OXFORD

*Briefings in Bioinformatics*, 2023, 1–17

https://doi.org/10.1093/bib/bbac578

**Problem Solving Protocol**

用于药物–靶标相互作用预测的元路径聚合异构图神经网络

# Metapath-aggregated heterogeneous graph neural network for drug–target interaction prediction

Mei Li, Xiangrui Cai, Sihan Xu and Hua Ji

Corresponding author. Sihan Xu, 38 Tongyan Road, Jinnan District, Tianjin 300350, P.R. China. E-mail: xusihan@nankai.edu.cn

## Abstract

Drug–target interaction (DTI) prediction is an essential step in drug repositioning. A few graph neural network (GNN)-based methods have been proposed for DTI prediction using heterogeneous biological data. However, existing GNN-based methods only aggregate information from directly connected nodes restricted in a drug-related or a target-related network and are incapable of capturing high-order dependencies in the biological heterogeneous graph. In this paper, we propose a metapath-aggregated heterogeneous graph neural network (MHGNN) to capture complex structures and rich semantics in the biological heterogeneous graph for DTI prediction. Specifically, MHGNN enhances heterogeneous graph structure learning and high-order semantics learning by modeling high-order relations via metapaths. Additionally, MHGNN enriches high-order correlations between drug-target pairs (DTPs) by constructing a DTP correlation graph with DTPs as nodes. We conduct extensive experiments on three biological heterogeneous datasets. MHGNN favorably surpasses 17 state-of-the-art methods over 6 evaluation metrics, which verifies its efficacy for DTI prediction. The code is available at https://github.com/Zora-LM/MHGNN-DTI.

**Keywords:** Drug–target interaction prediction, heterogeneous graph, graph neural network, metapath

## Introduction

Drug–target interaction (DTI) prediction can significantly accelerate drug repositioning where around 75% drugs can be repositioned [1]. Traditional experimental identification of DTIs is time-consuming and costly. Recently, machine learning (ML)-based approaches for DTI prediction have gained a lot of attention in academia and industry. However, most existing works only employ chemical and genomic data for DTI prediction [2–5], while it neglect pharmacological and phenotypic information, such as diseases and side-effects [6].

The interaction of a drug with one or more targets can generate multiple therapeutic effects or may produce unexpected side-effects [6]. These drug actions generally reflect the binding activities of drugs to targets. Specifically, a drug is initially designed and optimized mostly for the main indication as a particular disease of interest is the starting point of drug discovery, whereas a drug is likely to bind to multiple proteins (so called off-targets), which can thus produce a variety of unexpected new therapeutic indications. For example, in the tragic case of Thalidomide, its strong antiangiogenic activity turns out to be useful for the treatment of multiple myeloma [7]. Besides, targets themselves are involved in multiple biological processes and relevant for other diseases, such as, tyrosine kinase ABL as the target for both Parkinson disease and cancers. Additionally, a drug would extend its use from the original to closely related diseases. For instance, antiangiogenic antibody bevacizumab expands its use from colon cancer to other solid cancers. The interrelations between bioentities (e.g. drugs, targets, diseases and side-effects) contain rich semantic information and offer a system-level understanding about DTIs. Thus, incorporating heterogeneous biological data can potentially contribute to DTI prediction and further benefit drug repositioning.

The complex interrelations between bioentities can be naturally formulated as multiple drug- and target-related networks (Figure 1A) or a heterogeneous graph (Figure 1B), where nodes are bioentities, i.e. drugs (D), targets (T), diseases (I) and side-effects (S), and edges are interactions, associations or similarities between these entities, for instance, target–target associations. DTI prediction is formulated as a link prediction problem (red dotted line in Figure 1B), that is whether there exists a link between a drug node and a target node. Motivated by the superiority of graph neural networks (GNNs) in graph analysis [8, 9], several GNN-based methods [10–14] have been proposed for DTI prediction using heterogeneous biological data, whereas they suffer from the following two limitations. Firstly, present works [10, 12–14] consider drug- and target-related networks separately (see Figure 1A) and thus neglect the diverse semantic relations across networks, e.g. *drug–target–disease*. Secondly, existing works, e.g. NeoDTI [10], are based on traditional GNNs, such as GCN [15] and GAT [16], while these GNNs are designed for homogenous graphs

**Mei Li** is currently a Ph.D. student in College of Computer Science, Nankai University, Tianjin, China. Her research interests include graph deep learning, machine learning and bioinformatics.

**Xiangrui Cai** received his Ph.D. degree in computer science of Nankai University, China in 2018. From June 2019, He is a lecturer in College of Computer Science, Nankai University, China. His research interests include medical data analysis, natural language processing and trusted AI.

**Sihan Xu** is an assistant professor in College of Cyber Science, Nankai University, Tianjin, China. Her research interests include AI security, deep learning testing, and software analysis.

**Hua Ji** is a professor of the Department of Computer Sience and the director of Trusted-AI Lab at the Nankai University, Tianjin, China. He received a Ph.D in Computer Science from Nanjng University, China in 1997. His research interest include safety system, trusted AI, formal methods and operating system.

**Received:** July 1, 2022. **Revised:** November 3, 2022. **Accepted:** November 26, 2022

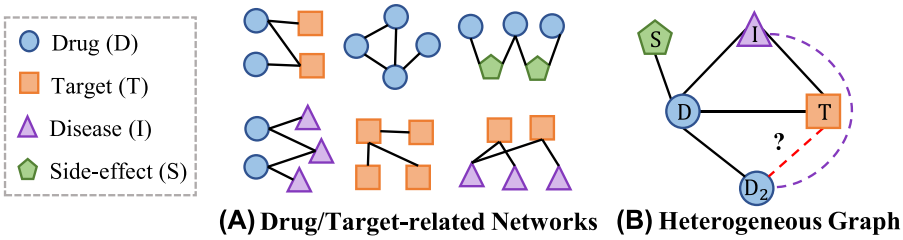


**Figure 1. (A)** Drug- and target-related networks. **(B)** Biological heterogeneous graph. Red dotted line denotes the link prediction between $D_2$ and $T$. Existing GNN-based methods for DTI prediction are incapable of capturing high-order dependencies in the biological heterogeneous graph, for instance, the information between $D_2$ and $I$ (purple dotted line). MHGNN enhances graph structure learning by modeling high-order relations via metapaths ($D_2DI$ and $D_2DTI$).

and restricted to immediate neighbor aggregation. Thus, existing methods obtain suboptimal bioentity representations with limited semantic information.

The heterogeneous GNNs have been intensively studied in various fields, including recommendation systems [17], and node classification and clustering [18–20]. Nevertheless, the effectiveness of heterogeneous GNNs on DTI prediction task has not been fully investigated. *Metapaths* are a widely used technique in heterogeneous graph embedding. A metapath describes a composite relation between two nodes and reveals structural correlations and semantics in a heterogeneous graph [21]. For example, the metapath $drug \xrightarrow{binds\ to} target \xrightarrow{causes} disease \xrightarrow{treated\ by} drug$ describes four semantic roles and three kinds of relationships among bioentities, which provides rich context information for bioentity representation learning. Besides, metapaths construct high-order relationships between a starting node and an ending node and specify biological semantic relations in certain substructures, i.e., *metagraphs*. For example, a drug and a target can be connected through different metapaths: (a) $drug \xrightarrow{similar\ to} drug \xrightarrow{binds\ to} target$, (b) $drug \xrightarrow{binds\ to} target \xrightarrow{causes} disease \xrightarrow{associates\ to} target$. Metapaths (a) indicates that a drug tends to bind to a target to which other structurally similar drugs bind. Metapath (b) indicates that a drug associates with a target that can cause similar diseases with the target to which the drug binds. Thus, we can capture high-order relationships with metagraphs.

In this paper, we propose a metapath-aggregated heterogeneous graph neural network (**MHGNN**) to simultaneously capture high-order dependencies in the biological heterogeneous graph and high-order associations between drug-target pairs (DTPs) for DTI prediction. Specifically, we design MHGNN with two major components, representation learning module and DTI prediction module. The representation learning module is devised as dual channels to learn drug representations and target representations, respectively. We first apply node type-specific transformations to project bioentity features to the same feature dimension, which eliminates heterogeneity originated from raw data. Then, in each channel, we learn drug (target) representations equipped with rich semantic information from each metapath using the GAT [16]. Moreover, in the DTI prediction module, we construct a DTP correlation graph with DTPs as nodes and apply two GCN layers to further explore the high-order relationships between DTPs. The key idea is that DTPs with common (similar) drugs or targets are strongly associated. The adjacency matrix of the constructed DTP correlation graph is mined from the learned drug representations and target representations. Each element of the adjacency matrix reflects the similarity between DTPs.

从学习到的药物表示和靶标表示中挖掘构建的DTP相关图的邻接矩阵。

The contributions of our work are summarized as follows.

- We propose a novel metapath-aggregated GNN to explore both high-order dependencies in the biological heterogeneous graph and high-order associations between DTPs for DTI prediction. Specifically, we construct a dual-channel GNN model, MHGNN, to learn drug and target representations. We further establish a DTP correlation graph to make the most of the potential relations between DTPs.
- To evaluate MHGNN, we extend a well-known biological heterogeneous dataset [22] by collecting DTPs from the latest released resources. The extended dataset contains newly discovered DTIs, providing more samples for training and evaluation. The dataset is available at https://github.com/Zora-LM/MHGNN-DTI.
- We conduct extensive experiments on both the public and extended datasets, comparing MHGNN with 17 state-of-the-art methods. The results show that MHGNN outperforms state-of-the-art baseline by a good margin on DTI prediction over six evaluation metrics.

## Related work

In this section, we present works that are closely related to this work, including DTI prediction and GNN.

### DTI prediction

The DTI prediction is often modeled as a binary classification problem of predicting whether the interaction exists for a DTP. Nevertheless, in reality, labeled DTPs are very limited and quite expensive to obtain, which restricts models from learning comprehensive patterns of DTPs. Heterogeneous biological data provide a multi-perspective view for modeling relationships of drugs and targets. Existing works for DTI prediction using biological heterogeneous data can be generally categorized into similarity-based methods, knowledge graph (KG)-based methods and GNN-based methods.

The similarity-based methods [22–25] hold a basic assumption of 'guilt-by-association', namely, similar drugs tend to bind with similar targets, and vice versa. They perform DTI prediction in a three-step strategy of (i) drug- and target-related similarity matrix computation using Jaccard distance, (ii) drug and target representation learning via ML-based models, e.g. random walk and autoencoder (AE), and (iii) DTI prediction using multiple multilayer perceptrons, inductive matrix completion or decision trees. These methods are deficient in modeling complex data correlations and neglect the fact that chemically dissimilar drugs can still bind to the same target [6]. Additionally, they separate feature learning from the prediction task, resulting in suboptimal

化学性质不同的药物仍然可以与相同的靶标结合

solutions for DTI prediction as the learned representations of drugs and targets may deviate from the prediction task.

The KG-based techniques associate bioentity node relations with shallow embeddings based on knowledge graph embedding models, e.g. TransE [26], DistMult [27] and ComplEx [28]. For instance, Mohamed *et al.* [29] and Zhang *et al.* [30] customized ComplEx to identify DTIs. KGE_NFM [31] integrated representations of drugs and targets learned from DistMult through the neural factorization machine (NFM) [32] to realize DTI prediction, whereas these methods are deficient in modeling composition relations in the biological heterogeneous graph.

The GNN-based methods [10, 12–14] formulate heterogeneous biological data as multiple drug- and target-related networks. Then, they extract drug and target representations using a GCN [15]- or GAT [16]-based model. Instead of viewing bioentities as nodes, DTI-MGNN [11] regarded DTPs as nodes and constructed a topology graph based on DTIs and a feature graph based on drug and target features to explore topological structure and semantic information in DTP representations. A few works incorporate metapaths into GNNs to extract semantic information. For example, IMCHGAN [13] and SGCL-DTI [14] mined drug (target) representations from metapath-based neighbors. Nevertheless, the metapaths used in IMCHGAN and SGCL-DTI were extracted from each individual network. Also, neither IMCHGAN nor SGCL-DTI leveraged the context information along metapaths, which benefits the improvements of DTI prediction performance as discussed in Section 5.5.

Different from the above methods, MHGNN is capable of modeling complex bioentity interrelations and exploiting high-order context dependencies in biological heterogeneous graph and higher order associations between DTPs.

## Graph neural network

GNNs adopt the message passing mechanism, where each node representation is updated by aggregating the messages from its local neighbor representations [33]. GraphSAGE [34], GCN [15] and GAT [16] are classical GNN models. Especially, GAT generalizes attention mechanisms into graph data by assigning neighboring nodes with different importance at the feature aggregation step. However, these GNN models are devised for homogeneous graphs and cannot be applied to heterogeneous graphs with various types of nodes and edges (links). To overcome this challenge, some heterogeneous graph embedding methods incorporate metapaths into GNNs.

For instance, HAN [18] and MAGNN [19] incorporated graph attention and metapath techniques to learn node representations for node classification and node clustering with two major information aggregation procedures, namely, intra-metapath aggregation and inter-metapath aggregation. In the intra-metapath aggregation step, HAN adopted the attention mechanism to aggregate information for a center node from its metapath-based neighbors. In the inter-metapath aggregation step, HAN aggregated representations for the center node from different metapaths with weights. MAGNN extended from HAN. The major difference is that in the intra-metapath aggregation step, MAGNN took all node representations along a metapath instance into consideration for information aggregation and attention coefficient computation. However, neither HAN nor MAGNN can be directly applied to DTI prediction task as they are established for node classification and clustering. Also, neither of them modeling the correlations between node pairs.

# Preliminary

This section presents formal definitions of the terms used in this paper. Graphical explanations are illustrated in Figure 2.

**Definition 1.1.** (Heterogeneous graph). A heterogeneous graph is defined as a directed graph $G = (V, E, \mathcal{A}, \mathcal{R})$, where $V, E, \mathcal{A}, \mathcal{R}$ represent node set, edge set, node type set and edge type set, respectively, and $|\mathcal{A}| + |\mathcal{R}| > 2$.

In this paper, a biological heterogeneous graph is constructed to model the interrelations between bioentities. The nodes are bioentities, such as drug (D), target (T), disease (I), side-effect (S), gene ontology (GO) and edges are interactions between bioestities, for example D-T, D-D, D-I, D-S, T-T, T-D, T-I, I-D, I-T, S-D and T-GO. Figure 2 presents a heterogeneous graph with 4 node types and 10 edge types.

**Definition 1.2.** (Metapath). A metapath $\Phi$ is defined as a path in the form of $\Phi = A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} A_{l+1}$ (abbreviated as $A_1A_2\cdots A_{l+1}$), which describes a composite relation $R = R_1 \circ R_2 \circ \cdots \circ R_l$ between node types $A_1$ and $A_{l+1}$, where $\circ$ denotes the composition operator on relations.

Different metapaths imply different semantics. Particularly, DID indicates a disease can be treated by two drugs, while DTD suggests that two drugs share the same target.

**Definition 1.3.** (Metapath instance). Given a metapath $\Phi$, there exist multiple metapath instances $\phi$ for each node following the schema of $\Phi$.

In Figure 2C, there are six metapath instances for drug node $D_1$ given metapath DID (i.e. $D_1I_2D_1$, $D_1I_2D_2$, $D_1I_2D_1$, $D_1I_2D_3$, $D_1I_3D_1$ and $D_1I_3D_3$), and two metapath instances for target node $T_1$ given metapath TDIT (i.e. $T_1D_1I_3T_3$ and $T_1D_1I_2T_4$), respectively.

**Definition 1.4.** (Metapath-based neighbors). Given a metapath $\Phi$, the metapath-based neighbors $\mathcal{N}_v^{\Phi}$ of node $v$ is defined as the set of nodes connecting to $v$ via metapath instances of $\Phi$. A neighbor node connected by two different metapath instances is regarded as two different nodes in $\mathcal{N}_v^{\Phi}$. Also, $\mathcal{N}_v^{\Phi}$ consists of $v$ itself.

As shown in Figure 2D, the metapath-based neighbors of $D_1$ including $D_1$ (×3, itself), $D_2$ (×1) and $D_3$ (×2) given the metapath DID.

**Definition 1.5.** (Metagraph). Given a metapath $\Phi$ in a heterogeneous graph $G$, the metagraph $G_v^{\Phi}$ of node $v$ is a subgraph of $G$, and it is built by all the metapath-based neighbors $u \in \mathcal{N}_v^{\Phi}$, where the intermediate nodes along metapath instances are converted to edge information.

# Method

In this section, we present the details of MHGNN. It includes two key building blocks, the drug/target representation learning module and the DTI prediction module. The overall framework of MHGNN is illustrated in Figure 3.

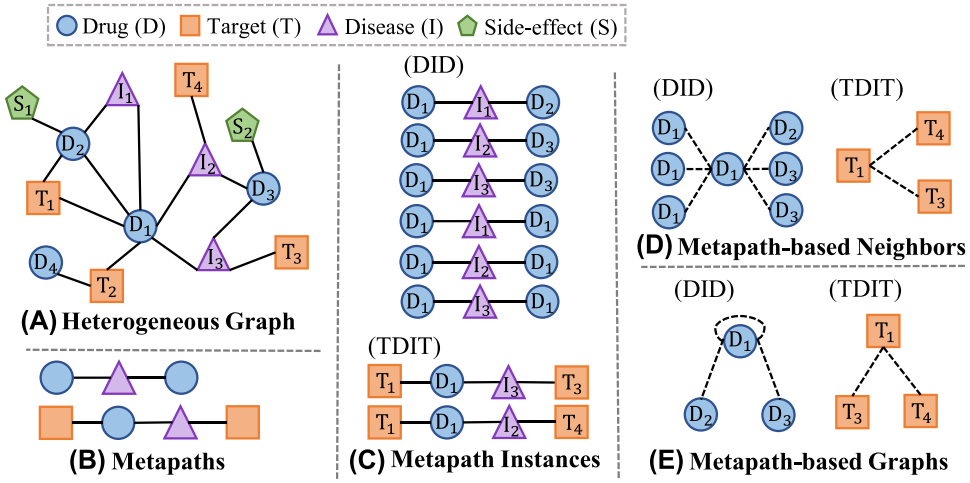


**Figure 2.** An illustration of the terms defined in Section 3. (C) – (E) are obtained from (A) based on the schema of (B) given drug node $D_1$.

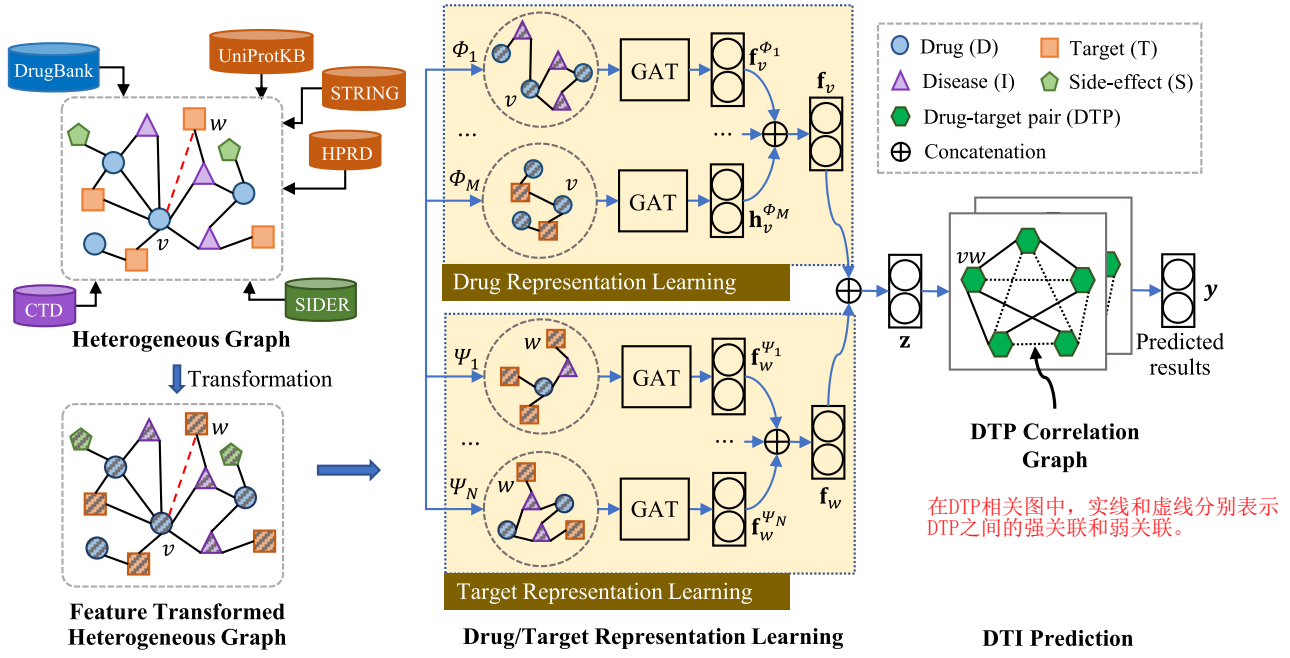


**Figure 3.** The overall framework of MHGNN. There are *M* metapaths for drug representation learning and *N* metapaths for target representation learning, respectively. Each metapath specify biological semantic relations in a certain substructure. In DTP correlation graph, the real lines and the dotted lines indicate strong associations and weak associations between DTPs, respectively.

从各种数据库中收集数据，基于收集的数据构建异构图，分别以药物节点和靶标节点为中心构建元路径

## Metapath construction

The heterogeneous biological data are collected from diverse public database resources [35–39]. Based on these data, we construct the biological heterogeneous graph (see Figure 3) as definition 1, in which the nodes of drugs, targets, diseases and side-effects have different traits. We apply dual channels to learn representations for drugs and targets, so that their specific properties can be maintained. Then, we construct metapaths for drug nodes and target nodes. Specifically, for drug nodes, we restrict the metapaths to starting and ending at drug nodes. So do for target nodes. In this way, the generated metagraphs can be processed with traditional GNNs. The maximum metapath length is restricted to 5 for both drug nodes and target nodes. Such a metapath length is semantically enough for capturing the structural information. Long metapaths increase computation complexity and memory consumption and entail misleading information as well. Accordingly, for the heterogeneous graph shown in Figure 3, we obtain 10 metapaths for each drug node (i.e. DD, DTD, DID, DSD, DTTD, DTDTD, DIDID, DSDSD, DTITD, DITID) and 8 metapaths for each target node (i.e. TT, TDT, TIT, TDDT, TITIT, TDTDT, TDIDT, TIDIT).

## Drug/target representation learning

**Feature transformation** To process the nodes of drugs, targets, diseases and side-effects in a unified framework, we first initialize node embeddings to one-hot encoding. Then, we apply node type-specific transformation $\mathbf{W}_o$ to project them into the same feature dimension

$$\mathbf{f}'_v = \mathbf{W}_o \cdot \mathbf{x}_v, \tag{1}$$

where $\mathbf{f}'_v$ and $\mathbf{x}_v$ are the transformed and raw feature vectors of node $v$, respectively, and $o \in \{D, T, I, S\}$ refers to node type. Feature transformation addresses heterogeneity that originates from raw node features and hence facilitates the following processes.

**Message passing** Let $\Phi$ be a metapath of drug node $v$ in the biological heterogeneous graph. Then, we obtain the metagraph $G_v^\Phi$ composed of numerous metapath instances $\phi_{vu}$ ($u \in \mathcal{N}_v^\Phi$). The messages transmit along each $\phi_{vu}$ to drug node $v$.

$$\mathbf{f}_{\phi_{vu}} = f_\theta(\phi_{vu}) = f_\theta(\{\mathbf{f}_i', \forall i \in \{\phi_{vu}\}\}), \tag{2}$$

where $\mathbf{f}_{\phi_{vu}}$ is the encoded representation of $\phi_{vu}$, and $f_\theta(\cdot)$ is the message transmission function with parameters $\theta$. Notice, for each metapath instance, we can consider only its two end nodes as in HAN [18] or all its context nodes as in MAGNN [19]. For the latter, multiple operations can be adopted to encode context node information, including average, linear, max-pooling and RotatE [40]. In Section 5.5, we compare the DTI prediction performances under different message transmission manners. We find that the context information along metapaths benefits DTI prediction improvements.

**Message update** We update the features of drug node $v$ using the GAT [16]. The basic idea is that different metapath instances contribute differently to $v$. To enhance model expressiveness and stabilize the learning process, we further employ the multi-head attention strategy by performing $K$ independent attentions and then concatenating their results.

$$\begin{aligned}
\mathbf{f}_v^\Phi &= \Big\Vert_{k=1}^{K} \sigma\left(\sum_{u \in \mathcal{N}_v^\Phi} [\alpha_{\phi_{vu}}]_k \cdot \mathbf{f}_{\phi_{vu}}\right), \\
e_{\phi_{vu}} &= \delta(\mathbf{a}_\Phi^{\mathrm{T}} \cdot [\mathbf{f}_v' \| \mathbf{f}_{\phi_{vu}}]), \\
\alpha_{\phi_{vu}} &= \mathrm{softmax}(e_{\phi_{vu}}) = \frac{\exp(e_{\phi_{vu}})}{\sum_{u' \in \mathcal{N}_v^\Phi} \exp(e_{\phi_{vu'}})},
\end{aligned} \tag{3}$$

where $\|$ denotes the concatenation operation, $[\alpha_{\phi_{vu}}]_k$ is the normalized attention coefficient of metapath instance $\phi_{vu}$ to drug node $v$ at the $k$th attention head, $\mathbf{a}_\Phi^{\mathrm{T}}$ is the learnable attention vector for metapath $\Phi$, $\sigma(\cdot)$ is the ELU [41] and $\delta$ is the LeakyReLU. From Eq. (2) to Eq. (3), we can observe that $\mathbf{f}_v^\Phi$ is embedded with metapath-specific structural and semantic information. This is a critical property for modeling the biological processes that a drug is involved in.

**Feature integration** Following above procedures, for drug node $v \in V_D$ with $M$ metapaths, $\Omega_D = \{\Phi_1, \Phi_2, \cdots, \Phi_M\}$, we obtain $M$ metapath-specific vector representations for it, i.e. $\{\mathbf{f}_v^{\Phi_1}, \mathbf{f}_v^{\Phi_2}, \cdots, \mathbf{f}_v^{\Phi_M}\}$. Each $\mathbf{f}_v^\Phi$ exhibits one aspect of semantic information embedded in drug node $v$. Then, we integrate these metapath-specific vector representations into one vector with the feature fusion function $g_\omega(\cdot)$, where $\omega$ is the learning parameters. In our work, we adopt the concatenation operation.

$$\mathbf{f}_v = g_\omega(\{\mathbf{f}_v^{\Phi_1}, \mathbf{f}_v^{\Phi_2}, \cdots, \mathbf{f}_v^{\Phi_M}\}) = \Big\Vert_{m=1}^{M} \mathbf{f}_v^{\Phi_m}. \tag{4}$$

Notice that $g_\omega(\cdot)$ can be other feature fusion operations, such as attention, average and max-pooling. In Section 5.5, we compare the DTI prediction performances under different feature fusion operations.

Similarly, for target node $w \in V_T$ with $N$ metapaths, $\Omega_T = \{\Psi_1, \Psi_2, \cdots, \Psi_N\}$, we extract $N$ metapath-specific vector representations, i.e. $\{\mathbf{f}_w^{\Psi_1}, \mathbf{f}_w^{\Psi_2}, \cdots, \mathbf{f}_w^{\Psi_N}\}$. By concatenating these vectors, we obtain its final feature vector $\mathbf{f}_w$. The graphical illustration of drug and target representation learning process is presented in Figure 3.

## DTI prediction

For each DTP, we obtain its representation $\mathbf{z}$ through concatenating the drug representation and the target representation, that is $\mathbf{z} = \mathbf{f}_v \| \mathbf{f}_w$, where $v \in V_D$, $w \in V_T$. Accordingly, given a batch of $B$ DTPs, we obtain a representation matrix $\mathbf{Z}$, where $\mathbf{Z}_b$ denotes the feature vector of the $b$th DTP. Considering that similar drugs are prone to interact with similar targets, and vice versa, therefore, DTPs with the common (similar) drugs or targets have stronger associations than these without. We build a DTP correlation graph $G_{DTP} = (\mathbf{Z}, \mathbf{A})$ to discover the potential relationships between DTPs, where each node is a DTP (refer to Figure 3). The adjacency matrix $\mathbf{A}$ is constructed from DTP representations.

$$\mathbf{A} = \mathrm{softmax}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}), \tag{5}$$

where $\mathbf{A}$ is symmetric, and $\mathbf{A}_{ij}$ denotes the normalized similarity between the $i$th and the $j$th DTPs in a batch. The greater the $\mathbf{A}_{ij}$ value is, the stronger the association is between the $i$th and the $j$th DTPs. Since the size of $\mathbf{A}$ is equal to the batch size used for model training/evaluation, MHGNN has no risk of data size explosion.

Next, we feed $\mathbf{Z}$ into two GCN layers. The outputs of the second GCN layer are viewed as the prediction results.

$$\mathbf{y} = \mathbf{A}(\mathrm{ReLU}(\mathbf{A}\mathbf{Z}\mathbf{W}_1))\mathbf{W}_2, \tag{6}$$

where $\mathbf{W_1}$ and $\mathbf{W}_2$ are learning parameters in the first GCN layer and the second GCN layer, respectively.

The binary cross entropy (BCE) loss is utilized to evaluate the differences between the predicted and the ground-truth DTIs. It is formulated as

$$L_{BCE} = -\frac{1}{B}\sum_{b=1}^{B}[y_b \log(y_b') + (1 - y_b)\log(1 - y_b')], \tag{7}$$

where $B$ is the number of training DTPs in a batch, $y_b = 1$ if the interaction of the $b$th DTP exists, otherwise, $y_b = 0$, and $y_b'$ is the predicted interaction probability of the $b$th DTP, which is normalized with the softmax function.

# Experiment

## Dataset

该模型在三个生物异构数据库中进行了评估

MHGNN is evaluated on three biological heterogeneous datasets. The first one (denoted as *Hetero-A*) is introduced in DTINet [22], consisting of 708 drugs (D), 1512 targets (T), 5603 diseases (I) and 4192 side-effects (S), and six connections, i.e. D-T, D-D, D-I, D-S, T-T and T-I. More details about Hetero-A can be referred to [22].

Over the last decade, a large number of novel DTIs as well as other interactions have been discovered, but they are not fully explored in [22]. Furthermore, the incompleteness of positive samples (e.g. DTIs) not only induces errors in data modeling process but also hides a great risk of false negatives during model evaluation, leading to unknown bias between predictions and the actual results. In light of this, we extend Hetero-A from a list of latest released resources [35–39]. Particularly, we collect 2214 unique drugs from DrugBank (Version 5.1.8) [35]. All the drugs are small molecule drugs and marked 'annotated' and 'approved'. The HPRD database has not been updated since 2010. Hence, we gather targets from UniProtKB (Release 2021_04) [39], wherein we focus on protein targets in *Homo sapiens (Human)* and filter out targets that are not marked 'Reviewed', obtaining 1968 targets.

从DrugBank中收集带有注释和批准字样的药物分子，2214
从UniProtKB中收集人类靶标，并删除没有标记“已审核”的靶标，1968

从CTD收集了7205种疾病类型
从SIDER收集了3935种副作用类型



**Table 1.** Statistics of datasets Hetero-A and Hetero-B. The last column lists the ratios of Hetero-B to Hetero-A

| Types | Items | Hetero-A | | Hetero-B Hetero-B是作者自己构建的数据库 | | B/A |
|---|---|---|---|---|---|---|
| | | Numbers | Resources | Numbers | Resources | |
| Node | Drug (D) | 708 | DrugBank (Version 3.0) [43] | 2,214 | DrugBank (Version 5.1.8) [35] | 3.13 |
| | Target (T) | 1,152 | HPRD (Release 9) [44] | 1,968 | UniProtKB (Release 2021_04) [39] | 1.71 |
| | Disease (I) | 5,603 | CTD (2013) [45] | 7,205 | CTD (2021) [38] | 1.29 |
| | Side-effect (S) | 4,192 | SIDER (Version 2) [46] | 3,935 | SIDER (Version 4) [37] | 0.94 |
| Edge | D-T | 1,923 | DrugBank (Version 3.0) [43] | 8,750 | DrugBank (Version 5.1.8) [35] | 4.55 |
| | D-D | 10,036 | DrugBank (Version 3.0) [43] | 1,091,870 | DrugBank (Version 5.1.8) [35] | 108.80 |
| | D-I | 199,214 | CTD (2013) [45] | 542,970 | CTD (2021) [38] | 2.73 |
| | D-S | 80,164 | SIDER (Version 2) [46] | 104,629 | SIDER (Version 4) [37] | 1.31 |
| | T-T | 7,363 | HPRD (Release 9) [44] | 456,592 | STRING (Version 11.0) [36] | 62.01 |
| | T-I | 1,596,745 | CTD (2013) [45] | 2,922,064 | CTD (2021) [38] | 1.83 |
| **Metapaths** | | DD, DTD, DID, DSD, DTTD, DTDTD, DIDID, DSDSD, DTITD, DITID TT, TDT, TIT, TDDT, TITIT, TDTDT, TDIDT, TIDIT | | | | |

根据药物的SMILES序列计，算了药物对之间的Tanimoto得分，该得分作为药物对结构相似性得分。

根据蛋白质序列计，计算了靶标对之间的Smith-Waterman得分，该得分作为靶标对的相似性得分。

**Table 2.** Statistics of dataset Hetero-C

| Types | Items | Numbers | Resources |
|---|---|---|---|
| Node | D | 1,094 | DrugBank (Version 4.0) [47] |
| | T | 1,556 | DrugBank and DrugCentral [48] |
| | Su | 738 | DrugBank |
| | St | 881 | PubMed [49] |
| | S | 4,063 | SIDER (Version 4) [37] |
| | GO | 4.098 | EMBL-EBI [50] |
| Edge | D-T | 11,819 | |
| | D-Su | 20,798 | |
| | D-St | 133,880 | |
| | D-S | 122,792 | |
| | T-GO | 35,980 | |
| **Metapaths** | | DD, DTD, DSuD, DStD, DSD, DTTD TT, TDT, T(GO)T, TDDT | |

**Note**: D, T, Su, St, S and GO denote drug, target, substituent, chemical structure, side-effect and GO terms, respectively.

We assemble 7205 diseases from CTD (2021) [38] and 3935 side-effects from SIDER [37], respectively. Note that, we only collect side-effects labeled with PT (preferred term). The interactions of D-T and D-D are obtained from DrugBank (Version 5.1.8), the associations D-I and T-I are compiled from CTD (2021), the D-S associations are gathered from SIDER (Version 4.1) [37] and the P-P interactions are collected from STRING (Version 11.5) [36], wherein P-P interactions include direct (physical) and indirect (functional) interactions. We denote this extended heterogeneous dataset as *Hetero-B*, which is much larger than Hetero-A. Note that, Hetero-B is collected from October 1 to 15, 2021.

The last dataset (denoted as *Hetero-C*) with 1094 drugs (D) and 1556 targets (T) is introduced in [42]. Besides, it also contains four kinds of heterogeneous information, including 738 drug substitutes (Su), 881 drug chemical structures (St), 4063 side-effects (S) and 4098 gene ontology (GO) terms. These heterogeneous bioentities formulate 5 connections, namely, D-T, D-Su, D-St, D-S and T-GO.

The statistical information and resources about hetero-A and Hetero-B are summarized in Table 1, and these about Hetero-C are summarized in Table 2.

**Redundancy check of drugs and targets.** High redundant data set could impair model generalization ability. To this end, we check the redundancy of drugs and targets, where the redundancy is measured by their similarity score distributions. Particularly, we calculate the Tanimoto score [51] from drug SMILES (simplified molecular-input line-entry system) strings and use it as the structural similarity for each drug–drug pair (DDP). We compute the normalized Smith–Waterman score [52] from protein sequences and use it as the similarity of each target–target pair (TTP).

Figure 4 shows the similarity score distributions of DDPs and TTPs on datasets Hetero-A, Hetero-B and Hetero-C. We see that the majority of drug–drug similarity scores are around zeros. Specifically, about 95.87% in Hetero-A, 88.30% in Hetero-B and 83.88% in Hetero-C of DDPs hold similarity scores smaller than 0.5. For the similarities of TTPs, only about 2.83% in Hetero-A and 0.18% in Hetero-B of TTPs possess similarity scores larger than 0.5, and the majority of the scores are close to zeros. However, in dataset Hetero-C, 36.89% TTPs are with similarity scores more than 0.5. Therefore, we can conclude that the redundancies of drugs and targets in both Hetero-A and Hetero-B are negligible, whereas Hetero-C contains a high redundancy in terms of targets.

## Evaluation metric

We employ six evaluation metrics, including the precision score, the recall score, the F1 score, the Matthews correlation coefficient (MCC), the area under the curve (AUC) and the area under the precision-recall curve (AUPR). In binary classification, we obtain the count of true negatives (TN), false negatives (FN), true positives (TP) and false positives (FP). The precision is the ratio TP / (TP + FP). It reflects the ability of the classifier not to label as positive a sample that is negative. The recall is the ratio TP / (TP + FN). It indicates the ability of the classifier to find all the positive samples. The F1 score is interpreted as a harmonic mean of the precision and recall, that is F1 = 2 ∗ (precision ∗ recall) / (precision + recall). The MCC measures the quality of binary classification and is generally regarded as a balanced measure with a correlation coefficient value between −1 and +1. Apart from the MCC, other metrics are with a value between 0 and 1.

## Implementation detail 实现细节

For MHGNN, we conduct a 10-fold cross validation, where the training and evaluation process of MHGNN is conducted 10 times independently, including the metapath construction process. We consider all known DTIs as positive samples and randomly select the same number of unknown drug–target pairs as negative samples. For each fold, 90% positive samples and negative samples are randomly selected as the training set, and the remaining samples are used as the test set. The Adam is employed as the optimizer with weight decay rate $1 \times 10^{-5}$ and learning rate selected from

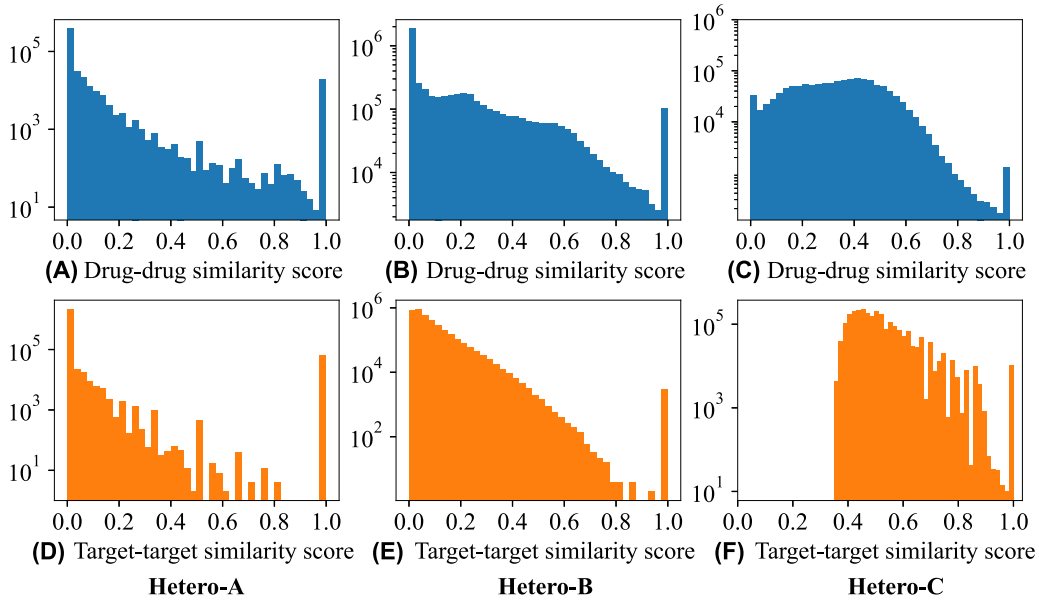


**Figure 4.** Drug-drug similarity score distributions and target-target similarity score distributions on datasets Hetero-A, Hetero-B and Hetero-C. Notice that, the vertical axes are set to the logarithmic scale. The bar at 1.0 of the X-axis in each plot denotes the self-similarity distributions of drugs and targets.

$\{1 \times 10^{-6}, 1 \times 10^{-5}\}$. MHGNN is trained with 200 epochs, together with the early stopping strategy. The dimension of the projected and hidden features is set to 64. The batch size is set to 256. The number of the attention head $K$ is set to 8.

## Baseline

MHGNN is compared with 17 baselines, including 5 similarity-based methods (DTINet [22], deepDTnet [23], MEDTI [25], NEDTP [24] and MultiDTI [53]), 6 KG-based methods (TransE [26], DistMult [27], ComplEx [28], KGE_NFM [31], TransE-NFM and ComplEx-NFM), 4 GNN-based methods (NeoDTI [10], IMCHGAN [13], DTI-MGNN [11] and SGCL-DTI [14]) and 2 heterogenous graph embedding methods (HAN [18] and MAGNN [19]).

**DTINet** [22] extracts drug and target representations from drug- and target-related similarity networks through random walk with restart (RWR) with a dimension reduction scheme. Then, the inductive matrix completion (IMC) is adopted to predict novel DTIs.

**deepDTnet** [23] converts drug- and target-related similarity networks into drug and target representation matrices via RWR and positive pointwise mutual information (PPMI) techniques. Afterwards, AE models are applied to each matrix to learn drug (target) features separately.

**MEDTI** [25] is similar to deepDTnet [23]. The difference is that in MEDTI, multiple drug (target) representation matrices are combined to learn drug (target) features using AE.

**NEDTP** [24] extracts drug and target features through the Word2Vec algorithm. Then, it constructs the gradient boosting decision tree (GBDT) to predict new DTIs based on learned features.

**MultiDTI** [53] exploits drug and target representations from sequences using convolutional layers with residual connections, respectively. After that, it maps the representations of drugs, targets, side effects and disease nodes in the heterogeneous network into a common space.

**KGE_NFM** [31] first utilizes DistMult [27] to extract low-dimensional representations for bioentities. Then, it integrates the learned representations of drugs and targets through the neural factorization machine (NFM) [32]. Afterwards, the feed-forward neural network is employed to yield DTI prediction. We obtain **ComplEx-NFM** by replacing DistMult in KGE_NFM with ComplEx since Mohamed *et al.* [29] and Zhang *et al.* [30] do not release their code. Similarly, we obtain **TransE-NFM**. Accordingly, TransE, ComplEx and DistMult are models that remove NFM.

**NeoDTI** [10] integrates drug (target) representations learned from different relation networks using GCNs. Then, IMC is adopted to predict novel DTIs.

**IMCHGAN** [13] adopts a two-level neural attention mechanism approach to learn drug and target feature representations from the DTI heterogeneous network, respectively. Subsequently, the learned features are fed into IMC to obtain the prediction results of DTIs.

**SGCL-DTI** [14] first extracts drug (target) representations from metapath-based neighbors using the GCN. Next, it constructs a topology graph and a semantic graph with DTPs as nodes. Then, the contrastive learning strategy is employed to refine drug and target representations from these two graphs.

**HAN** [18] incorporates metapaths and attention mechanisms for node classification and clustering tasks. It aggregates two end nodes information along each metapath instance. Then, it applies the attention strategy to fuse features for each center node from intra-metapath level and inter-metapath level. We customize HAN for DTI prediction by developing a model with two channels, one for drug representation learning and the other for target representation learning, where each channel is based on HAN. Afterwards, the outputs of these two channels are concatenated and then feed into the prediction module with two linear layers with 1024 and 2 neurons.

**MAGNN** [19] is an extension of HAN. Instead of only aggregating two end node information, it takes all node information along each metapath instance into consideration. We adjust MAGNN for DTI prediction with the same operation as mentioned in HAN.

For HAN and MAGNN, we set the dimension of the attention vector in the feature fusion to 128. Other parameters are set to the same as these in MHGNN. For fair comparison, all these baselines adopt the same training set and test set as MHGNN. For each

**Table 3.** Ablation studies on dataset Hetero-A

| Groups | Methods | AUC$_{\pm std}$ | AUPR$_{\pm std}$ | Precision$_{\pm std}$ | Recall$_{\pm std}$ | F1$_{\pm std}$ | MCC$_{\pm std}$ |
|---|---|---|---|---|---|---|---|
| Instance-level | $Var_{nb}$ | $0.9437_{\pm 0.012}$ | $0.9282_{\pm 0.017}$ | $0.8152_{\pm 0.024}$ | $0.9595_{\pm 0.017}$ | $0.8812_{\pm 0.019}$ | $0.7535_{\pm 0.037}$ |
| | $Var_{avg}$ | $0.9559_{\pm 0.014}$ | $0.9161_{\pm 0.028}$ | $0.8998_{\pm 0.039}$ | $0.9287_{\pm 0.057}$ | $0.9126_{\pm 0.033}$ | $0.8263_{\pm 0.063}$ |
| | $Var_{linear}$ | $\mathbf{0.9833}_{\pm 0.007}$ | $\mathbf{0.9780}_{\pm 0.009}$ | $0.7950_{\pm 0.017}$ | $\mathbf{1.0000}_{\pm 0.000}$ | $0.8857_{\pm 0.011}$ | $0.7678_{\pm 0.022}$ |
| | $Var_{max}$ | $0.9727_{\pm 0.011}$ | $0.9552_{\pm 0.026}$ | $\mathbf{0.9316}_{\pm 0.021}$ | $\mathbf{0.9886}_{\pm 0.009}$ | $\mathbf{0.9590}_{\pm 0.009}$ | $\mathbf{0.9174}_{\pm 0.019}$ |
| | $Var_{rot}$ | $0.9682_{\pm 0.012}$ | $0.9436_{\pm 0.026}$ | $0.8806_{\pm 0.032}$ | $0.9844_{\pm 0.009}$ | $0.9293_{\pm 0.020}$ | $0.8561_{\pm 0.042}$ |
| Semantic-level | $Var^2_{avg}$ | $0.9831_{\pm 0.003}$ | $0.9749_{\pm 0.009}$ | $0.7500_{\pm 0.000}$ | $\mathbf{1.0000}_{\pm 0.000}$ | $0.8571_{\pm 0.000}$ | $0.7071_{\pm 0.000}$ |
| | $Var^2_{max}$ | $0.9590_{\pm 0.019}$ | $0.9272_{\pm 0.036}$ | $0.5952_{\pm 0.113}$ | $\mathbf{1.0000}_{\pm 0.000}$ | $0.7401_{\pm 0.087}$ | $0.2845_{\pm 0.331}$ |
| | $Var_{attn}$ | $\mathbf{0.9862}_{\pm 0.005}$ | $\mathbf{0.9837}_{\pm 0.007}$ | $0.8095_{\pm 0.179}$ | $0.9974_{\pm 0.005}$ | $0.8812_{\pm 0.124}$ | $0.6787_{\pm 0.392}$ |
| | $Var_{concat}$ | $0.9727_{\pm 0.011}$ | $0.9552_{\pm 0.026}$ | $\mathbf{0.9316}_{\pm 0.021}$ | $0.9886_{\pm 0.009}$ | $\mathbf{0.9590}_{\pm 0.009}$ | $\mathbf{0.9174}_{\pm 0.019}$ |
| Decision-level | $Var^2_{linear}$ | $0.9556_{\pm 0.009}$ | $0.9449_{\pm 0.011}$ | $0.9089_{\pm 0.029}$ | $0.7967_{\pm 0.085}$ | $0.8452_{\pm 0.040}$ | $0.7232_{\pm 0.048}$ |
| | $Var_{gcn}$ | $\mathbf{0.9727}_{\pm 0.011}$ | $\mathbf{0.9552}_{\pm 0.026}$ | $\mathbf{0.9316}_{\pm 0.021}$ | $\mathbf{0.9886}_{\pm 0.009}$ | $\mathbf{0.9590}_{\pm 0.009}$ | $\mathbf{0.9174}_{\pm 0.019}$ |

**Table 4.** Ablation studies on dataset Hetero-B

| Groups | Methods | AUC$_{\pm std}$ | AUPR$_{\pm std}$ | Precision$_{\pm std}$ | Recall$_{\pm std}$ | F1$_{\pm std}$ | MCC$_{\pm std}$ |
|---|---|---|---|---|---|---|---|
| Instance-level | $Var_{nb}$ | $0.9942_{\pm 0.002}$ | $0.9930_{\pm 0.003}$ | $0.8868_{\pm 0.034}$ | $0.9979_{\pm 0.002}$ | $0.9387_{\pm 0.019}$ | $0.8766_{\pm 0.038}$ |
| | $Var_{avg}$ | $\mathbf{0.9983}_{\pm 0.007}$ | $\mathbf{0.9984}_{\pm 0.002}$ | $0.9414_{\pm 0.014}$ | $0.9920_{\pm 0.003}$ | $\mathbf{0.9661}_{\pm 0.011}$ | $\mathbf{0.9316}_{\pm 0.018}$ |
| | $Var_{linear}$ | $0.9982_{\pm 0.001}$ | $0.9983_{\pm 0.001}$ | $0.8660_{\pm 0.016}$ | $\mathbf{1.0000}_{\pm 0.000}$ | $0.9281_{\pm 0.009}$ | $0.8553_{\pm 0.019}$ |
| | $Var_{max}$ | $0.9964_{\pm 0.002}$ | $0.9945_{\pm 0.003}$ | $\mathbf{0.9888}_{\pm 0.009}$ | $0.9279_{\pm 0.029}$ | $0.9570_{\pm 0.012}$ | $0.9194_{\pm 0.020}$ |
| | $Var_{rot}$ | $0.9980_{\pm 0.001}$ | $0.9980_{\pm 0.001}$ | $0.9843_{\pm 0.015}$ | $0.9327_{\pm 0.033}$ | $0.9572_{\pm 0.011}$ | $0.9194_{\pm 0.018}$ |
| Semantic-level | $Var^2_{avg}$ | $0.9955_{\pm 0.000}$ | $0.9958_{\pm 0.000}$ | $\mathbf{1.0000}_{\pm 0.000}$ | $0.8777_{\pm 0.000}$ | $0.9349_{\pm 0.000}$ | $0.8844_{\pm 0.000}$ |
| | $Var^2_{max}$ | $\mathbf{0.9977}_{\pm 0.001}$ | $\mathbf{0.9978}_{\pm 0.001}$ | $0.6762_{\pm 0.130}$ | $\mathbf{1.0000}_{\pm 0.000}$ | $0.7992_{\pm 0.094}$ | $0.5504_{\pm 0.223}$ |
| | $Var_{attn}$ | $0.9945_{\pm 0.001}$ | $0.9948_{\pm 0.001}$ | $1.0000_{\pm 0.000}$ | $0.5851_{\pm 0.239}$ | $0.7086_{\pm 0.198}$ | $0.6471_{\pm 0.192}$ |
| | $Var_{concat}$ | $0.9964_{\pm 0.002}$ | $0.9945_{\pm 0.003}$ | $0.9888_{\pm 0.009}$ | $0.9279_{\pm 0.029}$ | $\mathbf{0.9570}_{\pm 0.012}$ | $\mathbf{0.9194}_{\pm 0.020}$ |
| Decision-level | $Var^2_{linear}$ | $0.9397_{\pm 0.009}$ | $0.9222_{\pm 0.014}$ | $0.9111_{\pm 0.021}$ | $0.6814_{\pm 0.098}$ | $0.7746_{\pm 0.057}$ | $0.6362_{\pm 0.056}$ |
| | $Var_{gcn}$ | $\mathbf{0.9964}_{\pm 0.002}$ | $\mathbf{0.9945}_{\pm 0.003}$ | $\mathbf{0.9888}_{\pm 0.009}$ | $\mathbf{0.9279}_{\pm 0.029}$ | $\mathbf{0.9570}_{\pm 0.012}$ | $\mathbf{0.9194}_{\pm 0.020}$ |

method, we report the average value and the standard deviation over each evaluation metric.

## Ablation study

To validate the rationality of each component of MHGNN, we conduct experiments on the following variants of MHGNN. The experimental results of these variants on dataset Hetero-A, Hetero-B and Hetero-C are reported in Tables 3– 5.

**$Var_{nb}$** only considers the metapath-based neighbors for message passing. **$Var_{avg}$** calculates the element-wise mean of all node vectors along each metapath instance $\phi_{vu}$. **$Var_{linear}$** is an extension of $Var_{avg}$. It applies the linear transformation on the mean of all node vectors along each metapath instance $\phi_{vu}$. **$Var_{max}$** takes the element-wise max of all node vectors along each metapath instance $\phi_{vu}$. **$Var_{rot}$** considers the relational rotation [40] as metapath instance encoder. **$Var^2_{avg}$** computes the element-wise mean of all metapah-specific vectors for each drug and target node. **$Var^2_{max}$** takes the element-wise max of all metapah-specific vectors. **$Var_{concat}$** is the proposed one that concatenates all metapah-specific vectors of a drug (target) node into one vector. **$Var_{attn}$** aggregates metapah-specific vectors of a drug (target) node into one vector with the attention mechanism. **$Var^2_{linear}$** employs two consecutive linear layers in the DTI prediction module. **$Var_{gcn}$** is the proposed one with GCN layers in the DTI prediction module. For clarity, as shown in Tables 3– 5, we categorize above variants into three groups, instance-level, semantic-level and decision-level.

From Tables 3 to 5, we see that, in the instance-level, the overall performances of $Var_{rot}$, $Var_{max}$, $Var_{avg}$ and $Var_{linear}$ outperform $Var_{nb}$ over all evaluation metrics on all these three datasets. It demonstrates the importance of node contexts along metapaths for feature learning.

Comparing the results in the semantic-level, we observe that on dataset Hetero-A, $Var^2_{avg}$, $Var^2_{attn}$ and $Var_{concat}$ outperform $Var^2_{max}$ in five of six eveluation metrics. On Hetero-B, different variants show competitive results in terms of AUC and AUPR. Over the indicators of the F1 score and the MCC, $Var^2_{avg}$ and $Var_{concat}$ exceed $Var^2_{max}$ and $Var^2_{attn}$ by at least 0.13. On Hetero-C, different variants show competitive results in terms of AUC, AUPR and the recall score. Over the indicators of the precision score, the F1 score and the MCC, $Var^2_{avg}$, $Var_{concat}$ and $Var^2_{attn}$ show superior performances over $Var^2_{max}$.

In the decision level, $Var_{gcn}$ exceeds $Var_{linear}$ over all evaluation metrics on datasets Hetero-A, Hetero-B and Hetero-C. Particularly, $Var_{gcn}$ achieves significant gains by 0.0567, 0.0723, 0.0777, 0.2465, 0.1824 and 0.2832 over $Var^2_{linear}$ over the metrics of AUC, AUPR, the precision score, the recall score, the F1 score and the MCC on dataset Hetero-B, respectively. The results indicate that the exploration of potential relations between DTPs does help for improving DTI prediction performance. Additionally, we observe that, overall, the standard deviations computed on Hetero-B are much smaller than those computed on Hetero-A. It shows the augmentation of positive samples enhances model robustness for DTI prediction.

## Metapath combination analysis

To explore the impact of metapaths on DTI prediction, we conduct experiments on different metapath combinations of drugs and targets with average, linear, max-pooling, the relational rotation [40] or the neighboring operation as the metapath instance encoder. The results on Hetero-A, Hetero-B and Hetero-C are depicted in Figures 5– 7. Note that, G1 = {DD, DTD, TT, TDT} only employs drug and target information. G2 ∼ G4 contain metapaths of drugs and targets with the maximum length of 3, 4 and 5,

**Table 5.** Ablation studies on dataset Hetero-C.

| Groups | Methods | AUC$_{\pm std}$ | AUPR$_{\pm std}$ | Precision$_{\pm std}$ | Recall$_{\pm std}$ | F1$_{\pm std}$ | MCC$_{\pm std}$ |
|---|---|---|---|---|---|---|---|
| Instance-level | $Var_{nb}$ | 0.9884$_{\pm 0.020}$ | 0.9751$_{\pm 0.050}$ | 0.8766$_{\pm 0.111}$ | **1.0000**$_{\pm 0.0000}$ | 0.9298$_{\pm 0.075}$ | 0.8482$_{\pm 0.184}$ |
| | $Var_{avg}$ | 0.9982$_{\pm 0.001}$ | 0.9982$_{\pm 0.001}$ | 0.9658$_{\pm 0.011}$ | 0.9936$_{\pm 0.006}$ | **0.9795**$_{\pm 0.004}$ | **0.9588**$_{\pm 0.008}$ |
| | $Var_{linear}$ | 0.9978$_{\pm 0.000}$ | 0.9977$_{\pm 0.000}$ | 0.9234$_{\pm 0.002}$ | **1.0000**$_{\pm 0.000}$ | 0.9602$_{\pm 0.001}$ | 0.9202$_{\pm 0.002}$ |
| | $Var_{max}$ | 0.9977$_{\pm 0.001}$ | 0.9958$_{\pm 0.004}$ | **0.9976**$_{\pm 0.003}$ | 0.8736$_{\pm 0.039}$ | 0.9310$_{\pm 0.021}$ | 0.8788$_{\pm 0.032}$ |
| | $Var_{rot}$ | **0.9987**$_{\pm 0.000}$ | **0.9986**$_{\pm 0.000}$ | 0.9456$_{\pm 0.009}$ | 0.9999$_{\pm 0.000}$ | 0.9720$_{\pm 0.005}$ | 0.9440$_{\pm 0.009}$ |
| Semantic-level | $Var^2_{avg}$ | 0.9985$_{\pm 0.000}$ | 0.9985$_{\pm 0.000}$ | 0.9350$_{\pm 0.016}$ | 0.9994$_{\pm 0.001}$ | 0.9661$_{\pm 0.008}$ | 0.9320$_{\pm 0.016}$ |
| | $Var^2_{max}$ | **0.9991**$_{\pm 0.000}$ | **0.9991**$_{\pm 0.000}$ | 0.8418$_{\pm 0.171}$ | **1.0000**$_{\pm 0.000}$ | 0.9031$_{\pm 0.118}$ | 0.7394$_{\pm 0.370}$ |
| | $Var_{attn}$ | 0.9985$_{\pm 0.001}$ | 0.9985$_{\pm 0.001}$ | 0.9364$_{\pm 0.014}$ | 0.9999$_{\pm 0.000}$ | 0.9670$_{\pm 0.007}$ | 0.9340$_{\pm 0.015}$ |
| | $Var_{concat}$ | 0.9987$_{\pm 0.000}$ | 0.9986$_{\pm 0.000}$ | **0.9456**$_{\pm 0.009}$ | 0.9999$_{\pm 0.000}$ | **0.9720**$_{\pm 0.005}$ | **0.9440**$_{\pm 0.009}$ |
| Decision-level | $Var^2_{linear}$ | 0.9743$_{\pm 0.005}$ | 0.9700$_{\pm 0.005}$ | 0.8827$_{\pm 0.010}$ | 0.9681$_{\pm 0.007}$ | 0.9234$_{\pm 0.008}$ | 0.8434$_{\pm 0.016}$ |
| | $Var_{gcn}$ | **0.9987**$_{\pm 0.000}$ | **0.9986**$_{\pm 0.000}$ | **0.9456**$_{\pm 0.009}$ | **0.9999**$_{\pm 0.000}$ | **0.9720**$_{\pm 0.005}$ | **0.9440**$_{\pm 0.009}$ |

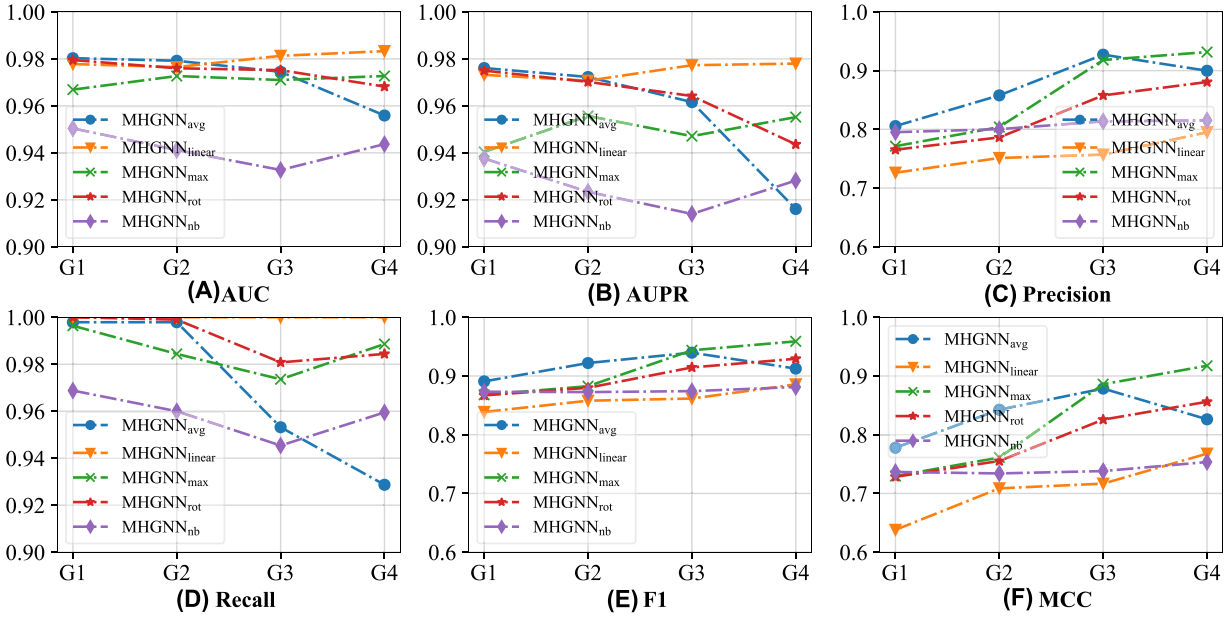


**Figure 5.** The DTI prediction performance of MHGNN on Hetero-A with different metapath combinations.

respectively. For example, G2 is with the maximum metapath length of 3, G2 = {DD, DTD, DID, DSD, TT, TDT, TIT} for Hetero-A and Hetero-B and G2 = {DD, DTD, DSuD, DStD, TT, TDT, T(GO)T} for Hetero-C.

We observe that, on Hetero-A (Figure 5), the overall performances of MHGNN continually increase with long metapaths over metrics of the precision score, the F1 score and the MCC. Generally, the results of MHGNN$_{max}$, MHGNN$_{linear}$, MHGNN$_{avg}$ and MHGNN$_{rot}$ are superior to these of MHGNN$_{nb}$ over AUC and AUPR and the recall score. Also, the results of MHGNN$_{max}$, MHGNN$_{avg}$ and MHGNN$_{rot}$ over the precision score, the F1 score and the MCC are generally higher than these of MHGNN$_{nb}$. On dataset Hetero-C (Figure 7), we find that the performances of MHGNN$_{linear}$, MHGNN$_{avg}$ and MHGNN$_{rot}$ improve with long metapaths over different evaluation metrics. Contrarily, the performances of MHGNN$_{nb}$ decline greatly with the increase of metapath lengths. Additionally, MHGNN$_{max}$ is significantly affected by metapath lengths in terms of the recall score. The results imply that the high-order dependency in the biological heterogeneous graph can benefit DTI prediction on the condition that the context information along metapaths are appropriately encoded.

Comparing the results on Hetero-B (Figure 5) with these on Hetero-A (Figure 6), we find that MHGNN$_{max}$, MHGNN$_{linear}$, MHGNN$_{avg}$ and MHGNN$_{rot}$ are stable at high performances and insensitive to different metapaths. It indicates that the augmentation of positive samples enhances model robustness once again.

## Parameter analysis

We investigate the performances of MHGNN with different parameter settings and report the results on Hetero-A and Hetero-B in figures from Figures 8 to 11.

- **Number of attention heads** *K*. We first explore the performance of MHGNN with different number of attention heads. According to the results in Figure 8, we find that the number of attention heads greatly affects the performances of MHGNN over the metrics of the precision score, the recall score, the F1-score and the MCC. MHGNN achieves the highest results over these metrics with 8 attention heads. With fewer attention heads, MHGNN presents high false negative errors. The results indicate that more attention heads enhance model robustness and model expressiveness. From Figure 9, we observe that MHGNN shows high performances over above metrics. This largely benefits from the augmentation of positive interactions between bioentities. Over AUC and AUPR, we see that MHGNN yields high

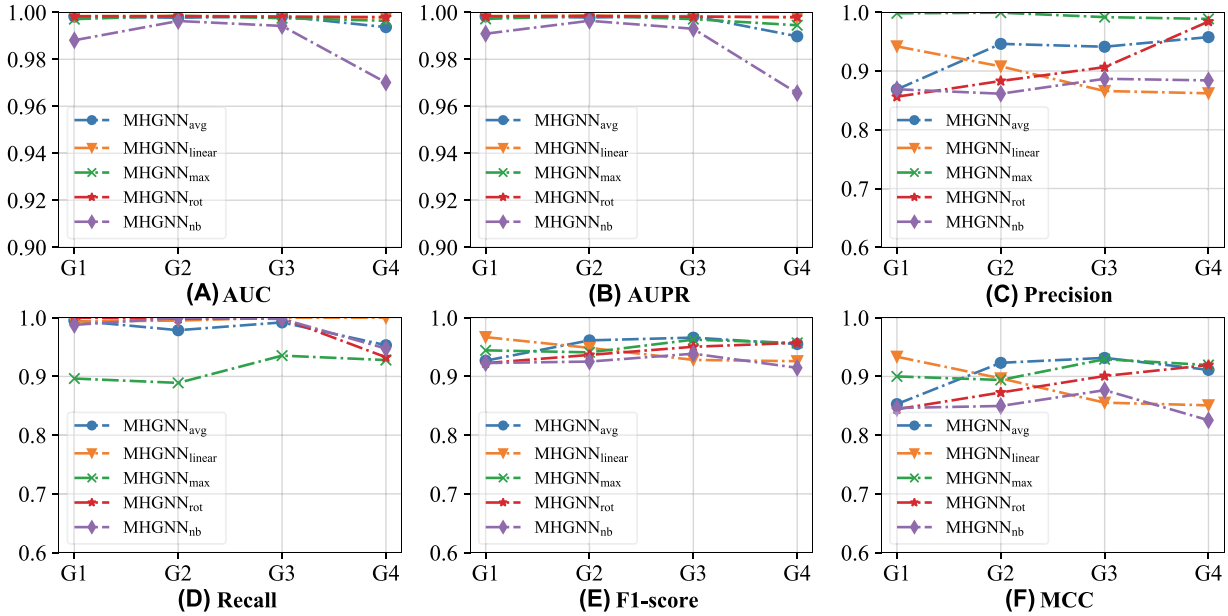


**Figure 6.** The DTI prediction performance of MHGNN on Hetero-B with different metapath combinations.

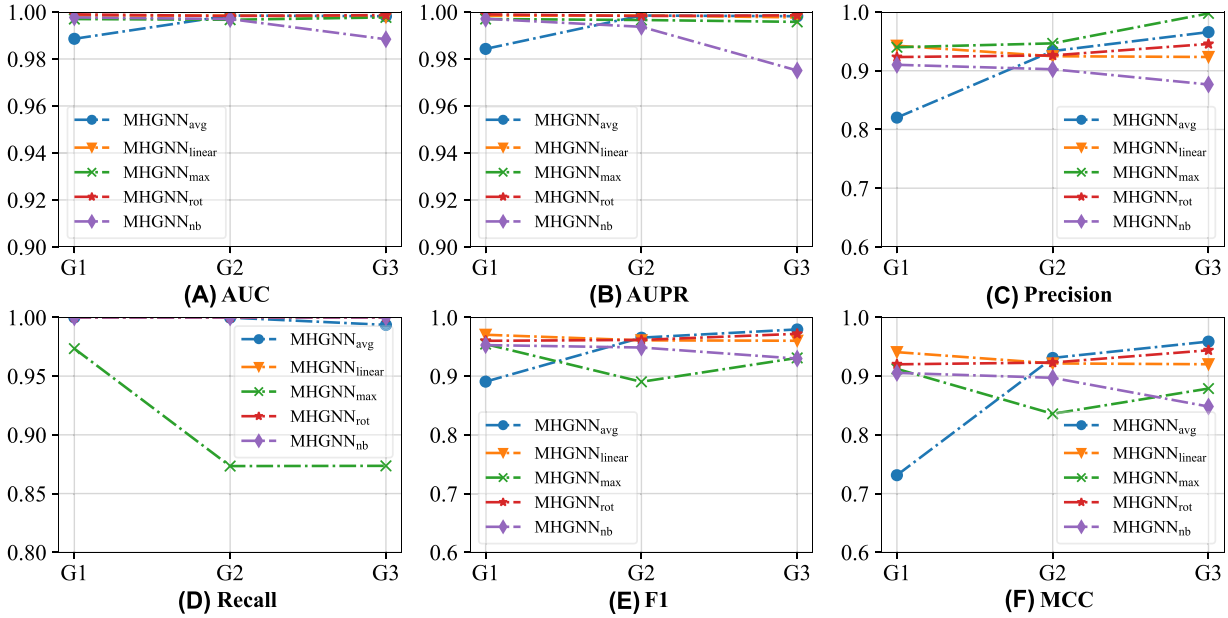


**Figure 7.** The DTI prediction performance of MHGNN on Hetero-C with different metapath combinations.

performances with different number of attention heads on both dataset Hetero-A and dataset Hetero-B.

- **Dimension of embeddings.** The projected and hidden embeddings are set to the same dimension. We vary the dimension from 32 to 256 to analyze DTI prediction performances. The results over different metrics on dataset Hetero-A are illustrated in Figure 10, and these on dataset Hetero-B are in Figure 11. From Figure 10, we see that the values of AUC and AUPR continually decrease with the increase of feature dimensions. The precision score first increases at dimension 64 and then remains stable. Over the recall score, the F1 score and the MCC, MHGNN reaches peak results with dimension 64. On Hetero-B, MHGNN obtains the results near 1 over AUC and AUPR on different feature dimensions. Over the recall score, the F1 score and the MCC, the results of MHGNN first increase at dimension 64. Afterwards, the results decreases sharply with higher dimensions.

For the rest of experiments, we adopt the metapath combination G4 for dataset Hetero-A and metapath combination G3 for dataset Hetero-B. This is because dataset Hetero-B contains much more interactions, and metapath combination G3 is sufficient enough for MHGNN to achieve promising results (see Figure 6). For Hetero-C, we use metapath combination G3. We employ the max-pooling operation as the metapath instance encoder for datasets Hetero-A and Hetero-B, and the relational rotation operation as the metapath instance encoder for dataset Hetero-C. We utilize 8 attention heads and set hidden features with the dimension of 64 for all datasets.

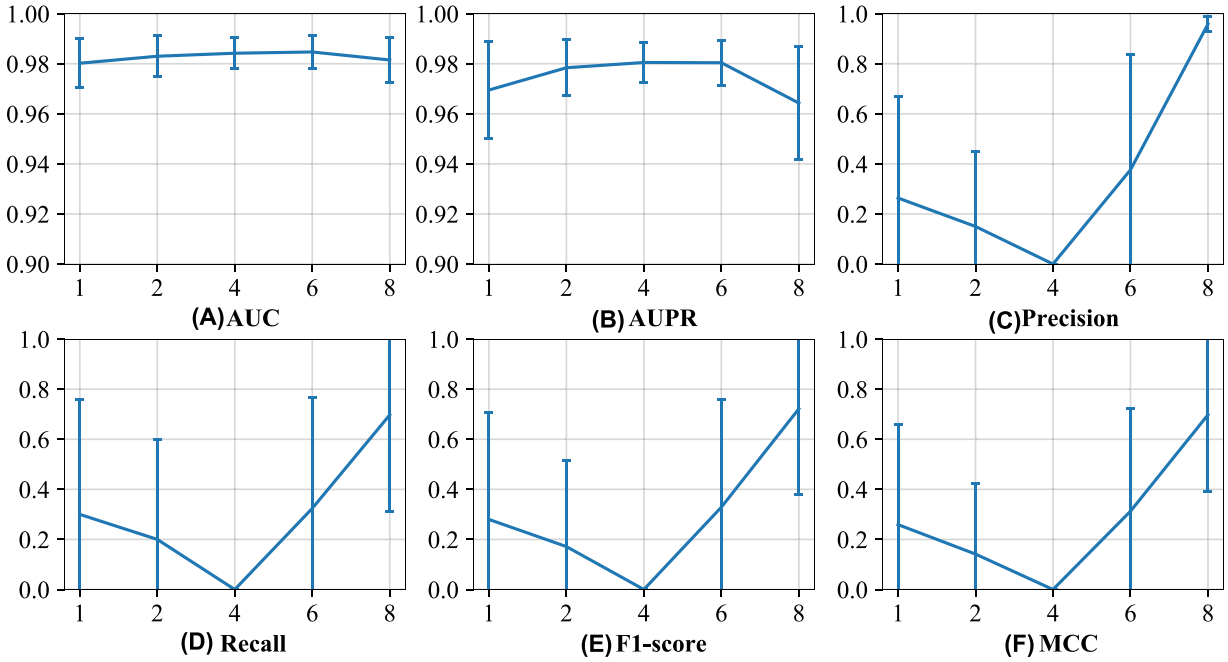


**Figure 8.** Impacts of different number of attention heads on Hetero-A.

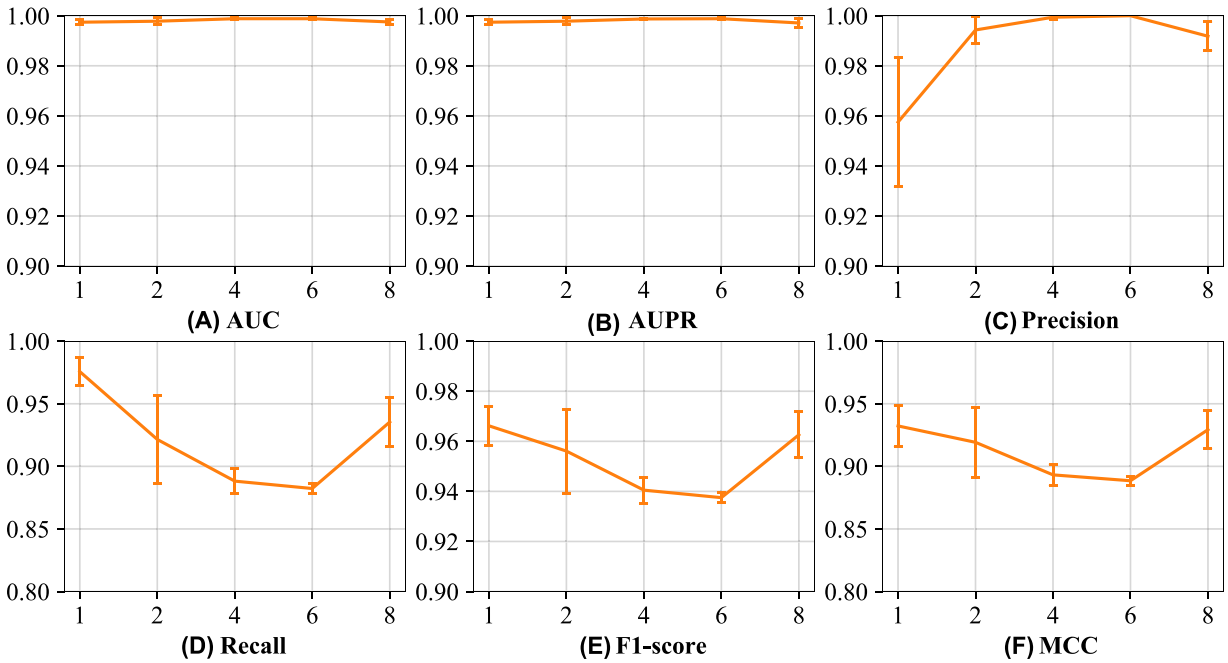


**Figure 9.** Impacts of different number of attention heads on Hetero-B.

## Comparison with baselines

Comparison results of MHGNN with baselines on datasets Hetero-A, Hetero-B and Hetero-C are presented in Tables 6–8. It is noticed that MHGNN yields the highest performances from five of six evaluation metrics over all baselines on Hetero-A. On dataset Hetero-B, MHGNN significantly exceeds all baselines with improvements of 0.0766, 0.0725, 0.0791, 0.0612, 0.1177 and 0.2327 over metrics of AUC, AUPR, the precision score, the recall score, the F1 score and the MCC. On dataset Hetero-C, MHGNN surpasses all baselines over all evaluation metrics as well. The results demonstrate the superiority of MHGNN in DTI prediction. Compared with the similarity-based methods, MHGNN is capable of capturing high-order relationships in the biological heterogeneous graph and predicting DTIs in an end-to-end manner. However, the similarity-based methods are deficient in modeling complex data correlations. Additionally, they separate feature learning from the prediction task, resulting in suboptimal solutions for DTI prediction as the learned representations of drugs and targets may deviate from the prediction task. Compared with the KG-based methods, MHGNN enables the learning of deep representations of drugs and targets.

Similar to our work, IMCHGAN adopts metapaths and GAT to predict DTIs. Nevertheless, it is limited to metapaths in each individual network, which cannot capture high-order semantic

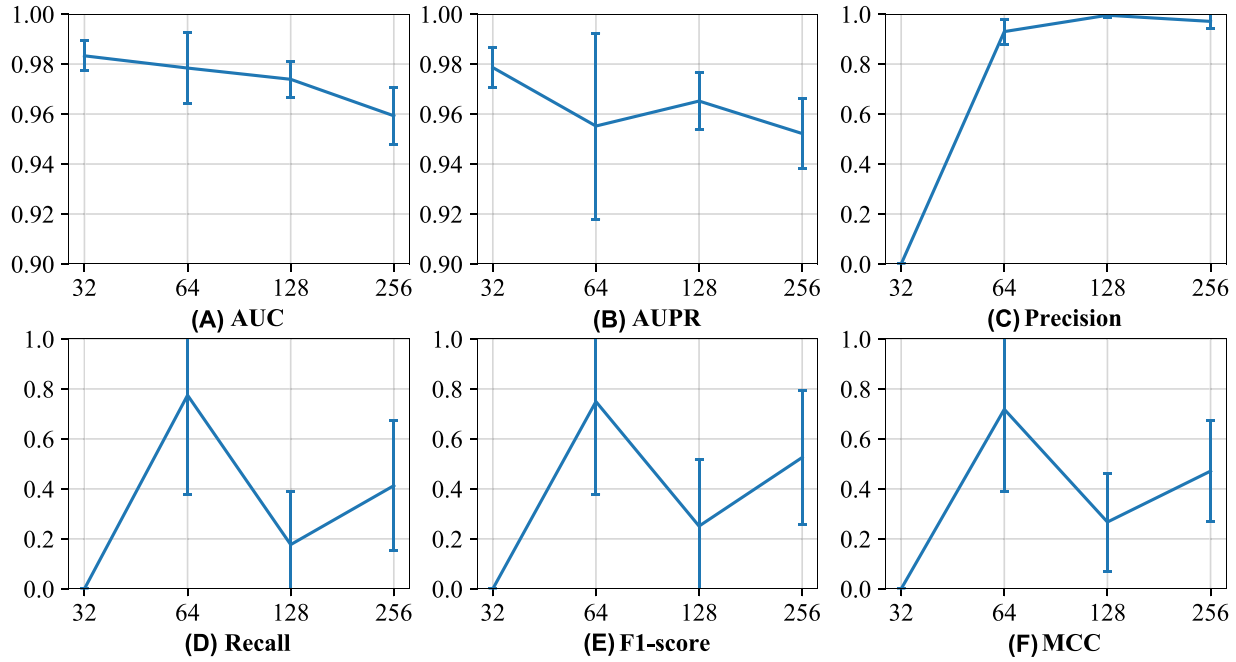


**Figure 10.** Impacts of different hidden feature dimensions on Hetero-A.

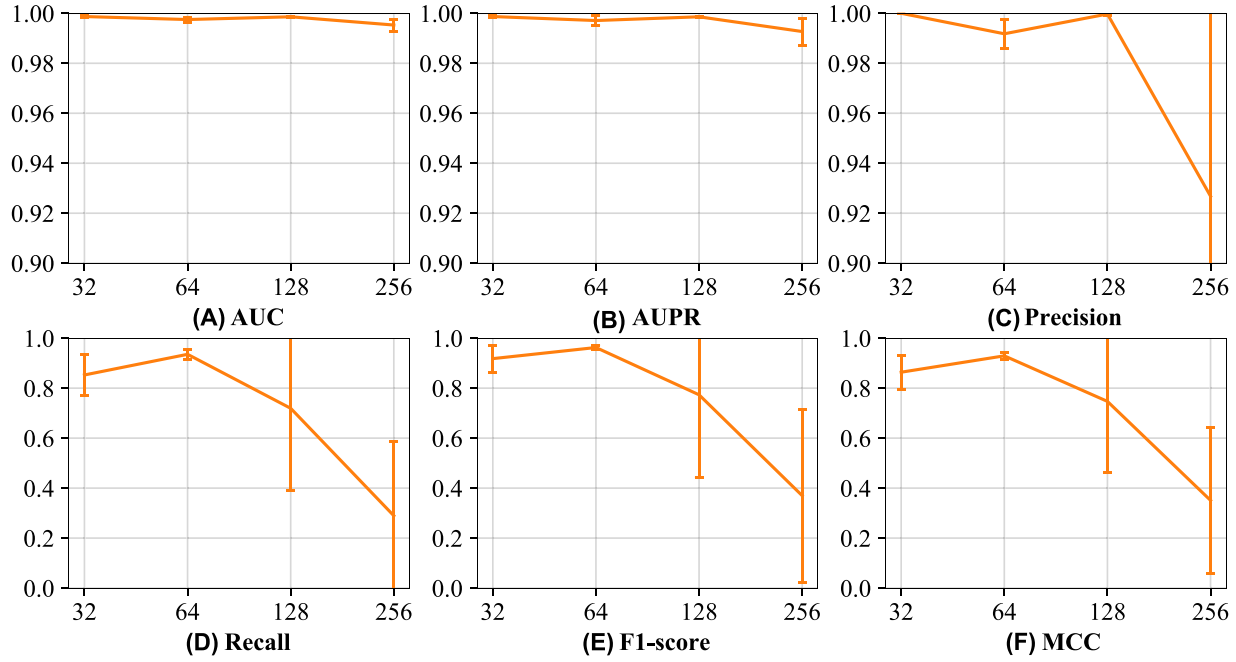


**Figure 11.** Impacts of different hidden feature dimensions on Hetero-B.

information inherent in biological heterogeneous network. Furthermore, IMCHGAN utilizes the inductive matrix completion to predict DTIs and neglects the high-order correlations between DTPs. SGCL-DTI is the concurrent work of MHGNN. It follows the same metapath construction strategy as IMCHAN. In SGCL-DTI, the topology graph and the semantic graph built with DTPs as nodes are highly overlapping. The former is obtained by assigning labels 1 between DTPs with common drugs or targets, and 0 otherwise. The latter is extracted from the most similar DTPs for each DTP. Since DTPs with common drugs or targets are prone to share higher similarities than these without. Hence, these two graphs are highly redundant. Moreover, both IMCHGAN and SGCL-DTI violate the principle that the labels of DTPs in the test set cannot be seen during the training process as they use labels of DTPs in the test set for metapath construction during the training process.

DTI-MGNN constructs DTP graph with DTPs as nodes. However, it assigns DTPs with hard associations (0/1). Also, it separates featuring learning from DTI prediction task. Contrarily, MHGNN assigns different weights to different DTP associations and optimizes the feature learning module and the DTI prediction module in an end-to-end manner. Additionally, comparing MHGNN with HAN and MAGNN, we can conclude that directly applying heterogenous graph embedding methods to DTI prediction task cannot obtain promising results due to domain gap and task gap.

**Table 6.** Comparison of MHGNN with 17 methods on dataset Hetero-A. The highest and the second highest results over each measurement are in bold and underlined, respectively.

| Methods | | AUC$_{\pm std}$ | AUPR$_{\pm std}$ | Precision$_{\pm std}$ | Recall$_{\pm std}$ | F1$_{\pm std}$ | MCC$_{\pm std}$ |
|---|---|---|---|---|---|---|---|
| **Similarity -based** | DTNet | 0.8838$_{\pm 0.022}$ | 0.9024$_{\pm 0.017}$ | 0.8677$_{\pm 0.024}$ | 0.7369$_{\pm 0.024}$ | 0.7967$_{\pm 0.020}$ | 0.6316$_{\pm 0.036}$ |
| | deepDTNet | 0.7688$_{\pm 0.040}$ | 0.7744$_{\pm 0.040}$ | 0.6970$_{\pm 0.031}$ | 0.6938$_{\pm 0.044}$ | 0.6952$_{\pm 0.036}$ | 0.3928$_{\pm 0.065}$ |
| | MEDTI | 0.7421$_{\pm 0.038}$ | 0.7428$_{\pm 0.029}$ | 0.6796$_{\pm 0.030}$ | 0.7036$_{\pm 0.038}$ | 0.6913$_{\pm 0.033}$ | 0.3726$_{\pm 0.064}$ |
| | NEDTP | 0.9115$_{\pm 0.021}$ | 0.919$_{\pm 0.021}$ | 0.8698$_{\pm 0.023}$ | 0.7993$_{\pm 0.037}$ | 0.8328$_{\pm 0.029}$ | 0.6822$_{\pm 0.050}$ |
| | MultiDTI | 0.8951$_{\pm 0.016}$ | 0.9126$_{\pm 0.015}$ | 0.8253$_{\pm 0.127}$ | 0.7882$_{\pm 0.106}$ | 0.7898$_{\pm 0.031}$ | 0.6007$_{\pm 0.098}$ |
| **KG-based** | TransE | 0.7435$_{\pm 0.035}$ | 0.7974$_{\pm 0.025}$ | 0.6496$_{\pm 0.042}$ | 0.7051$_{\pm 0.035}$ | 0.6759$_{\pm 0.036}$ | 0.3242$_{\pm 0.082}$ |
| | TransE-NFM | 0.9247$_{\pm 0.014}$ | 0.9286$_{\pm 0.014}$ | 0.8770$_{\pm 0.019}$ | 0.8201$_{\pm 0.024}$ | 0.8473$_{\pm 0.016}$ | 0.7065$_{\pm 0.030}$ |
| | DistMult | 0.7436$_{\pm 0.035}$ | 0.7975$_{\pm 0.025}$ | 0.6492$_{\pm 0.043}$ | 0.7046$_{\pm 0.033}$ | 0.6754$_{\pm 0.036}$ | 0.3232$_{\pm 0.083}$ |
| | KGE_NFM | 0.9250$_{\pm 0.014}$ | 0.9285$_{\pm 0.015}$ | 0.8839$_{\pm 0.025}$ | 0.8164$_{\pm 0.026}$ | 0.8484$_{\pm 0.017}$ | 0.7110$_{\pm 0.032}$ |
| | ComplEx | 0.7436$_{\pm 0.035}$ | 0.7977$_{\pm 0.025}$ | 0.6515$_{\pm 0.042}$ | 0.7051$_{\pm 0.033}$ | 0.6769$_{\pm 0.036}$ | 0.3272$_{\pm 0.081}$ |
| | ComplEx-NFM | 0.9242$_{\pm 0.013}$ | 0.9265$_{\pm 0.017}$ | 0.8784$_{\pm 0.019}$ | 0.8180$_{\pm 0.029}$ | 0.8469$_{\pm 0.021}$ | 0.7066$_{\pm 0.039}$ |
| **GNN -based** | NeoDTI | 0.8690$_{\pm 0.052}$ | 0.8821$_{\pm 0.047}$ | 0.7959$_{\pm 0.074}$ | 0.7904$_{\pm 0.038}$ | 0.7918$_{\pm 0.050}$ | 0.5825$_{\pm 0.116}$ |
| | IMCHGAN | 0.8862$_{\pm 0.021}$ | 0.9000$_{\pm 0.021}$ | 0.8747$_{\pm 0.023}$ | 0.7349$_{\pm 0.046}$ | 0.7982$_{\pm 0.033}$ | 0.6384$_{\pm 0.053}$ |
| | DTI-MGNN | 0.8969$_{\pm 0.026}$ | 0.8940$_{\pm 0.025}$ | 0.8725$_{\pm 0.028}$ | 0.8133$_{\pm 0.027}$ | 0.8418$_{\pm 0.025}$ | 0.6960$_{\pm 0.049}$ |
| | SGCL-DTI | 0.7462$_{\pm 0.108}$ | 0.7910$_{\pm 0.099}$ | 0.7415$_{\pm 0.097}$ | 0.628$_{\pm 0.133}$ | 0.6768$_{\pm 0.115}$ | 0.4193$_{\pm 0.186}$ |
| | HAN | 0.9484$_{\pm 0.003}$ | 0.9478$_{\pm 0.002}$ | 0.7902$_{\pm 0.021}$ | 0.9271$_{\pm 0.010}$ | 0.8528$_{\pm 0.008}$ | 0.6908$_{\pm 0.018}$ |
| | MAGNN | 0.9531$_{\pm 0.011}$ | 0.9410$_{\pm 0.018}$ | **0.9469**$_{\pm 0.025}$ | 0.5376$_{\pm 0.090}$ | 0.6803$_{\pm 0.081}$ | 0.5619$_{\pm 0.066}$ |
| | MHGNN | **0.9727**$_{\pm 0.011}$ | **0.9552**$_{\pm 0.026}$ | 0.9316$_{\pm 0.021}$ | **0.9886**$_{\pm 0.009}$ | **0.9590**$_{\pm 0.009}$ | **0.9174**$_{\pm 0.019}$ |

**Table 7.** Comparison of MHGNN with 17 methods on dataset Hetero-B. The highest and the second highest results over each measurement are in bold and underlined, respectively.

| Methods | | AUC$_{\pm std}$ | AUPR$_{\pm std}$ | Precision$_{\pm std}$ | Recall$_{\pm std}$ | F1$_{\pm std}$ | MCC$_{\pm std}$ |
|---|---|---|---|---|---|---|---|
| **Similarity -based** | DTNet | 0.8239$_{\pm 0.026}$ | 0.8378$_{\pm 0.023}$ | 0.7900$_{\pm 0.013}$ | 0.6727$_{\pm 0.024}$ | 0.7265$_{\pm 0.019}$ | 0.4996$_{\pm 0.030}$ |
| | deepDTNet | 0.8129$_{\pm 0.010}$ | 0.8165$_{\pm 0.099}$ | 0.7492$_{\pm 0.079}$ | 0.7234$_{\pm 0.077}$ | 0.7360$_{\pm 0.078}$ | 0.4816$_{\pm 0.153}$ |
| | MEDTI | 0.6870$_{\pm 0.018}$ | 0.7153$_{\pm 0.021}$ | 0.6375$_{\pm 0.025}$ | 0.6250$_{\pm 0.012}$ | 0.6311$_{\pm 0.017}$ | 0.2687$_{\pm 0.044}$ |
| | NEDTP | 0.9175$_{\pm 0.007}$ | 0.9220$_{\pm 0.006}$ | 0.8693$_{\pm 0.006}$ | 0.8059$_{\pm 0.013}$ | 0.8364$_{\pm 0.009}$ | 0.6867$_{\pm 0.015}$ |
| | MultiDTI | 0.8783$_{\pm 0.018}$ | 0.8883$_{\pm 0.015}$ | 0.8447$_{\pm 0.041}$ | 0.7647$_{\pm 0.052}$ | 0.8004$_{\pm 0.022}$ | 0.6254$_{\pm 0.038}$ |
| **KG-based** | TransE | 0.6720$_{\pm 0.018}$ | 0.6777$_{\pm 0.019}$ | 0.6427$_{\pm 0.015}$ | 0.6254$_{\pm 0.017}$ | 0.6339$_{\pm 0.015}$ | 0.2778$_{\pm 0.029}$ |
| | TransE-NFM | 0.9161$_{\pm 0.008}$ | 0.9124$_{\pm 0.009}$ | 0.8584$_{\pm 0.012}$ | 0.8212$_{\pm 0.020}$ | 0.8393$_{\pm 0.013}$ | 0.6864$_{\pm 0.023}$ |
| | DistMult | 0.7441$_{\pm 0.012}$ | 0.7677$_{\pm 0.013}$ | 0.6514$_{\pm 0.010}$ | 0.7011$_{\pm 0.013}$ | 0.6753$_{\pm 0.010}$ | 0.3268$_{\pm 0.020}$ |
| | KGE_NFM | 0.9149$_{\pm 0.008}$ | 0.9151$_{\pm 0.009}$ | 0.8526$_{\pm 0.012}$ | 0.8246$_{\pm 0.019}$ | 0.8382$_{\pm 0.012}$ | 0.6825$_{\pm 0.023}$ |
| | ComplEx | 0.7440$_{\pm 0.012}$ | 0.7676$_{\pm 0.013}$ | 0.6519$_{\pm 0.010}$ | 0.7015$_{\pm 0.013}$ | 0.6757$_{\pm 0.009}$ | 0.3278$_{\pm 0.019}$ |
| | ComplEx-NFM | 0.9161$_{\pm 0.009}$ | 0.9167$_{\pm 0.011}$ | 0.8524$_{\pm 0.012}$ | 0.8261$_{\pm 0.020}$ | 0.8389$_{\pm 0.014}$ | 0.6834$_{\pm 0.027}$ |
| **GNN -based** | NeoDTI | 0.9072$_{\pm 0.031}$ | 0.9078$_{\pm 0.032}$ | 0.8176$_{\pm 0.055}$ | 0.8387$_{\pm 0.023}$ | 0.8272$_{\pm 0.035}$ | 0.6488$_{\pm 0.081}$ |
| | IMCHGAN | 0.8706$_{\pm 0.021}$ | 0.8863$_{\pm 0.018}$ | 0.8510$_{\pm 0.018}$ | 0.7427$_{\pm 0.027}$ | 0.7930$_{\pm 0.020}$ | 0.6179$_{\pm 0.034}$ |
| | DTI-MGNN | 0.8306$_{\pm 0.034}$ | 0.8474$_{\pm 0.027}$ | 0.9097$_{\pm 0.027}$ | 0.4945$_{\pm 0.123}$ | 0.6278$_{\pm 0.129}$ | 0.4966$_{\pm 0.082}$ |
| | SGCL-DTI | 0.6945$_{\pm 0.014}$ | 0.7649$_{\pm 0.007}$ | 0.7706$_{\pm 0.025}$ | 0.5269$_{\pm 0.016}$ | 0.6255$_{\pm 0.013}$ | 0.3895$_{\pm 0.027}$ |
| | HAN | 0.8532$_{\pm 0.013}$ | 0.7988$_{\pm 0.018}$ | 0.7693$_{\pm 0.014}$ | 0.8667$_{\pm 0.018}$ | 0.8149$_{\pm 0.010}$ | 0.6116$_{\pm 0.021}$ |
| | MAGNN | 0.9198$_{\pm 0.058}$ | 0.8995$_{\pm 0.070}$ | 0.9089$_{\pm 0.064}$ | 0.5450$_{\pm 0.123}$ | 0.6754$_{\pm 0.111}$ | 0.5392$_{\pm 0.122}$ |
| | MHGNN | **0.9964**$_{\pm 0.002}$ | **0.9945**$_{\pm 0.003}$ | **0.9888**$_{\pm 0.009}$ | **0.9279**$_{\pm 0.029}$ | **0.9570**$_{\pm 0.012}$ | **0.9194**$_{\pm 0.020}$ |

## Cold-start analysis

In real-world applications, there is usually little information about the interaction between a drug and a target, which greatly degrades model generalization capability. To analyse the robustness of MHGNN with respect to cold start DTPs, we conduct the cold start analysis. Following SGCL-DTI [14], we select DTPs with no more than $L$ relationships as cold-start ones, where $L \in \{0, 3, 5\}$ for dataset Hetero-A and $L \in \{3, 5, 8\}$ for dataset Hetero-B. The results in Table 9 show that MHGNN can still achieve promising performances in cold-start scenarios.

## Case study

As shown in Figure 4, the original datasets contain homologous targets and similar drugs, which may cause the inflated evaluation performance. To investigate the robustness of MHGNN, we further conduct tests by removing positive DTIs with high homology from the training set and then predicting them in the test phase. We remove DTIs with high homology in the following five cases as [22]: (i) removing DTIs involving homologous proteins (sequence identity scores > 40%); (ii) removing DTIs with similar drugs (Tanimoto coefficients > 60%); (iii) removing DTIs with the drugs sharing similar side-effects (Jaccard similarity scores > 60%); (iv) removing DTIs DTIs with the drugs or proteins associated with similar diseases (Jaccard similarity scores > 60%) and (v) removing DTIs with either similar drugs (Tanimoto coefficients > 60%) or homologous proteins (sequence identity scores > 40%). The test results on dataset Hetero-A over the above cases (Table 10) show that MHGNN is robust against the removal of homologous proteins or similar drugs in the training data. We randomly select two drugs and show their predicted targets in Table 11.

**Table 8.** Comparison of MHGNN with 17 methods on dataset Hetero-C. The highest and the second highest results over each measurement are in bold and underlined, respectively.

| Methods | | $AUC_{\pm std}$ | $AUPR_{\pm std}$ | $Precision_{\pm std}$ | $Recall_{\pm std}$ | $F1_{\pm std}$ | $MCC_{\pm std}$ |
|---|---|---|---|---|---|---|---|
| **Similarity-based** | DTNet | $0.8672_{\pm 0.015}$ | $0.8649_{\pm 0.013}$ | $0.8186_{\pm 0.014}$ | $0.7483_{\pm 0.024}$ | $0.7818_{\pm 0.018}$ | $0.5849_{\pm 0.031}$ |
| | deepDTNet | $0.7894_{\pm 0.018}$ | $0.7682_{\pm 0.020}$ | $0.7111_{\pm 0.013}$ | $0.7287_{\pm 0.029}$ | $0.7197_{\pm 0.021}$ | $0.4330_{\pm 0.035}$ |
| | MEDTI | $0.7170_{\pm 0.007}$ | $0.6884_{\pm 0.008}$ | $0.6535_{\pm 0.007}$ | $0.6787_{\pm 0.009}$ | $0.6658_{\pm 0.008}$ | $0.3191_{\pm 0.015}$ |
| | NEDTP | $0.9146_{\pm 0.005}$ | $0.9076_{\pm 0.006}$ | $0.8655_{\pm 0.006}$ | $0.8142_{\pm 0.008}$ | $0.8390_{\pm 0.005}$ | $0.6889_{\pm 0.009}$ |
| | MultiDTI | $0.8805_{\pm 0.016}$ | $0.8739_{\pm 0.016}$ | $0.7930_{\pm 0.066}$ | $0.8129_{\pm 0.078}$ | $0.7971_{\pm 0.032}$ | $0.5957_{\pm 0.065}$ |
| **KG-based** | TransE | $0.6328_{\pm 0.021}$ | $0.6503_{\pm 0.028}$ | $0.5773_{\pm 0.012}$ | $0.6560_{\pm 0.016}$ | $0.6142_{\pm 0.013}$ | $0.1774_{\pm 0.028}$ |
| | TransE-NFM | $0.9059_{\pm 0.007}$ | $0.8991_{\pm 0.007}$ | $0.8490_{\pm 0.009}$ | $0.8233_{\pm 0.023}$ | $0.8357_{\pm 0.010}$ | $0.6772_{\pm 0.013}$ |
| | DistMult | $0.9207_{\pm 0.008}$ | $0.9195_{\pm 0.008}$ | $0.8634_{\pm 0.009}$ | $0.8193_{\pm 0.015}$ | $0.8407_{\pm 0.009}$ | $0.6906_{\pm 0.017}$ |
| | KGE_NFM | $0.9251_{\pm 0.005}$ | $0.9123_{\pm 0.007}$ | $0.8462_{\pm 0.013}$ | $0.8618_{\pm 0.013}$ | $0.8538_{\pm 0.007}$ | $0.7051_{\pm 0.015}$ |
| | ComplEx | $0.7978_{\pm 0.011}$ | $0.8040_{\pm 0.011}$ | $0.7038_{\pm 0.011}$ | $0.7099_{\pm 0.014}$ | $0.7068_{\pm 0.012}$ | $0.4111_{\pm 0.022}$ |
| | ComplEx-NFM | $0.9253_{\pm 0.006}$ | $0.9121_{\pm 0.008}$ | $0.8481_{\pm 0.015}$ | $0.8628_{\pm 0.022}$ | $0.8551_{\pm 0.009}$ | $0.7084_{\pm 0.016}$ |
| **GNN -based** | NeoDTI | $0.9160_{\pm 0.010}$ | $0.8981_{\pm 0.012}$ | $0.8365_{\pm 0.010}$ | $0.8605_{\pm 0.010}$ | $0.8483_{\pm 0.008}$ | $0.6925_{\pm 0.017}$ |
| | IMCHGAN | $0.7836_{\pm 0.032}$ | $0.7911_{\pm 0.031}$ | $0.7566_{\pm 0.046}$ | $0.6651_{\pm 0.096}$ | $0.7014_{\pm 0.060}$ | $0.4520_{\pm 0.059}$ |
| | DTI-MGNN | $0.8988_{\pm 0.003}$ | $0.8966_{\pm 0.004}$ | $0.8324_{\pm 0.021}$ | $0.8239_{\pm 0.027}$ | $0.8274_{\pm 0.004}$ | $0.6578_{\pm 0.003}$ |
| | SGCL-DTI | $0.7394_{\pm 0.015}$ | $0.7888_{\pm 0.011}$ | $\underline{0.8864}_{\pm 0.051}$ | $0.4265_{\pm 0.087}$ | $0.5679_{\pm 0.069}$ | $0.4299_{\pm 0.036}$ |
| | HAN | $0.9217_{\pm 0.019}$ | $0.9006_{\pm 0.028}$ | $0.7670_{\pm 0.028}$ | $0.9602_{\pm 0.007}$ | $0.8525_{\pm 0.018}$ | $0.6898_{\pm 0.040}$ |
| | MAGNN | $\underline{0.9629}_{\pm 0.004}$ | $\underline{0.9562}_{\pm 0.005}$ | $0.8667_{\pm 0.008}$ | $\underline{0.9498}_{\pm 0.007}$ | $\underline{0.9063}_{\pm 0.007}$ | $\underline{0.8074}_{\pm 0.014}$ |
| | MHGNN | $\mathbf{0.9987}_{\pm 0.000}$ | $\mathbf{0.9986}_{\pm 0.000}$ | $\mathbf{0.9456}_{\pm 0.009}$ | $\mathbf{0.9999}_{\pm 0.000}$ | $\mathbf{0.9720}_{\pm 0.005}$ | $\mathbf{0.9440}_{\pm 0.009}$ |

**Table 9.** Cold-start studies on datasets Hetero-A and Hetero-B. 0, 3, 5 and 8 denote the cold-start DTPs with no more than 0, 3, 5 and 8 relationships, respectively.

| Hetero-A | | | | | | | Hetero-B | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **No.** | **AUC** | **AUPR** | **Precision** | **Recall** | **F1** | **MCC** | **No.** | **AUC** | **AUPR** | **Precision** | **Recall** | **F1** | **MCC** |
| 0 | 1.0000 | 1.0000 | 1.0000 | 0.3333 | 0.5000 | 0.4472 | 3 | 0.9765 | 0.9762 | 1.0000 | 0.6684 | 0.8013 | 0.7085 |
| 3 | 0.9754 | 0.9465 | 0.9507 | 0.9650 | 0.9578 | 0.9151 | 5 | 0.9952 | 0.9913 | 0.9935 | 1.0000 | 0.9967 | 0.9935 |
| 5 | 0.9699 | 0.9436 | 0.9063 | 0.9564 | 0.9307 | 0.8589 | 8 | 0.9753 | 0.9745 | 0.8964 | 1.0000 | 0.9454 | 0.8904 |

**Table 10.** Case studies on dataset Hetero-A.

| Cases | AUC | AUPR | Precision | Recall | F1 | MCC |
|---|---|---|---|---|---|---|
| C1 | 0.9993 | 0.9993 | 0.9218 | 1.0 | 0.9593 | 0.9185 |
| C2 | 0.9501 | 0.8688 | 0.8079 | 0.9993 | 0.8935 | 0.7841 |
| C3 | 0.9316 | 0.8649 | 0.8629 | 0.9869 | 0.9207 | 0.8388 |
| C4 | 0.9386 | 0.8462 | 0.9166 | 0.9442 | 0.9302 | 0.8587 |
| C5 | 0.9104 | 0.7975 | 0.8159 | 0.9904 | 0.8947 | 0.7851 |

**Table 11.** Examples of case studies on dataset Hetero-A.

| DrugBank ID: Generic name | UniProt ID: Gene name | Result |
|---|---|---|
| DB00162: Vitamin A | P10745: RBP3 | True |
| DB00162: Vitamin A | P09455: BP1 | True |
| DB00162: Vitamin A | P00352: ALDH1A1 | True |
| DB00162: Vitamin A | Q92781: RDH5 | True |
| DB00162: Vitamin A | P12271: RLBP1 | True |
| DB00162: Vitamin A | O95237: LRAT | True |
| Accuracy | | 100% |
| DB00210: Adapalene | P28702: RXRB | True |
| DB00210: Adapalene | P13631: RARG | True |
| DB00210: Adapalene | P10826: RARB | True |
| DB00210: Adapalene | P19793: RXRA | True |
| DB00210: Adapalene | P10276: RARA | True |
| DB00210: Adapalene | P48443: RXRG | True |
| Accuracy | | 100% |

## Model generalization ability analysis

In reality, the verified interactions between drugs and targets are quite sparse, and the number of negative DTIs is much greater than that of positive DTIs. Aiming at analysing the generalization ability of MHGNN with imbalanced data distributions, we implement experiments on dataset Hetero-B under the positive and negative ratios of 1:1, 1:2, 1:5 and 1:10, respectively. From the results in Figure 12, we see that MHGNN have high generalization ability.

## Independent test analysis

In order to further validate the effectiveness of the proposed MHGNN, we conduct independent test analyses on dataset Hetero-B. We randomly split all positive DTIs into six partitions and select one as the independent test set. For the rest five partitions, at each run, we choose one as the validation set and regard the others as the training set. Additionally, for each set, we randomly select the same number of unknown DTPs as the number of positive DTIs as the negative samples. We compare MHGNN with 17 methods and report the comparison results in Table 12. We see that MHGNN yields the highest results over all methods in terms of all the evaluation metrics. Besides, MHGNN achieves comparable results to the case of 10-fold cross validation (Table 7). The comparison results demonstrate the effectiveness of MHGNN in DTI prediction.

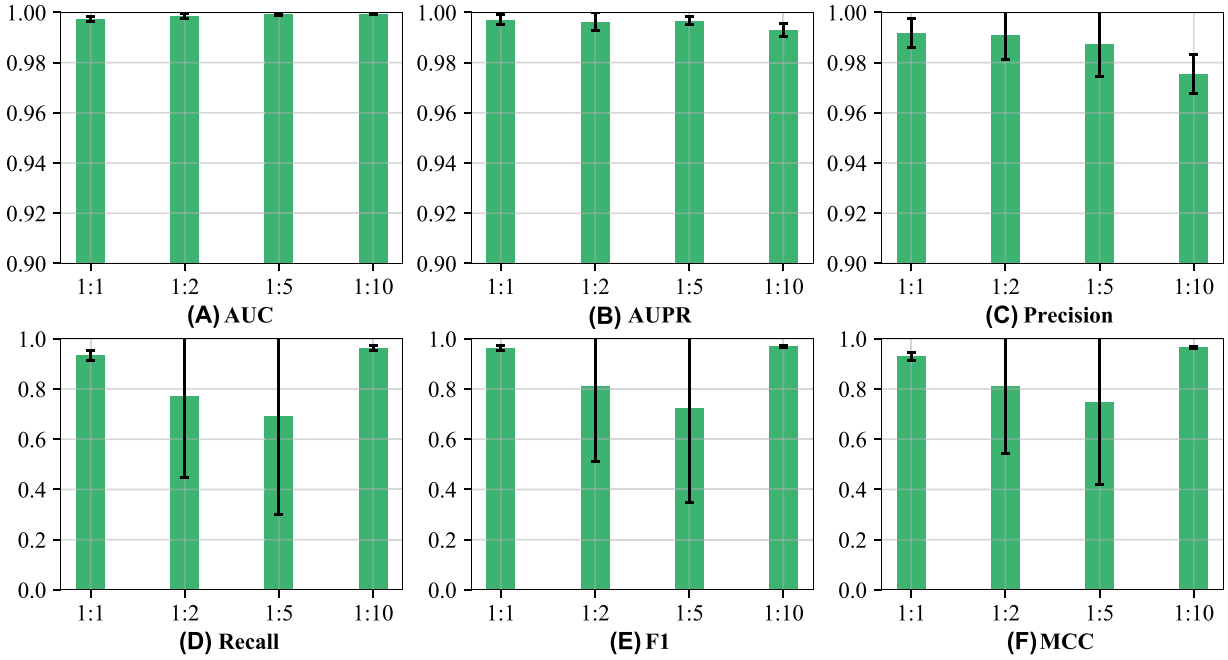


**Figure 12.** MHGNN generalization ability analysis on Hetero-B with different positive and negative ratios.

**Table 12.** Independent test analysis of MHGNN against 17 methods on dataset Hetero-B. The highest and the second highest results over each measurement are in bold and underlined, respectively.

| Methods | | $AUC_{\pm std}$ | $AUPR_{\pm std}$ | $Precision_{\pm std}$ | $Recall_{\pm std}$ | $F1_{\pm std}$ | $MCC_{\pm std}$ |
|---|---|---|---|---|---|---|---|
| **Similarity -based** | DTNet | $0.8251_{\pm 0.030}$ | $0.8400_{\pm 0.024}$ | $0.7854_{\pm 0.021}$ | $0.6766_{\pm 0.042}$ | $0.7267_{\pm 0.033}$ | $0.4973_{\pm 0.051}$ |
| | deepDTNet | $0.8514_{\pm 0.009}$ | $0.8563_{\pm 0.009}$ | $0.7886_{\pm 0.011}$ | $0.7634_{\pm 0.011}$ | $0.7758_{\pm 0.010}$ | $0.5591_{\pm 0.020}$ |
| | MEDTI | $0.7032_{\pm 0.008}$ | $0.7306_{\pm 0.008}$ | $0.6541_{\pm 0.014}$ | $0.6233_{\pm 0.004}$ | $0.6382_{\pm 0.007}$ | $0.2937_{\pm 0.021}$ |
| | NEDTP | $0.9075_{\pm 0.002}$ | $0.9128_{\pm 0.002}$ | $0.8578_{\pm 0.001}$ | $0.7933_{\pm 0.007}$ | $0.8243_{\pm 0.004}$ | $0.6637_{\pm 0.006}$ |
| | MultiDTI | $0.8569_{\pm 0.015}$ | $0.8677_{\pm 0.012}$ | $0.8344_{\pm 0.046}$ | $0.7212_{\pm 0.047}$ | $0.7710_{\pm 0.010}$ | $0.5808_{\pm 0.023}$ |
| **KG-based** | TransE | $0.6611_{\pm 0.004}$ | $0.6707_{\pm 0.004}$ | $0.6006_{\pm 0.003}$ | $0.6923_{\pm 0.011}$ | $0.6432_{\pm 0.005}$ | $0.2347_{\pm 0.007}$ |
| | TransE-NFM | $0.9322_{\pm 0.004}$ | $0.9202_{\pm 0.004}$ | $0.8674_{\pm 0.008}$ | $0.8384_{\pm 0.009}$ | $0.8526_{\pm 0.004}$ | $0.7106_{\pm 0.008}$ |
| | DistMult | $0.8202_{\pm 0.004}$ | $0.8211_{\pm 0.004}$ | $0.7384_{\pm 0.005}$ | $0.7337_{\pm 0.004}$ | $0.7361_{\pm 0.004}$ | $0.4738_{\pm 0.008}$ |
| | KGE_NFM | $0.9415_{\pm 0.002}$ | $\underline{0.9336}_{\pm 0.002}$ | $0.8716_{\pm 0.004}$ | $0.8578_{\pm 0.009}$ | $0.8646_{\pm 0.003}$ | $0.7315_{\pm 0.004}$ |
| | ComplEx | $0.7849_{\pm 0.004}$ | $0.7869_{\pm 0.007}$ | $0.7066_{\pm 0.001}$ | $0.6981_{\pm 0.005}$ | $0.7023_{\pm 0.003}$ | $0.4083_{\pm 0.005}$ |
| | ComplEx-NFM | $\underline{0.9431}_{\pm 0.005}$ | $0.9335_{\pm 0.005}$ | $0.8790_{\pm 0.007}$ | $0.8567_{\pm 0.011}$ | $\underline{0.8677}_{\pm 0.008}$ | $\underline{0.7391}_{\pm 0.014}$ |
| **GNN -based** | NeoDTI | $0.9061_{\pm 0.011}$ | $0.9067_{\pm 0.011}$ | $0.8370_{\pm 0.016}$ | $0.8256_{\pm 0.015}$ | $0.8312_{\pm 0.015}$ | $0.6648_{\pm 0.031}$ |
| | IMCHGAN | $0.8762_{\pm 0.006}$ | $0.8966_{\pm 0.006}$ | $0.8547_{\pm 0.012}$ | $0.7515_{\pm 0.009}$ | $0.7997_{\pm 0.007}$ | $0.6282_{\pm 0.014}$ |
| | DTI-MGNN | $0.8269_{\pm 0.080}$ | $0.8438_{\pm 0.058}$ | $0.7959_{\pm 0.090}$ | $0.7036_{\pm 0.105}$ | $0.7419_{\pm 0.087}$ | $0.5221_{\pm 0.157}$ |
| | SGCL-DTI | $0.5516_{\pm 0.144}$ | $0.5320_{\pm 0.144}$ | $0.5084_{\pm 0.010}$ | $0.7270_{\pm 0.384}$ | $0.6471_{\pm 0.033}$ | $0.0441_{\pm 0.025}$ |
| | HAN | $0.9086_{\pm 0.003}$ | $0.8818_{\pm 0.004}$ | $0.8087_{\pm 0.008}$ | $\underline{0.8745}_{\pm 0.013}$ | $0.8402_{\pm 0.005}$ | $0.6700_{\pm 0.010}$ |
| | MAGNN | $0.9085_{\pm 0.004}$ | $0.8747_{\pm 0.008}$ | $\underline{0.8871}_{\pm 0.010}$ | $0.4537_{\pm 0.060}$ | $0.5979_{\pm 0.052}$ | $0.4537_{\pm 0.039}$ |
| | MHGNN | $\mathbf{0.9893}_{\pm 0.004}$ | $\mathbf{0.9799}_{\pm 0.012}$ | $\mathbf{0.9633}_{\pm 0.004}$ | $\mathbf{0.9495}_{\pm 0.006}$ | $\mathbf{0.9563}_{\pm 0.003}$ | $\mathbf{0.9134}_{\pm 0.005}$ |

## Conclusion

In this paper, we have proposed MHGNN to capture complex structures and rich semantics in the biological heterogeneous graph for DTI prediction. MHGNN is designed to be a dual channel architecture to lean drug and target representations, respectively. Each channel is based on graph attention and metapath techniques. We further build a DTP correlation graph with DTPs as nodes to exploit the high-order correlations between DTPs. We have conducted comprehensive experiments on three biological heterogeneous datasets, and the results validate the effectiveness of MHGNN for DTI prediction. For the future work, we aim to develop a more flexible solution to capture high-order dependencies in biological heterogeneous graph for DTI prediction as the acquisition of metapaths requires expert knowledge and is inflexible.

**Key Points**

- A novel metapath-aggregated graph neural network is proposed to explore both high-order dependencies in the biological heterogeneous graph and high-order associations between DTPs for DTI prediction.
- A heterogeneous biological dataset is extended from the latest released resources. It contains newly

discovered DTIs, providing more samples for training and evaluation.
- MHGNN surpasses 17 state-of-the-art methods over six evaluation metrics with a good margin, which verifies its efficacy for DTI prediction.

## Funding

National Natural Science Foundation of China (62002178).

## References


1. Nosengo N. Can you teach old drugs new tricks? *Nature News* 2016;**534**(7607):314.
2. Gao KY, Fokoue A, Luo H, *et al.* Interpretable drug target prediction using deep neural representation. In *IJCAI*, International Joint Conferences on Artificial Intelligence Organization, California, USA, 2018. pp. 3371–7.
3. Huang K, Xiao C, Glass LM, *et al.* Moltrans: molecular interaction transformer for drug-target interaction prediction. *Bioinformatics* 2021;**37**(6):830–6.
4. Nguyen T, Le H, Quinn TP, *et al.* Graphdta: predicting drug–target binding affinity with graph neural networks. *Bioinformatics* 2021;**37**(8):1140–7.
5. Chu Y, Shan X, Chen T, *et al.* Dti-mlcd: predicting drug-target interactions using multi-label learning with community detection method. *Brief Bioinform* 2021;**22**(3):bbaa205.
6. Adasme MF, Parisi D, Sveshnikova A, *et al.* Structure-based drug repositioning: potential and limits. *Semin Cancer Biol* 2021;**68**:192–8.
7. Tanoli Z, Seemab U, Scherer A, *et al.* Exploration of databases and methods supporting drug repurposing: a comprehensive survey. *Brief Bioinform* 2021;**22**(2):1656–78.
8. Liu Z, Nguyen T-K, Fang Y. Tail-gnn: Tail-node graph neural networks. In *SIGKDD*, Association for Computing Machinery, New York, NY, United States, 2021. pp. 1109–19.
9. Hao J, Lei C, Efthymiou V, *et al.* Medto: Medical data to ontology matching using hybrid graph neural networks. In *SIGKDD*, Association for Computing Machinery, New York, NY, United State, 2021. pp. 2946–54.
10. Wan F, Hong L, Xiao A, *et al.* Neodti: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics* 2019;**35**(1):104–11.
11. Li Y, Qiao G, Wang K, *et al.* Drug–target interaction predication via multi-channel graph neural networks. *Brief Bioinform* 2021;**23**:bbab346.
12. Peng J, Wang Y, Guan J, *et al.* An end-to-end heterogeneous graph representation learning-based framework for drug–target interaction prediction. *Brief Bioinform* 2021;**22**(5):bbaa430.
13. Li J, Wang J, Lv H, *et al.* Imchgan: inductive matrix completion with heterogeneous graph attention networks for drug-target interactions prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**19**;1–1.
14. Li Y, Qiao G, Gao X, *et al.* Supervised graph co-contrastive learning for drug–target interaction prediction. *Bioinformatics* 2022;**38**(10):2847–54.
15. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In *ICLR*, OpenReview.net, 2016.
16. Veličković P, Cucurull G, Casanova A, *et al.* Graph attention networks. In *ICLR*, OpenReview.net, 2018.
17. Zhang C, Song D, Huang C, *et al.* Heterogeneous graph neural network. In *SIGKDD*, Association for Computing Machinery, New York, NY, United States, 2019. pp. 793–803.
18. Wang X, Ji H, Shi C, *et al.* Heterogeneous graph attention network. In *WWW*, Association for Computing Machinery, New York, NY, United States, 2019. pp. 2022–32.
19. Xinyu F, Zhang J, Meng Z, *et al.* Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. In *WWW*, Association for Computing Machinery, New York, NY, United States, 2020. pp. 2331–41.
20. Zhao J, Wang X, Shi C, *et al.* Network schema preserved heterogeneous information network embedding. In *IJCAI*, International Joint Conferences on Artificial Intelligence Organization, California, USA, 2020.
21. Dong Y, Chawla NV, Swami A. metapath2vec: Scalable representation learning for heterogeneous networks. In *SIGKDD*, Association for Computing Machinery, New York, NY, United States, 2017. 135–44.
22. Luo Y, Zhao X, Zhou J, *et al.* A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun* 2017;**8**(1):1–13.
23. Zeng X, Zhu S, Weiqiang L, *et al.* Target identification among known drugs by deep learning from heterogeneous networks. *Chem Sci* 2020;**11**(7):1775–97.
24. An Q, Liang Y. A heterogeneous network embedding framework for predicting similarity-based drug-target interactions. *Brief Bioinform* 2021;**22**(6):bbab275.
25. Shang Y, Gao L, Zou Q, *et al.* Prediction of drug-target interactions based on multi-layer network representation learning. *Neurocomputing* 2021;**434**:80–9.
26. Bordes A, Usunier N, Garcia-Duran A, *et al.* Translating embeddings for modeling multi-relational data. *NeurIPS* 2013;**26**: 1–9.
27. Lu J, Sun J, Wang Y, *et al.* Heterogeneous graph convolutional network integrates multi-modal similarities for drug-target interaction prediction. *BIBM*, IEEE, 2021;137–40.
28. Trouillon T, Welbl J, Riedel S, *et al.* Complex embeddings for simple link prediction. *ICML*, JMLR.org, 2016;2071–80.
29. Mohamed SK, Nounu A, Nováček V. Drug target discovery using knowledge graph embeddings. *SAC*, Association for Computing Machinery, New York, NY, United States, 2019;11–8.
30. Zhang S, Lin X, Zhang X. Discovering dti and ddi by knowledge graph with mhrw and improved neural network. *BIBM*, IEEE, 2021;588–93.
31. Ye Q, Hsieh C-Y, Yang Z, *et al.* A unified drug-target interaction prediction framework based on knowledge graph and recommendation system. *Nat Commun* 2021;**12**(1):1–12.
32. He X, Chua T-S. Neural factorization machines for sparse predictive analytics. In *SIGIR*, Association for Computing Machinery, New York, NY, United States, 2017.
33. Gilmer J, Schoenholz SS, Riley PF, *et al.* Neural message passing for quantum chemistry. In *ICML*, JMLR.org40. OpenReview.net, 2017. pp. 1263–72.
34. Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. *NeurIPS* 2017;**30**:1025–35.
35. Wishart DS, Feunang YD, Guo AC, *et al.* Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Res* 2018;**46**(D1):D1074–82.
36. Szklarczyk D, Gable AL, Nastou KC, *et al.* The string database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 2021;**49**(D1):D605–12.

37. Kuhn M, Letunic I, Jensen LJ, *et al.* The sider database of drugs and side effects. *Nucleic Acids Res* 2016;**44**(D1):D1075–9.
38. Davis AP, Grondin CJ, Johnson RJ, *et al.* Comparative toxicogenomics database (ctd): update 2021. *Nucleic Acids Res* 2021;**49**(D1):D1138–43.
39. UniProt Consortium. Uniprot: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;**47**(D1):D506–15.
40. Sun Z, Deng Z-H, Nie J-Y, *et al.* Rotate: Knowledge graph embedding by relational rotation in complex space. In *ICLR*, 2019.
41. Clevert D-A, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus). In *ICLR*, OpenReview.net, 2016.
42. Zheng Y, Peng H, Zhang X, *et al.* Predicting drug targets from heterogeneous spaces using anchor graph hashing and ensemble learning. In *IJCNN*, IEEE, Piscataway, NJ, USA, pp. 1–7, 2018.
43. Knox C, Law V, Jewison T, *et al.* Drugbank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res* International Business Machines Corp., 2010; **39**(suppl_1):D1035–41.
44. Keshava Prasad TS, Goel R, Kandasamy K, *et al.* Human protein reference database-2009 update. *Nucleic Acids Res* 2009; **37**(suppl_1):D767–72.
45. Davis AP, Murphy CG, Johnson R, *et al.* The comparative toxicogenomics database: update 2013. *Nucleic Acids Res* 2013;**41**(D1): D1104–14.
46. Kuhn M, Campillos M, Letunic I, *et al.* A side effect resource to capture phenotypic effects of drugs. *Nucleic Acids Res* 2010; **6**(1):343.
47. Law V, Knox C, Djoumbou Y, *et al.* Drugbank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 2014;**42**(D1): D1091–7.
48. Ursu O, Holmes J, Knockel J, *et al.* Drugcentral: online drug compendium. *Nucleic Acids Res* 2016;gkw993.
49. Chen B, Wild D, Guha R. Pubchem as a source of polypharmacology. *J Chem Inf Model* 2009;**49**(9):2044–55.
50. Gene Ontology Consortium. The gene ontology (go) database and informatics resource. *Nucleic Acids Res* 2004;**32**(suppl_1): D258–61.
51. Tanimoto TT. *Elementary mathematical theory of classification and prediction*. 1958;**45**.
52. Yamanishi Y, Araki M, Gutteridge A, *et al.* Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008;**24**(13): i232–40.
53. Zhou D, Zhijian X, Li WT, *et al.* Multidti: drug–target interaction prediction based on multi-modal representation learning to bridge the gap between new chemical entities and known heterogeneous network. *Bioinformatics* 2021;**37**(23): 4485–92.