Information Sciences 613 (2022) 507–523

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Hierarchical graph representation learning for the prediction of drug-target binding affinity

Zhaoyang Chu [a,1], Feng Huang [a,1], Haitao Fu [a], Yuan Quan [a], Xionghui Zhou [a], Shichao Liu [a,b,*], Wen Zhang [a,b,*]

[a] College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
[b] Agricultural Bioinformatics Key Laboratory of Hubei Province, Hubei Engineering Technology Research Center of Agricultural Big Data, Key Laboratory of Smart Animal Farming Technology, Ministry of Agriculture, Huazhong Agricultural University, Wuhan 430070, China

ARTICLE INFO

Article history:
Received 21 February 2022
Received in revised form 5 September 2022
Accepted 10 September 2022
Available online 17 September 2022

Keywords:
Graph representation learning
Binding affinity prediction
Hierarchical graph
Coarse-to-fine fusion
Drug discovery

ABSTRACT

Computationally predicting drug-target binding affinity (DTA) has attracted increasing attention due to its benefit for accelerating drug discovery. Currently, numerous deep learning-based prediction models have been proposed, often with a biencoder architecture that commonly focuses on how to extract expressive representations for drugs and targets but overlooks modeling explicit drug-target interactions. However, known DTA can provide underlying knowledge about how the drugs interact with targets that is beneficial for predictive accuracy. In this paper, we propose a novel hierarchical graph representation learning model for DTA prediction, named HGRL-DTA. The main contribution of our model is to establish a hierarchical graph learning architecture to integrate the coarse- and fine-level information from an affinity graph and drug/target molecule graphs, respectively, in a well-designed coarse-to-fine manner. In addition, we design a similarity-based representation inference method to infer coarse-level information when it is unavailable for new drugs or targets under the cold start scenario. Comprehensive experimental results under four scenarios across two benchmark datasets indicate that HGRL-DTA outperforms the state-of-the-art models in almost all cases.

1. Introduction

Drug discovery is an extremely time- and finance-consuming process that aims to locate a compound that can bind a given target protein to prevent the growth of the related disease. Compared to drug-target interactions containing information about whether a drug binds to a target, drug-target binding affinity (DTA) measured by dissociation constant (K_d), inhibition constant (K_i), or the half maximal inhibitory concentration (IC_{50}) provides richer information on the strength of the interaction of a drug-target pair and can be viewed as an important indicator for rapidly screening desired candidate drugs. Hence, efficient and accurate prediction of DTA by computational methods instead of identifying it by demanding experimental assays can accelerate drug discovery.

Traditional physics-based approaches, such as computational molecular docking and molecular dynamics simulation, still suffer from the challenge of scoring function design or computing resource consumption [42]. A promising remedy is the

* Corresponding author at: College of Informatics, Huazhong Agricultural University, Wuhan 430070, China (W. Zhang, S. Liu).
E-mail addresses: zhangwen@mail.hzau.edu.cn (W. Zhang), scliu@mail.hzau.edu.cn (S. Liu).
[1] The first two authors made equal contributions.

https://doi.org/10.1016/j.ins.2022.09.043
0020-0255/

data-driven regression model. Earlier attempts of classic machine learning models, including but not limited to random forest, logistic regression, support vector machine, Kronecker regularized least square, and gradient boosting, have not produced enough generalization on DTA prediction [8,26,35] due to their overreliance on complicated feature engineering that often heavily demands expert domain knowledge. Currently, deep learning for modeling DTA has become increasingly relevant, as it is able to capture hidden complex information that is hard to abstract according to human experience [9] and achieve remarkable success in predictive accuracy. The existing applications of deep learning in DTA prediction can be predominantly divided into two categories: structure- and nonstructure-based models [38].

Structure-based methods focus on the usage of spatial structure information of the protein-ligand complex. Most of them voxelized the complex as a 3D-grid representation by a set of physical-based descriptors and then applied 3D convolutional neural networks (CNNs) [10,29,36]. Several recent methods have developed geometry-oriented graph neural networks (GNNs) to address distance encoding or the protein-ligand complex molecule graphs [18,50]. Although these structure-based methods have achieved relatively high predictive performance, there is an apparent limitation: they do not work when the 3D structure of the complex is unavailable. In practice, the accurate 3D structure of the protein-ligand complex or even the protein is usually difficult to obtain, and predicting it by molecular docking or dynamics simulation remains challenging [11,12].

In the past decade, several nonstructure-based models have emerged to overcome the abovementioned limitation. The vast majority of them have a biencoder architecture similar to a Siamese network, where a pair of encoders separately learns representations for drugs and targets, followed by several fully connected (FC) layers to map the concatenated representations of drug-target pairs into prediction scores [9,23,25]. With different forms of input encoding, the backbone networks are distinct. Generally, both drug and target can be represented as 1D sequence data that are the simplified molecular-input line-entry system (SMILES) strings for drugs and amino acid sequences for targets, and some backbones commonly used in language models have been adequately studied for DTA prediction, such as CNN [1,25,31], recurrent neural network (RNN) [11], long short-term memory (LSTM) [47] and transformer [15]. Additionally, they can be converted into 2D graph data comprising molecular graphs with atoms as nodes and bonds as edges for drugs and residue-residue contact maps for targets, and then many GNN backbones are applied to DTA prediction [9,24]. The different combinations and hybrids of these encoders are assembled into various deep learning models. However, these models lose predictive accuracy, as they isolatedly cope with drugs and targets, ignoring realistic drug-target interactions whose benefit to DTA prediction has been demonstrated in some works [8,15,33].

In this work, we propose to model the DTA data as a hierarchical graph, also called a graph of graphs with inspiration from [3,7,41,43], where a set of graphs serve as nodes and constitute a graph. As shown in Fig. 1, in our constructed hierarchical graph, the coarse-level affinity graph consists of drug nodes, target nodes, and affinity weight edges; meanwhile, drug and target nodes are represented as molecular graphs, called fine-level graphs. We propose a novel hierarchical graph representation learning model for structure-free DTA prediction, named HGRL-DTA, to integrate coarse- and fine-level information in the hierarchical graph. Specifically, we adopt simple yet effective graph convolutional networks (GCNs) [13] as backbone encoders to learn coarse-level representations containing known DTA information (i.e., intermolecular interaction) and fine-level representations implying intramolecular structures. Then, the coarse-level information is mixed into fine-level representations in a well-designed coarse-to-fine manner to make full use of known DTA information. The fine-level representation enhanced by coarse-level information is then further refined by GCN encoders and makes up drug and target representations through readout operations. Finally, the representations for a drug and a target are concatenated together to featurize the drug-target pair, and the prediction scores are produced by a multilayer perceptron (MLP) taking the drug-target features as input. Moreover, to solve the cold start problem, we design a similarity-based representation inference method to deduce the coarse-level information for the new drug or target node based on external drug-drug similarities and target-target similarities. The advantages of our model lie in twofold aspects: 1) our model retains the biencoder

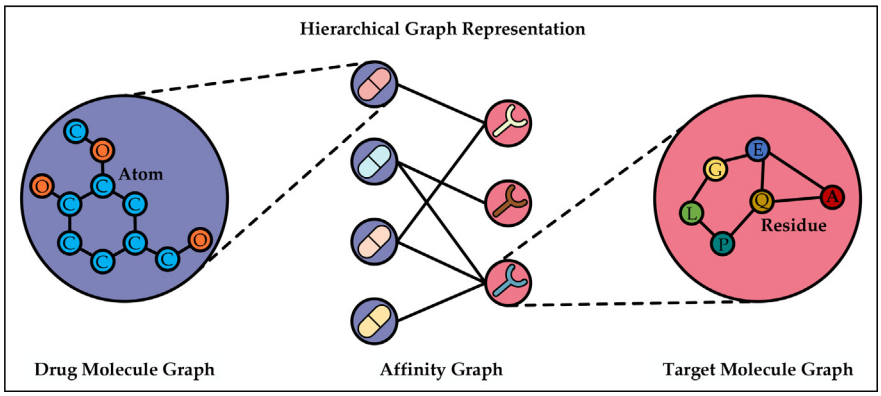

Fig. 1. An illustration of the hierarchical graph representation for DTA.

architecture if the coarse-level GCN-block and coarse-to-fine fusion strategy are neglected, and doing so will maintain the benefits of the in-depth representation extraction from drug and target molecule structure information; 2) our designed coarse-to-fine strategy can adaptively mix coarse-level information into the fine-level representation to enhance its expressive power. Experimental results demonstrate that even a simple GCN that uses our devised information fusion manner to effectively integrate known DTA information can beat the state-of-the-art methods in almost all cases. The main contributions of our work are summarized as follows:

- We bring about a new perspective in modeling DTA that represents DTA as a hierarchical graph.
- We propose a novel hierarchical graph representation learning model for DTA prediction, named HGRL-DTA. HGRL-DTA simulates intermolecular interactions and models intramolecular structures; moreover, it integrates hierarchical information in a well-designed coarse-to-fine manner.
- We design a similarity-based representation inference method to complement the generalization of HGRL-DTA on predicting affinity values for the new drugs or targets, where the coarse-level information of the new drug or target unavailable from the model is substituted with the aggregation of that from its most similar known drugs or targets.
- Extensive experiments under four experimental scenarios on two benchmark datasets are conducted to evaluate the performance of HGRL-DTA. Compared with several state-of-the-art methods, HGRL-DTA achieves significantly better performance in almost all cases.

2. Related work

In line with the focus of our work, we first detail recent advances in nonstructure-based deep learning for DTA prediction and then review some related literature about hierarchical graph representation learning.

2.1. Nonstructure-based deep learning for DTA prediction

Nonstructure-based deep learning methods aim to attribute DTA prediction to the intrinsic molecular nature of drugs and targets. Consequently, most of these methods present a biencoder architecture in which a pair of encoders are applied to learn latent representations for drugs and targets. DeepDTA [25] adopts a pair of CNNs to learn representations from drug SMILES strings and target protein residue sequences. It uses a simple concatenation operation to obtain drug-target pair representations that are then passed through an FC network to make DTA predictions. Based on DeepDTA, AttentionDTA [49] replaces the concatenation operation with an attention mechanism. With the advances of GNNs in molecular modeling, they have also been utilized in DTA prediction for handling drug or target molecule graphs. GraphDTA [23] retains the usage of CNNs for the target sequence and investigated 4 versions of GNNs for drug molecule graphs. DGraphDTA [9] uses the residue contact map of the protein as the target molecule graph and harnesses a pair of GNNs to handle drug and target graphs. Many later works modified these models by designing encoders with more complex architectures [11,46,47], imposing additional regularizers [12,17,19] or mixing multiview feature inputs [20,28,31]. For example, MGraphDTA [46] devises a multiscale GNN for drug encoding and a multiscale CNN for target encoding; MONN [17] considers an additional task for predicting pairwise atom-residue interactions; and DeepGS [20] encodes two input modalities of drugs by a graph attention network (GAT) and a bidirectional gated recurrent unit (BiGRU). These models held a similar architecture comprised of isolated encoders and ignored modeling the interaction between them. However, bridging the communication between the drug encoder and the target encoder is more rational and interpretable. Some works have attempted to provide interpretable prediction by attention links between drug and target encoders [1,24,16]. However, these attention weights stem from black-box optimization, which may uncontrolledly lead to contradictions with reality. DeepAffinity+ [12] and MONN [17] use extra supervised tasks to induce interpretability but inevitably demand real or imprecise predicted structure information in the training stage, although they remain structure-free in the testing stage.

For almost all models, known DTA only acts as supervisory signals for the target task. By contrast, known DTA has been used to manually compile input features of drugs and targets for some machine learning-based methods [8,33]. Few deep learning studies have investigated the benefit of known DTA information for final prediction. BERT-GCN [15] treats observed drug-target interactions as a graph where the nodes are attributed by the embeddings produced by a pair of pretrained BERT models. Then, a GCN was deployed on this attributed graph to simulate information interaction between drugs and targets. It finally adopted a previous FC network-based predictor to make DTA predictions. If stacking the pretrained BERTs, BERT-GCN actually revises the previous biencoder architecture by inserting a GCN before the decoder. Compared to some previous unsteerable attention-based interaction modules, its GCN directly models realistically meaningful drug-target interactions. However, there appear to be two deficiencies: 1) the GCN could be easily overfitted and perform suboptimally; 2) it readily oversmoothes the drug and target information, which could make the representations undistinguishable. Hence, how to more effectively integrate known DTA information remains challenging.

2.2. Hierarchical graph representation learning

Recent years have witnessed the powerful expressivity of GNNs in graph representation learning. Several works observe that some real-life data can be modeled as a graph representing interactions among a set of graph-structured entities [3,41],

such as drug-drug interactions (DDIs), drug-target interactions (DTIs) and protein-protein interactions (PPIs). BiGNN [3] and GoGNN [41] learn the graph-level representations for the graph-structured entities by a GNN and use another GNN to handle the entity interaction graph in which the learned representations serve as initial node features. This model architecture can naturally be extended to DTA prediction. However, the bottom GNN for molecular modeling suffers from troublesome training, such as inadequate training due to gradient vanishing, for which it may not be able to adequately capture the molecular structure. Moreover, it cannot work under the inductive settings that expect it to provide predictions for new nodes unseen in the entity interaction graph at the training stage. Accordingly, how to design effective information interactions between the entity graphs (i.e., molecular graph in this paper) and the interaction graph (i.e., affinity graph in this work) is a key challenge in hierarchical graph learning.

3. Preliminaries

This section introduces some core definitions for the convenience of describing and formulating our proposed method and target task.

Definition 3.1 (*Affinity Graph*). An affinity graph is a weighted graph $\mathscr{G} = \{\mathscr{V}, \mathscr{E}, \mathscr{W}\}$ depicting drug-target binding relations, where $\mathscr{V}$ is the node set containing M drugs and N targets (i.e., $|\mathscr{V}| = M + N$), $\mathscr{E}$ is the set of edges representing drug-target pairs, and $\mathscr{W}$ is the set of edge weights measuring the relative binding strength of the corresponding drug-target pairs. More specifically, the edge weights are assigned according to the normalized drug-target binding affinities that are scaled to $[0, 1]$ by min-max normalization, where a nonzero value denotes known affinity. Let $\mathbf{A} \in [0, 1]^{|\mathscr{V}|\times|\mathscr{V}|}$ be the adjacency matrix of $\mathscr{G}$.

Definition 3.2 (*Molecular Graph*). One drug molecule can be viewed as a graph with atoms as nodes and covalent bonds as edges, while one target molecule can also be formulated as a residue contact graph with residues as nodes and their contacts as edges [9]. Let $\mathscr{G}_u = \{\mathscr{V}_u, \mathscr{E}_u\}$ denote the molecular graph of $u \in \mathscr{V}$. Note that u is a drug or a target.

Definition 3.3 (*Hierarchical Graph*). The graph with molecular graphs as nodes and molecular interactions as edges is called the hierarchical graph, as it consists of a coarse-level affinity graph representing intermolecular relations and fine-level molecular graphs containing fine intramolecular structure information. Let $\mathscr{H} = \{\{\mathscr{G}_u\}_{u\in\mathscr{V}}, \mathscr{E}, \mathscr{W}\}$ denote the hierarchical graph.

Definition 3.4 (*Drug-Target Binding Affinity Prediction*). Given the hierarchical graph $\mathscr{H}$ and the observed drug-target binding affinity matrix $\mathbf{Y} \in \mathbb{R}_{\geqslant 0}^{M\times N}$, our goal of predicting drug-target binding affinities is to train a hierarchical graph representation learning framework $\Theta(\mathscr{H}, \mathbf{Y}; \omega)$ to recover the unobserved entries (i.e., zeros) in $\mathbf{Y}$, where ω is the trainable parameter.

4. Model framework

In this section, we introduce the proposed hierarchical graph representation learning model for drug-target binding affinity prediction, named HGRL-DTA. HGRL-DTA builds information propagation and fusion from the coarse level to the fine level over the hierarchical graph. By doing so, each atom or residue is able to know about its molecule assignment and how the molecule interacts with other molecules, which leads to refined drug and target representations. Concretely, we first utilize two GCN-based blocks: one including a GCN encoder deployed on the affinity graph $\mathscr{G}$ encoding drug-target binding affinity information into coarse-level drug/target representations and the other containing a pair of GCN encoders operating on drug and target molecule graphs to learn fine-level atom/residue representations from molecular structural information. Then, we enhance the fine-level representations by a coarse-to-fine information flow from nodes in $\mathscr{G}$ to atoms/residues inside the corresponding molecular graphs and further refine the enhanced representations by another pair of GCN encoders followed by readouts to obtain final representations of drugs and targets. Finally, the concatenation of a pair of representations from one drug and one target is taken as the representation of the drug-target pair and fed into an MLP to make binding affinity predictions. Fig. 2 illustrates the framework of our proposed HGRL-DTA model.

4.1. Fine-level graph representation learning on molecular graphs

As in previous work [9,24] that capitalized on graph neural networks deployed over molecular graphs to encode fine atom/residue-level chemistry and structure information into informative representations of drugs and targets, we still expect their dominance over drug-target binding affinity prediction, conforming to the consensus that structure is able to determine function. We simply adopt GCNs as the backbone of modeling the molecular graphs.

Mathematically, for a molecular graph $\mathscr{G}_u = \{\mathscr{V}_u, \mathscr{E}_u\}$, the iterative feed-forward graph convolution layer can be denoted as:

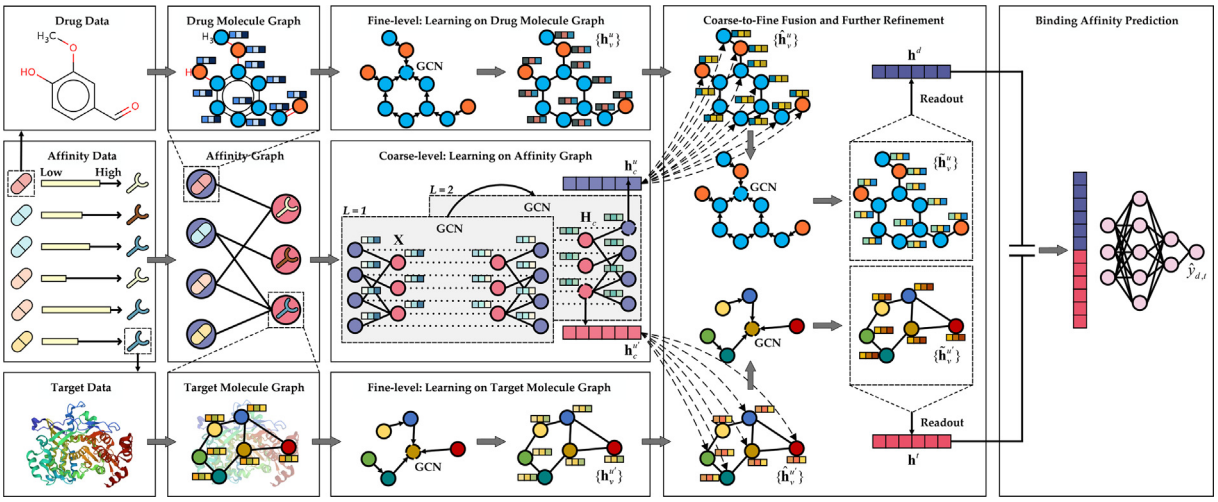

Fig. 2. Overview of HGRL-DTA.

$$\mathbf{h}_v^{(l)} = \mathrm{ReLU}\left(\sum_{v' \in \mathcal{N}(v) \cup \{v\}} \frac{1}{\sqrt{d_v d_{v'}}} \mathbf{h}_{v'}^{(l-1)} \mathbf{W}^{(l)}\right) \tag{1}$$

where $\mathrm{ReLU}(\cdot)$ denotes the rectified linear unit (ReLU) activation function, $\mathbf{h}_v^{(l)}$ denotes the hidden state at the l-th layer for node $v \in \mathscr{V}_u, \mathcal{N}(v)$ is the neighborhood set of node $v, d_v = 1 + |\mathcal{N}(v)|$ is the degree of node v in graph $\mathscr{G}_u$ with a self-loop, and $\mathbf{W}^{(l)}$ is the weight parameter at the l-th layer. Note that we use the self-loop trick to maintain the information of the node itself when it receives the neighbors' message at each layer. By the message passing scheme of GCNs that iteratively aggregates localized information, each atom or residue can be aware of its surrounding structure and obtain a high-level representation that contains intramolecular local structure information.

In this paper, we utilize the physics and chemistry attribution of the node as the initial embedding $\mathbf{h}_v^{(0)}$ and two isolated GCNs in this block with one shared for all drug molecules and the other for target molecules. We denote the fine-level representation of atom/residue v in the molecular graph $\mathscr{G}_u$ output as $\mathbf{h}_v^u = g_u\left(\mathscr{G}_u, \mathbf{h}_v^{(0)}\right)$, where g_u represents the GCN model in this block.

4.2. Coarse-level graph representation learning on the affinity graph

As mentioned before, modeling drug-target interactions is believed to benefit drug-target binding affinity prediction. In this block, we build a GCN on the affinity graph to learn discriminable node representations. Due to the nature of GCNs, node representations encode information about how drugs and targets interact with each other.

Formally, given the affinity graph $\mathscr{G} = \{\mathscr{V}, \mathscr{E}, \mathscr{W}\}$ with its adjacency matrix $\mathbf{A}$, we renormalize the adjacency matrix by Laplacian normalization:

$$\widehat{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \tag{2}$$

where $\mathbf{D}$ denotes a diagonal matrix with diagonal elements $\mathbf{D}_{i,i} = \sum_j \mathbf{A}_{i,j}$. Then, taking an initial graph signal $\mathbf{X}$ as inputs, the GCN outputs drug and target representations:

$$\mathbf{H}_c = g_c(\mathscr{G}, \mathbf{X}) = \mathrm{ReLU}\left(\widehat{\mathbf{A}}\ \mathrm{ReLU}\left(\widehat{\mathbf{A}}\mathbf{X}\mathbf{W}_c^{(1)}\right)\mathbf{W}_c^{(2)}\right) \tag{3}$$

where $\mathbf{H}_c$ with $|\mathscr{V}|$ rows stores drug and target representations, g_c represents the GCN model in this block, and $\mathbf{W}_c^{(1)}$ and $\mathbf{W}_c^{(2)}$ are trainable parameters. This encoder captures affinity relationships of directly connected drug-target pairs (i.e., 1-hop neighbors) at the first layer and recognizes the potential similar patterns at the second layer by aggregating the information of other drugs/targets linked to the same target/drug node (i.e., 2-hop neighbors) for each drug/target node, as the empirical assumption says that similar drugs are more likely to interact with the same target and vice versa.

There seems to be a plight in the training process of GCNs of interest in node representations where oversmoothing leads to indistinguishable representations. To alleviate it, we introduce a regularization technique, DropEdge [32], which randomly removes a certain number of edges from the input graph at each training epoch. We denote the coarse-level representation of $u \in \mathscr{V}$ as $\mathbf{h}_c^u$, that is, the u-th row of $\mathbf{H}_c$.

4.3. Coarse-to-fine information fusion and further representation refinement

In this block, we first design a coarse-to-fine manner to mix coarse- and fine-level information aiming at enhancing the fine-level representations instead of spoiling their leading role in the final prediction task. Then, the enhanced representations are further refined by a pair of GCNs operating on drug and target molecule graphs. This process is illustrated in Fig. 3.

Given the node $u \in \mathscr{V}$ in the affinity graph $\mathscr{G}$ with the corresponding molecular graph $\mathscr{G}_u$, we transfer the coarse-level representation $\mathbf{h}_c^u$ for $u \in \mathscr{V}$ in $\mathscr{G}$ into the fine-level representation of each atom/residue (e.g., $\mathbf{h}_v^u$ for $v \in \mathscr{V}_u$) in $\mathscr{G}_u$. Formally, the enhanced representation $\hat{\mathbf{h}}_v^u$ for the atom/residue v can be defined as:

$$\hat{\mathbf{h}}_v^u = [\mathbf{h}_v^u \oplus f_c(\mathbf{h}_c^u)] || [\mathbf{h}_v^u \ominus f_c(\mathbf{h}_c^u)] \tag{4}$$

where $||$ denotes the concatenation operation, f_c is an MLP in this coarse-to-fine fusion module to harmonize the coarse-level representation with the fine-level representation, and $\oplus$ and $\ominus$ denote element-wise addition and subtraction operations, respectively. Through this design, each atom/residue can know which molecule it belongs to and how the molecule interacts with other molecules, as the coarse-level representations learned by the most expressive GCN that models the realistic drug-target interactions are distinguishable. Furthermore, the fine-level representation can adaptively receive coarse-level information. This is because there is an adaptive trade-off between the addition and subtraction operations when the enhanced representation is squashed and refined in some subsequent neural layers. It is worth mentioning that we adopt two MLPs for drug and target nodes to keep the coarse-level representation aware of node type.

Next, we reuse GCNs such as Eq. (1) to further refine the enhanced representations. There are two reasons for this design: 1) smoothing and reconciling the adaptive hierarchical information fusion; 2) avoiding blurring fine molecule structure information in the enhanced representations and further emphasizing the decisive importance of molecular structure to the prediction task. Specifically, a pair of GCNs deployed on drug and target molecule graphs take the enhanced representations as inputs and output the final representation for atoms or residues. We denote the final representation for v in $\mathscr{G}_u$ as $\tilde{\mathbf{h}}_v^u = g_r\left(\mathscr{G}_u, \hat{\mathbf{h}}_v^u\right)$, where g_r represents the GCN model in this refinement module.

Finally, the readout of the model that generates final drug/target representations from their atom/residue representations is given by:

$$\mathbf{h}^u = f_r\left(\frac{1}{|\mathscr{V}_u|}\sum_{v\in\mathscr{V}_u}\tilde{\mathbf{h}}_v^u\right) \tag{5}$$

where f_r denotes an MLP in this readout module. Similar to the above design, two MLPs are used for drugs and targets. We denote the final representations for drug d and target t as $\mathbf{h}^d$ and $\mathbf{h}^t$.

4.4. Drug-target binding affinity prediction

In this study, binding affinity prediction is a regression task. The concatenation of the drug representation $\mathbf{h}^d$ and the target representation $\mathbf{h}^t$ is treated as the representation of the drug-target pair (d, t). Then, we feed the drug-target pair representation into an MLP to obtain the prediction affinity score $\hat{y}_{d,t}$:

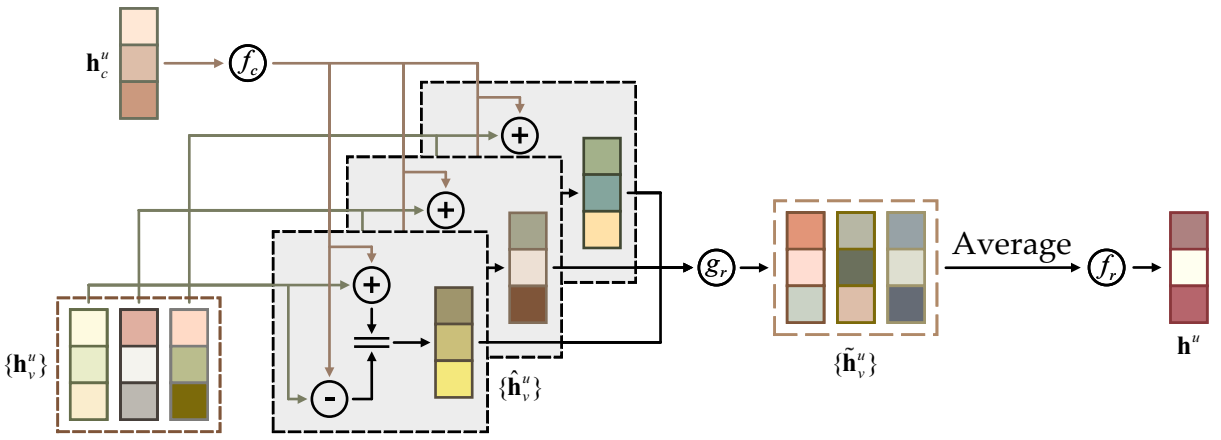

Fig. 3. Pipeline of coarse-to-fine information fusion and further representation refinement.

$$\hat{y}_{d,t} = f_p\left(\mathbf{h}^d \,||\, \mathbf{h}^t\right) \tag{6}$$

where f_p is an MLP with three fully connected linear layers. We use the mean squared error (MSE) as the loss function as follows:

$$\mathscr{L} = \frac{1}{|\mathscr{T}|}\sum_{(d,t)\in\mathscr{T}}\left(y_{d,t} - \hat{y}_{d,t}\right)^2 \tag{7}$$

where $\mathscr{T} \subset \mathscr{E}$ is the sampled training set and $y_{d,t}$ denotes the ground-truth affinity value.

4.5. Prediction for new drugs or targets

Our proposed HGRL-DTA can make full use of known drug-target binding affinity information, but it fails to directly predict binding affinities involving new drugs or targets (i.e., binding affinities between new drugs/targets and known targets/drugs or between new drugs and new targets) as the GCN on the affinity graph cannot be naturally generalized to inductive settings.

To solve this problem, we devise a **similarity-based representation inference method** to speculate the coarse-level information of those drugs or targets unobserved in the training stage based on external drug-drug similarities or target-target similarities, which is inspired by [48], and the inferred coarse-level information is used for coarse-to-fine information fusion in the test stage. In detail, after fixing our HGRL-DTA model trained on the observed hierarchical graph $\mathscr{H} = \{\{\mathscr{G}_u\}_{u\in\mathscr{V}}, \mathscr{E}, \mathscr{W}\}$, for a new molecule $\mathscr{G}_{\hat{u}}$ with $\hat{u} \notin \mathscr{V}$, we construct its coarse-level information $\mathbf{z}_c^{\hat{u}}$ based on the information from the set of molecules $\mathscr{N}_{\hat{u}} \subset \mathscr{V}$ that contains the sim$K$ most similar molecules ranked according to the external similarities:

$$\mathbf{z}_c^{\hat{u}} = \sum_{u\in\mathscr{N}_{\hat{u}}} \mathrm{sim}(u,\hat{u}) \cdot f_c(\mathbf{h}_c^u) \tag{8}$$

where $\mathrm{sim}(u,\hat{u})$ denotes the normalized external similarity between molecule $\hat{u}$ and observed molecule u. Then, we perform the coarse-to-fine fusion by:

$$\widehat{\mathbf{h}}_v^{\hat{u}} = \left[\mathbf{h}_v^{\hat{u}} \oplus \mathbf{z}_c^{\hat{u}}\right] || \left[\mathbf{h}_v^{\hat{u}} \ominus \mathbf{z}_c^{\hat{u}}\right] \tag{9}$$

Finally, the enhanced representation $\hat{\mathbf{h}}_v^{\hat{u}}$ is fed into the fixed subsequent layers to make predictions.

5. Experiments

In this section, we first present the datasets, evaluation metrics, and experimental settings used in our experiments. Then, we compare the HGRL-DTA model with several state-of-the-art methods under four experimental scenarios. Finally, we conduct further and deeper analyses on HGRL-DTA with ablation studies, parameter analysis, visualization analysis, computational complexity analysis, and case studies.

5.1. Datasets

To evaluate the performance of our proposed model on the binding affinity prediction task, two classic benchmark datasets, the Davis dataset [5] and the KIBA dataset [37], were chosen as benchmark datasets in our experiments. We introduce their details as follows:

- **Davis.** The Davis dataset contains 68 unique drugs and 442 unique targets, with 30,056 kinase dissociation constant K_d values as drug-target affinities. [8] [8] converted the K_d values into log space as $pK_d = -\log_{10}\left(K_d/10^9\right)$. The preprocessed Davis dataset is filled with affinities ranging from 5.0 to 10.8, where the boundary value 5.0 is regarded as the true negative drug-target pair that either has very weak binding affinities or is not detected in the wet lab experiment. The Davis dataset provides the SMILES strings of the drugs and the protein sequences of the targets.
- **KIBA.** The KIBA dataset introduces KIBA scores as drug-target affinities, based on the integration of kinase inhibitor bioactivities from various sources, such as K_i, K_d, and IC_{50} [37]. The dataset originally consists of 52,498 drugs and 467 targets with 246,088 affinities. [8] [8] filtered it to comprise 118,254 affinities between 2,111 unique drugs and 229 unique targets with at least 10 affinities of each drug or target. The preprocessed KIBA dataset contains affinity values ranging from 0.0 to 17.2, and NaN values indicate that there are no experimental values for corresponding drug-target pairs. The KIBA dataset also contains drug SMILES strings and target protein sequences.

In our experiments, the raw datasets were preprocessed to obtain the input graphs for the proposed model. In the following, we introduce the construction process of the input graphs, i.e., the affinity graph and the molecular graph defined in Section 3:

- **Affinity Graph.** We constructed the affinity graph by addressing the known affinities as weighted edges and encoding each node (i.e., drug or target) as a multidimensional binary feature vector, which consists of two kinds of information: one-hot encoding of the node type (i.e., either drug-type or target-type) and one-hot encoding of the neighbor nodes (i.e., row vector in the connectivity matrix of the affinity graph). It should be noted that each target connects with too many (on average 518 and up to 1,452) drugs in the constructed affinity graph on the KIBA dataset. However, it has been theoretically and empirically proven that nodes with high degrees are more likely to suffer from oversmoothing in multilayer GNN-based models [4]. To handle this issue, we selectively dropped edges out from the affinity graph in the data preprocessing phase, for which only the top$K$ highest affinity edges related to each target were preserved and other connected ones were removed. Meanwhile, we performed the same operation for each drug. Note that we selectively removed affinities only when conducting experiments on the KIBA dataset.
- **Molecular Graph.** We transformed the SMILES strings of drugs into their corresponding graphs with the open-source cheminformatics software RDKit [14]. A group of atomic features adopted from DeepChem [30] were used as the initial drug molecule graph signals. Similarly, an open-source and highly efficient protein structure prediction approach, Pconsc4 [22], was adopted in our work to generate target molecule graphs for mining useful topological information hidden in protein sequences. The Pconsc4 approach transforms the protein sequences of targets into their corresponding contact maps, i.e., residue-residue association matrices, whose entries are the Euclidean distance-based contacts. In this contact map, there exists a contact between two atoms if the Euclidean distance between them is less than a specified threshold [44]. We set the threshold as 0.5 according to the previous study DGraphDTA [9]. A set of residue features extracted by DGraphDTA [9] was used as the initial attributions of the residues in the target molecule graph.

5.2. Evaluation metrics

In this study, we followed previous works [9,25] and employed four classic metrics to evaluate the performance of the models: mean squared error (MSE), concordance index (CI), r_m^2, and Pearson correlation coefficient (PCC). For each model, we reported the mean and the standard deviation (std) of these metrics across ten runs with different random seeds.

MSE is a common measure metric for regression tasks that measures the distinction between the real value and the predicted value. MSE is defined in Eq. (10).

$$\mathrm{MSE} = \frac{1}{T}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2 \tag{10}$$

where T denotes the size of the test set and y_i and $\hat{y}_i$ denote the real value and the predicted value of the i-th test sample, respectively.

CI measures the order concordance between true values and predicted values. This metric ranges from 0 to 1, where a value closer to 1 denotes a better result. The CI formula is given as follows:

$$\mathrm{CI} = \frac{1}{Z}\sum_{y_i > y_j} h(\hat{y}_i - \hat{y}_j), h(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0.5, & \text{if } x = 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{11}$$

where Z represents the normalization constant and $h(\cdot)$ is the step function.

The r_m^2 metric involved in DeepDTA [25] is used to evaluate the external predictive potential of quantitative structure-activity relationship (QSAR) models. A model with a larger r_m^2 value for the test set is regarded as a more acceptable model. We calculate the r_m^2 metric as follows:

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \tag{12}$$

where r^2 and r_0^2 are the squared correlation coefficients with and without intercept, respectively.

The PCC measures the linear correlation between true values and predicted values, which ranges from -1 to 1, where 1, -1, and 0 indicate complete correlation, reversed correlation, and no correlation, respectively. The PCC is calculated through Eq. 13.

$$\mathrm{PCC} = \frac{cov(\hat{y}, y)}{\sigma(\hat{y})\sigma(y)} \tag{13}$$

where $cov(\hat{y}, y)$ is the covariance between the predicted value $\hat{y}$ and the real value y, and $\sigma(\cdot)$ indicates the standard deviation.

5.3. Experimental settings

Following previous works [9,25], we divided each dataset into training and test sets in a 5:1 ratio. When conducting experiments, we trained the proposed model and comparison methods on the training set and evaluated them on the test set. We conducted 5-fold cross validation (5-CV) on the training set to select the best hyperparameters from fixed ranges for our proposed model. For all the comparison methods, the hyperparameters were set as the optimal values provided in the corresponding studies [9,23,25,49].

In this study, to comprehensively verify the generalization and robustness of the models, we considered the following four experimental scenarios [26]:

- **S1:** Entries in the drug-target matrix **Y** are randomly selected for testing.
- **S2:** Row vectors in the drug-target matrix **Y** are randomly selected for testing.
- **S3:** Column vectors in the drug-target matrix **Y** are randomly selected for testing.
- **S4:** The intersection set of the row vectors in scenario **S2** and the column vectors in scenario **S3** is selected for testing, and their nonintersectional parts are used for neither training nor testing.

In scenario **S1**, both the drugs and the targets of the test drug-target pairs can be observed in the training phase. As the most widely used experimental scenario in previous studies, scenario **S1** assumes that some known drug-target affinities are randomly masked, and our aim is to infer these masked affinities using the known affinities. Compared to scenario **S1**, scenarios **S2** and **S3** have recently attracted more attention in real-world applications, where only part of the drug/target information is available during the training phase and the models need to predict affinities for new drugs/targets without any known affinities. Scenario **S4** corresponds to the most challenging case in computational works, which aims to predict affinities between unknown drugs and targets.

We implemented our proposed model with PyTorch 1.4.0 [27] and PyTorch Geometric 1.7.0 [6]. We ran HGRL-DTA on our workstation with 2 Intel(R) Xeon(R) Gold 6146 3.20 GHz CPUs, 128 GB RAM, and 2 NVIDIA 1080 Ti GPUs. Table 1 summarizes the hyperparameter settings of HGRL-DTA. The source code is freely available at https://github.com/Zhaoyang-Chu/HGRL-DTA.

5.4. Comparison with state-of-the-art methods

To demonstrate the superiority of the proposed model, we conducted experiments to compare our approach with the following state-of-the-art methods:

- **DeepDTA** [25] employs CNNs to learn 1D drug and target representations from drug SMILES strings and target protein sequences.
- **AttentionDTA** [49] utilizes 1D CNNs to learn sequence representations of drugs and targets and an attention mechanism to find the weight relationships between drug subsequences and protein subsequences.
- **GraphDTA** [23] models drugs as molecular graphs to capture the bond information between atoms by GNNs and leverages CNNs to learn 1D representations of target proteins. This method has four variants based on different GNN backbones: GCN [13], GAT [40], GIN [45], and GAT-GCN that combines the former two. We selected the variants performing best on the Davis dataset and the KIBA dataset for comparison.
- **DGraphDTA** [9] constructs target molecule graphs from the corresponding protein sequences via a protein structure prediction method and applies GNNs to mine structural information hidden in drug molecule graphs and target molecule graphs.

Table 1
Model configuration of HGRL-DTA.

Hyperparameter	Value	Hyperparameter	Value
Learning rate	0.0005	g_c	{512, 256}
Batch size	512	g_u for drug	{78}
Epochs under **S1**	2000	g_u for target	{54}
Epochs under **S2**, **S3**, **S4**	200	f_c for drug	{1024, 78}
DropEdge rate	0.2	f_c for target	{1024, 54}
Dropout rate	0.1	g_r for drug	{156, 312}
sim$K$ for drug (simK_d)	2	g_r for target	{108, 216}
sim$K$ for target (simK_t)	7	f_r for drug	{1024, 128}
top$K$ for drug	40	f_r for target	{1024, 128}
top$K$ for target under **S1**, **S3**	150	f_p	{1024, 512, 1}
top$K$ for target under **S2**, **S4**	90	-	-

Note: $\{\cdot\}$ denotes the layer dimension of neural networks.

- **MGraphDTA** [46] argues against the existing shallow GNNs and builds a superdeep GNN with 27 stacked layers to encode the multiscale molecular structures of drugs. For targets, this method applies a multiscale convolutional neural network to extract their multiscale sequence features.
- **BERT-GCN** [15] adopts a pair of pretrained BERT models to produce sequence representations for drugs and targets and then models the drug-target binding affinities through a GCN to update drug/target representations. In this method, the BERT representations of drugs and targets are utilized as initial node representations for the GCN. It should be noted that this method is limited by the transductive nature of GCN and thus cannot predict binding affinities involving new drugs or targets. For that, we only consider its predictive performance in scenario **S1**. Since the source code of BERT-GCN is not available, we implemented a variant model referring to the idea of this method for comparison, i.e., **BERT-GCN-V** replaces the pair of BERT models with a pair of GCN encoders deployed on the drug molecule graphs and the target molecule graphs.

Table 2 shows the performances of our proposed HGRL-DTA and the state-of-the-art methods on the two benchmark datasets under four experimental scenarios. According to the experimental results, we can observe that the proposed HGRL-DTA model achieves the best performance compared to the state-of-the-art methods in most cases, which demonstrates the generalization and robustness of our model. In the four experimental scenarios, over the best baseline models, we achieve 19.0% (**S1**), 3.1% (**S2**), and 2.4% (**S4**) improvements in MSE on the Davis dataset and 3.1% (**S1**), 5.2% (**S2**), 11.5% (**S3**), and 10.4% (**S4**) improvements in MSE on the KIBA dataset. Across the four metrics, most of the results show that HGRL-DTA has obtained the optimal or suboptimal standard deviation compared with other baseline methods, which demonstrates the relative stability of our proposed predictive model.

Among all baselines, the performances of the sequence-based methods (i.e., DeepDTA and AttentionDTA) are relatively poor due to the inadequate exploitation of the molecular chemical structures. This indicates that simply modeling drugs as SMILES strings and targets as protein sequences is not sufficient to capture the intrinsic properties of molecules. By contrast, the graph-based models (i.e., GraphDTA, DGraphDTA, and MGraphDTA) represent molecules as molecular graphs to take advantage of their chemical structural information, which produces better predictive performance. However, these graph-based models mainly focus on encoding molecular structures but ignore the topological affinity relationships between drugs and targets, which may be the reason why they are slightly inferior to our proposed model in most cases. In scenario S3, it is worth mentioning that although MGraphDTA performs slightly better than HGRL-DTA in terms of MSE, r_m^2, and PCC on the Davis dataset, it consists of more complex structures, which may capture deeper molecular structural information to some extent, and we will study it in future work to further improve the prediction performance. Compared with state-of-the-

Table 2
Performances of HGRL-DTA and comparison methods on the two benchmark datasets.

	Architecture	Davis				KIBA			
		MSE↓ (std)	CI↑ (std)	r_m^2↑ (std)	PCC↑ (std)	MSE↓ (std)	CI↑ (std)	r_m^2↑ (std)	PCC↑ (std)
S1	DeepDTA	0.245 (0.014)	0.888 (0.004)	0.665 (0.015)	0.842 (0.004)	0.181 (0.007)	0.868 (0.004)	0.711 (0.021)	0.864 (0.003)
	AttentionDTA	0.233 (0.006)	0.889 (0.002)	0.676 (0.020)	0.845 (0.004)	0.150 (0.002)	0.883 (0.001)	0.760 (0.018)	0.888 (0.001)
	GraphDTA	0.243 (0.005)	0.887 (0.002)	0.685 (0.016)	0.839 (0.003)	0.148 (0.006)	0.891 (0.001)	0.730 (0.015)	0.895 (0.001)
	DGraphDTA	0.216 (0.003)	0.900 (0.001)	0.686 (0.015)	0.857 (0.002)	0.132 (0.002)	0.902 (0.001)	0.800 (0.011)	0.903 (0.001)
	MGraphDTA	0.225 (0.003)	0.889 (0.006)	0.708 (0.008)	0.850 (0.001)	0.129 (0.001)	0.903 (0.001)	**0.805** (0.006)	0.903 (0.001)
	BERT-GCN-V	0.205 (0.002)	0.881 (0.002)	0.689 (0.005)	0.866 (0.001)	0.170 (0.001)	0.876 (0.001)	0.696 (0.006)	0.873 (0.001)
	HGRL-DTA	**0.166** (0.002)	**0.911** (0.002)	**0.751** (0.006)	**0.892** (0.001)	**0.125** (0.001)	**0.906** (0.001)	0.789 (0.017)	**0.907** (0.001)
S2	DeepDTA	0.985 (0.114)	0.548 (0.045)	0.027 (0.022)	0.126 (0.109)	0.494 (0.070)	0.747 (0.012)	0.337 (0.026)	0.623 (0.023)
	AttentionDTA	0.869 (0.053)	0.642 (0.028)	0.079 (0.024)	0.289 (0.048)	0.506 (0.018)	0.744 (0.005)	0.298 (0.015)	0.618 (0.006)
	GraphDTA	0.801 (0.038)	0.659 (0.015)	0.160 (0.019)	0.416 (0.022)	0.475 (0.047)	0.753 (0.002)	**0.382** (0.007)	0.652 (0.002)
	DGraphDTA	0.818 (0.012)	0.646 (0.006)	0.114 (0.005)	0.356 (0.010)	0.458 (0.008)	0.754 (0.002)	0.362 (0.012)	0.622 (0.004)
	MGraphDTA	0.907 (0.033)	0.599 (0.022)	0.082 (0.021)	0.298 (0.044)	0.469 (0.049)	0.752 (0.002)	0.366 (0.016)	0.638 (0.005)
	HGRL-DTA	**0.776** (0.019)	**0.684** (0.007)	**0.163** (0.015)	**0.422** (0.018)	**0.434** (0.007)	**0.757** (0.003)	0.370 (0.010)	**0.653** (0.003)
S3	DeepDTA	0.552 (0.086)	0.729 (0.017)	0.258 (0.029)	0.523 (0.028)	0.732 (0.197)	0.676 (0.016)	0.273 (0.026)	0.587 (0.033)
	AttentionDTA	0.436 (0.017)	0.787 (0.018)	0.304 (0.022)	0.588 (0.027)	0.529 (0.039)	0.693 (0.008)	0.254 (0.024)	0.592 (0.022)
	GraphDTA	0.860 (0.083)	0.666 (0.012)	0.134 (0.014)	0.379 (0.018)	0.469 (0.089)	0.710 (0.005)	0.388 (0.013)	0.627 (0.009)
	DGraphDTA	0.445 (0.019)	0.788 (0.009)	0.289 (0.016)	0.558 (0.017)	0.364 (0.010)	0.718 (0.007)	0.429 (0.022)	0.671 (0.009)
	MGraphDTA	**0.359** (0.007)	0.813 (0.008)	**0.415** (0.008)	**0.681** (0.005)	0.483 (0.055)	0.674 (0.008)	0.342 (0.007)	0.617 (0.011)
	HGRL-DTA	0.383 (0.010)	**0.816** (0.008)	0.375 (0.018)	0.621 (0.012)	**0.322** (0.014)	**0.741** (0.004)	**0.502** (0.016)	**0.729** (0.007)
S4	DeepDTA	0.767 (0.091)	0.508 (0.057)	0.009 (0.012)	0.015 (0.098)	0.700 (0.075)	0.627 (0.009)	0.140 (0.017)	0.401 (0.025)
	AttentionDTA	0.679 (0.021)	0.554 (0.030)	0.005 (0.008)	0.036 (0.062)	0.609 (0.021)	0.629 (0.007)	0.143 (0.015)	0.407 (0.022)
	GraphDTA	0.988 (0.096)	0.569 (0.017)	0.020 (0.006)	0.141 (0.020)	0.676 (0.113)	0.641 (0.003)	0.149 (0.007)	0.404 (0.009)
	DGraphDTA	0.658 (0.026)	0.569 (0.008)	0.031 (0.005)	0.180 (0.015)	0.594 (0.022)	0.632 (0.009)	0.148 (0.013)	0.403 (0.019)
	MGraphDTA	0.764 (0.020)	0.507 (0.021)	0.001 (0.001)	0.011 (0.036)	0.660 (0.094)	0.627 (0.007)	0.152 (0.012)	0.418 (0.018)
	HGRL-DTA	**0.642** (0.016)	**0.602** (0.009)	**0.044** (0.005)	**0.215** (0.013)	**0.532** (0.008)	**0.642** (0.004)	**0.207** (0.009)	**0.491** (0.010)

Note: The best score in each column is in **bold** and the second best score is underlined in every scenario.

art models, HGRL-DTA can capture the fine-level information hidden inside the intrinsic molecular properties and the coarse-level information derived from the topological affinity relationships synergistically and fuse such hierarchical information to enhance the representations of drugs and targets, which significantly facilitates the performance of predicting drug-target binding affinities.

Notably, BERT-GCN-V is similar to our approach in that it also fuses fine-level molecular structural information and coarse-level binding affinity information but exhibits relatively poor predictive performance. This may be because BERT-GCN-V takes a fine-to-coarse information fusion manner, in which the fine-level drug/target representations encoding the molecular structure are smoothed out in the subsequent coarse-level GCN encoder, which spoils the dominant role of the fine-level molecular structural information in the drug-target binding affinity prediction task. In contrast to BERT-GCN-V, we design a coarse-to-fine information fusion manner to enhance the fine-level molecular representations with the coarse-level binding affinity information, after which we reuse the GCN on molecular graphs to refine the enhanced representations to avoid blurring the fine-level molecular structural information. Such results show that our coarse-to-fine information fusion manner is more reasonable and effective.

In addition, we can observe that all the models perform best in scenario **S1**, have relatively poor performance in scenarios **S2** and **S3**, and perform worst in scenario **S4**. With more unknown drugs or targets in the four experimental scenarios, the predictive performance of the models significantly declines. Different from scenario **S1**, scenarios **S2**, **S3**, and **S4** test the generalization and robustness of the models for new drugs or targets, which is another necessary measurement of performance evaluation. Through the similarity-based representation inference method, HGRL-DTA infers coarse-level information of new drugs/targets using the learned representations of known drugs/targets, which can make full use of the known affinity and similarity information to improve the generalization and robustness of the model. As illustrated in Table 2, in scenarios **S2**, **S3**, and **S4**, the proposed HGRL-DTA model obtains the best performance in most cases, which indicates that HGRL-DTA is more generalizable and robust than baseline methods when only part of the drug/target information is known.

5.5. Ablation studies

To investigate the important factors that impact the predictive capacity of our model, we conducted ablation studies with the following variants of HGRL-DTA in scenario **S1**:

- **HGRL-DTA without the coarse-level affinity graph** (w/o CAG) learns only fine-level representations on molecular graphs without the GCN deployed on the coarse-level affinity graph. Note that this variant keeps the same number of GCN iterations on the molecular graph as HGRL-DTA, where GCN iterations in the representation refinement procedure of HGRL-DTA are also considered.
- **HGRL-DTA without fine-level molecular graphs** (w/o FMG) only applies a GCN encoder to the coarse-level affinity graph without considering the fine-level molecular graphs. The MLP-based predictor is directly applied with the input of the coarse-level representations of drugs and targets for the binding affinity prediction task.
- **HGRL-DTA without weighted affinities** (w/o WA) addresses the affinity graph as an unweighted graph, which only considers binary interaction relationships instead of continuous affinities.
- **HGRL-DTA-L** is a variant implementation of our idea on late coarse-to-fine information fusion, which moves the information fusion procedure behind the molecular representation readout operation. Note that this variant keeps the same number of GCN iterations on the molecular graph as HGRL-DTA.

Fig. 4 compares HGRL-DTA with its four variants on the two benchmark datasets. Overall, the proposed HGRL-DTA outperforms other variants, which demonstrates the effectiveness of the hierarchical graph learning architecture. In detail, HGRL-DTA (w/o CAG) and HGRL-DTA (w/o FMG) have the most significant performance gaps with HGRL-DTA. These results suggest that coarse-level and fine-level components contribute the most to our model and removing either component will severely undermine its predictive performance. In addition, HGRL-DTA (w/o WA) performs worse than HGRL-DTA since it only constructs the affinity graph using binary interaction relationships, which loses more realistic information hidden in continuous affinities. We also observe that HGRL-DTA-L is not far from HGRL-DTA on MSE, CI, and PCC and even produces better performance on the r_m^2 metric, which suggests that coarse-to-fine information fusion is beneficial to drug-target binding affinity prediction independent of its position in the model. In our experiments, we mainly focus on the MSE metric and select models and hyperparameters based on the 5-CV result of it rather than of r_m^2; hence, the HGRL-DTA framework is chosen as our main model.

5.6. Parameter analysis

To further validate the effectiveness of the similarity-based representation inference method for inferring new drugs or targets, we analyzed the impacts of two major hyperparameters simK_d and simK_t used in this method.

We conducted the parameter study experiment on the Davis dataset by changing the hyperparameters simK_d and simK_t from 1 to 8 while keeping other hyperparameters fixed as default settings. We tested simK_d under scenario **S2**, where the drug is unseen, and simK_t under scenario **S3**, where the target is unknown. To analyze the effect of the inferred representa-

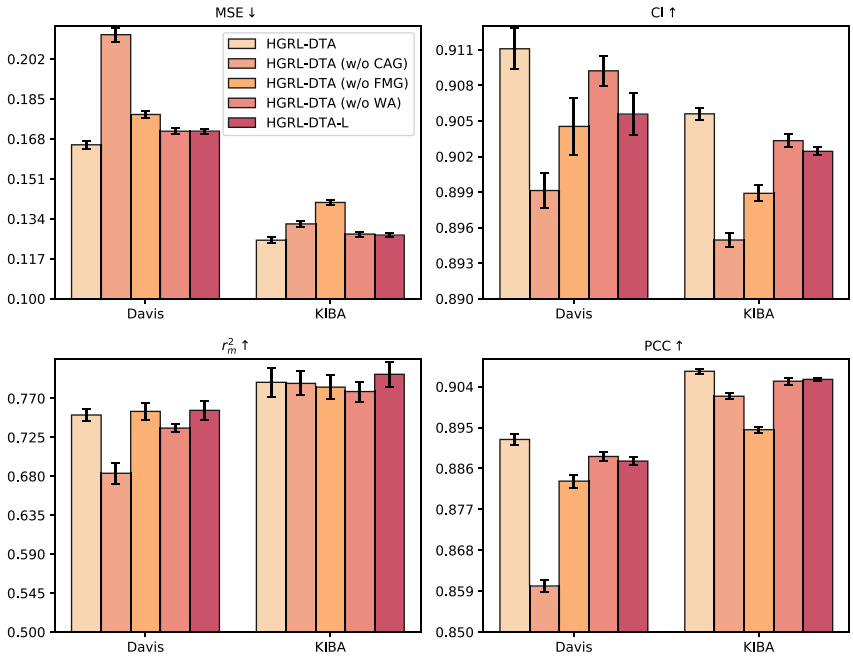

Fig. 4. Results of ablation experiments.

tions of new drugs or targets (i.e., nodes absent in the affinity graph) more directly, we observe the performance variation of HGRL-DTA w/o FMG.

As shown in Fig. 5, the similarity-based representation inference method influences the predictive performance of HGRL-DTA w/o FMG by changing simK_d and simK_t. We can see that the model performs best when sim$K_d = 2$ and sim$K_t = 7$. With the increase in simK_d or simK_t, aggregating more representations of known drugs/targets to infer unseen drugs/targets can encode more useful information, which leads to dramatic performance improvements. When simK_d or simK_t reaches its optimal value, the performance begins to decline because the aggregation of too many representations may introduce redundant and noisy information that can harm the predictive capacity. Furthermore, the nonzero choices of simK_d and simK_t demonstrate the importance of utilizing the similarity-based representation inference method to infer new drugs or targets in our method.

5.7. Visualization analysis

In this subsection, we designed an additional experiment to explore the representation power of the proposed model from the view of the representations of drug-target pairs.

To simplify the discussion, we divided drug-target pairs into two clusters through predefined affinity thresholds provided in previous studies [8,37], where the pK_d value of 7 and the KIBA score of 12.1 were selected as thresholds for the Davis dataset and the KIBA dataset, respectively. Drug-target pairs with affinities below the predefined threshold are classified as weak-affinity pairs and those above as strong-affinity pairs. It should be noted that such division was conducted on the test sets of the two benchmark datasets in scenario **S1**. We preserved the trained HGRL-DTA model and then extracted the representations before the final prediction layer for drug-target pair samples in the test set.

This experimental analysis is based on an empirical assumption that drug-target pairs are expected to be as close as possible in the same cluster and as far as possible in different clusters in the pair representation space. The performance of clustering of drug-target pair representations is positively associated with the representation power of the models. To evaluate the clustering performance of the drug-target pair representations extracted from various models, we chose three classical metrics, including silhouette coefficient (SC), Calinski-Harabasz index (CHI), and Davies-Bouldin index (DBI), with reference to the previous study [2].

SC measures how close each sample in a cluster is to samples in the neighboring clusters, which ranges from −1 to 1. The higher (the closer to 1) the SC metric is, the more separated the clusters are. The SC formula is given as:

$$\mathrm{SC} = \frac{1}{T}\sum_{i=1}^{T}\frac{\hat{d}_i - \bar{d}_i}{\max\left(\bar{d}_i, \hat{d}_i\right)} \tag{14}$$

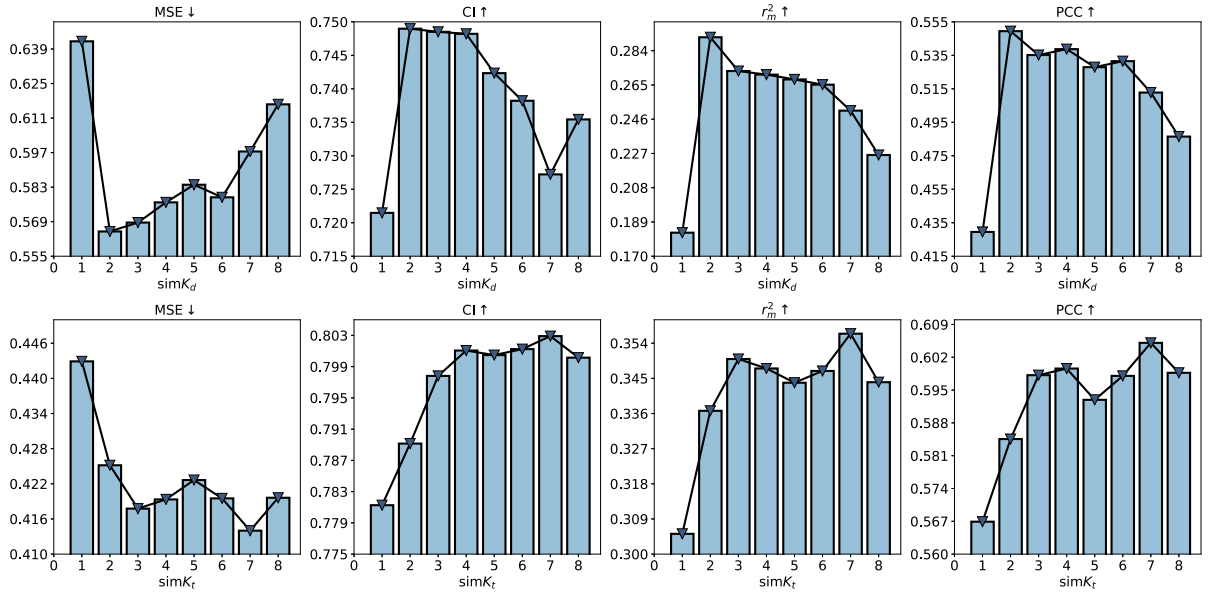

Fig. 5. Parameter study of simK_d and simK_t for inferring new drug or target.

where $\bar{d}_i$ denotes the mean distance between the i-th sample and other samples in the same cluster and $\hat{d}_i$ indicates the mean distance between the i-th sample and all samples in the next nearest cluster.

CHI is the ratio of the sum of between-clusters dispersion and of within-cluster dispersion for all clusters, where the dispersion is defined as the sum of distances squared. A higher CHI score indicates better clustering performance. CHI is formulated as:

$$\mathrm{CHI} = \frac{tr(B_k)}{tr(W_k)} \times \frac{T-k}{k-1} \tag{15}$$

where k denotes the number of clusters and $tr(B_k)$ and $tr(W_k)$ are the traces of the between group dispersion matrix and the within-cluster dispersion matrix, respectively, defined by:

$$W_k = \sum_{p=1}^{k} \sum_{\mathbf{x} \in C_p} (\mathbf{x} - \mathbf{c}_p)(\mathbf{x} - \mathbf{c}_p)^T, B_k = \sum_{p=1}^{k} n_p (\mathbf{c}_p - \mathbf{c})(\mathbf{c}_p - \mathbf{c})^T \tag{16}$$

where C_p is the set of the vectors of samples in the p-th cluster, $\mathbf{c}_p$ and $\mathbf{c}$ denote the centers of the p-th cluster and the test set, respectively, and n_p indicates the number of samples in the p-th cluster.

DBI signifies the average similarity between each cluster C_p and its most similar cluster C_q. A lower DBI score indicates better separation between the clusters. The DBI formula is defined as:

$$\mathrm{DBI} = \frac{1}{k} \sum_{p=1}^{k} \max_{p \neq q} \frac{\bar{d}_p + \bar{d}_q}{d_{p,q}} \tag{17}$$

where $\bar{d}_p$ denotes the average distance between each sample of the p-th cluster and the centroid of that cluster and $d_{p,q}$ denotes the distance between the centroids of the p-th and q-th clusters.

Table 3 reports the clustering performance of drug-target pair representations of our model and baselines on the two benchmark datasets. As we can see, compared with baseline methods, our HGRL-DTA model achieves the best and second-best cluster performance on the Davis and KIBA datasets, respectively. Moreover, to analyze the drug-target pair representations more intuitively, we sampled weak-affinity pairs and strong-affinity pairs with a ratio of 1:1 from the test set of the KIBA dataset and projected their representations into 2D space using t-distributed stochastic neighbor embedding (t-SNE) [39] for visualization. As illustrated in Fig. 6, HGRL-DTA and MGraphDTA can well distinguish weak-affinity pairs (red) and strong-affinity pairs (blue); DeepDTA, GraphDTA, and DGraphDTA recognize most of the strong-affinity pairs; and AttentionDTA and BERT-GCN-V differentiate part of the drug-target pairs. These results indicate that HGRL-DTA allows more delicate drug-target pair representations, which leads to better performance for binding affinity prediction.

Table 3
Clustering performance of drug-target pair representations.

Architecture	Davis			KIBA		
	SC↑	CHI↑	DBI↓	SC↑	CHI↑	DBI↓
DeepDTA	0.585	3122.789	0.730	0.305	4325.479	1.711
AttentionDTA	0.303	592.239	1.728	0.176	2200.669	2.470
GraphDTA	0.615	2751.677	0.917	0.313	5589.479	1.593
DGraphDTA	0.643	2506.929	0.906	0.353	4034.783	1.968
MGraphDTA	**0.667**	4119.269	0.640	**0.501**	**17203.114**	**0.857**
BERT-GCN-V	0.350	805.878	1.518	0.227	1926.217	2.681
HGRL-DTA	0.639	**4330.756**	**0.587**	0.410	10385.635	1.194

Note: The best score in each column is in **bold** and the second best score is underlined.

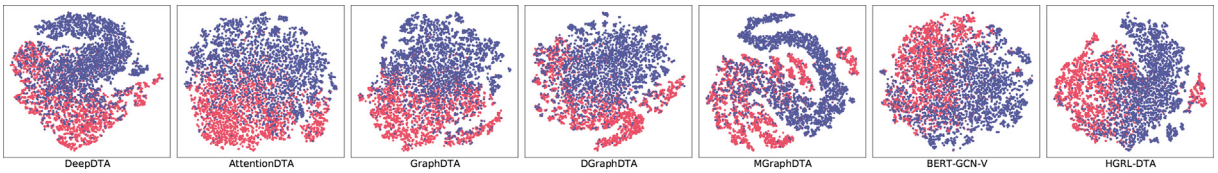

Fig. 6. Visualization of drug-target pair representations. Red: weak-affinity pair. Blue: strong-affinity pair.

5.8. Computational complexity analysis

In this subsection, we discuss the time complexity of the proposed HGRL-DTA model and compare the running time of different algorithms.

For the convenience of analysis, we define the number of drugs and targets by K, the maximum number of edges in the molecular graph by E, and the maximum value of vector dimensions by D. The key parts of our model are the computations of GCNs on the molecular graph and the affinity graph. Regarding the GCN on the molecular graph, the cores are the fine-level graph representation learning and representation refinement components based on Eq. (1), which require $\mathcal{O}\left(3\left(KED + KD^2\right)\right)$ flops in total. In addition, the computation by the coarse-level graph representation learning module in Eq. (3) requires $\mathcal{O}\left(2\left(K^2D + KD^2\right)\right)$. Overall, the upper bound of the computational complexity of our model is $\mathcal{O}(K(2K + 3E + 5D)D)$.

Table 4 shows the training time of different algorithms on the two benchmark datasets under scenario **S1**. We can observe that the graph-based models generally consume more time to reach convergence during the training stage than the sequence-based models. Among the graph-based models, we also find that the models modeling proteins as molecular graphs (i.e., DGraphDTA, BERT-GCN-V, and HGRL-DTA), and the model consisting of superdeep GNN layers (i.e., MGraphDTA) have a longer training time than GraphDTA. Moreover, compared with DGraphDTA, HGRL-DTA adds a GCN component deployed on the affinity graph, but which does not result in a significant increase in training time. This is because the size of the affinity graph is much smaller than that of molecular graphs, which makes the runtime of the model mainly dependent on the computational time consumed by the GCN deployed on molecular graphs. In addition, after the first training of our model reaches convergence, it can be directly used in subsequent applications without retraining. It is worth mentioning that the running time is not our main focus in this study.

5.9. Case studies

To validate the effectiveness of our proposed HGRL-DTA model in realistic scenarios, we conducted a case study of drug repurposing by predicting and ranking the binding affinity scores between commercially existing drugs and SARS-CoV-2 3C-like protease ($3CL^{Pro}$). The 3C-like protease plays a crucial role in coronavirus replication and is a considerable therapeutic target for diseases caused by coronaviruses, including COVID-19. Drug repurposing for the $3CL^{Pro}$ target facilitates the identification of inhibitors capable of suppressing virus spread and the development of new vaccines.

Following the previous study [16], we selected 82 antiviral drugs, 1 antiparasitic drug (Ivermectin), and 1 unrelated drug (Aspirin) for our experiments and collected their SMILES strings from the PubChem database. We obtained the amino acid sequence of $3CL^{Pro}$ from the Protein Data Bank (PDB) database. Then, the HGRL-DTA model pretrained on the KIBA dataset takes the SMILES strings of these 84 existing drugs and the amino acid sequence of $3CL^{Pro}$ as input and returns a list of ranked drug candidates according to the predicted affinity scores. Note that the case study was conducted under scenario **S4**, where the chosen drugs and the $3CL^{Pro}$ target were unseen for HGRL-DTA in the training stage.

Table 4
Training time of different algorithms.

Dataset	DeepDTA	AttentionDTA	GraphDTA	DGraphDTA	MGraphDTA	BERT-GCN-V	HGRL-DTA
Davis	700s	9974s	1944s	26336s	60732s	25058s	26548s
KIBA	2300s	22676s	10942s	95836s	173982s	61084s	65814s

Table 5
Ranking results of repurposing drugs for SARS-CoV2 3C-like protease.

Rank	Drug	PubMed ID	Rank	Drug	PubMed ID
1	Docosanol	-	10	Doravirine	-
2	Cobicistat	32671131	11	Vicriviroc	-
3	Methisazone	32818545	12	Grazoprevir	33790352
4	Nevirapine	-	13	Pleconaril	-
5	Delavirdine	-	14	Raltegravir	33790352
6	Ritonavir	32297571	15	Hydroxychloroquine	32373993
7	Glecaprevir	32441299	75	Aspirin	33417877
8	Adefovir	-	83	Ivermectin	33662102
9	Chloroquine	32020029	-	-	-

Note: For several drugs in our predicted list, we provide the PubMed IDs of the publications that studied their effects on COVID-19. PubMed (https://pubmed.ncbi.nlm.nih.gov/) is a free search engine accessing primarily biomedical literature.

Table 5 reports the ranking results of the top 15 repurposing antiviral drugs and two counterexample drugs (i.e., Ivermectin and Aspirin). These results show that 8 recommendations out of the 15 have been confirmed to have inhibitory or therapeutic effects against COVID-19 by many studies. In contrast, Aspirin and Ivermectin have weak affinities with the $3CL^{Pro}$ target, ranking 75 and 83 out of 87, respectively. This is consistent with the existing research results that the two drugs do not have apparent associations with the treatment of COVID-19 [21,34]. The results of the case study demonstrate the ability of HGRL-DTA to repurpose drugs in realistic scenarios.

6. Conclusion

In this paper, we propose a novel hierarchical graph representation learning model to learn the representations of drugs and targets for better drug-target binding affinity prediction. Our model can synergistically capture the coarse- and fine-level information from intermolecular interactions and intramolecular structures involving drugs/target molecules and integrate such hierarchical graph information in a well-designed coarse-to-fine manner. To generalize our model to the cold start situation, we design a similarity-based representation inference method to deduce the coarse-level information for new drugs or targets. Extensive experiments under four different scenarios have demonstrated that integrating coarse- and fine-level information into the representations of drugs and targets can significantly improve the predictive capacity of the models. We also find experimental evidence suggesting that the coarse-to-fine manner is beneficial for the integration of the hierarchical graph information, and the similarity-based representation inference method is an effective strategy to infer coarse-level representations for new drugs or targets. In the future, we will extend the proposed method to other biological entity association prediction tasks with hierarchical graph architecture, e.g., drug-drug interaction (DDI) prediction and protein-protein interaction (PPI) prediction.

CRediT authorship contribution statement

Zhaoyang Chu: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft. **Feng Huang:** Validation, Formal analysis, Investigation, Visualization, Writing - original draft, Writing - review & editing. **Haitao Fu:** Writing - review & editing, Visualization, Validation. **Yuan Quan:** Writing - review & editing, Investigation. **Xionghui Zhou:** Writing - review & editing, Formal analysis. **Shichao Liu:** Writing - review & editing, Supervision, Conceptualization, Resources. **Wen Zhang:** Writing - review & editing, Supervision, Conceptualization, Resources, Project administration, Funding acquisition.

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.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No.62072206, 62102158), the Fundamental Research Funds for the Central Universities (Grant No.2662022JC004, 2662021JC008), 2021 Foshan support project for promoting the development of university scientific and technological achievements service industry(zc03040000014), Huazhong Agricultural University Scientific & Technological Self-innovation Foundation. The funders have no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

References

[1] N. Aleb, A Mutual Attention Model for Drug Target Binding Affinity Prediction, IEEE/ACM Transactions on Computational Biology and Bioinformatics (2021).
[2] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J.M. Pérez, I. Perona, An extensive comparative study of cluster validity indices, Pattern Recognition 46 (2013) 243–256.
[3] Y. Bai, K. Gu, Y. Sun, W. Wang, Bi-Level Graph Neural Networks for Drug-Drug Interaction Prediction, arXiv preprint arXiv:2006.14002 (2020).
[4] M. Chen, Z. Wei, Z. Huang, B. Ding, Y. Li, Simple and Deep Graph Convolutional Networks, in: H.D. III, A. Singh (Eds.), Proceedings of the 37th International Conference on Machine Learning, volume 119 of Proceedings of Machine Learning Research, PMLR, 2020, pp. 1725–1735. https://proceedings.mlr.press/v119/chen20v.html.
[5] M.I. Davis, J.P. Hunt, S. Herrgard, P. Ciceri, L.M. Wodicka, G. Pallares, M. Hocker, D.K. Treiber, P.P. Zarrinkar, Comprehensive analysis of kinase inhibitor selectivity, Nature biotechnology 29 (2011) 1046–1051.
[6] M. Fey, J.E. Lenssen, Fast Graph Representation Learning with PyTorch Geometric, in: ICLR 2019 Workshop on Representation Learning on Graphs and Manifolds, 2019.
[7] W. Guan, H. Wen, X. Song, C. Wang, C.-H. Yeh, X. Chang, L. Nie, Partially Supervised Compatibility Modeling, IEEE Transactions on Image Processing 31 (2022) 4733–4745.
[8] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, M. Ester, SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines, Journal of Cheminformatics 9 (2017).
[9] M. Jiang, Z. Li, S. Zhang, S. Wang, X. Wang, Q. Yuan, Z. Wei, Drug-target affinity prediction using graph neural network and contact maps, RSC Adv. 10 (2020) 20701–20712.
[10] J. Jiménez, M. Škalič, G. Martínez-Rosell, G. De Fabritiis, KDEEP: Protein-Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks, Journal of Chemical Information and Modeling 58 (2018) 287–296, PMID:29309725.
[11] M. Karimi, D. Wu, Z. Wang, Y. Shen, DeepAffinity: interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks, Bioinformatics 35 (2019) 3329–3338.
[12] M. Karimi, D. Wu, Z. Wang, Y. Shen, Explainable Deep Relational Networks for Predicting Compound-Protein Affinities and Contacts, Journal of Chemical Information and Modeling 61 (2021) 46–66, PMID:33347301.
[13] T.N. Kipf, M. Welling, Semi-Supervised Classification with Graph Convolutional Networks, in: in: International Conference on Learning Representations (ICLR), 2017.
[14] G. Landrum, RDKit: Open-Source Cheminformatics Software (2016).
[15] M. Lennox, N. Robertson, B. Devereux, Modelling Drug-Target Binding Affinity using a BERT based Graph Neural network, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021, pp. 4348–4353, https://doi.org/10.1109/EMBC46164.2021.9629695.
[16] M. Li, Z. Lu, Y. Wu, Y. Li, BACPI: a bi-directional attention neural network for compound-protein interaction and binding affinity prediction, Bioinformatics 38 (2022) 1995–2002.
[17] S. Li, F. Wan, H. Shu, T. Jiang, D. Zhao, J. Zeng, MONN: A Multi-objective Neural Network for Predicting Compound-Protein Interactions and Affinities, Cell Systems 10 (2020) 308–322.e11.
[18] S. Li, J. Zhou, T. Xu, L. Huang, F. Wang, H. Xiong, W. Huang, D. Dou, H. Xiong, Structure-aware Interactive Graph Neural Networks for the Prediction of Protein-Ligand Binding Affinity, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021.
[19] T. Li, X. Zhao, L. Li, Co-VAE: Drug-target binding affinity prediction by co-regularized variational autoencoders, IEEE Transactions on Pattern Analysis and Machine Intelligence (2021), 1–1.
[20] X. Lin, K. Zhao, T. Xiao, Z. Quan, Z.-J. Wang, P.S. Yu, DeepGS: Deep Representation Learning of Graphs and Sequences for Drug-Target Binding Affinity Prediction, in: ECAI, 2020.
[21] E. López-Medina, P. López, I.C. Hurtado, D.M. Dávalos, O. Ramirez, E. Martínez, J.A. Díazgranados, J.M. Oñate, H. Chavarriaga, S. Herrera, B. Parra, G. Libreros, R. Jaramillo, A.C. Avendaño, D.F. Toro, M. Torres, M.C. Lesmes, C.A. Rios, I. Caicedo, Effect of Ivermectin on Time to Resolution of Symptoms Among Adults With Mild COVID-19: A Randomized Clinical Trial, JAMA 325 (2021) 1426–1435.
[22] M. Michel, D. Menéndez Hurtado, A. Elofsson, PconsC4: fast, accurate and hassle-free contact predictions, Bioinformatics 35 (2018) 2677–2679.
[23] T. Nguyen, H. Le, T.P. Quinn, T. Nguyen, T.D. Le, S. Venkatesh, GraphDTA: predicting drug-target binding affinity with graph neural networks, Bioinformatics 37 (2020) 1140–1147.
[24] T.M. Nguyen, T. Nguyen, T.M. Le, T. Tran, GEFA: Early fusion approach in drug-target affinity prediction, IEEE/ACM transactions on computational biology and bioinformatics (2021).
[25] H. Öztürk, A. Özgür, E. Ozkirimli, DeepDTA: deep drug-target binding affinity prediction, Bioinformatics 34 (2018) i821–i829.
[26] T. Pahikkala, A. Airola, S. Pietilä, S. Shakyawar, A. Szwajda, J. Tang, T. Aittokallio, Toward more realistic drug-target interaction predictions, Briefings in Bioinformatics 16 (2014) 325–337.
[27] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Advances in Neural Information Processing Systems 32, Curran Associates Inc, 2019, pp. 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.
[28] Y. Pu, J. Li, J. Tang, F. Guo, DeepFusionDTA: drug-target binding affinity prediction with information fusion and hybrid deep-learning ensemble model, IEEE/ACM Transactions on Computational Biology and Bioinformatics (2021), 1–1.
[29] M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri, D.R. Koes, Protein-Ligand Scoring with Convolutional Neural Networks, Journal of Chemical Information and Modeling 57 (2017) 942–957, PMID:28368587.
[30] B. Ramsundar, P. Eastman, P. Walters, V. Pande, Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More, O'Reilly Media, 2019.
[31] A.S. Rifaioglu, R. Cetin Atalay, D. Cansen Kahraman, T. Dogan, M. Martin, V. Atalay, MDeePred: novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery, Bioinformatics 37 (2020) 693–704.
[32] Y. Rong, W. Huang, T. Xu, J. Huang, DropEdge: Towards Deep Graph Convolutional Networks on Node Classification, in: in: International Conference on Learning Representations (ICLR), 2020.

[33] X. Ru, X. Ye, T. Sakurai, Q. Zou, NerLTR-DTA: drug-target binding affinity prediction based on neighbor relationship and learning to rank, Bioinformatics Btac048 (2022).
[34] H.M. Salah, J.L. Mehta, Meta-analysis of the effect of aspirin on mortality in COVID-19, The American journal of cardiology 142 (2021) 158.
[35] P.A. Shar, W. Tao, S. Gao, C. Huang, B. Li, W. Zhang, M. Shahen, C. Zheng, Y. Bai, Y. Wang, Pred-binding: large-scale protein-ligand binding affinity prediction, Journal of Enzyme Inhibition and Medicinal Chemistry 31 (2016) 1443–1450, PMID:26888050.
[36] M.M. Stepniewska-Dziubinska, P. Zielenkiewicz, P. Siedlecki, Development and evaluation of a deep learning model for protein-ligand binding affinity prediction, Bioinformatics 34 (2018) 3666–3674.
[37] J. Tang, A. Szwajda, S. Shakyawar, T. Xu, P. Hintsanen, K. Wennerberg, T. Aittokallio, Making Sense of Large-Scale Kinase Inhibitor Bioactivity Data Sets: A Comparative and Integrative Analysis, Journal of Chemical Information and Modeling 54 (2014) 735–743, PMID:24521231.
[38] M. Thafar, A.B. Raies, S. Albaradei, M. Essack, V.B. Bajic, Comparison Study of Computational Prediction Tools for Drug-Target Binding Affinities, Frontiers in Chemistry 7 (2019) 782.
[39] L. Van der Maaten, G. Hinton, Visualizing Data using t-SNE, Journal of Machine Learning Research 9 (2008) 2579–2605.
[40] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph Attention Networks, in: International Conference on Learning Representations, 2018. https://openreview.net/forum?id=rJXMpikCZ.
[41] H. Wang, D. Lian, Y. Zhang, L. Qin, X. Lin, GoGNN: Graph of Graphs Neural Network for Predicting Structured Entity Interactions, in: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI'20, 2021a.
[42] K. Wang, R. Zhou, Y. Li, M. Li, DeepDTAF: a deep learning method to predict protein-ligand binding affinity, Briefings in Bioinformatics 22 (2021). Bbab072.
[43] Y. Wang, Y. Min, X. Chen, J. Wu, Multi-View Graph Contrastive Representation Learning for Drug-Drug Interaction Prediction, Association for Computing Machinery, New York, NY, USA, 2021, pp. 2921–2933, 10.1145/3442381.3449786.
[44] Q. Wu, Z. Peng, I. Anishchenko, Q. Cong, D. Baker, J. Yang, Protein contact prediction using metagenome sequence data and residual neural networks, Bioinformatics 36 (2019) 41–48.
[45] K. Xu, W. Hu, J. Leskovec, S. Jegelka, How Powerful are Graph Neural Networks?, in: International Conference on Learning Representations, 2019. https://openreview.net/forum?id=ryGs6iA5Km.
[46] Z. Yang, W. Zhong, L. Zhao, C. Yu-Chian Chen, MGraphDTA: deep multiscale graph neural network for explainable drug-target binding affinity prediction, Chem. Sci. 13 (2022) 816–833.
[47] W. Yuan, G. Chen, C.Y.-C. Chen, FusionDTA: attention-based feature polymerizer and knowledge distillation for drug-target binding affinity prediction, Briefings in Bioinformatics 23 (2021). Bbab506.
[48] W. Zhang, Z. Li, W. Guo, W. Yang, F. Huang, A Fast Linear Neighborhood Similarity-Based Network Link Inference Method to Predict MicroRNA-Disease Associations, IEEE/ACM Transactions on Computational Biology and Bioinformatics 18 (2021) 405–415.
[49] Q. Zhao, F. Xiao, M. Yang, Y. Li, J. Wang, AttentionDTA: prediction of drug-target binding affinity using attention model, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019, pp. 64–69, https://doi.org/10.1109/BIBM47256.2019.8983125.
[50] J. Zhou, S. Li, L. Huang, H. Xiong, F. Wang, T. Xu, H. Xiong, D. Dou, Distance-aware molecule graph attention network for drug-target binding affinity prediction, arXiv preprint arXiv:2012.09624 (2020).