Table 2 that MGPred is significantly better than the other comparison methods (p-value < 0.05).

From the results, GFPred ranked second, taking advantage of the similarity of multiple types of drugs and learning the topological representations of drug (disease) nodes in different networks. This shows that the use of information from multiple heterogeneous networks contributes to improved performance. CBpred ranked third, considering the path information of the node pairs. Although the AUROC value of LRSSL is slightly better than that of MBiRW, the AUPRC value is significantly lower than that of MBiRW. One possible reason may be that the former takes advantage of different drug similarities, whereas the latter considers only a type of similarity. SCMFDD and HGBI are based on non-negative matrix factorization and information flow, respectively. Their AUROC and AUPRC values are similar, but their performance was poor. It may be that they do not consider multiple kinds of drug similarities and various bi-layer networks. The improvement in the MGPred performance is mainly due to the use of multiple types of drug similarity used to establish multiple bi-layer networks. In addition, we deeply learn the multiscale neighboring topologies of various networks and the attributes of r_i-d_j node pairs.

From Figure 5, we can see the recall rate for under different top-k values. Because we have learned multi-scale neighboring topologies and node pair attributes, our model is significantly better than the other six models under different top-k values. When $k = 30$, MGPred exhibited the highest recall rate (83.6%). GFPred is second, at 75.2%. CBPred ranked third, at 68.8%. When the

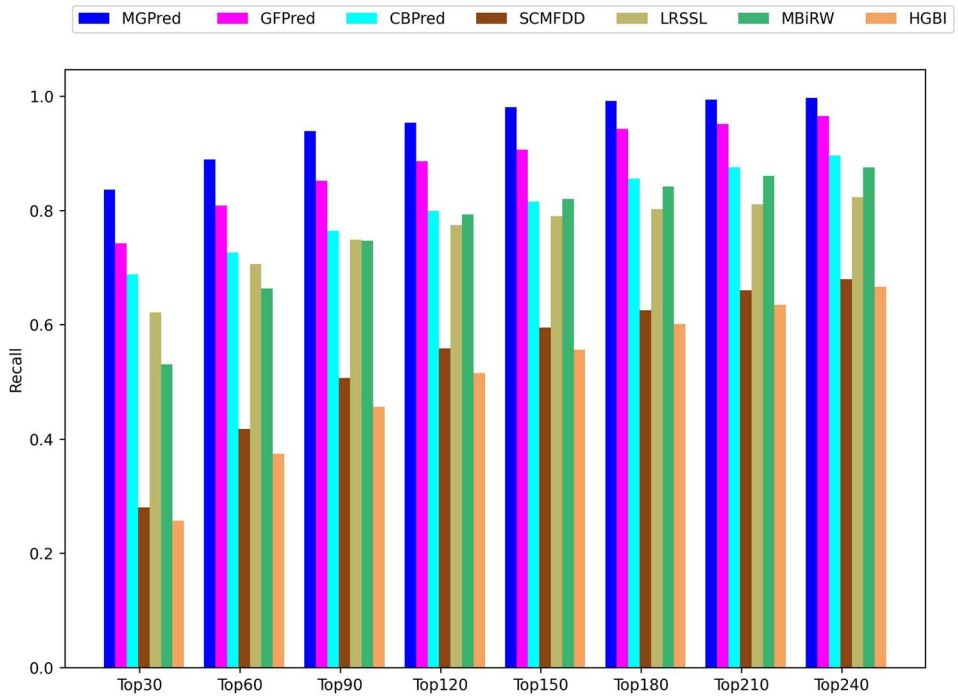

Figure 5. Recall rates of all the methods at different *k* cutoffs.

top-*k* values range from 60 to 120, our model still has the highest recall rate, which is 88.9%, 93.9% and 95.4%, respectively, and GFPred ranks second, with recall rates of 80.79%, 85.2% and 88.6%, respectively. CBPred was slightly higher than LRSSL, and the corresponding recall rates of CBPred were 72.6%, 76.4% and 79.9%, whereas the recall rates of LRSSL are 70.6%, 74.9% and 77.4%, respectively. Within the top-*k* values range of 30 to 120, the recall rates of MBiRW are 62.1%, 70.6%, 74.89% and 77.4%, and the recall rates of SCMFDD are 27.9%, 41.7%, 50.7% and 55.8%, respectively. HGBI have the lowest recall rates of 25.6%, 37.3%, 45.6% and 51.6%, respectively.

Case studies on five drugs

To prove the capacity of MGPred to predict drug–disease associations, we carry out a case study of five drugs: ampicillin, acetaminophen, etoposide, clonidine and sertraline. The predicted association scores between each drug and the candidate diseases are ranked in descending order. We also collect the top-10 candidate diseases associated with each drug for analysis. The results are shown in Table 3.

The Comparative Toxicogenomics Database (CTD) contains important information on the drugs and their effects on human diseases, hand-curated from the published literature [47]. DrugBank covers the clinical data and attributes information of the drugs [48]. In Table 3, 37 candidates were recorded in the CTD and six are contained in the DrugBank database, which shows that these candidate diseases are more likely to be associated with the corresponding drugs.

The largest clinical trial database, ClinicalTrials.gov, is operated by NLM. It contains information regarding on medical studies on human volunteers. PubChem is supported by the NIH and includes the chemical structures of small organic molecules and information regarding their biological activities. A total of 31 candidates are recorded in ClinicalTrials.gov. This suggests that the corresponding drug–disease association is supported by clinical trials. A total of 16 candidates are covered by PubChem [49]. It indicates there is a higher possibility that these drugs are associated with the corresponding diseases. Two candidate diseases are labeled 'literature' demonstrating that the effect of the corresponding drug on the disease is supported in the literature.

Hand-collated drug–disease associations and potential associations inferred from the literature are covered in the CTD. Two candidates (postoperative complications and pseudomonas infections) are supported in the CTD literature, indicating that they are more likely to be related to the corresponding drugs. Among the 50 total associations, one association is not confirmed and is labeled as 'unconfirmed'. In summary, the case studies of five drugs further prove that MGPred has the ability to discover drug-related diseases.

Prediction of novel drug–disease associations

We use the trained MGPred to conduct the association predictions. The 30 potential candidates for each drug are listed in supplementary table ST1.

Conclusions

The MGPred model was proposed to formulate and encode multiple similarities and associations between drugs and diseases, as well as multi-scale neighboring

Table 3. Top 10 candidate diseases related to each of five drugs.

Drug name	Rank	Disease name	Evidence	Rank	Disease name	Evidence
	1	Osteoarthritis	CTD/PubChem/ClinicalTrials	6	Spondyli-tis,Ankylosing	CTD/ClinicalTrials
	2	Inflammation	CTD/PubChem/ClinicalTrials	7	Gout	CTD/ClinicalTrials
acetaminophen	3	Dysmenorrhea	CTD/Drugbank/ClinicalTrials	8	Premenstrual Syndrome	Drugbank
	4	Arthritis,Rheumatoid	Drugbank/ClinicalTrials	9	Photophobia	CTD
	5	Arthritis,Juvenile Rheumatoid	CTD/PubChem/ClinicalTrials	10	Macular Edema	CTD/PubChem
	1	Breast Neoplasms	PubChem/ClinicalTrials	6	Precursor Cell Lymphoblastic Leukemia-Lymphoma	CTD/PubChem/Clini-calTrials
etoposide	2	Klebsiella Infections	unconfirmed	7	Escherichia coli Infections	CTD
	3	Lymphoma	CTD/ClinicalTrials	8	Sarcoma	PubChem/ClinicalTrials
	4	Bronchitis	CTD	9	Urinary Tract Infections	CTD
	5	Proteus Infections	CTD	10	Testicular Neoplasms	CTD/PubChem/Clini-calTrials
	1	Septicemia	PubChem/ClinicalTrials	6	Pseudomonas Infections	inferred candicate by 1 literature
	2	Klebsiella Infections	CTD/PubChem	7	Pneumonia, Bacterial	PubChem/ClinicalTrials
ampicillin	3	Skin Diseases, Infectious	DrugBank/ClinicalTrials	8	Wound Infection	ClinicalTrials
	4	Meningitis, Bacterial	CTD	9	Proteus Infections	PubChem
	5	Gram-Negative Bacterial Infections	DrugBank/PubChem/ClinicalTrials	10	Postoperative Complications	inferred candicate by 1 literature
	1	Depressive Disorder	CTD/ClinicalTrials	6	Autistic Disorder	CTD/ ClinicalTrials
	2	Attention Deficit Disorder with Hyperactivity	CTD	7	Phobic Disorders	CTD/ClinicalTrials
sertraline	3	Panic Disorder	CTD/DrugBank/ClinicalTrials	8	Tobacco Use Disorder	CTD
	4	Anxiety Disorders	CTD/ClinicalTrials	9	Bipolar Disorder	CTD/ClinicalTrials
	5	Pain	CTD/ClinicalTrials	10	Substance-Related Disorders	CTD/ClinicalTrials
	1	Asthma	CTD	6	Muscle Rigidity	CTD
	2	Spasm	CTD/ClinicalTrials	7	Rhinitis, Allergic, Perennial	CTD
clonidine	3	Pain	CTD/ClinicalTrials	8	Low Back Pain	CTD/ClinicalTrials
	4	Hypertension, Malignant	CTD/DrugBank/ ClinicalTrials	9	Hypotension, Orthostatic	CTD/PubChem/Clini-calTrials
	5	Angina Pectoris	CTD	10	Muscle Spasticity	CTD/PubChem/Clini-calTrials

topologies to predict disease-related drugs. A framework based on multiple graph convolutional autoencoders and multi-receptive field CNNs was established to learn the multi-scale neighboring topologies and node-pair-level attributes of drugs and diseases. The scale attention mechanism was proposed to assign more weight to the features of more important scales. The AUC and AUPR of MGPred are consistently higher than those of other drug–disease association prediction methods. The top-k candidates for MGPred have more realistic drug–disease associations. Comparisons with the latest six methods and five drug case studies demonstrate the powerful capacity of MGPred to identify potential drug-related diseases.

Key Points

- Three drug–disease bi-layer networks were built, which helped to extract the drug similarity, disease similarity and their associations from multi-sourced data. These networks were also conducive to further learning the topological representations of drugs and diseases from different perspectives.
- A new framework based on multiple graph convolutional autoencoders and CNN with multiple receptive fields was built to encode and integrate the two-level representations of drugs and diseases. Neighboring topological representations revealed multi-scale neighbor topologies of drugs and diseases in different heterogeneous networks, and node pair attribute representations revealed

the deep connection between drugs and diseases that were co-associated by r_i-d_j.
- Neighboring topologies of different scales contribute differently to the prediction of predicting drug-related diseases. We proposed scale-level attention to distinguish their contributions.
- The improvement in the predictive performance of MGPred was confirmed through a comparison with six advanced methods, the recall rates in top-ranked candidates and case studies of five drugs.

Supplementary data

Supplementary data are available at *Briefings in Bioinformatics* online.

Data availability

The dataset about this study is available at https://github.com/pingxuan-hlju/MGpred.

Funding

This work was supported by the Natural Science Foundation of China (62172143, 61972135), the Natural Science Foundation of Heilongjiang Province (LH2019A029), the China Postdoctoral Science Foundation (2020M670939, 2019M650069), and Heilongjiang Postdoctoral Scientific Research Staring Foundation (BHLQ18104).

References

1. Neuberger A, Oraiopoulos N, Drakeman DL, *et al.* Renovation as innovation: is repurposing the future of drug discovery research? *Drug Discov Today* 2019;**24**(1):1–3.
2. Sinha S, Vohora D, *et al.* Drug discovery and development: An overview. *Pharmaceutical medicine and translational clinical research* 2018;**2018**:19–32.
3. Pushpakom S, Iorio F, Eyers PA, *et al.* Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* 2019;**18**(1):41–58.
4. Zhao K, So HC, *et al.* Drug repositioning for schizophrenia and depression/anxiety disorders: A machine learning approach leveraging expression data. *IEEE J Biomed Health Inform* 2018;**23**(3):1304–15.
5. Ceddia G, Pinoli P, Ceri S, *et al.* Matrix factorization-based technique for drug repurposing predictions. *IEEE J Biomed Health Inform* 2020;**24**(11):3162–72.
6. Ashburn TT, Thor KB, *et al.* Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 2004;**3**(8):673–83.
7. Peng Y, Wang M, Xu Y, *et al.* Drug repositioning by prediction of drug's anatomical therapeutic chemical code via network-based inference approaches. *Brief Bioinform* 2020;**22**(2):2058–72.
8. Padhy BM, Gupta YK, *et al.* Drug repositioning: re-investigating existing drugs for new therapeutic indications. *J Postgrad Med* 2011;**57**(2):153–60.
9. Nosengo N, *et al.* Can you teach old drugs new tricks? *Nature News* 2016;**534**(7607):314.
10. Alfedi G, Luffarelli R, Cond I, *et al.* Drug repositioning screening identifies etravirine as a potential therapeutic for friedreich's ataxia. *Mov Disord* 2019;**34**(3):323–34.
11. Pritchard JE, Omara TA, Glubb DM, *et al.* Enhancing the Promise of Drug Repositioning through Genetics. *Front Pharmacol* 2017;**8**:896.
12. Wang Y, Chen S, Deng N, *et al.* Drug Repositioning by Kernel-Based Integration of Molecular Structure, Molecular Activity, and Phenotype Data. *Plos One* 2013;**8**(11):e78518.
13. Zhang W, Yue X, Lin W, *et al.* Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC bioinformatics* 2018;**19**(1):1–12.
14. Zhang W, Huang F, Yue X, *et al.* Prediction of drug-disease associations and their effects by signed network-based nonnegative matrix factorization. *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 2018;**2018**:798–802.
15. 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.
16. 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.
17. Liu H, Song Y, Guan J, *et al.* Inferring new indications for approved drugs via random walk on drug-disease heterogenous networks. *BMC bioinformatics* 2016;**17**(17):269–77.
18. 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.
19. Wang W, Yang S, Zhang X, *et al.* Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* 2014;**30**(20):2923–30.
20. 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**(8):1187–96.
21. Wang L, Wang Y, Hu Q, *et al.* Systematic Analysis of New Drug Indications by Drug-Gene-Disease Coherent Subnetworks. *CPT Pharmacometrics Syst Pharmacol* 2015;**3**(11):1–9.
22. Yang J, Li Z, Fan X, *et al.* Drug-disease association and drug-repositioning predictions in complex diseases using causal inference-probabilistic matrix factorization. *J Chem Inf Model* 2014;**54**(9):2562–9.
23. Wang MN, You ZH, Li LP, *et al.* WGMFDDA: A Novel Weighted-Based Graph Regularized Matrix Factorization for Predicting Drug-Disease Associations. *2020 IEEE International Conference on Intelligent Computing Methodologies (ICIC)* 2020;**2020**:542–51.
24. Luo H, Wang J, Li M, *et al.* Computational drug repositioning with random walk on a heterogeneous network. *IEEE/ACM Trans Comput Biol Bioinform* 2018;**16**(6):1890–900.
25. Chen H, Zhang Z, Peng W, *et al.* miRDDCR: a miRNA-based method to comprehensively infer drug-disease causal relationships. *Sci Rep* 2017;**7**(1):1–9.
26. Yu L, Su R, Wang B, *et al.* Prediction of novel drugs for hepatocellular carcinoma based on multi-source random walk. *IEEE/ACM Trans Comput Biol Bioinform* 2016;**14**(4):966–77.
27. Huang YF, Yeh HY, Soo VW, *et al.* Inferring drug-disease associations from integration of chemical, genomic and phenotype data using network propagation. *BMC Med Genomics* 2013;**6**(3):1–14.
28. 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.

29. Xuan P, Cui H, Shen T, *et al.* HeteroDualNet: A Dual Convolutional Neural Network With Heterogeneous Layers for Drug-Disease Association Prediction via Chou's Five-Step Rule. *Front Pharmacol* 2019;**10**:1301.
30. Jiang HJ, You ZH, Huang YA, *et al.* Predicting drug-disease associations via sigmoid kernel-based convolutional neural networks. *J Transl Med* 2019;**17**(1):1–11.
31. Jiang HJ, Huang YA, You ZH, *et al.* Predicting Drug-Disease Associations via Using Gaussian Interaction Profile and Kernel-Based Autoencoder. *Biomed Res Int* 2019;**2019**: 1–11.
32. Xuan P, Ye Y, Zhang T, *et al.* Convolutional neural network and bidirectional long short-term memory-based method for predicting drug-disease associations. *Cell* 2019;**8**(7):705.
33. Yu Z, Huang F, Zhao X, *et al.* Predicting drug-disease associations through layer attention graph convolutional network. *Brief Bioinform* 2021;**22**(4):bbaa243.
34. Xuan P, Gao L, Sheng N, *et al.* Graph Convolutional Autoencoder and Fully-Connected Autoencoder with Attention Mechanism Based Method for Predicting Drug-Disease Associations. *IEEE J Biomed Health Inform* 2021;**25**(5):1793–804.
35. Gao L, Cui H, Zhang T, *et al.* Prediction of drug-disease associations by integrating common topologies of heterogeneous networks and specific topologies of subnets. *Brief Bioinform* 2022;**23**(1):bbab467.
36. Wang B, Lyu X, Qu J, *et al.* GNDD: A Graph Neural Network-Based Method for Drug-Disease Association Prediction. *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 2019;**2019**:1253–5.
37. Jiang HJ, Huang YA, You ZH, *et al.* SAEROF: an ensemble approach for large-scale drug-disease association prediction by incorporating rotation forest and sparse autoencoder deep neural network. *Sci Rep* 2020;**10**(1):1–11.
38. Xuan P, Cao Y, Zhang T, *et al.* Drug repositioning through integration of prior knowledge and projections of drugs and diseases. *Bioinformatics* 2019;**35**(20):4108–19.
39. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004;**32**(suppl_1):D267–70.
40. Kim S, Thiessen PA, Bolton EE, *et al.* PubChem substance and compound databases. *Nucleic Acids Res* 2016;**44**(D1):D1202–13.
41. Mitchell A, Chang HY, Daugherty L, *et al.* The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* 2015;**43**(D1):D213–21.
42. Consortium U.P. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 2020;**2020**(D1):D1.
43. Cui Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* 2010;**26**(13):1644–50.
44. Ding H, Takigawa I, Mamitsuka H, *et al.* Similarity-based machine learning methods for predicting drug Ctarget interactions: a brief review. *Brief Bioinform* 2014;**15**(5):734–47.
45. Perlman L, Gottlieb A, Atias N, *et al.* Combining drug and gene similarity measures for drug-target elucidation. *J Comput Biol* 2011;**18**(2):133–45.
46. 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**(13):1644–50.
47. Davis AP, Grondin CJ, Johnson RJ, *et al.* Comparative toxicogenomics database (CTD): update 2021. *Nucleic Acids Res* 2021;**49**(D1):D1138–43.
48. Wishart DS, Feunang YD, An CG, *et al.* DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**(D1):D1074–82.
49. Kim S, Thiessen PA, Bolton EE, *et al.* PubChem substance and compound databases. *Nucleic Acids Res* 2016;**44**(D1):D1202–13.