



REDDA: Integrating multiple biological relations to heterogeneous graph neural network for drug-disease association prediction

Yaowen Gu^{a,1}, Si Zheng^{a,b,1}, Qijin Yin^c, Rui Jiang^c, Jiao Li^{a,*}

^a Institute of Medical Information (IMI), Chinese Academy of Medical Sciences and Peking Union Medical College (CAMS & PUMC), Beijing, 100020, China

^b Institute for Artificial Intelligence, Department of Computer Science and Technology, BNRIst, Tsinghua University, Beijing, 100084, China

^c Ministry of Education Key Laboratory of Bioinformatics, Bioinformatics Division at the Beijing National Research Center for Information Science and Technology, Center for Synthetic and Systems Biology, Department of Automation, Tsinghua University, Beijing, 100084, China



ARTICLE INFO

Keywords:

Drug repositioning
Drug-disease association prediction
Heterogeneous graph neural network
Topological subnet Topological subnet

ABSTRACT

Computational drug repositioning is an effective way to find new indications for existing drugs, thus can accelerate drug development and reduce experimental costs. Recently, various deep learning-based repurposing methods have been established to identify the potential drug-disease associations (DDA). However, effective utilization of the relations of biological entities to capture the biological interactions to enhance the drug-disease association prediction is still challenging. To resolve the above problem, we proposed a heterogeneous graph neural network called REDDA (Relations-Enhanced Drug-Disease Association prediction). Assembled with three attention mechanisms, REDDA can sequentially learn drug/disease representations by a general heterogeneous graph convolutional network-based node embedding block, a topological subnet embedding block, a graph attention block, and a layer attention block. Performance comparisons on our proposed benchmark dataset show that REDDA outperforms 8 advanced drug-disease association prediction methods, achieving relative improvements of 0.76% on the area under the receiver operating characteristic curve (AUC) score and 13.92% on the precision-recall curve (AUPR) score compared to the suboptimal method. On the other benchmark dataset, REDDA also obtains relative improvements of 2.48% on the AUC score and 4.93% on the AUPR score. Specifically, case studies also indicate that REDDA can give valid predictions for the discovery of -new indications for drugs and new therapies for diseases. The overall results provide an inspiring potential for REDDA in the *in silico* drug development. The proposed benchmark dataset and source code are available in <https://github.com/gu-yao-wen/REDDA>.

1. Introduction

The traditional wet-experiment-guided drug discovery is a time-consuming and high-risk process [1]. Recently, it has become increasingly difficult to identify potential therapeutic entities with novel chemical structures. The total cost of developing a new drug range from 3.2 to 27.0 billion dollars, take over 5.8–15.2 years and only achieve a success rate of 6.2% [2–4]. Thus, computational methods with cheaper and labor-saving solutions can accelerate the drug discovery, and have attracted increasing interests for both pharmaceutical industry and academic research communities [5,6]. For instance, there have been successful applications in drug property prediction [7–9], drug-target interaction assessment [8,10,11], and drug sensitivity prediction [12,

13], etc. Computational drug repositioning methods focus on determining the new indications for drugs [14], thus reducing the unnecessary cost and improving success rate of drug development [15]. A variety of promising applications show that the role of computational drug repositioning in drug development cannot be ignored [16–20].

The existing computational drug repositioning methods can be approximately divided into 4 categories [21]: classical machine learning approaches, network propagation approaches, matrix factorization/completion approaches, and deep learning approaches.

Classical machine learning approaches take known drug-disease association pairs as positive labels to convert drug repositioning into a binary classification problem and further adopt the drug and disease information as input features to train machine learning classifiers. For

* Corresponding author.

E-mail address: li.jiao@imicams.ac.cn (J. Li).

¹ The first two authors contribute equally to the paper.

instance, Gao et al. proposed a Laplacian regularized least squares algorithm combined with a similarity kernel fusion method to predict the drug-disease association, which called DDA-SKF [22]. Network propagation approaches construct drug-disease heterogeneous networks and use network-based algorithms (e.g., random walk) to predict the drug-disease association probabilities [6]. For example, Luo et al. proposed a method called MBiRW, which used similarity measurers to construct a drug-disease heterogeneous network and adopted the bi-random walk algorithm to predict potential drug-disease associations [23]. Matrix factorization/completion approaches model drug repositioning as a recommendation system, thus recommending new drugs/indications based on prior information such as known drug-disease associations. Zhang et al. proposed a similarity constrained matrix factorization method called SCMFDD for drug-disease association prediction. It maps the drug and disease features to low-rank spaces for solving constraint optimization [24]. Yang et al. took drug-disease association prediction as a noisy matrix completion problem and developed a bounded nuclear norm regularization (BNNR) method for it [25]. Yang et al. proposed a multi-similarities for bilinear matrix factorization (MSBMF) to extract effective drug and disease representations, which could be used to infer missing drug-disease associations [26]. With ongoing development in recent years, these types of computational drug repositioning methods have gained competitive performance. However, some crucial shortcomings still limit the achievement of higher accuracy and the utilization of practical scenarios, such as the high dependence on the quality of input features in machine learning approaches, the representation bias for nodes with high degrees on heterogeneous network in network propagation approaches, and the weak representation ability of drug-disease associations caused by linear multiplication in matrix factorization/completion approaches.

Deep learning approaches have been effectively applied in many biological domains, such as gene regulatory representation [27,28], single-cell omics analysis [29–32], drug efficacy prediction [33], etc. For drug repositioning, deep learning approaches use neural networks to model the interactions between drugs and diseases with high flexibility and scalability, which have been widely used and proven to be highly competitive compared to the above three approaches [34–41]. For instance, Zeng et al. integrated 10 drug-disease-related networks and trained a multimodal deep autoencoder on them to learn the high-order representations for drug repositioning, which is called deepDR [37]; Yu et al. established a graph convolutional network called LAGCN on a heterogeneous drug-disease network [34]; Meng et al. proposed a neighborhood and neighborhood interaction-based neural collaborative filtering approach called DRWBNCF for drug repositioning [36]. Zhang et al. designed a multi-scale topology learning method which integrated multiple drug-disease heterogeneous network and adopted random walk and attention mechanism for representation learning [42]. Xuan et al. proposed a graph autoencoder architecture with scale-level attention and convolutional neural network which called MGPPred [43].

As sufficiently advanced methods for modeling drug-disease associations, these approaches have provided a series of attractive methodologies for deep learning-based drug repositioning, such as the construction of the heterogeneous networks, the utilization of layer attention mechanism, and bilinear dot decoder, etc. However, the drug-disease associations cannot be simply integrated as an isolated biological system as the above studies have done, while ignoring other extensive biological interactions, such as drug-protein, protein-gene, gene-pathway, pathway-disease, etc. From our perspective, these external biological relations can be assembled in the drug-disease heterogeneous network and bring extra information for the simulations of drug therapeutic process, thus enhance the representation ability of the drug repositioning model. Nevertheless, these concerns have not been studied in depth.

To resolve these problems, we propose a benchmark dataset which can construct to a heterogeneous network with 5 entities (drug, protein, gene, pathway, and disease) and 10 relations (drug-drug, drug-protein,

protein-protein, gene-gene, gene-pathway, pathway-pathway, pathway-disease, disease-disease, and drug-disease) for drug repositioning. Furthermore, we also develop a promising drug repositioning method on the heterogeneous network, which we called **Relations-Enhanced Drug-Disease Association prediction (REDDA)**. The main contributions of this work are summarized as follows:

- We propose a large-scale benchmark dataset for drug repositioning. The benchmark contains 41,100 nodes and 1,008,258 edges with 5 biological entities (drug, protein, gene, pathway, and disease).
- We propose a deep learning-based method for drug repositioning, namely REDDA. It takes the heterogeneous graph neural network as the backbone and integrates 3 attention mechanisms to learn the node representations of the heterogeneous network and topological subnets.
- Comprehensive experiments demonstrate that REDDA outperforms several state-of-the-art algorithms. Ablation experiments indicate that the fusion of extra biological relations is beneficial for REDDA to predict drug-disease associations. Attention visualization analysis shows the importance of topological decomposition, graph-level aggregation, and layer-level aggregation in REDDA.

2. Materials and methods

2.1. Dataset

As the existing benchmarks lack the biological entities and their relations, we construct a drug-disease association benchmark, including 5 entities (drug, protein, gene, pathway, and disease) and 10 relations (drug-drug, drug-protein, protein-protein, protein-gene, gene-gene, gene-pathway, pathway-pathway, pathway-disease, disease-disease, and drug-disease) as these biological entities and relations have been proved to contribute to drug repositioning [44–46]. We assemble the Fdataset [47], Cdataset [23], and additional data downloaded from KEGG [48] as our drug-disease association data, which contains 894 drugs, 454 diseases, and 2704 confirmed drug-disease associations. Then, we adopt the biological relation data from DrugBank [49], CTD [50], KEGG [48], STRING [51], and UniProt [52], integrating them into a large-scale benchmark including 41,100 nodes and 1,008,258 edges. The descriptions of our proposed benchmark dataset are listed in **Table 1**. The Venn diagrams and other details of our proposed benchmark dataset can be founded in **Fig. S1** and **Table S1**.

To emphasize the robustness and stability of REDDA, we also reorganize a public drug-disease association benchmark dataset called B-dataset for evaluation, which is used in Ref. [44] and originally proposed in Ref. [24], including 269 drugs, 598 diseases, but 18,416

Table 1
Summary of our proposed drug-disease association benchmark dataset.

Dataset	Data Type				
	Entities				
Number	Drugs	Proteins	Genes	Pathways	Diseases
Resource	894	18,877	20,561	314	454
Number	DrugBank,	KEGG,	CTD	KEGG,	KEGG,
Resource	Reference	Uniprot		CTD	Reference
Interactions					
Drug-Drug					
Drug-Protein					
Protein-Protein					
Protein-Gene					
Gene-Pathway					
Pathway-Pathway					
Pathway-Disease					
Disease-Disease					
Associations					
Drug-Protein					
Protein-Gene					
Gene-Pathway					
Pathway-Disease					
Drug-Disease					
Number	14,291	201,382	712,546	1669	7199
Resource	Sim Comput	STRING	CTD	KEGG	Sim Comput
Associations					
Drug-Protein					
Protein-Gene					
Gene-Pathway					
Pathway-Disease					
Drug-Disease					
Number	4397	18,545	25,995	19,530	2704
Resource	DrugBank	KEGG,	CTD'	CTD	KEGG,
		Uniprot			Reference

drug-disease associations. Moreover, we collect the drug-protein, protein-disease associations, and protein-protein interactions data from DrugBank, STRING, and CTD databases. The detailed descriptions of the B-dataset are listed in Table 2.

2.2. Construction of the heterogeneous network

2.2.1. Drug-drug similarities

As previous research has proven the benefit of computing drug-drug similarities as the drug-drug interaction probabilities in the drug-disease heterogeneous network [34–36], we compute the chemical structure similarities in our method by converting the SMILES sequences to 2048-bit ECFP4 fingerprints [53] and calculate the pair-wise similarities of drug fingerprints using some similarity measurements. Given two drugs i and j , their fingerprints are represented as x_i and x_j , the Jaccard index can be calculated by:

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

Similarly, the Tanimoto similarity can be calculated by:

$$S_{ij}^R = \frac{x_i \cdot x_j}{x_i^2 + x_j^2 - x_i \cdot x_j} \quad (2)$$

The Dice index can be calculated by:

$$S_{ij}^R = \frac{2|x_i \cap x_j|}{|x_i| + |x_j|} \quad (3)$$

The model performance comparisons for different similarity measurements can be shown in Table S2. Consequently, we adopt the Tanimoto similarity as the optimal method for calculating drug-drug similarities. Also, considering the computing complexity of the heterogeneous network, we simplify the continual drug-drug similarities to binarization values. Given a drug i and its similarities with another drug $S_i^R \in [0, 1]$, we convert the topk similarities (here $k = 15$) S_{ik}^R to 1, and the last ones are converted to 0 which means no existing interaction. After the conversion, the $S^R \in \{0, 1\}$ are gained as the final drug-drug interactions.

2.2.2. Disease-disease similarities

To adopt disease-disease interactions, medical subject headings (MeSH) identifiers are used to calculate semantic similarities as interaction probabilities. Similar to previous research [34,35,44,54], the MeSH identifier of a disease can be represented as a hierarchical directed acyclic graph (DAG). Given a disease d , the DAG can be represented as $\text{DAG}(d) = (\mathcal{N}(d), \mathcal{E}(d))$, where $\mathcal{N}(d)$ denotes the set of nodes including d and its ancestral nodes, and $\mathcal{E}(d)$ denotes the parent-child relation links among $\mathcal{N}(d)$. The semantic contribution of a node $n \in \mathcal{N}(d)$ for d can be formulated as:

$$C_d(n) = \begin{cases} 1, & \text{if } n = d \\ \max\{\Delta.C_d(n') | n' \in \text{children of } n\}, & \text{otherwise} \end{cases} \quad (4)$$

where δ is a contribution factor. The overall semantic contribution of d is $DV(d) = \sum_{n \in \mathcal{N}(d)} C_d(n)$. Then, the disease-disease similarity of d_i and d_j can be measured by the number of common ancestral nodes and the semantic contribution proportion of these ancestral nodes in $DV(d_i)$ and $DV(d_j)$, which can be formulated as:

$$S_{ij}^D = \frac{\sum_{n \in \mathcal{N}(d_i) \cap \mathcal{N}(d_j)} (C_{d_i}(n) + C_{d_j}(n))}{DV(d_i) + DV(d_j)} \quad (5)$$

Similar to the operation on drug similarity, $S^D \in \{0, 1\}$ are acquired as the final disease-disease interactions by top15 filtering.

2.2.3. Construction of heterogeneous network

The heterogeneous network is constructed by our proposed benchmarks. Given an entity type set $\mathcal{O} = \{\text{o}|R, P, G, M, D\}$ (representing drug, protein, gene, pathway, and disease, respectively), then the heterogeneous network can be regarded as a heterogeneous graph $\mathcal{G}^\theta = (\mathcal{V}^\theta, \mathcal{E}^\theta)$, where $\mathcal{V}^\theta = \{\mathcal{V}^R, \mathcal{V}^P, \mathcal{V}^G, \mathcal{V}^M, \mathcal{V}^D\}$ is the node set with 5 types, and $\mathcal{E}^\theta = \{\mathcal{E}^{R-R}, \dots, \mathcal{E}^{o_1-o_2}, \dots, \mathcal{E}^{D-D}\}$ is the edge set of \mathcal{G}^θ . The unweighted and undirected heterogeneous network can be divided into an adjacency matrix A^θ and the node feature matrix H^θ . Based on the collected relations in our benchmark, the adjacency matrix $A^\theta \in \mathbb{R}^{(N^R+N^P+N^G+N^M+N^D) \times (N^R+N^P+N^G+N^M+N^D)}$ can be defined as:

$$A^\theta = \begin{bmatrix} S^R & A^{R-P} & 0 & 0 & A^{R-D} \\ (A^{R-P})^T & A^{P-P} & A^{P-G} & 0 & 0 \\ 0 & (A^{P-G})^T & A^{G-G} & A^{G-M} & 0 \\ 0 & 0 & (A^{G-M})^T & A^{M-M} & A^{M-D} \\ (A^{R-D})^T & 0 & 0 & (A^{M-D})^T & S^D \end{bmatrix} \quad (6)$$

We also take the computed similarities S^R and S^D to represent the chemical structure pairwise features and disease classification pairwise features, so $H^\theta \in \mathbb{R}^{(N^R+N^D) \times (N^D+N^R)}$ can be initialized and represented as:

$$H^\theta = \begin{bmatrix} H^R \\ H^D \end{bmatrix} = \begin{bmatrix} S^R & 0 \\ 0 & S^D \end{bmatrix} \quad (7)$$

2.3. Backbone model and mechanism

2.3.1. Heterogeneous graph convolutional network (HeteroGCN)

In REDDA, a HeteroGCN is used as the backbone model, which learns the node representations in a series of homogeneous graphs and then aggregates the neighbor nodes' embedding with different node types to accomplish heterogeneous graph learning and node embedding updating.

Given a homogeneous graph $\mathcal{G}^{\theta-\theta}$ whose node type is θ , the node representations \hat{H}^θ can be learned by a classic GCN in $\mathcal{G}^{\theta-\theta}$. The computation in the GCN layer can be formulated as:

$$\hat{H}^\theta = GCN(A^{\theta-\theta}, H^\theta, W^\theta) \quad (8)$$

where $A^{\theta-\theta}$ is the adjacency matrix of the $\theta-\theta$ interaction network, H^θ is the input node embedding, and W^θ is the trainable parameter matrix in GCN. Specifically, the propagation of the GCN layer can be represented as:

$$GCN(A, H, W) = \sigma(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}HW) \quad (9)$$

where $D = \text{diag}\left(\sum_j A_{ij}\right)$ is the degree matrix of $\mathcal{G}^{\theta-\theta}$ and $\sigma =$

Table 2
Summary of B-dataset.

Dataset	Data Type	Entities	Proteins	Diseases
Number Resource	Drugs	6040	598	
Resource	DrugBank, Reference	Uniprot		Reference
	Interactions			
Number Resource	Drug-Drug	592,926	357,604	
	Sim Comput	STRING		Sim Comput
	Associations			
Number Resource	Drug-Protein	17,631	18,416	
	DrugBank	CTD		Reference

$$\begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases}$$

is the PReLU activation function.

After generating the homogeneous-graph-level node representations, a sum aggregation method is traversed and executed on all nodes in the heterogeneous graph to acquire the heterogeneous-graph-level node representations. Given a node i , the aggregation method sums the embeddings of its neighbor nodes $j \in \mathcal{N}_i^o$ to calculate the final embedding of i , which can be formulated as:

$$\tilde{H}_i = \sum_{n \in \mathcal{O}} \sum_{j \in \mathcal{N}_i^n} \hat{H}_j^n \quad (10)$$

2.3.2. Attention mechanism

Different blocks in REDDA focus on capturing different representations for identifying the biological relations. The attention mechanism can aggregate multi-source representations and adjust their importance weights dynamically. Therefore, we adopt the attention mechanism in REDDA inspired by Ref. [55] to gather different representations. Given K multi-source representations for node i , we first calculate the importance weight w_i^k for each source k by measuring the non-linear similarity of the h_i^k and an attention vector q , which can be formulated as:

$$w_i^k = \frac{1}{K} \sum_{i \in K} q^T \text{Linear}(h_i^k) \quad (11)$$

Then we normalize the w_i^k and acquired the attention coefficient $\alpha_i^k \in (0, 1)$ by a Softmax function:

$$\alpha_i^k = \frac{\exp(w_i^k)}{\sum_{i=1}^K \exp(w_i^k)} \quad (12)$$

The final representation \hat{H}_i for node i can be calculated by a weighted sum:

$$\hat{H}_i = \sum_{k=1}^K \alpha_i^k H_i^k \quad (13)$$

2.4. The architecture of REDDA

In this section, we introduce the architecture of REDDA for drug-disease association prediction, which is shown in Fig. 1. The overall

REDDA computation flow is described in Algorithm 1.

Algorithm 1. The REDDA algorithm.

2.4.1. Node embedding block

In REDDA, a node embedding block is adopted as the first module, which includes a feature transformation layer and two heterogeneous GCN (HeteroGCN) layers. First, taking the node feature matrix $H^o \in \mathbb{R}^{(N^R+N^P+N^G+N^M+N^D) \times K}$ and heterogeneous network \mathcal{G}^o as input, the REDDA deploy a feature transformation layer to map node representations with different node types into the same hidden space. The feature transformation layer can be formulated as:

$$H^o(0) = \begin{bmatrix} H^{R(0)} \\ H^{P(0)} \\ H^{G(0)} \\ H^{M(0)} \\ H^{D(0)} \end{bmatrix} = \begin{bmatrix} \text{Linear}(H^R) \\ \text{Initialize}(\mathcal{V}^P) \\ \text{Initialize}(\mathcal{V}^G) \\ \text{Initialize}(\mathcal{V}^M) \\ \text{Linear}(H^D) \end{bmatrix} \quad (14)$$

where the *Linear* function is a linear layer for the feature dimension alignment and *Initialize* function is used to allocate initial features for protein, gene, and pathway nodes.

For the node embedding updating, we use a HeteroGCN (described in section 2.3.1) to propagate adjacent biological network information. Given the l -th layer, the output $H^o(l)$ can be formulated as:

$$\tilde{H}^o(l) = \text{HeteroGCN}(A^o, \hat{H}^o(l-1), W^o(l)) \quad (15)$$

2.4.2. Topological subnet embedding block

Although the node embedding block captures the global adjacency relation among the heterogeneous network, the node representations are biased because of the unbalanced number of different associations (e.g., the number of pathway-disease associations is over 22 times the number of drug-disease associations). Therefore, we propose a topological subnet embedding block to learn the node embedding guided by different association information.

First, the topological subnet embedding block decomposes the heterogeneous network into 5 subnets that each of them include 2 node types and 3 relation types (e.g., the subnet, which comprises drug-drug

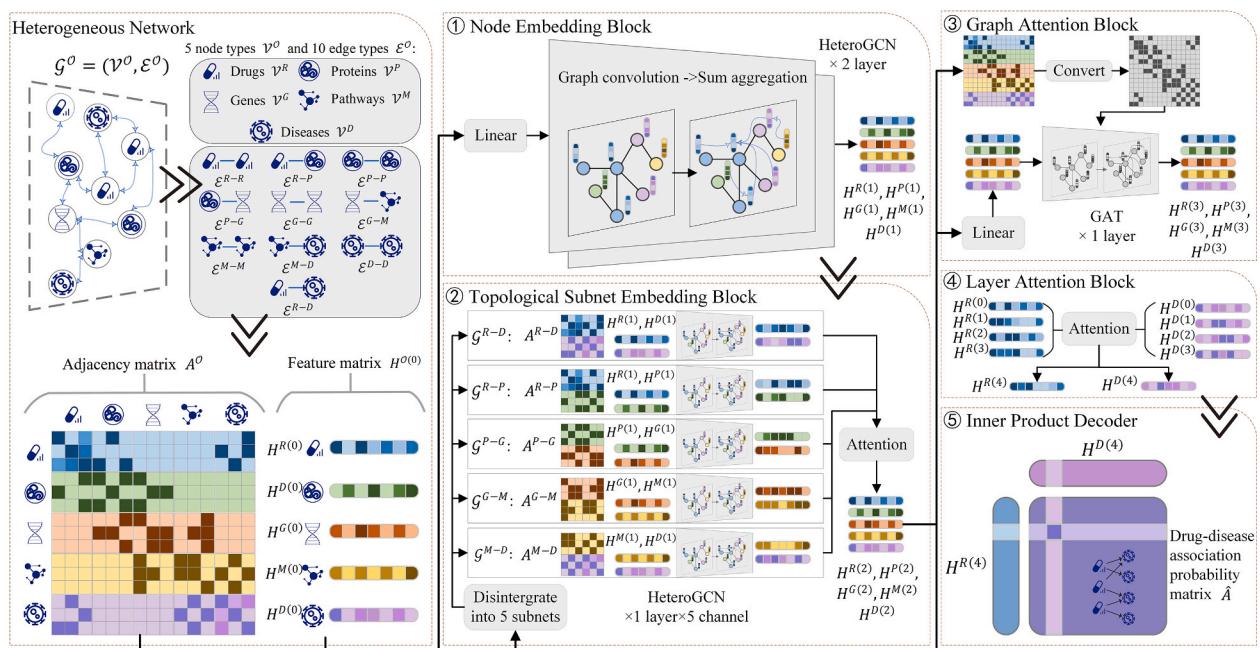


Fig. 1. The architecture of REDDA.

Algorithm 1: The REDDA algorithm

Input: The node type $\sigma \in \{R, P, G, M, D\}$; the heterogeneous network $\mathcal{G}^{\mathcal{O}} = \{\mathcal{V}^{\mathcal{O}}, \mathcal{E}^{\mathcal{O}}\}$; the adjacency matrix of heterogeneous network $A^{\mathcal{O}}$ as the **Equation 4** formulated; the initialized node feature matrix $H^{\mathcal{O}}$ as the **Equation 5** formulated; dataset $\mathcal{D} = A^{R-D}$; number of training epochs T .

Output: Reconstructed drug-disease association probability matrix \hat{A} .

- 1 Initialize the trainable parameters in MODDA ;
- 2 Split dataset $\mathcal{D} = \{A_{train}^{R-D}, A_{test}^{R-D}\}$, delete existing drug-disease associations A_{test}^{R-D} from $A^{\mathcal{O}}$;
- 3 **for** training epochs $i = 1, \dots, T$ **do**
- 4 Compute the initial embedding $H^{\mathcal{O}(0)}$ by **Equation 11** ;
- 5 Compute the embedding $H^{\mathcal{O}(\text{NodeEmb})}$ in node embedding block by **Equation 12**;
- 6 Decompose the heterogeneous network to 5 topological subnets
 $\mathcal{G}^{R-D}, \mathcal{G}^{R-P}, \mathcal{G}^{P-G}, \mathcal{G}^{G-M}, \mathcal{G}^{M-D}$;
- 7 **for** topological subnets $\mathcal{G} = \mathcal{G}^{R-D}, \dots, \mathcal{G}^{G-M}$ **do**
- 8 Construct a HeteroGCN for each subnet and compute the node embedding $H^{\sigma_1}, H^{\sigma_2}$;
- 9 **endfor**
- 10 Aggregate H^{σ} and compute the embedding in topological subnet embedding block
 $H^{\mathcal{O}(\text{SubnetEmb})}$ by **Equation 15** ;
- 11 Convert $\mathcal{G}^{\mathcal{O}}$ to a homogeneous graph $\mathcal{G}^{\mathcal{O}(\text{Homo})}$ and compute the embedding $H^{\mathcal{O}(\text{GraphEmb})}$ in graph attention block by **Equation 16-17** on $\mathcal{G}^{\mathcal{O}(\text{Homo})}$;
- 12 Extract drug embedding $H^{R(0)}, H^{R(\text{NodeEmb})}, H^{R(\text{SubnetEmb})}, H^{R(\text{GraphEmb})}$ and disease embedding $H^{D(0)}, H^{D(\text{NodeEmb})}, H^{D(\text{SubnetEmb})}, H^{D(\text{GraphEmb})}$;
- 13 Compute the embedding $H^{R(\text{LayerAttn})}, H^{D(\text{LayerAttn})}$ in layer attention block by **Equation 18-19**;
- 14 Reconstruct the drug-disease association probability matrix \hat{A} by **Equation 20** ;
- 15 Update parameters by optimizing **Equation 21** ;
- 16 **endfor**
- 17 Output \hat{A} .

interactions, disease-disease interactions, and drug-disease associations). In this way, the topological subnets can be constructed without unavailable associations in our heterogeneous network such as the drug-pathway associations. Given the heterogeneous network \mathcal{G} , the topological decomposition can be formulated as:

$$\mathcal{G} = \{\mathcal{G}^{R-D}, \mathcal{G}^{R-P}, \mathcal{G}^{P-G}, \mathcal{G}^{G-M}, \mathcal{G}^{M-D}\} \quad (16)$$

Given 2 kinds of nodes θ_1 and θ_2 , the adjacency matrix $A^{\theta_1-\theta_2} \in \mathbb{R}^{(N^{\theta_1}+N^{\theta_2}) \times (N^{\theta_2}+N^{\theta_1})}$ of the subnet $\mathcal{G}^{\theta_1-\theta_2}$ can be formulated as:

$$A^{\theta_1-\theta_2} = \begin{bmatrix} A^{\theta_1-\theta_1} & A^{\theta_1-\theta_2} \\ (A^{\theta_1-\theta_2})^T & A^{\theta_2-\theta_2} \end{bmatrix} \quad (17)$$

Then, each subnet is assembled with a HeteroGCN to learn subnet-level node embedding. Consequently, each node v^{θ} will acquire two embedding. For node θ_1 , the node embedding is generated from the subnets $\mathcal{G}^{\theta_1-\theta_2}$ and $\mathcal{G}^{\theta_1-\theta_3}$ (e.g., the drug embedding includes representations from drug-disease subnet and drug-protein subnet). Moreover, the attention mechanism (described in section 2.3.3) is adopted to dynamically aggregate node embedding, and the attention coefficients represent the contributions of different subnets learned from REDDA. The above computation can be formulated as:

$$\hat{H}^{\theta_1} = \text{Attention} \left(\begin{bmatrix} \text{HeteroGCN}(A^{\theta_1-\theta_2}, H^{\theta_1}, W^{\theta_1-\theta_2}) \\ \text{HeteroGCN}(A^{\theta_1-\theta_3}, H^{\theta_1}, W^{\theta_1-\theta_3}) \end{bmatrix} \right) \quad (18)$$

2.4.3. Graph attention block

After generating the subnet-level node embedding, we consider reconstructing the whole biological network \mathcal{G} and updating the node embedding with a graph attention layer. In the graph attention block, \mathcal{G} is reconstructed as a homogeneous graph so that the graph attention

network can iteratively run on the whole graph. Given two adjoining nodes i and j , the attention weight α_{ij} can be calculated as:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a^T [WH_i WH_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(a^T [WH_i WH_k]))} \quad (19)$$

where \mathcal{N}_i is the node set which is the neighbors of i ; H_i and H_j are the embeddings of node i and j ; $W \in \mathbb{R}^{1 \times K}$ and $a^T \in \mathbb{R}^{2 \times K}$ are the learnable parameter matrixes; and $\text{LeakyReLU} = \begin{cases} x, x \geq 0 \\ 0.001 \cdot x, x < 0 \end{cases}$ is the activation function.

To enhance the numerical stability of the graph attention network, we use multi-head attention to generate the output of node embeddings. Setting K attention heads that are randomly initialized, the output node embedding \hat{H}_i with multi-head attention can be formulated as:

$$\hat{H}_i = \left| \sum_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W^k H_j \right) \right| \quad (20)$$

where W^k is the learnable parameter matrix of the k -th attention heads, and σ is the PReLU activation function.

2.4.4. Layer attention block

To alleviate the over-smoothing of the model and connect the low-order/high-order representations, we use a layer attention block to aggregate each embedding from the previous blocks by attention mechanism (described in section 2.3.3). The final embedding of drug nodes $H^{R(\text{LayerAttn})} \in \mathbb{R}^{N^R \times K}$ and disease nodes $H^{D(\text{LayerAttn})} \in \mathbb{R}^{N^D \times K}$ can be formulated as:

$$H^{R(\text{Layer Attn})} = \text{Attention} \left(\begin{bmatrix} H^{R(0)} \\ H^{R(\text{NodeEmb})} \\ H^{R(\text{SubnetEmb})} \\ H^{R(\text{GraphAttn})} \end{bmatrix} \right) \quad (21)$$

$$H^{D(\text{Layer Attn})} = \text{Attention} \left(\begin{bmatrix} H^{D(0)} \\ H^{D(\text{NodeEmb})} \\ H^{D(\text{SubnetEmb})} \\ H^{D(\text{GraphAttn})} \end{bmatrix} \right) \quad (22)$$

2.4.5. Bilinear inner product decoder

To reconstruct the drug-disease association matrix, we use an inner product decoder with a linear layer, which is formulated as:

$$\hat{A} = f(H^R, H^D) = \text{sigmoid}(H^R W (H^D)^T) \quad (23)$$

where $W \in \mathbb{R}^{K \times K}$ is the trainable parameter matrix and \hat{A} is the reconstructed drug-disease association matrix.

2.5. Optimization

We use a weighted cross-entropy loss function to balance the impact of different categories and help REDDA focus on the known drug-disease associations. Given N^R drugs and N^D diseases in a heterogeneous network, the known/unknown drug-disease associations are labeled as S^+ and S^- , then the loss function can be formulated as:

$$\text{Loss} = -\frac{1}{N^R \times N^D} \left(\gamma \sum_{(i,j) \in S^+} \log \hat{A}_{ij} + \sum_{(i,j) \in S^-} (1 - \log \hat{A}_{ij}) \right) \quad (24)$$

where $\gamma = \frac{|S^+|}{|S^+| + |S^-|}$ is the balance weight, $|S^+|$ and $|S^-|$ are the number of known/unknown drug-disease associations in the training set, and \hat{A}_{ij} is the predicted probability of drug i and disease j .

Following previous studies [35], we use the Adam optimizer for model optimization and initialize the trainable parameters in each layer by Xavier [56]. Moreover, the dropout layer and batch normalization layer are also adopted to inhibit overfitting. To enhance the stability of the training process, a cyclic learning scheduler is also used to dynamically adjust the learning rate during the training process.

2.6. Experimental settings

To estimate the performance of REDDA and other drug repositioning algorithms, we execute 10-fold cross-validations on each model and repeated them 10 times to decrease the random error caused by data splitting.

Considering the label imbalance in drug-disease association prediction which need to focus more on the predictions of confirmed associations, we use multiple metrics to evaluate the performance of REDDA and other proposed models, including AUC, AUPR, F1-Score, Accuracy, Recall, Specificity, and Precision which are calculated by the process proposed by Ref. [34].

2.7. Hyper-parameter setting

For the model-related hyper-parameters in REDDA, our proposed REDDA algorithm is assembled by a sequential architecture (HeteroGCN -> HeteroGCN- > GAT) in the encoder and a fully connected layer in the decoder. Each layer has 128 hidden units, a dropout rate of 0.4, and batch normalization. For the training-related hyper-parameters in REDDA, we set the maximum and minimum learning rates to 0.01 and 0.001 in the cyclic learning rate scheduler and the number of training epochs to 4000.

We used a grid search to determine these above hyper-parameters throughout the experiments. The results of the grid search

experiments are shown in Figs. S2–S3.

3. Results

3.1. Comparison with state-of-the-art approaches

We compared REDDA with 8 drug-disease association prediction methods to demonstrate the effectiveness of our model, including SCMFDD [24], MBiRW [23], NIMGCN [57], HINGRL-Node2Vec-RF, HINGRL-DeepWalk-RF [44], LAGCN [34], and DRWBNCF [36]. The details for the construction of these baseline methods can be found in supplementary materials.

The performance results of 10-fold cross-validations on our proposed benchmark dataset are shown in Table 3 and Fig. 2, while the statistical results of the difference between these performance metrics of REDDA to those of 8 baseline methods are listed in Tables S3–S4. These results indicated that our proposed REDDA outperformed the other 8 baseline methods in terms of the majority of metrics including AUC, AUPR, F1-Score, and Recall, on which REDDA achieved relative improvements of 0.44%, 26.14%, 18.99%, and 26.69% compared to the suboptimal methods, while the statistical results also proved the significance of these improvements (p -values < 0.001). Focusing on the accuracy, specificity, and precision metrics, the performance of our method is still comparable. It should also be noted that compared to the other HeteroGCN-based methods, such as NIMGCN, LAGCN, and DRWBNCF, our proposed REDDA obtained desirable improvements on the major metrics (13.97% on AUC, 26.14% on AUPR and 18.99% on F1-Score compared to the suboptimal method). These results tended to indicate the effective use of HeteroGCN in REDDA. Compared to HINGRL-Node2Vec-RF and HINGRL-DeepWalk-RF, which used extra biological relations as well and can be regarded as the suboptimal methods among the 8 models, our method still achieved noticeable relative improvements (0.44% on AUC, 84.23% on AUPR, and 71.28% on F1-Score). We also calculated the Recall@k metric adopted by some previous studies [36,37], which means the ratio of correctly predicted drug-disease associations in the Top k predictions. The Recall@10,000 curves shown in Fig. 2 indicated that REDDA achieved the best Recall@10,000 values compared to 8 methods (Top3 results: 66.68% in REDDA; 51.63% in DRWBNCF; and 49.78% in HINGRL-Node2Vec-RF).

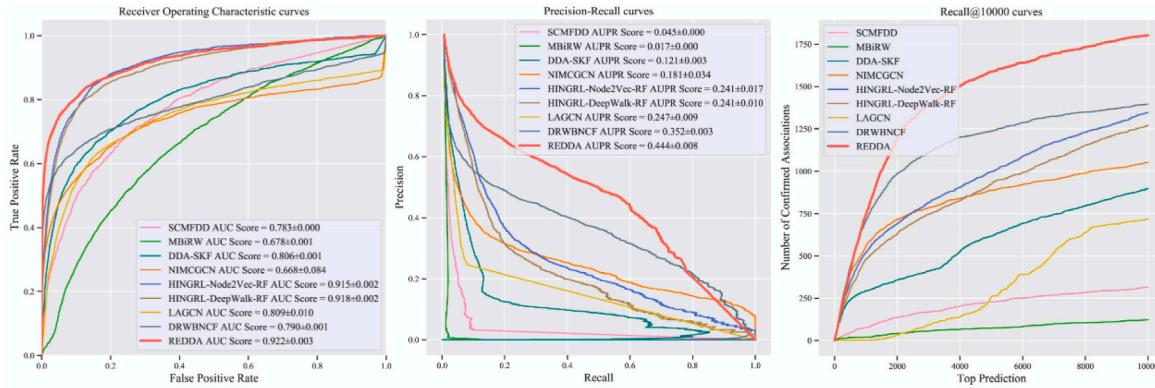
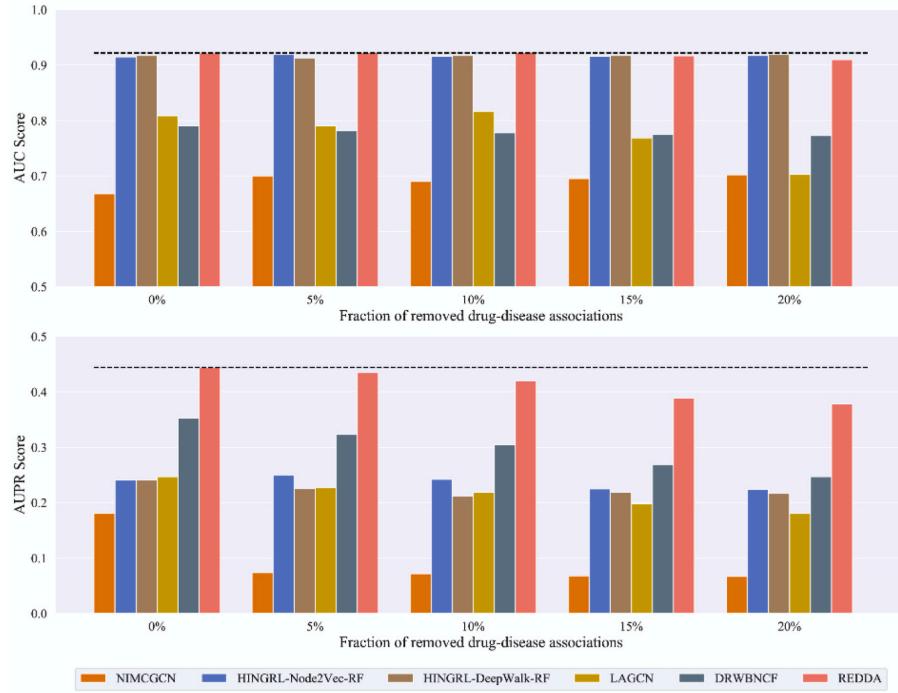
When facing a sparse biological network where the known drug-disease associations are insufficient, the methods need to learn from numerous unknown drug-disease associations and external biological information to maintain stability and robustness. As the previous study mentioned [34], we randomly removed part of known drug-disease associations in our benchmark datasets at a ratio $r \in \{0\%, 5\%, 10\%, 15\%, 20\%\}$, and trained REDDA and other 5 models (NIMGCN, HINGRL-Node2Vec-RF, HINGRL-DeepWalk-RF, LAGCN, and DRWBNCF) on them. The AUC and AUPR results of 10-fold cross-validations on these partially removed datasets are shown in Fig. 3. The results showed that our proposed REDDA maintained the state-of-the-art performance among the 6 methods. More inspiringly, REDDA keeps stable AUPR scores while some other methods got significant decreases (e.g., NIMCGCN, LAGCN, and DRWBNCF). These results demonstrated that REDDA was relatively insensitive to the proportion of known drug-disease associations of the training set, implying a potential of REDDA for sparse biological network-based drug repositioning.

We also accomplished a performance comparison experiment on the B-dataset to emphasize the solidarity of REDDA on public benchmark. The performance results and curves of REDDA and 4 baseline methods (NIMCGCN, HINGRL-DeepWalk-RF, LAGCN, and DRWBNCF) are shown in Table 4 and Fig. 4, while the p -values of performance comparisons are listed in Tables S5–S6. We observed that REDDA also performed best among the 5 methods. In each metric comparison, REDDA respectively achieved a relative improvement of 2.48%, 4.93%, 4.35%, 1.14%, 0.18%, 1.20%, and 8.06% compared to the suboptimal method

Table 3

Performance comparison of 9 methods on our proposed benchmark dataset.

Methods	AUC	AUPR	F1-Score	Accuracy	Recall	Specificity	Precision
SCMFDD	0.783 ± 0.000	0.045 ± 0.000	0.094 ± 0.000	0.989 ± 0.000	0.136 ± 0.000	0.993 ± 0.000	0.072 ± 0.000
MBiRW	0.678 ± 0.001	0.017 ± 0.000	0.036 ± 0.000	0.922 ± 0.005	0.218 ± 0.013	0.020 ± 0.000	0.927 ± 0.005
DDA-SKF	0.806 ± 0.001	0.121 ± 0.003	0.165 ± 0.002	0.985 ± 0.000	0.226 ± 0.007	0.990 ± 0.001	0.130 ± 0.003
NIMGCN	0.668 ± 0.084	0.181 ± 0.034	0.260 ± 0.036	0.993 ± 0.000	0.197 ± 0.033	0.998 ± 0.001	0.388 ± 0.056
HINGRL-Node2Vec-RF	0.915 ± 0.002	0.241 ± 0.017	0.289 ± 0.008	0.992 ± 0.000	0.242 ± 0.015	0.997 ± 0.000	0.360 ± 0.024
HINGRL-DeepWalk-RF	0.918 ± 0.002	0.241 ± 0.010	0.283 ± 0.012	0.992 ± 0.001	0.248 ± 0.030	0.997 ± 0.001	0.341 ± 0.045
LAGCN	0.809 ± 0.010	0.247 ± 0.009	0.223 ± 0.019	0.983 ± 0.002	<u>0.356 ± 0.012</u>	0.988 ± 0.002	0.163 ± 0.019
DRWBNCF	0.790 ± 0.001	<u>0.352 ± 0.003</u>	<u>0.416 ± 0.004</u>	0.994 ± 0.000	0.347 ± 0.017	0.998 ± 0.000	0.524 ± 0.036
REDDA	0.922 ± 0.003	0.444 ± 0.008	0.495 ± 0.009	0.994 ± 0.000	<u>0.451 ± 0.024</u>	<u>0.998 ± 0.000</u>	<u>0.552 ± 0.035</u>

Footnote: The best results in each column are in **bold faces** and the second-best results are underlined.**Fig. 2.** The AUC, AUPR, and Recall@10,000 curves of 9 methods on our proposed benchmark dataset.**Fig. 3.** The AUC and AUPR scores of 6 methods on 5 partially removed datasets.

(LAGCN), while the statistical results also emphasized these remarkable improvements. Moreover, the Recall@10,000 value of REDDA is significantly outperforming those of 4 baseline methods (Top3 results: 35.93% in REDDA; 32.89% in LAGCN, and 32.25% in DRWBNCF).

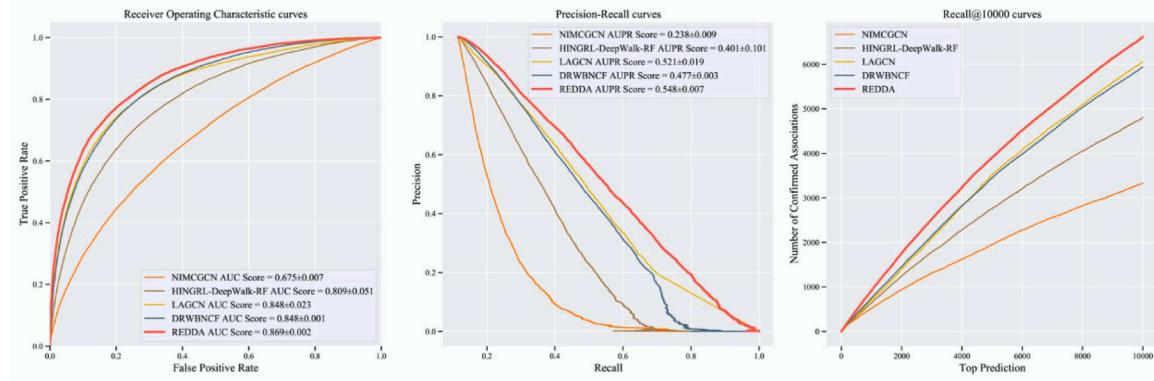
These inspiring results implied that the high-order representations of drug-disease associations can benefit from fusing biological relations (e.

., drug-protein associations and pathway-disease associations). Moreover, the way we utilized this information in REDDA (topological subnet embedding block and graph attention block) may also partially account for the considerable performance improvement.

Table 4

Performance comparison of 5 methods on B-dataset.

Methods	AUC	AUPR	F1-Score	Accuracy	Recall	Specificity	Precision
NIMGCN	0.675 ± 0.007	0.238 ± 0.009	0.295 ± 0.007	0.764 ± 0.012	0.430 ± 0.013	0.807 ± 0.015	0.224 ± 0.009
HINGRL-DeepWalk-RF	0.809 ± 0.051	0.401 ± 0.101	0.439 ± 0.067	0.843 ± 0.032	0.529 ± 0.040	0.883 ± 0.031	0.379 ± 0.080
LAGCN	0.848 ± 0.023	0.521 ± 0.019	0.506 ± 0.012	0.874 ± 0.004	0.564 ± 0.019	0.914 ± 0.004	0.459 ± 0.012
DRWBNCF	0.848 ± 0.001	0.477 ± 0.003	0.490 ± 0.002	0.868 ± 0.005	0.555 ± 0.021	0.908 ± 0.009	0.440 ± 0.015
REDDA	0.869 ± 0.002	0.548 ± 0.007	0.528 ± 0.004	0.884 ± 0.004	0.565 ± 0.013	0.925 ± 0.005	0.496 ± 0.013

Footnote: The best results in each column are in **bold faces** and the second-best results are underlined.**Fig. 4.** Performance comparison of 5 methods on B-dataset.

3.2. Ablation study

To demonstrate the importance and necessity of the architecture of REDDA, we designed a series of ablation experiments by proposing and testing some variants of REDDA that can be regarded as simplified REDDA. The details of these variants of REDDA are listed as follows:

- REDDA-Base: The REDDA only includes the node embedding block and bilinear dot decoder.
- REDDA-w/o BR: Basic REDDA architecture, but the input is a drug-disease heterogeneous network (without protein, gene, and pathway relations involved).
- REDDA-w/o TSE: The REDDA without the topological subnet embedding block.
- REDDA-w/o GAT: The REDDA without the graph attention block.
- REDDA-w/o LA: The REDDA without the layer attention block.

The performance comparison results of REDDA and its 5 simplified versions on our proposed benchmark dataset are listed in [Table 5](#). Compared to these simplified models, REDDA achieved the optimal AUC score, AUPR score, and the suboptimal F1-Score, which can be regarded as the best-performed model. Among the simplified models, the REDDA-Base and REDDA-w/o BR both under-performed than REDDA, indicating that a well-designed model structure and an integration of biological relations both can contribute to better performance in drug repositioning tasks; Compared REDDA-w/o TSE and REDDA-w/o GAT to REDDA-

Base, respectively, the performance results implied that the topological subnet embedding block and graph attention block both can benefit the model performance, which tended to efficiently utilize the biological information. The REDDA-w/o LA performed worst due to the deep neural network, which required the residual architecture like layer attention block to resolve potential overfitting. These results demonstrated the reasonability of the REDDA structure.

3.3. Attention visualization analysis

In REDDA, 3 attention mechanisms are adopted to enhance the representation: subnet attention, layer attention, and graph attention mechanisms. In this section, we used attention visualization analysis to discuss these attention mechanisms.

3.3.1. Topological subnet attention

Subnet Attention is used for weighted aggregating the representations from different topological subnets. The auto-learned attention coefficients in REDDA are shown in [Fig. 5A](#). The results showed that each topological subnet played a role in generating the node representations. Particularly, the integration of biological relations such as \mathcal{G}^{R-P} and \mathcal{G}^{M-D} , had profound impacts on learning drug and disease representations in topological subnet embedding block, which got attention coefficients rather than 0.6. It also cannot be ignored that the attention coefficients of topological subnets for the same node are different but not extreme (compared to the extremely unbalanced number of associations such as drug-disease and drug-protein associations), which tended to show that the topological subnet embedding block can partly prevent the generation of biased node representations.

3.3.2. Graph attention

Graph attention coefficients represent the importance of the edges which are connected to a certain node and other neighbor nodes. We plotted the distributions of attention coefficient of each edge type for each node type (shown in [Fig. 5B](#)), which indicated that the graph attention block aggregated the node representations with varied attention coefficients. Moreover, in drug and disease nodes representations,

Table 5
Performance comparison of REDDA and 5 simplified methods.

Methods	AUC	AUPR	F1-Score
REDDA-Base	0.887 ± 0.004	0.357 ± 0.000	0.421 ± 0.004
REDDA-w/o BR	0.914 ± 0.000	0.425 ± 0.002	0.475 ± 0.002
REDDA-w/o TSE	0.889 ± 0.001	0.359 ± 0.016	0.418 ± 0.010
REDDA-w/o GAT	0.893 ± 0.001	0.441 ± 0.000	0.500 ± 0.002
REDDA-w/o LA	0.713 ± 0.007	0.021 ± 0.003	0.042 ± 0.007
REDDA	0.922 ± 0.003	0.444 ± 0.008	0.495 ± 0.009

Footnote: The best results in each column are in **bold faces** and the second-best results are underlined.

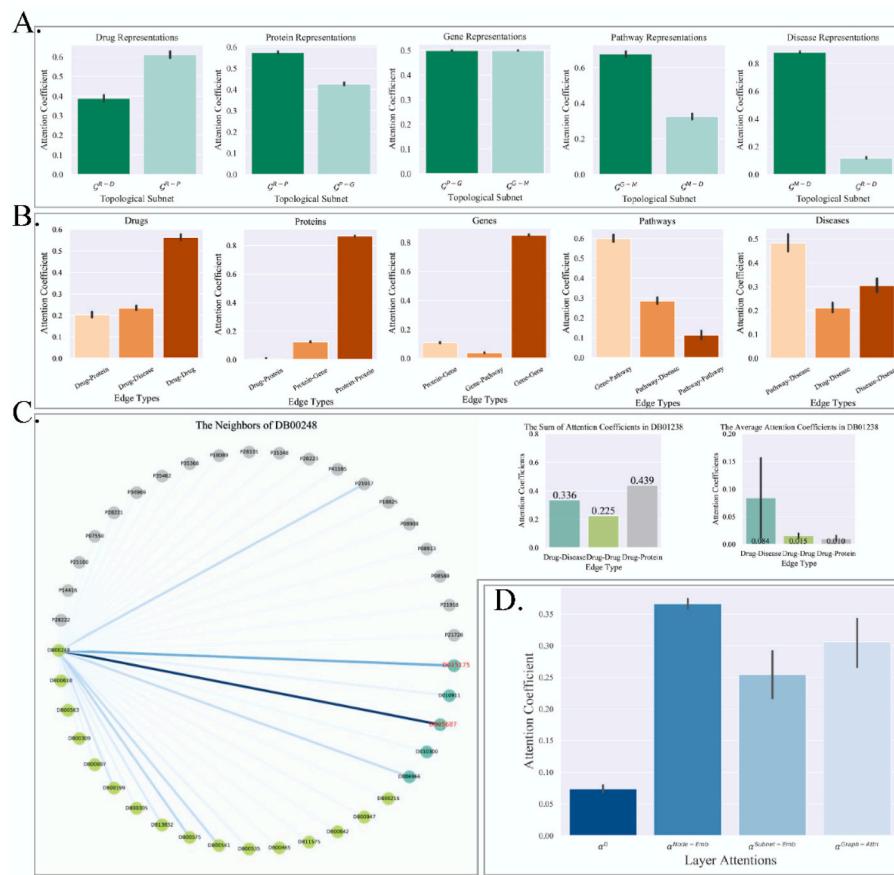


Fig. 5. Attention visualization analysis of REDDA. A. Attention coefficients for 5 topological subnets in the node representations in topological subnet embedding block. B. Attention coefficients for 10 edge types in the node representations in graph attention block. C. Attention coefficients for the neighbors of node DB00248 in graph attention block. D. Attention coefficients for the node representations from 4 blocks in REDDA.

the drug-protein associations and pathway-disease associations also occupied portions of attention coefficients, which indicated the integration of extra biological nodes and relations provided additional valid information for the node representation learning.

We also selected a drug node DB00248 (generic name: cabergoline) as a case and visualized the attention coefficients of its edges, which are shown in Fig. 5C. In the neighbors of DB00248 which is used for the treatment of hyperprolactinemia, REDDA gave more attention to the fewer associated disease neighbors. Moreover, REDDA can also recognize the most related nodes to DB00248, such as the target diseases galactorrhea (MeSH ID: D005687) and prolactinoma (MeSH ID: D015175).

3.3.3. Layer attention

We adopted the residual structure by aggregating representations with a layer attention mechanism. The attention coefficients represented the importance contributions of different blocks to the final drug/disease representations (shown in Fig. 5D). The results showed that the proposed computational blocks in REDDA all occupied portions of attention coefficients in the layer attention block. Compared to the low-order ones, the high-order representations in the topological subnet embedding block and graph attention block tended to focus on different views of nodes and provided considerable contributions to the final node representations (the sum of attention coefficient ≈ 0.55).

The above all attention visualizations demonstrated that REDDA can selectively identify the important information in the heterogeneous network for drug-disease association predictions and generated node representations by weighted aggregations on the subnet/graph/layer levels with varied attentions.

3.4. Case study

To demonstrate the ability of REDDA for discovering new indications and new therapies, we trained a REDDA model using all known drug-disease associations and predicted the unknown drug-disease associations for known drugs/diseases. As for verifications, we used two reliable public databases (ClinicalTrials and CTD) to emphasize the accuracy of our predictions. Moreover, the experimental results reported in references were also additional evidence to verify the ability to discover new indications/therapies of MOODA.

3.4.1. Discovery of new indications

We selected two drugs to investigate the ability of REDDA in discovering new indications. Except for the known drug-disease associations, the Top10 positive predictions of these drugs are listed in [Table 6](#), while the details of predicted associated drugs are listed in [Table S6](#). Among the tested drugs, Ifosfamide (DrugBank ID: DB01181) is an alkylating and immunosuppressive agent which is widely used as a chemotherapeutics for tumor treatment, including the treatment of non-Hodgkin's lymphoma, small cell lung cell, etc. Among the Top10 positive predictions of REDDA, 9 of the associated diseases are tumors and have been confirmed by databases. Dexamethasone (DrugBank ID: DB01234) is a glucocorticoid that is used for the treatment of inflammatory conditions. In [Tables 6](#) and [8](#) of the Top10 associated diseases have been confirmed by reliable sources or clinical trials. These pieces of evidence indicate that REDDA can learn from limited biological information and identify confirmed indications that not have been collected in the training dataset.

Furthermore, as for the discovery of new potential indications, one of

Table 6

The Top10 REDDA-predicted new indications.

Drug	Disease	MeSH ID	Evidence (Database)	Evidence (PMID)
Dexamethasone	Nasal Polyps	D009298	CTD	24917907
	Inflammatory	D015212	ClinicalTrials/	29109767
	Bowel Diseases		CTD	
	Rhinitis, Allergic	D065631	CTD	17762268
	Spondylarthritis	D025241	CTD	N/A
	Otitis Media	D010033	ClinicalTrials	24093464
	Erythema	D004892	N/A	N/A
	Multiforme			
	Pityriasis Rosea	D017515	N/A	N/A
	Adrenal	D000312	ClinicalTrials/	14764770
	Hyperplasia,		CTD	
	Congenital			
Ifosfamide	Brain Neoplasms	D001932	ClinicalTrials/	31346902
			CTD	
	Crohn Disease	D003424	ClinicalTrials/	17635367
			CTD	
	Mycosis	D009182	ClinicalTrials/	3130316
	Fungoides		CTD	
	Colitis, Ulcerative	D003093	CTD	N/A
	Lymphoma,	D008224	ClinicalTrials/	12736225
	Follicular		CTD	
	Lymphoma, T-Cell, Peripheral	D016411	ClinicalTrials/	33147935
	Lymphoma, Mantle-Cell	D020522	ClinicalTrials	20038221
	Tuberculosis	D014376	N/A	N/A
	Burkitt	D002051	ClinicalTrials/	12181251
	Lymphoma		CTD	
	Hepatitis C	D006526	CTD	N/A
	Hyperthyroidism	D006980	CTD	N/A
	Carcinoma, Non-Small-Cell Lung	D002289	ClinicalTrials/	10582135
			CTD	

the predicted associated diseases of dexamethasone, erythema multiforme, is non-confirmed by databases. However, related studies have mentioned that two rare skin diseases, called Stevens-Johnson Syndrome (SJS) and Toxic Epidermal Necrolysis (TEN), are recognized as the descendants of erythema multiforme in MeSH categories, can be treated by some corticosteroids including dexamethasone [58,59]. Meanwhile, associated with tissue damage and necroptosis, the Annexin A1 receptor which is mediated by dexamethasone and exists in our heterogeneous network, is also a candidate marker in SJS/TEN reported by a previous study [60]. Overall, these indirect shreds of evidence enhance the reliability of the new unconfirmed potential indication discovered by REDDA.

Table 7

The Top10 REDDA-predicted new therapies.

Disease	Drug	DrugBank ID	Evidence (Database)	Evidence (PMID)
Breast Neoplasms	Zoledronic acid	DB00399	ClinicalTrials/CTD	29082518
	Norethisterone	DB00717	N/A	N/A
	Estrone	DB00655	CTD	12796390
	Leuprolide	DB00007	ClinicalTrials/CTD	29747931
	Testosterone cypionate	DB13943	ClinicalTrials	26160683
	Metformin	DB00331	ClinicalTrials/CTD	31164151
	Gonadorelin	DB00644	N/A	32006118
	Prednisone	DB00635	ClinicalTrials/CTD	27052658
	Sirolimus	DB00877	ClinicalTrials/CTD	32335491
	Risedronate acid	DB00884	ClinicalTrials/CTD	25792492
	Everolimus	DB01590	ClinicalTrials/CTD	34224367
	Flurandrenolide	DB00846	N/A	N/A
Brain Neoplasms	Ranimustine	DB13832	N/A	28214639
	Mitoxantrone	DB01204	CTD	12450040
	Betamethasone	DB00443	CTD	869982
	Difluprednate	DB06781	N/A	N/A
	Etretinate	DB00926	N/A	N/A
	Pacitaxel	DB01229	ClinicalTrials/CTD	21692650
	Clobetasol propionate	DB01013	N/A	N/A
	Hydrocortisone	DB00741	ClinicalTrials/CTD	34214336

3.4.2. Discovery of new therapies

We selected two diseases (breast neoplasms and brain neoplasms) to investigate the ability of REDDA in discovering new therapies. Same as the details in the 3.4.1 Section, the Top10 positive predictions results are listed in Table 7 and the details are listed in Table S8. In the Top10 results of breast neoplasms (MeSH ID: D001943), 8 of the Top10 predicted drug-disease associations are experimentally confirmed by databases. Among them, 5 of these associated drugs are hormones (norethisterone, estrone, leuprolide, testosterone cypionate, and gonadorelin), which are highly related to breast neoplasms. Focusing on the results of brain neoplasms (MeSH ID: D001932), 5 of the Top10 results are confirmed by databases. Among them, 4 associated drugs are anti-neoplastic agents and 5 are corticosteroids. These pieces of evidence demonstrate that our proposed REDDA can also predict the confirmed therapies for existing diseases. Meanwhile, comparing the associated drugs of breast neoplasms to those of brain neoplasms, it was inspiring to observe that REDDA tended to selectively generate the prediction based on the characteristic of the disease. These results demonstrate that REDDA has the potential to identify therapies by knowledge-aware prediction.

For the discovery of new therapies which are not confirmed by databases, the predicted gonadorelin for breast neoplasms is a synthetic gonadotropin-releasing hormone (GnRH) and also a GnRH agonist (GnRHa). Some previous studies indicated that GnRHa can be used for the treatment of early breast neoplasms, infertility caused by breast neoplasms and metastatic male breast neoplasms [61–63], etc. These shreds of evidence emphasize the capacity of gonadorelin for the treatment of breast neoplasms, which have not been mentioned by previous studies or related databases. Moreover, a predicted drug for brain neoplasms, ranimustine has been used in the treatment of chronic myelogenous leukemia, Peripheral T-cell lymphoma, and glioblastoma multiforme [64–66]. Meanwhile, the Top3 drugs which are the most similar to ranimustine in our heterogeneous network (carmustine, lomustine, and streptozocin) can all be used for the treatment of brain neoplasms, which may account for the high association probability of ranimustine predicted by REDDA. Other 2 predicted associated drugs (clobetasol propionate and etretinate) for brain neoplasms are also confirmed for the treatment of other cancers or cancer-related complications [67,68], which provided potentials for the therapies of brain neoplasms.

In a nutshell, the case studies prove that REDDA can learn from the large-scale heterogeneous network to recognize the drug-disease associations which are unknown in the training set but confirmed in the real-world. Meanwhile, REDDA can also give confident predictions to some

unconfirmed but potential drug-disease associations, implying that REDDA has a noticeable ability for discovering new indications/therapies for existing drugs/diseases of REDDA.

4. Conclusion

In this study, we propose a benchmark dataset for drug-disease association prediction. Much larger than a single drug-disease heterogeneous network, the constructed comprehensive heterogeneous network contains 5 biological entities and 10 relations. Moreover, to enhance the effectiveness of these extra biological relations on improving the performance of computational drug repositioning, a graph learning method (REDDA) on a heterogeneous network is proposed for predicting drug-disease associations, which is designed and constructed on the basis of HeteroGCNs, attention mechanisms, and topological disintegrations. Unlike to other methods which can only be used in isolated drug-disease heterogeneous network, or calculating multi-view similarities using extra biological relationships, our proposed REDDA can be directly trained on a heterogeneous network with abundant biological relations, which can enhance the representations of drugs and diseases, and decrease the secondary transformation loss of biological information. Extensive experiments emphasized the effectiveness of the fused extra biological information and the superiority of REDDA compared to various baseline methods.

Although REDDA has achieved desirable performance. Some limitations still need to be in-depth considered. Firstly, more biological associations should be further concerned which are inaccessible in our heterogeneous network, such as the protein-disease associations, drug-miRNA associations, miRNA-disease associations, etc. Secondly, for the model structure, more advanced backbone heterogeneous neural networks should be involved in REDDA where we only use the classical HeteroGCN in this version. In the future, we will concentrate on integrating larger benchmark datasets with multiple data sources to spur the isolated drug repositioning to large-scale biological network aided drug repositioning. Meanwhile, the model interpretability in the computational drug repositioning method is another concern of us.

In conclusion, our proposed REDDA can enhance the ability to predict drug-disease associations by using external biological information and dynamically aggregating the drug/disease representations. With a promising capacity, it may be used in computational drug repositioning for accelerating the discovery of positive entities and reducing the cost of time-consuming wet-lab experiments.

Funding

This work was supported by Chinese Academy of Medical Sciences (Grant No. 2021-I2M-1-056), Fundamental Research Funds for the Central Universities (Grant No. 3332022144), National Key R&D Program of China (Grant No. 2016YFC0901901 and Grant No. 2017YFC0907503) and the National Natural Science Foundation of China (Grant No. 81601573).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank all anonymous reviewers for their constructive advice.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2022.106127>.

[org/10.1016/j.combiomed.2022.106127](https://doi.org/10.1016/j.combiomed.2022.106127).

References

- [1] H.C.S. Chan, H. Shan, T. Dahoun, et al., Advancing drug discovery via artificial intelligence, *Trends Pharmacol. Sci.* 40 (2019) 592–604.
- [2] V. Prasad, S. Mailankody, Research and development spending to bring a single cancer drug to market and revenues after approval, *JAMA Intern. Med.* 177 (2017) 1569–1575.
- [3] J.A. DiMasi, H.G. Grabowski, R.W. Hansen, Innovation in the pharmaceutical industry: new estimates of R&D costs, *J. Health Econ.* 47 (2016) 20–33.
- [4] C.H. Wong, K.W. Siah, A.W. Lo, Estimation of clinical trial success rates and related parameters, *Biostatistics* 20 (2019) 273–286.
- [5] J. Vamathevan, D. Clark, P. Czodrowski, et al., Applications of machine learning in drug discovery and development, *Nat. Rev. Drug Discov.* 18 (2019) 463–477.
- [6] K. Yang, X. Zhao, D. Waxman, et al., Predicting drug-disease associations with heterogeneous network embedding, *Chaos* 29 (2019), 123109.
- [7] Y. Gu, S. Zheng, J. Li, CurrMG: a curriculum learning approach for graph based molecular property prediction, in: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021, pp. 2686–2693.
- [8] Q. Ye, C.Y. Hsieh, Z. Yang, et al., A unified drug-target interaction prediction framework based on knowledge graph and recommendation system, *Nat. Commun.* 12 (2021) 6775.
- [9] Y. Gu, S. Zheng, Z. Xu, et al., An efficient curriculum learning-based strategy for molecular graph learning, *Briefings Bioinf.* 23 (2022) bbac099.
- [10] W. Kong, X. Tu, W. Huang, et al., Prediction and optimization of NaV1.7 sodium channel inhibitors based on machine learning and simulated annealing, *J. Chem. Inf. Model.* 60 (2020) 2739–2753.
- [11] T. Li, X. Zhao, L. Li, Co-VAE: drug-target binding affinity prediction by co-regularized variational autoencoders, *IEEE Trans. Pattern Anal. Mach. Intell.* (2021) 1, 1.
- [12] Q. Liu, Z. Hu, R. Jiang, et al., DeepCDR: a hybrid graph convolutional network for predicting cancer drug response, *Bioinformatics* 36 (2020) i911–i918.
- [13] F. Zhang, M. Wang, J. Xi, et al., A novel heterogeneous network-based method for drug response prediction in cancer cell lines, *Sci. Rep.* 8 (2018) 3355.
- [14] J. Li, S. Zheng, B. Chen, et al., A survey of current trends in computational drug repositioning, *Briefings Bioinf.* 17 (2016) 2–12.
- [15] H. Xue, J. Li, H. Xie, et al., Review of drug repositioning approaches and resources, *Int. J. Biol. Sci.* 14 (2018) 1232–1244.
- [16] K. Mohamed, N. Yazdanpanah, A. Saghazadeh, et al., Computational drug discovery and repurposing for the treatment of COVID-19: a systematic review, *Bioorg. Chem.* 106 (2021), 104490.
- [17] G. Fahimian, J. Zahiri, S.S. Arab, et al., RepCOOL: computational drug repositioning via integrating heterogeneous biological networks, *J. Transl. Med.* 18 (2020) 375.
- [18] J.I. Traylor, H.E. Sheppard, V. Ravikumar, et al., Computational drug repositioning identifies potentially active therapies for chordoma, *Neurosurgery* 88 (2021) 428–436.
- [19] L. Bai, M.K.D. Scott, E. Steinberg, et al., Computational drug repositioning of atorvastatin for ulcerative colitis, *J. Am. Med. Inf. Assoc.* 28 (2021) 2325–2335.
- [20] C. Budak, V. Mençik, V. Gider, Determining similarities of COVID-19 - lung cancer drugs and affinity binding mode analysis by graph neural network-based GEFA method, *J. Biomol. Struct. Dyn.* (2021) 1–13.
- [21] H. Luo, M. Li, M. Yang, et al., Biomedical data and computational models for drug repositioning: a comprehensive review, *Briefings Bioinf.* 22 (2021) 1604–1619.
- [22] C.-Q. Gao, Y.-K. Zhou, X.-H. Xin, et al., DDA-SKF, Predicting drug-disease associations using similarity kernel fusion, *Front. Pharmacol.* 12 (2022) 784171, 784171.
- [23] H. Luo, J. Wang, M. Li, et al., Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm, *Bioinformatics* 32 (2016) 2664–2671.
- [24] W. Zhang, X. Yue, W. Lin, et al., Predicting drug-disease associations by using similarity constrained matrix factorization, *BMC Bioinf.* 19 (2018) 233.
- [25] M. Yang, H. Luo, Y. Li, et al., Drug repositioning based on bounded nuclear norm regularization, *Bioinformatics* 35 (2019) i455–i463.
- [26] M. Yang, G. Wu, Q. Zhao, et al., Computational drug repositioning based on multi-similarities bilinear matrix factorization, *Briefings Bioinf.* 22 (2021) bbaa267.
- [27] Q. Cao, Z. Zhang, A.X. Fu, et al., A unified framework for integrative study of heterogeneous gene regulatory mechanisms, *Nat. Mach. Intell.* 2 (2020) 447–456.
- [28] W. Zeng, J. Xin, R. Jiang, et al., Reusability report: compressing regulatory networks to vectors for interpreting gene expression and genetic variants, *Nat. Mach. Intell.* 3 (2021) 576–580.
- [29] Q. Liu, S. Chen, R. Jiang, et al., Simultaneous deep generative modeling and clustering of single cell genomic data, *Nat. Mach. Intell.* 3 (2021) 536–544.
- [30] X. Chen, S. Chen, S. Song, et al., Cell type annotation of single-cell chromatin accessibility data via supervised Bayesian embedding, *Nat. Mach. Intell.* 4 (2022) 116–126.
- [31] P. Zeng, J. Wangwu, Z. Lin, Coupled co-clustering-based unsupervised transfer learning for the integrative analysis of single-cell genomic data, *Briefings Bioinf.* (2021) 22.
- [32] X. Huang, Y. Huang, Cellsnp-lite: an efficient tool for genotyping single cells, *Bioinformatics* 37 (23) (2021) 4569–4571.
- [33] J. Zhu, J. Wang, X. Wang, et al., Prediction of drug efficacy from transcriptional profiles with deep learning, *Nat. Biotechnol.* 39 (2021) 1444–1452.

- [34] Z. Yu, F. Huang, X. Zhao, et al., Predicting drug-disease associations through layer attention graph convolutional network, *Briefings Bioinf.* (2021) 22.
- [35] L. Cai, C. Lu, J. Xu, et al., Drug repositioning based on the heterogeneous information fusion graph convolutional network, *Briefings Bioinf.* (2021) 22.
- [36] Y. Meng, C. Lu, M. Jin, et al., A weighted bilinear neural collaborative filtering approach for drug repositioning, *Briefings Bioinf.* 23 (2) (2022), bbab581.
- [37] X. Zeng, S. Zhu, X. Liu, et al., deepDR: a network-based deep learning approach to *in silico* drug repositioning, *Bioinformatics* 35 (2019) 5191–5198.
- [38] P. Xuan, Y. Ye, T. Zhang, et al., Convolutional Neural Network and Bidirectional Long Short-Term Memory-Based Method for Predicting, vol. 8, *Drug-Disease Associations*, 2019. Cells.
- [39] H. Liu, W. Zhang, Y. Song, et al., HNet-DNN: inferring new drug-disease associations with deep neural network based on heterogeneous network features, *J. Chem. Inf. Model.* 60 (2020) 2367–2376.
- [40] P. Xuan, L. Gao, N. Sheng, et al., Graph convolutional autoencoder and fully-connected autoencoder with attention mechanism based method for predicting drug-disease associations, *IEEE J Biomed Health Inform* 25 (2021) 1793–1804.
- [41] M. Coşkun, M. Koyutürk, Node similarity based graph convolution for link prediction in biological networks, *Bioinformatics* 37 (2021) 4501–4508.
- [42] H. Zhang, H. Cui, T. Zhang, et al., Learning multi-scale heterogenous network topologies and various pairwise attributes for drug-disease association prediction, *Briefings Bioinf.* 23 (2022) bbac009.
- [43] P. Xuan, X. Meng, L. Gao, et al., Heterogeneous multi-scale neighbor topologies enhanced drug-disease association prediction, *Briefings Bioinf.* 23 (2022) bbac123.
- [44] B.W. Zhao, L. Hu, Z.H. You, et al., HINGRL: predicting drug-disease associations with graph representation learning on heterogeneous information networks, *Briefings Bioinf.* 23 (2022).
- [45] J. Huang, J. Chen, B. Zhang, et al., Evaluation of gene-drug common module identification methods using pharmacogenomics data, *Briefings Bioinf.* (2021) 22.
- [46] J. Wang, Z. Wu, Y. Peng, et al., Pathway-based drug repurposing with DPNetinfer: a method to predict drug-pathway associations via network-based approaches, *J. Chem. Inf. Model.* 61 (2021) 2475–2485.
- [47] A. Gottlieb, G.Y. Stein, E. Ruppin, et al., PREDICT: a method for inferring novel drug indications with application to personalized medicine, *Mol. Syst. Biol.* 7 (2011) 496.
- [48] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* 28 (2000) 27–30.
- [49] D.S. Wishart, C. Knox, A.C. Guo, et al., DrugBank: a knowledgebase for drugs, drug actions and drug targets, *Nucleic Acids Res.* 36 (2008) D901–D906.
- [50] A.P. Davis, C.J. Grondin, R.J. Johnson, et al., Comparative toxicogenomics database (CTD): update 2021, *Nucleic Acids Res.* 49 (2021) D1138–d1143.
- [51] C. von Mering, M. Huynen, D. Jaeggi, et al., STRING: a database of predicted functional associations between proteins, *Nucleic Acids Res.* 31 (2003) 258–261.
- [52] UniProt: the universal protein knowledgebase, *Nucleic Acids Res.* 45 (2017) D158–d169.
- [53] D. Rogers, M. Hahn, Extended-connectivity fingerprints, *J. Chem. Inf. Model.* 50 (2010) 742–754.
- [54] D. Wang, J. Wang, M. Lu, et al., Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases, *Bioinformatics* 26 (2010) 1644–1650.
- [55] X. Wang, H. Ji, C. Shi, et al., Heterogeneous graph attention network, in: *The World Wide Web Conference*, 2019, pp. 2022–2032.
- [56] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256. JMLR Workshop and Conference Proceedings.
- [57] J. Li, S. Zhang, T. Liu, et al., Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction, *Bioinformatics* 36 (2020) 2538–2546.
- [58] B.R. Del Pozzo-Magana, A. Lazo-Langner, B. Carleton, et al., A systematic review of treatment of drug-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in children, *J Popul Ther Clin Pharmacol* 18 (2011) e121–e133.
- [59] S.H. Kardaun, M.F. Jonkman, Dexamethasone pulse therapy for Stevens-Johnson syndrome/toxic epidermal necrolysis, *Acta Derm. Venereol.* 87 (2007) 144–148.
- [60] R. Abe, Immunological response in Stevens-Johnson syndrome and toxic epidermal necrolysis, *J. Dermatol.* 42 (2015) 42–48.
- [61] N.A.J. Khan, M. Tirona, An updated review of epidemiology, risk factors, and management of male breast cancer, *Med. Oncol.* 38 (2021) 39.
- [62] E. Silvestris, M. Dellino, P. Cafforio, et al., Breast cancer: an update on treatment-related infertility, *J. Cancer Res. Clin. Oncol.* 146 (2020) 647–657.
- [63] D.A. Beyer, F. Amari, M. Thill, et al., Emerging gonadotropin-releasing hormone agonists, *Expert Opin. Emerg. Drugs* 16 (2011) 323–340.
- [64] M. Gkotzamanidou, C.A. Papadimitriou, Peripheral T-cell lymphoma: the role of hematopoietic stem cell transplantation, *Crit. Rev. Oncol. Hematol.* 89 (2014) 248–261.
- [65] M. Kameda, Y. Otani, T. Ichikawa, et al., Congenital glioblastoma with distinct clinical and molecular characteristics: case reports and a literature review, *World Neurosurg* 101 (2017) 817, e815–817.e814.
- [66] N. Kusaba, H. Yoshida, F. Ohkubo, et al., [Granulocyte-colony stimulating factor-producing myeloma with clinical manifestations mimicking chronic neutrophilic leukemia], *Rinsho Ketsueki* 45 (2004) 228–232.
- [67] A.G. Venkat, S. Areppalli, S. Sharma, et al., Local therapy for cancer therapy-associated uveitis: a case series and review of the literature, *Br. J. Ophthalmol.* 104 (2020) 703.
- [68] K. Yonekura, K. Takeda, N. Kawakami, et al., Therapeutic efficacy of etretinate on cutaneous-type Adult T-cell leukemia-lymphoma, *Acta Derm. Venereol.* 99 (2019) 774–776.