Prediction of multi-relational drug–gene interaction via Dynamic hyperGraph Contrastive Learning

Wen Tao, Yuansheng Liu, Xuan Lin, Bosheng Song and Xiangxiang Zeng

Corresponding author: College of Computer Science and Electronic Engineering, Hunan Univers ity, Changsha, 410082, Hunan, China.
E-mail: yuanshengliu@hnu.edu.cn

Abstract

Drug–gene interaction prediction occupies a crucial position in various areas of drug discovery, such as drug repurposing, lead discovery and off-target detection. Previous studies show good performance, but they are limited to exploring the binding interactions and ignoring the other interaction relationships. Graph neural networks have emerged as promising approaches owing to their powerful capability of modeling correlations under drug–gene bipartite graphs. Despite the widespread adoption of graph neural network-based methods, many of them experience performance degradation in situations where high-quality and sufficient training data are unavailable. Unfortunately, in practical drug discovery scenarios, interaction data are often sparse and noisy, which may lead to unsatisfactory results. To undertake the above challenges, we propose a novel Dynamic hyperGraph Contrastive Learning (DGCL) framework that exploits local and global relationships between drugs and genes. Specifically, graph convolutions are adopted to extract explicit local relations among drugs and genes. Meanwhile, the cooperation of dynamic hypergraph structure learning and hypergraph message passing enables the model to aggregate information in a global region. With flexible global-level messages, a self-augmented contrastive learning component is designed to constrain hypergraph structure learning and enhance the discrimination of drug/gene representations. Experiments conducted on three datasets show that DGCL is superior to eight state-of-the-art methods and notably gains a 7.6% performance improvement on the DGIdb dataset. Further analyses verify the robustness of DGCL for alleviating data sparsity and over-smoothing issues.

Keywords: drug–gene interaction; hypergraph; contrastive learning; graph neural networks

INTRODUCTION

Identification of interactions between drugs and genes allows the simultaneous discovery of new drugs and target candidates [1–3]. Unlike expensive conventional experiments [4], computational methods aim to detect drug–gene interactions (DGIs) accurately and cost-effectively [5–7], laying a cornerstone for drug discovery and the management of a wide range of diseases [8–12].

Nevertheless, previous studies on DGI prediction are mainly focused on binding interactions and overlook the other relation types of interactions, such as modulator and allosteric modulator [13]. Modulator interactions occur when a drug regulates or changes the activity of its target gene. Interactions of this class may not be relevant to the binding of a drug to a gene. In contrast, allosteric modulator interactions take place when drugs have impacts on their target genes via a binding site different from the natural ligand site. Therefore, it is vital to take into account multi-relational DGIs.

Recently, graph neural networks (GNNs) have shown prominent performance in DGI prediction owing to their excellent ability to express relational data [14]. In GNN-based methods, multi-hop convolutions are performed on the drug–gene bipartite graph, exploiting the initial graph structure to capture explicit drug–gene relations. However, such information harvested from a fixed graph structure is easily contaminated by noisy interactions stemming from inevitable false-positive samples [15–17].

While GNN-based models have offered state-of-the-art performance, most of them adhere to the supervised learning paradigm, in which model training requires sufficient DGI signals. As a consequence, they are potentially vulnerable to data sparsity, which is a pervasive issue in the drug discovery field [18]. Since only a limited number of DGIs are determined following a series of wet-lab experiments, GNN-based models may not be able to perform high-quality prediction in this scenario. Apart from the data sparsity issue, GNN-based methods struggle with the over-smoothing issue. The node embeddings tend to become indistinguishable as the number of layers increases [19, 20], leading to a decrease in the model performance.

In this paper, we present a novel Dynamic hyperGraph Contrastive Learning (DGCL) framework that injects local and global relations into the representations for DGI prediction. A hypergraph is composed of nodes and hyperedges which can connect two or more nodes [21, 22]. The complex global-level correlations of drugs and genes can be naturally modeled by hypergraphs. Differing from previous studies, DGCL distinguishes itself through its effective handling of sparse and noisy data, along with its capacity to mitigate the over-smoothing issue. Furthermore, our work not

only offers accurate DGI prediction but also facilitates the understanding of the mechanisms underlying drug interactions. Concretely, graph convolutions are applied to embed local structure information. As a complement to the explicit local topology-aware embeddings, a hypergraph structure learning module is designed to encode implicit global dependencies. Compared with the predefined hyperedges, our module optimizes the potential node relatedness dynamically during the training process to construct the adaptively learnable hypergraph structure, which is better aligned with the model's objective and more robust to the noisy information of original DGIs.

Although the learnable hypergraph structure models DGIs in a more capable and flexible manner, the superiority of the hypergraph comes with an overfitting issue. To overcome the challenge brought by dynamic hypergraph learning, the automatic learning of the hypergraph is constrained through contrastive learning in which representations of two views are aligned. In DGCL, we are naturally provided with two contrastive views: the original bipartite graph and the learned hypergraph. Instead of performing graph augmentation, which might corrupt pivotal interactions, essential information hidden in the original graph is retained with our self-augmented contrastive learning. Moreover, both the over-smoothing issue and the data sparsity issue can be alleviated by contrastive learning due to the discrimination enhancement of representations and the auxiliary self-supervised signals.

Our main contributions are summarized as follows:

- We propose a novel dynamic hypergraph contrastive learning framework named DGCL to mitigate the over-smoothing and data sparsity issues of GNN-based models. To the best of our knowledge, this is the first hypergraph-enhanced method for multi-relational DGI prediction.
- In the proposed DGCL, the implicit global dependency is modeled via global message passing on the dynamically learnable hypergraph, which is guided by the designed effective self-augmented contrastive learning paradigm with local and global dependency as two contrastive views.
- Extensive experiments conducted on three public datasets show that our approach consistently outperforms competitive baselines.

RELATED WORK

DGI prediction aims to determine the interactions of drug–gene pairs and has greatly contributed to the success of drug development. Here we mainly introduce the prevailing feature- and network-based methods.

Inspired by the success of artificial intelligence [23–25], most feature-based models leverage well-established deep learning architectures to generate representations of drugs and genes. For instance, DeepConv-DTI [26] utilizes fully-connected layers with ECFP [27] as inputs to encode drugs. The DeepDTA model [28] attempts to learn drug and target representations by employing convolutional neural networks (CNNs) [29] on drug SMILES strings and protein sequences. GraphDTA [30] converts SMILES into graphs and strives to generate representations of drugs by leveraging GNNs.

In addition to feature-based methods, several studies have investigated network-based approaches for DGI prediction, garnering considerable attention [31, 32]. These methods rely on the construction of a network to learn low-dimensional representations of drugs and genes. In the network, nodes represent

biological entities and edges represent relationships between entities. The ability of network-based methods to integrate heterogeneous data and model complex relationships has led to a promising performance [33]. Moreover, network-based models can achieve excellent performance under the setting where models are completely driven by the network topology while disregarding extra features. Hence, the unavailability of domain features is circumvented.

Nevertheless, limitations still exist in the methods mentioned above. Firstly, feature-based methods require careful feature engineering or dimensionality reduction techniques [34, 35]. Their satisfactory performance largely relies on the availability of high-quality domain features [36–38]. Meanwhile, some features are essential for DGI prediction but may be absent for most kinds of drugs/genes in large-scale datasets [39]. Secondly, network-based methods are incapable of distilling the long-range context, which is critical to understanding the underlying mechanisms of DGI and facilitating the discovery of new drugs/druggable genes. Lastly, the majority of existing feature-based and network-based methods only focus on binding interaction records while neglecting other potential relations among drugs and genes. In this study, we propose to model interaction types between drugs and genes by contrasting and integrating the graph-based local-level and hypergraph-based global-level representations.

METHODOLOGY
Overview of DGCL

The overall framework of our proposed DGCL is illustrated in Figure 1. At a high level, DGCL first constructs a bipartite graph, and then learns the locally aggregated embeddings of drugs/genes via the conventional graph-oriented message passing process. Meanwhile, DGCL figures out global semantic neighbors to complement the local structure information through the designed hypergraph neural network (HGNN) with dynamic hypergraph structure learning. Moreover, self-augmented graph contrastive learning is performed between the local topology and global semantic views to constrain the learned hypergraph structure. Based on the abovementioned well-designed modules, DGCL estimates the type of interaction between drug and gene with state-of-the-art performance.

Explicit local relation modeling

Given DGIs with M drugs and N genes, we first construct the original drug–gene bipartite graph which models the observed interactions. The explicit relation within drugs and genes is encoded by the node's local topology. Inspired by LightGCN [40], we adopt a simplified version of GCN [41] to capture the local dependency. As a common practice, we assign initial ID embedding vectors to drugs and genes in the raw bipartite graph. Specifically, we create two learnable parameter matrices defined as $\mathbf{E}^{(d)} \in \mathbb{R}^{M \times d}$ and $\mathbf{E}^{(g)} \in \mathbb{R}^{N \times d}$, where d denotes the embedding size. Unlike the conventional GCN, the activation function and feature transformation are discarded during the aggregation of embeddings from neighbors, as these may not provide benefits to the target task. The propagation process in layer l can be expressed as:

$$\mathbf{E}_{l-1} = [\mathbf{E}_{l-1}^{(d)}, \ \mathbf{E}_{l-1}^{(g)}], \tag{1}$$

$$\bar{\mathbf{E}}_l = p(\bar{\mathbf{A}})\mathbf{E}_{l-1}, \tag{2}$$

where $\mathbf{E}_{l-1}^{(d)}/\mathbf{E}_{l-1}^{(g)}$ represents either the initial embeddings $\mathbf{E}_0^{(d)}/\mathbf{E}_0^{(g)}$, or the aggregated embeddings of drugs/genes from the $(l-1)$-th
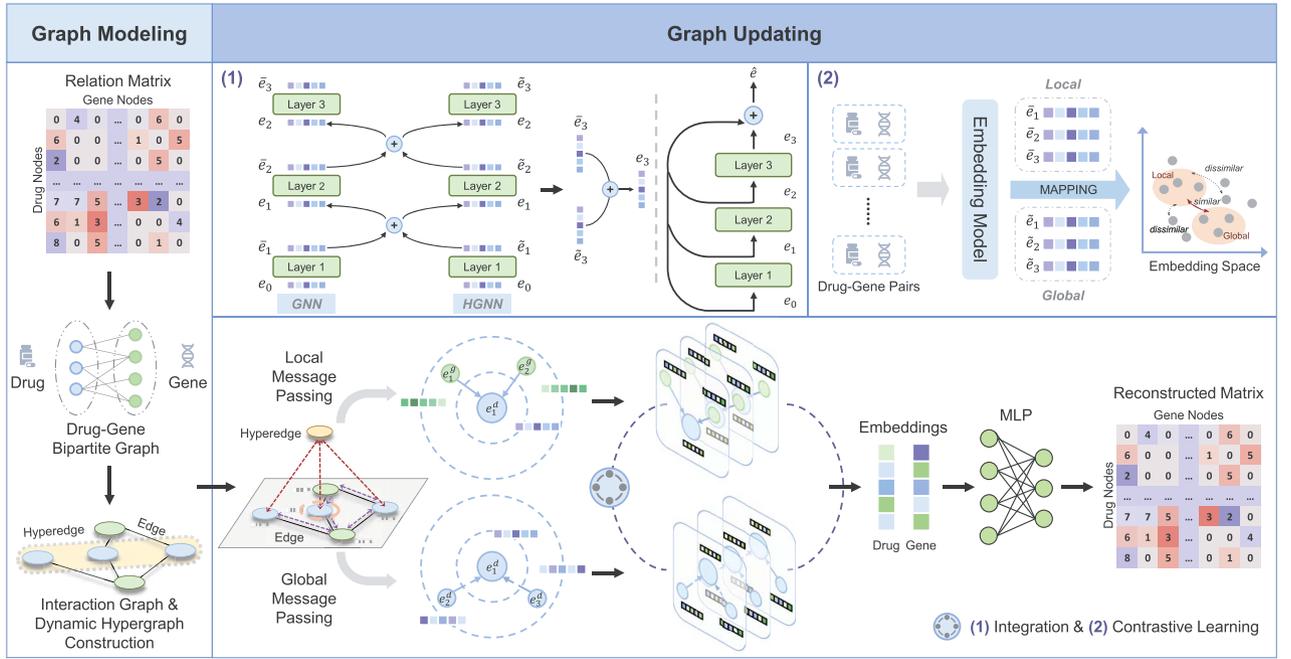
Figure 1. The overall framework of DGCL. The first frame describes the interaction graph and dynamic hypergraph construction process. The solid black lines represent pair-wise edges, which can only connect two nodes. The yellow rounded rectangle indicates a hyperedge, which is capable of connecting two or more nodes. In the second frame, each node communicates with its 1-hop neighbors through local message passing (dashed purple lines) by pair-wise edges (solid black lines). In addition, the long-range information is propagated through global message passing (dashed red lines) by utilizing the hyperedge (yellow node) as an intermediate hub. Then, the embeddings from each layer of GNN and HGNN are bridged by the integration and contrastive learning components, which are illustrated in the top left and top right parts of the second frame, respectively. With the final drug/gene embeddings that aggregate both local messages and non-local messages, the interaction types of given drug–gene pairs are predicted via multi-layer perceptron (MLP).

propagation layer. $p(\cdot)$ denotes the edge dropout operation which mitigates overfitting. $\bar{\mathbf{A}}$ is the normalized adjacency matrix calculated as

$$\bar{\mathbf{A}} = \bar{\mathbf{D}}^{-\frac{1}{2}}(\mathbf{A} + \mathbf{I}_{M+N})\bar{\mathbf{D}}^{-\frac{1}{2}}, \tag{3}$$

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{R} \\ \mathbf{R}^\top & 0 \end{bmatrix}, \tag{4}$$

where $\bar{\mathbf{D}}$ represents a $(M + N) \times (M + N)$ degree matrix whose diagonal entries are the degrees of nodes, and $\bar{\mathbf{D}}_{ii} = \sum_j (\mathbf{A} + \mathbf{I}_{M+N})_{ij}$. $\mathbf{I}_{M+N}$ denotes the identity matrix. $\mathbf{A}$ represents the adjacency matrix. During message passing, the self-connection operation is included by adding the identity matrix to the adjacency matrix. It integrates the $(l - 1)$th layer information of nodes, avoiding information dilution. $\mathbf{R} \in \mathbb{R}^{M \times N}$ denotes the binary matrix representing DGIs.

Dynamic hypergraph structure learning

Though existing graph-based models are capable of extracting local structural information of observed DGIs, the underlying correlations among drugs or genes are hard to be learned from the immutable graph structure. To empower DGI prediction with global structure learning, thus breaking the above-mentioned limitation, we propose dynamic hypergraph learning optimized along with model training to inject the structural information from the global perspective. In comparison with the pre-defined hypergraph, which may not be intrinsically adaptive to the prediction task, resulting in suboptimal performance, the proposed dynamic hypergraph structure is adjustable to different tasks and datasets.

We argue that it is sufficient to construct the hypergraph which reflects the implicit dependencies among drugs and genes on the fly according to the topological relationship of the initial bipartite graph and the task-specific supervised signals. Specifically, we perform the dynamic hypergraph structure learning based on two learnable adjacency matrices $\mathbf{H}^{(d)} \in \mathbb{R}^{M \times K}$ and $\mathbf{H}^{(g)} \in \mathbb{R}^{N \times K}$ denoting drug- and gene-hyperedge matrices, respectively. K denotes the number of hyperedges. However, when there is a huge number of hyperedges or drugs/genes, the cost of computing the hyperedge matrix for drugs/genes will skyrocket. For the sake of scaling to a larger number of hyperedges and drugs/genes, directly learning dense matrices may be impractical. To remedy this issue, we assume that nodes' connections with hyperedges are more likely akin if their local structures are similar. Based on the assumption, we factorize $\mathbf{H}^{(d)}$ and $\mathbf{H}^{(g)}$ as follows:

$$\mathbf{H}^{(d)} = \mathbf{E}^{(d)}\mathbf{W}^{(d)}, \quad \mathbf{H}^{(g)} = \mathbf{E}^{(g)}\mathbf{W}^{(g)}, \tag{5}$$

where $\mathbf{W}^{(d)}$ and $\mathbf{W}^{(g)} \in \mathbb{R}^{d \times K}$ denote the trainable parametric matrices for drug-hyperedge matrix and gene-hyperedge matrix, respectively. $\mathbf{E}^{(d)}$ and $\mathbf{E}^{(g)}$ are the embeddings of drugs and genes, respectively. In this way, we approximate the drug-hyperedge matrix and gene-hyperedge matrix with low-rank matrices to reduce the size of model parameters and avoid overfitting.

Implicit global relation modeling

To harvest the potential relationships among drugs and genes, we design the hypergraph message passing layer, which performs the embedding propagation upon the adaptive hypergraph. Following HGNN [22], we initially aggregate the embeddings of drugs/genes

to generate the embeddings of hyperedges. Subsequently, we compute the node embeddings by aggregating the information from hyperedges. In our hypergraph message passing layer, the messages of drugs and genes can be efficiently communicated in a manner of breaking the distance limitations.

To be specific, we formulate the hypergraph message passing on the learned hypergraph as follows:

$$\tilde{\mathbf{E}}_l^{(d)} = \mathbf{H}^{(d)}\mathbf{Z}^{(d)} = \mathbf{H}^{(d)}\mathbf{H}^{(d)\top}\mathbf{E}_{l-1}^{(d)}, \tag{6}$$

where $\mathbf{Z}^{(d)} \in \mathbb{R}^{K \times d}$ represents the hyperedge embeddings for drugs. $\tilde{\mathbf{E}}_l^{(d)} \in \mathbb{R}^{M \times d}$ denotes the embeddings of drugs learned from the dynamic hypergraph in layer l. By analogy with equation (6), we can generate the hypergraph-based gene embeddings $\tilde{\mathbf{E}}_l^{(g)} \in \mathbb{R}^{N \times d}$.

In a nutshell, our proposed hypergraph message passing process takes the drug and gene embeddings $\mathbf{E}_{l-1}^{(d)}$ and $\mathbf{E}_{l-1}^{(d)}$, and the learnable hypergraph adjacency matrices $\mathbf{H}^{(d)}$ and $\mathbf{H}^{(g)}$ as inputs to model the global property for DGI patterns.

Local-global self-augmented contrastive learning

The foregoing modules combine the learning of the dynamic hypergraph structure with the exploration of global relations to learn non-local messages over the entire graph, alleviating over-smoothing. However, it will lead to the overfitting problem when the hypergraph structure learning relies solely on the supervised signals. Drawing inspiration from the successful practice of contrastive learning [42, 43], we propose an effective contrastive learning paradigm to provide auxiliary self-supervised signals for constraining the learned hypergraph structure.

More precisely, our contrastive learning framework contrasts the local topology-aware embeddings of the original interaction graph with the global semantic-aware embeddings of the dynamic hypergraph. Compared with the common graph contrastive learning methods which generate two extra views by augmenting the original graph through random perturbation, our proposed self-augmented contrastive learning paradigm avoids misleading the information encoding of drugs and genes. In our method, the local-level and global-level embeddings serve as two naturally augmented views.

Based on InfoNCE [44], our proposed contrastive learning objective is formulated as follows:

$$\mathcal{L}_s^{(d)} = \sum_{i=1}^M \sum_{l=1}^L -\log \frac{\exp(\mathrm{sim}(\bar{\mathbf{e}}_{i,l}^{(d)}, \tilde{\mathbf{e}}_{i,l}^{(d)})/\tau)}{\sum_{i'=1}^M \exp(\mathrm{sim}(\bar{\mathbf{e}}_{i,l}^{(d)}, \tilde{\mathbf{e}}_{i',l}^{(d)})/\tau)}, \tag{7}$$

where $\mathrm{sim}(\cdot)$ represents computing the cosine similarity of local interaction-aware embedding $\bar{\mathbf{e}}_{i,l}^{(d)}$ and global interaction-aware embedding $\tilde{\mathbf{e}}_{i,l}^{(d)}$, and τ denotes the temperature hyperparameter which controls the penalties on hard negative samples. We treat two embeddings of the local learning view and global learning view from the same drug/gene as a positive pair, and generate negative samples using all the possible combinations of local and global embeddings from different drugs/genes. The distance between a positive pair is minimized while the distance between a negative pair is maximized. The object $\mathcal{L}_s^{(g)}$ for genes is defined in an analogous way. It is crucial to underscore that we apply a stop-gradient operation on $\bar{\mathbf{e}}_{i,l}^{(d)}/\bar{\mathbf{e}}_{i,l}^{(g)}$, which encodes the local graph structure. This operation makes the model more inclined to constrain the hypergraph. The above contrastive learning guides the local features to supervise the embedding learning based on global features, thus restricting the optimization of the implicit structure.

Integration

To deepen the integration of knowledge from local and global perspectives, the following operations are performed in each layer. Concretely, we aggregate local messages first, then propagate non-local embeddings. At last, we combine the local-global dependency embeddings to generate input embeddings:

$$\mathbf{e}_{i,l}^{(d)} = \bar{\mathbf{e}}_{i,l}^{(d)} + \tilde{\mathbf{e}}_{i,l}^{(d)}, \quad \mathbf{e}_{j,l}^{(g)} = \bar{\mathbf{e}}_{j,l}^{(g)} + \tilde{\mathbf{e}}_{j,l}^{(g)}, \tag{8}$$

where $\bar{\mathbf{e}}_{i,l}^{(d)}$ denotes the embedding of drug i aggregated from itself and its explicit neighbors in layer l. $\tilde{\mathbf{e}}_{i,l}^{(d)}$ is the lth layer embedding of drug i updated through the HGNN. We perform element-wise addition on $\bar{\mathbf{e}}_{i,l}^{(d)}$ and $\tilde{\mathbf{e}}_{i,l}^{(d)}$ to construct the fused embedding $\mathbf{e}_{i,l}^{(d)}$ of drug i. The input embedding $\mathbf{e}_{j,l}^{(g)}$ for gene j is calculated analogously. $\mathbf{e}_{i,l}^{(d)}$ and $\mathbf{e}_{j,l}^{(g)}$ will be utilized for local and global information extraction as inputs in the next layer.

The residual connections [45] are further employed to calculate the final drug/gene embeddings as follows:

$$\hat{\mathbf{e}}_i^{(d)} = \sum_{l=0}^L \mathbf{e}_{i,l}^{(d)}, \quad \hat{\mathbf{e}}_j^{(g)} = \sum_{l=0}^L \mathbf{e}_{j,l}^{(g)}. \tag{9}$$

This operation emphasizes the semantics of each layer's output and avoids the over-smoothing issue.

Prediction

The fully connected layers with the concatenation of embeddings from drug i and gene j as inputs are applied to estimate the probability of each drug–gene relation type:

$$\hat{y} = \mathrm{softmax}(\mathbf{W}_2\mathrm{ReLU}(\mathbf{W}_1(\hat{\mathbf{e}}_i^{(d)}||\hat{\mathbf{e}}_j^{(g)}) + b_1) + b_2), \tag{10}$$

where $\hat{y} \in \mathbb{R}^C$, with C as the number of interaction types. $\mathbf{W}_1$ and $\mathbf{W}_2$ are learnable weight matrices. b_1 and b_2 are bias vectors.

Formally, the objective function of DGI prediction is formulated as follows:

$$\mathcal{L}_p = -\sum_{c=1}^C y_c \log \hat{y}_c, \tag{11}$$

where $y_c \in \{0, 1\}$, indicating whether the class label c is the true label for the drug–gene pair. $\hat{y}_c$ denotes the predicted probability that the relation type of the DGI sample is type c.

Optimization

The goal of our model is to predict the interaction types between drugs and genes. We regard the contrastive learning task as an auxiliary task to jointly learn with the prediction task using a multi-task learning strategy. The combined objective is defined as follows:

$$\mathcal{L} = \mathcal{L}_p + \lambda_1(\mathcal{L}_s^{(d)} + \mathcal{L}_s^{(g)}) + \lambda_2\|\Theta\|_2^2, \tag{12}$$

where λ_1 and λ_2 are trade-off hyper-parameters. λ_1 controls the weight of the contrastive learning loss and λ_2 is the weight decay coefficient for the model-specific regularization term to prevent overfitting. Θ denotes the parameters of the model.

Table 1. The statistical details of the datasets

Datasets	Drugs	Genes	Interactions	Types
DrugBank	425	11 284	80 924	2
DGIdb	1185	1664	11 366	14
LINCS L1000	187	3769	20 610	2

EXPERIMENTS
Experimental setup

Datasets. Three multi-relational datasets are utilized in our experiments to evaluate the performance of our model and the baselines: DrugBank, DGIdb and LINCS L1000. These datasets are widely used, and they possess complementary strengths, enabling them to verify models from various perspectives. The statistical details of the datasets are shown in Table 1.

DrugBank [46]: The DrugBank dataset contains two types of relation about drugs regulating gene expression: upregulation and downregulation of genes (i.e., increased and decreased).

DGIdb [13]: A dataset mined from 41 data sources contains 14 types of DGI. The interaction types include ligand, activator, inhibitor, vaccine, cofactor, etc.

LINCS L1000 [47]: It profiles changes in gene expression and cellular processes that are perturbed by different drugs, providing two types of interactions: increased and decreased.

Compared models. Following [14], we compare our DGCL with two groups of competitive methods: matrix factorization (MF)-based methods (MC, GRALS and F-EAE) and GNN-based methods (GC-MC, sRGCNN, PinSage, IGMC and CoSMIG). The results of baselines are taken from [14].

- **MC** [48]: It proposes an algorithm for accurately completing missing entries in a partially observed matrix based on convex optimization.
- **GRALS** [49]: Graph Regularized Alternating Least Squares (GRALS) is a graph-regularized matrix factorization method that integrates structural information in the form of graphs.
- **F-EAE** [50]: A method for prediction of interactions between two or more sets of objects using deep neural networks. The interactions are represented as an exchangeable matrix or higher-dimensional tensor.
- **GC-MC** [51]: A graph auto-encoder framework that performs message passing on the bipartite graph which represents interaction data.
- **sRGCNN** [52]: It leverages the local stationarity structures of graphs and reduces the size of learnable parameters, by adopting graph convolutional neural networks and recurrent neural networks.
- **PinSage** [53]: A model utilizes random walks to sample neighbors and performs graph convolutions on the graph with node attributes.
- **IGMC** [54]: A matrix completion method that does not rely on side information. It involves two steps: extracting enclosing subgraphs and applying a GNN.
- **CoSMIG** [14]: It proposes a subgraph representation learning framework to predict the relation types of interactions through a communicative message passing mechanism.

Evaluation metrics. To evaluate the performance of multi-relational DGI prediction, we adopt the widely used accuracy score (ACC) as the evaluation metric, which measures the percentage of samples that have been predicted correctly.

Implementation details. For the implementation of our proposed framework, the model is trained with Adam [55] as the optimizer and a learning rate of $1e^{-3}$. The embedding dimension is set to 128. The batch size is set as 4,096. The number of graph convolution layers and hypergraph propagation layers is configured as 3. The weights λ_1 and λ_2 are searched from $\{1e^{-4}, 1e^{-3}, 1e^{-2}, 1e^{-1}\}$ and $\{1e^{-8}, 1e^{-7}, 1e^{-6}, 1e^{-5}\}$, respectively. The edge dropout ratio of the original bipartite graph is tuned from $\{0.25, 0.5, 0.75\}$. The temperature τ in contrastive learning is selected from $\{0.1, 0.3, 1, 3, 10\}$.

Overall performance

To verify the superior predictive performance of our model, we compare DGCL with other competitive methods. Table 2 summarizes the experimental results carried out on DrugBank and DGIdb datasets. The best result is marked in boldface and the second best result is underlined. Based on these results, we find the following observations and draw the main conclusions:

- Compared with the MF-based methods MC, GRALS and F-EAE, GNN-based methods (GC-MC, sRGCNN, PinSage, IGMC and CoSMIG) exhibit consistent superiority. This suggests that the embeddings of drugs and genes learned by the matrix factorization techniques may be inadequate to carry the characteristics of DGI types. For GNN-based methods, their robust prediction performance could be attributed to the effectiveness of exploiting relations in a local region by applying GNN to the bipartite graph. GNN endows these methods with the ability to capture explicit correlations between local neighbors, thereby enhancing DGI prediction.
- Our proposed DGCL consistently outperforms both MF-based and GNN-based methods. In comparison with the strongest baseline CoSMIG, DGCL has a 1.6% improvement with respect to ACC in DrugBank. For the DGIdb dataset, DGCL reaches a significant improvement over CoSMIG by 7.6% in terms of ACC. One of the reasons for the performance improvements is the injection of global messages. The other reason is that the embeddings generated from traditional GNN-based methods are easily misled by sparse and noisy datasets of DGI, and will become indistinguishable with the increase of propagation layers.
- The reasons for DGCL performing significantly better with the DGIdb dataset than with the DrugBank dataset can be attributed to two main factors. Firstly, the performance of all models on DGIdb surpasses that on DrugBank, indicating that the inherent complexity of the DGIdb dataset is lower than that of the DrugBank dataset. It is easier for models to capture the nonlinear relationships of DGIdb samples compared with DrugBank samples. Secondly, due to the relatively fewer and sparser samples in DGIdb compared to DrugBank, coupled with its imbalanced nature, DGCL's advantages in mitigating data sparsity and handling imbalanced datasets become more pronounced. As a result, DGCL demonstrates a significant improvement over other models on the DGIdb dataset.

Ablation studies

To investigate the effects of the prominent components in DGCL, we conduct ablation studies on DrugBank and DGIdb by analyzing the contribution of our proposed implicit global structure learning and self-augmented contrastive learning.

Table 2. Performance comparison with baselines on DrugBank and DGIdb datasets. '–' represents a vacant position, indicating that the result is not reported in the original paper

	Methods	DrugBank		DGIdb	
		Validation ACC	Test ACC	Validation ACC	Test ACC
MF-based	MC	–	0.518(0.013)	–	0.559(0.009)
	GRALS	–	0.532(0.021)	–	0.578(0.016)
	F-EAE	–	0.566(0.004)	–	0.623(0.003)
GNN-based	GC-MC	–	0.586(0.008)	–	0.601(0.005)
	sRGCNN	–	0.602(0.010)	–	0.689(0.007)
	PinSage	–	0.629(0.004)	–	0.713(0.005)
	IGMC	–	0.634(0.003)	–	0.803(0.006)
	CoSMIG	0.658(0.008)	0.678(0.003)	0.840(0.011)	0.852(0.012)
Proposed	DGCL	**0.674(0.003)**	**0.694(0.002)**	**0.920(0.001)**	**0.928(0.002)**

Table 3. Ablation study on main components of DGCL

Variants	DrugBank		DGIdb	
	Validation ACC	Test ACC	Validation ACC	Test ACC
$DGCL_{w/o\ CL}$	0.661(0.004)	0.672(0.002)	0.897(0.004)	0.911(0.002)
$DGCL_{w/o\ Hyper}$	0.651(0.002)	0.668(0.004)	0.866(0.003)	0.888(0.003)
DGCL	**0.674(0.003)**	**0.694(0.002)**	**0.920(0.001)**	**0.928(0.002)**

Table 3 presents the results of the following variants:

- $DGCL_{w/o\ CL}$: This is a DGCL variant in which we disable the self-augmented contrastive learning module. It relies on both local and global messages to capture more rich association relations among drugs and genes. However, its dynamic global neighbors are explored without the constraints of contrasting with local neighbors.
- $DGCL_{w/o\ Hyper}$: This is a variant of DGCL in which the implicit global structure learning component is removed. The embeddings of drugs/genes learned by this variant only aggregate information from local neighborhoods. The global messages over an entire graph are not collected by the model.

From Table 3, we can observe that the two variants without different components both degrade the performance of predicting DGI. This indicates that the removed components are of great significance for the performance improvements in modeling interaction types. Moreover, $DGCL_{w/o\ Hyper}$ exhibits inferior performance compared to $DGCL_{w/o\ CL}$, further emphasizing the benefits of exploiting implicit high-order relations through the dynamic hypergraph.

Performance of DGCL in alleviating over-smoothing

To validate that the proposed DGCL can alleviate the over-smoothing issue, we compute the Mean Average Distance (MAD) [56] values of the embeddings generated by DGCL and its two variants ($DGCL_{w/o\ Hyper}$ and $DGCL_{w/o\ CL}$). MAD is a quantitative metric that measures the global smoothness of a graph. As it is obtained by averaging the cosine distance between each node pair, smaller MAD values indicate more indistinguishable embeddings, i.e., the more obvious over-smoothing phenomenon. The MAD results are reported in Table 4. From the results, we can see that $DGCL_{w/o\ Hyper}$ and $DGCL_{w/o\ CL}$ have smaller MAD values than DGCL. Apart from this, the MAD values of our

proposed DGCL get close to 1, which means that DGCL is effective in over-smoothing alleviation. We consider the reasons for the above observations from two aspects. First, the integration of embeddings from local and global relation learning is help to avoid the situation of embeddings tending to converge to the same values. Second, the self-augmented contrastive learning module makes the drug/gene embeddings more discriminative by contrasting the local information encoded with GCN and global information encoded with HGNN.

Performance of DGCL in addressing data sparsity

We evaluate the robustness of our proposed DGCL in the face of sparse interaction data. Towards this end, we focus on drugs with less than 20 interactions. Then, we compare the prediction performance of DGCL, $DGCL_{w/o\ Hyper}$, $DGCL_{w/o\ CL}$ and the best baseline CoSMIG on this fraction of drugs. Figure 2 reports the results for performance on DrugBank and DGIdb. As shown in Figure 2, we can find that DGCL and its variants all perform better than CoSMIG. Meanwhile, the performance gap between DGCL and CoSMIG in sparse interaction data is larger than the gap in overall performance presented in Table 2. In addition, removing either the dynamic hypergraph learning component or the self-augmented contrastive learning component compromises the performance of DGCL in sparse data. These results verify that DGCL has the potential in handling sparse data by facilitating the embedding learning for tail drugs/genes. We ascribe this potential to the well-designed global message passing mechanism and contrastive learning module.

For tail drugs/genes, GNNs aggregate messages from a limited number of neighbors, potentially leading to biased or underrepresented representations. In contrast, DGCL leverages dynamic hypergraph construction to generate a larger set of neighbors for these tail nodes, effectively capturing long-range dependencies. DGCL can identify drugs/genes that appear to have similar representations and establish connections between them.

Table 4. MAD values of the drug/gene embeddings

	Variants	DrugBank		DGIdb	
		Validation	Test	Validation	Test
Drug	DGCL$_{w/o\ CL}$	0.978	0.978	0.974	0.977
	DGCL$_{w/o\ Hyper}$	0.678	0.606	0.927	0.931
	DGCL	**0.997**	**0.997**	**0.995**	**0.997**
Gene	DGCL$_{w/o\ CL}$	0.980	0.976	0.984	0.985
	DGCL$_{w/o\ Hyper}$	0.819	0.781	0.934	0.929
	DGCL	**0.998**	**0.999**	**0.999**	**0.999**
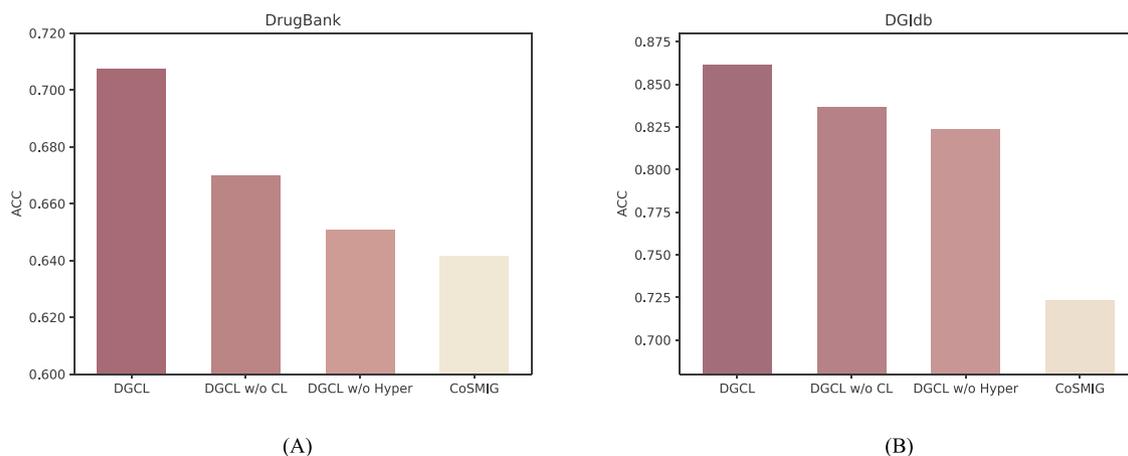
Figure 2. The performance of DGCL, DGCL's variants and CoSMIG on sparse data.

Consequently, DGCL enables the propagation of more meaningful messages through the hyperedges, reducing the reliance on their sparse neighbors. In addition to hyperedges, the contrastive learning module can provide auxiliary self-supervised signals to enhance the representation learning of tail drugs/genes.

Hyperparameter analysis

We further evaluate the influence of the following hyperparameters on DGCL's performance: the dimensionality of embeddings and the number of layers. The evaluation results on DrugBank and DGIdb are shown in Figure 3.

- **The dimensionality of embeddings**. We conduct experiments to validate whether the dimensionality of embeddings will exert an influence on performance. Concretely, we vary the dimensionality of embeddings in {16, 32, 64, 128}. From Figure 3, we can observe that the performance of DGCL increases with the higher dimensionality of embeddings. Such results are expected since the low dimensionality of embeddings is insufficient to preserve the information of drugs/genes.
- **The number of layers**. Experiments have shown that the optimal model performance is achieved when the number of propagation layers is 3. Moreover, we can see that the performance of DGCL is relatively stable under different configurations of layer number. This advantage can be ascribed to our proposed local and global aggregations, which can aggregate sufficient information from local and global neighbors through one propagation layer. In addition, the local and global aggregations also equip our model with the potential ability against the over-smoothing issue in deeper architectures.

External test on LINCS L1000

To confirm the generalization ability of our proposed DGCL, we perform the external evaluation on the LINCS L1000 dataset. In particular, we utilize DrugBank as the training set and LINCS L1000 as the testing set. The results of DGCL, DGCL$_{w/o\ Hyper}$, DGCL$_{w/o\ CL}$ and CoSMIG are presented in Figure 4. As can be seen from the results, DGCL achieves the best performance among all models and significantly outperforms the best baseline CoSMIG by 4.6% with respect to ACC. These observations demonstrate the generalization powers of DGCL for discovering potential interactions.

Visualization analysis

We visualize embeddings of the samples using t-distributed stochastic neighbor embedding (t-SNE) [57], aiming to test whether samples of the same relation would have similar embeddings. Figure 5(A) and (B) display the visualizations of the embeddings from DrugBank and DGIdb, respectively. It can be observed that the embeddings obtained from DGCL fall into distinct clusters which correspond to their interaction types in the two-dimensional space, indicating the excellent performance of our model in discriminating different kinds of DGI.

Imbalanced dataset performance analysis

In addition to accuracy, we also calculate sensitivity for DGCL and the best baseline CoSMIG on the DrugBank and DGIdb datasets to comprehensively evaluate predictive models from multiple perspectives. The sensitivity values for DGCL and CoSMIG on the DrugBank dataset are 0.693 and 0.677, respectively. On the DGIdb dataset, the sensitivity values for DGCL and CoSMIG are 0.815 and 0.713, respectively. As DGIdb is an imbalanced dataset, the
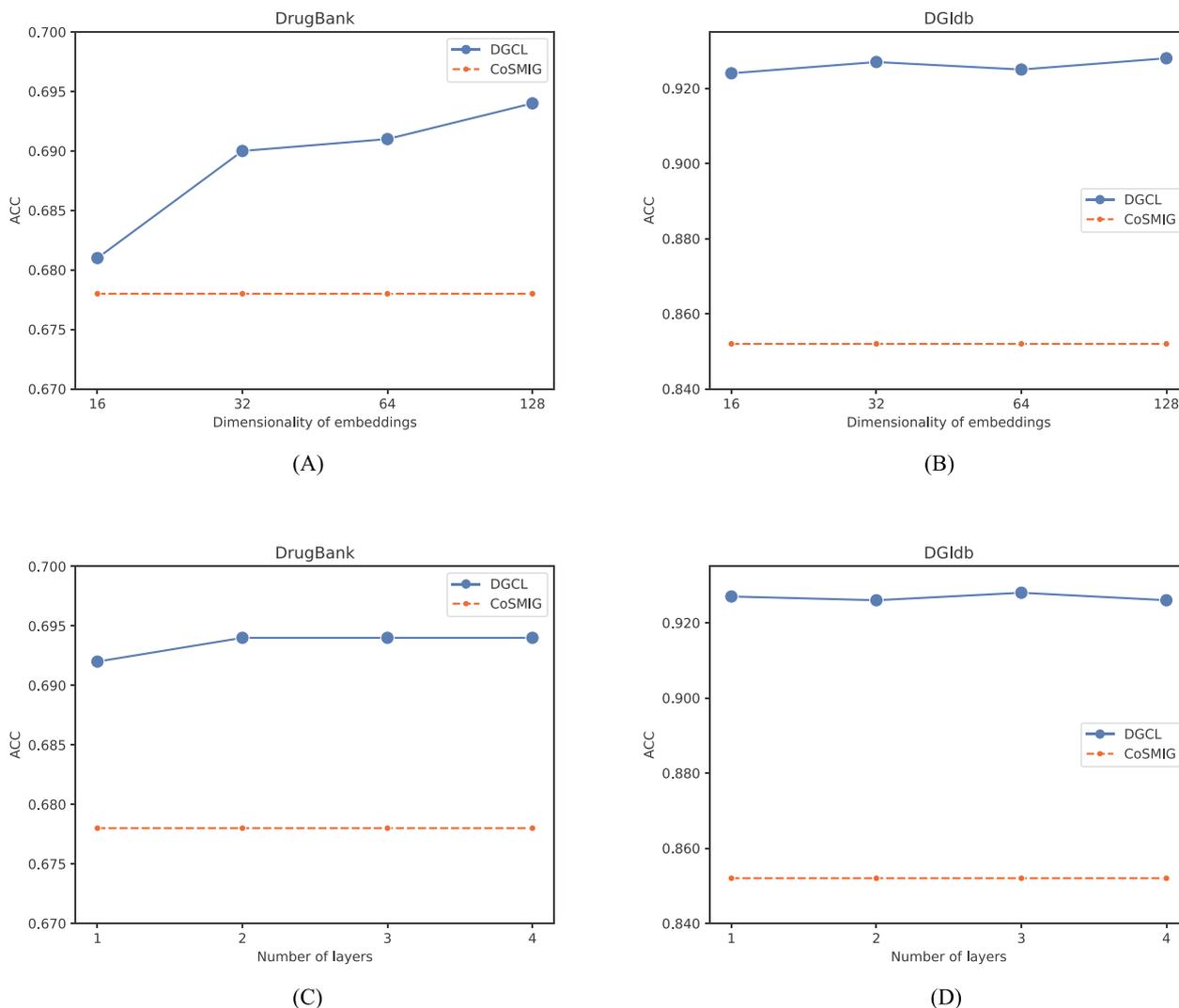
(A)

(B)

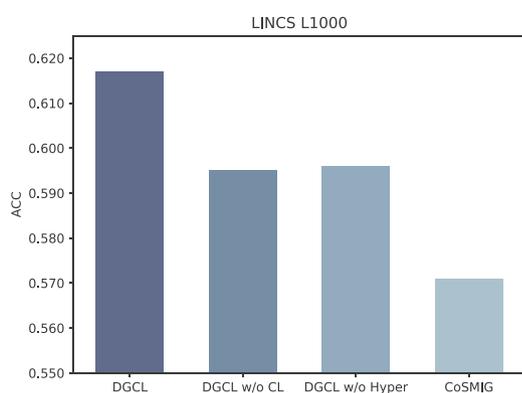(C)

(D)

Figure 3. The impact of embedding dimensionality and layer number.

Figure 4. Performance of DGCL on LINCS L1000 dataset.

DGCL identifies novel DGIs

To further validate DGCL's predictive capabilities, we conduct novel DGI prediction for the gene HMOX1 using DrugBank as the training set. HMOX1 is well-documented for its anti-inflammatory and anti-oxidant activities, constituting a protective mechanism that regulates the inflammatory response in lesional skin. Supplementary Table 1 provides a detailed presentation of the top-10 DGCL-predicted novel DGIs, including the canonical name of the drug, the predicted interaction, gene name and supporting literature references. These predictions are not present in the current datasets but have supporting literature references.

The results reveal that 70% of the top 10 drugs (7 out of 10) are substantiated by previous studies in the literature. Specifically, DGCL predicts that Estriol, a weak estrogen, could lead to an increase in HMOX1 expression. This prediction finds support in prior research indicating higher levels of HMOX1 expression in non-menopausal patients (with higher blood levels of estrogen) compared to menopausal patients [58]. Deficiencies in female sex hormones appear to be among the risk factors influencing the progression of psoriasis in women. Therefore, maintaining normal physiological levels, achieved through Estriol treatment, may potentially prevent or mitigate the disease.
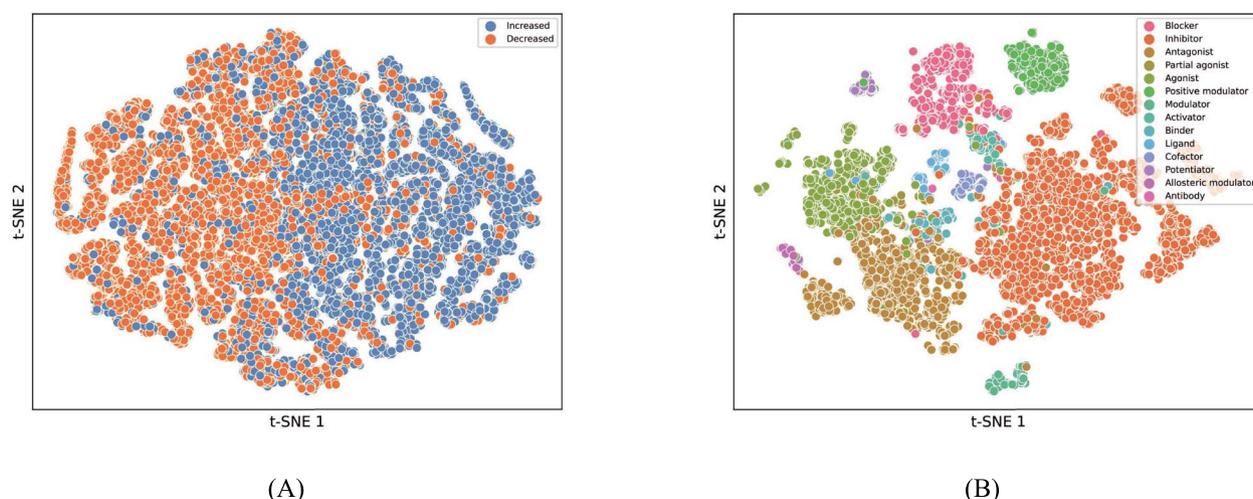
models may perform favorably on the majority classes but exhibit suboptimal performance on the minority class. If a class exhibits notably low performance, this can substantially influence the sensitivity score due to the equal contribution of each class to the final evaluation. According to the sensitivity values of the DGCL and CoSMIG models on the DGIdb dataset, it is evident that DGCL outperforms CoSMIG in handling imbalanced datasets.

Figure 5. Visualization of interaction embeddings. Interactions from DrugBank and DGIdb are illustrated in (**A**) and (**B**), respectively.

Application of DGCL to drug–drug interaction prediction

We rank all potential drug–drug interactions, which may lead to polypharmacy side effects, based on the embedding similarities of drug pairs. Subsequently, we verify the top 10 predicted drug–drug interactions using evidence from multiple sources. As demonstrated in Supplementary Table 2, while there is no interaction information between different drugs in the DrugBank dataset, 7 out of these 10 predicted drug–drug interactions have been confirmed. This shows our DGCL's capability to infer implicit correlations among drugs by generating similar embeddings solely based on DGI data.

CONCLUSION

In this work, we present a novel hypergraph-based framework, coined DGCL, to perform message aggregation in a local and global region for seeking better DGI modeling. In DGCL, we design the dynamic hypergraph structure learning and self-augmented contrastive learning to improve the prediction performance and robustness against over-smoothing and data sparsity issues. Through extensive experiments on several datasets, we have demonstrated the effectiveness and generalizability of the proposed DGCL. Our work shows that the adaptive hypergraph can offer promising prospects for an effective means of interaction modeling, motivating broad expansion to many other applications, such as drug–drug interaction prediction and protein–protein interaction prediction, in future explorations.

While our current investigation has centered on utilizing the drug–gene bipartite graph for multi-relational DGI prediction, it is important to highlight the good adaptability and flexibility that the DGCL framework offers. Depending on the specific task at hand, the drug–gene bipartite graph can be substituted with the alternative bipartite graph.

Key Points
- We present a general dynamic hypergraph contrastive learning framework for multi-relational DGI prediction, with remarkable effectiveness, robustness and generalization.
- The proposed model employs the cooperation of dynamic hypergraph structure learning and hypergraph message passing, which adaptively captures the implicit global dependency. Furthermore, the self-augmented contrastive learning module is designed to provide guidance for the dynamic learning of the hypergraph.
- The experiments on three datasets demonstrate that our model outperforms the state-of-the-art methods. In addition, further analyses are presented to justify the superiority of the proposed model.

DATA AND CODE AVAILABILITY

All data and source code of this study are available at https://github.com/wentao228/DGCL.

REFERENCES

1. Strittmatter SM. Overcoming drug development bottlenecks with repurposing: old drugs learn new tricks. *Nat Med* 2014;**20**(6):590–1.
2. Pritchard J-LE, O'Mara TA, Glubb DM. Enhancing the promise of drug repositioning through genetics. *Front Pharmacol* 2017;**8**:896.
3. Bagherian M, Sabeti E, Wang K, *et al.* Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. *Brief Bioinform* 2021;**22**(1):247–69.
4. Stachel SJ, Sanders JM, Henze DA, *et al.* Maximizing diversity from a kinase screen: identification of novel and selective pan-Trk inhibitors for chronic pain. *J Med Chem* 2014;**57**(13):5800–16.
5. Tsubaki M, Tomii K, Sese J. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics* 2019;**35**(2):309–18.

6. Zheng S, Li Y, Chen S, *et al*. Predicting drug–protein interaction using quasi-visual question answering system. *Nat Mach Intell* 2020;**2**(2):134–40.

7. Bai P, Miljković F, John B, Haiping L. Interpretable bilinear attention network with domain adaptation improves drug–target prediction. *Nat Mach Intell* 2023;**5**(2):126–36.

8. Wang M, Wei Z, Jia M, *et al*. Deep learning model for multi-classification of infectious diseases from unstructured electronic medical records. *BMC Med Inform Decis Mak* 2022;**22**(1):1–13.

9. Khan S, Imran A, Khan AA, *et al*. Systems biology approaches for the prediction of possible role of chlamydia pneumoniae proteins in the etiology of lung cancer. *PloS One* 2016;**11**(2):e0148530.

10. Khan S, Zaidi S, Alouffi AS, *et al*. Computational proteome-wide study for the prediction of escherichia coli protein targeting in host cell organelles and their implication in development of colon cancer. *ACS Omega* 2020;**5**(13):7254–61.

11. Li Y, Khan S, Chaudhary AA, *et al*. Proteome-wide screening for the analysis of protein targeting of *Chlamydia pneumoniae* in endoplasmic reticulum of host cells and their possible implication in lung cancer development. *Biocell* 2022;**46**(1): 87.

12. Wang Y, Imran A, Shami A, *et al*. Decipher the *Helicobacter pylori* protein targeting in the nucleus of host cell and their implications in gallbladder cancer: an insilico approach. *J Cancer* 2021;**12**(23): 7214.

13. Freshour SL, Kiwala S, Cotto KC, *et al*. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res* 2021;**49**(D1): D1144–51.

14. Rao J, Zheng S, Mai S, Yang Y. Communicative subgraph representation learning for multi-relational inductive drug-gene interaction prediction. In: blueDe Raedt L, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, International Joint Conferences on Artificial Intelligence, Vienna, Austria, pp. 3919–25, 2022. Main Track.

15. Rishton GM. Reactive compounds and in vitro false positives in HTS. *Drug Discov Today* 1997;**2**(9):382–4.

16. Sink R, Gobec S, Pecar S, Zega A. False positives in the early stages of drug discovery. *Curr Med Chem* 2010;**17**(34):4231–55.

17. Weng T, Zhou X, Li K, *et al*. Distributed approaches to butterfly analysis on large dynamic bipartite graphs. *IEEE Trans Parallel Distrib Syst* 2023;**34**(2):431–45.

18. Dara S, Dhamercherla S, Jadav SS, *et al*. Machine learning in drug discovery: a review. *Artif Intell Rev* 2022;**55**(3):1947–99.

19. Min Y, Wenkel F, Wolf G. Scattering GCN: overcoming over-smoothness in graph convolutional networks. *Adv Neural Inf Process Syst* 2020;**33**:14498–508.

20. Zhou K, Huang X, Li Y, *et al*. Towards deeper graph neural networks with differentiable group normalization. *Adv Neural Inf Process Syst* 2020;**33**:4917–28.

21. Zhou D, Huang J, Schölkopf B. Learning with hypergraphs: clustering, classification, and embedding. *Adv Neural Inf Process Syst* 2006;**19**:1601–8.

22. Feng Y, You H, Zhang Z, *et al*. Hypergraph neural networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press, Honolulu, Hawaii, USA, 2019;**33**:3558–65.

23. Patronov A, Papadopoulos K, Engkvist O. Has artificial intelligence impacted drug discovery? *Methods Mol Biol* 2022;**2390**:153–76.

24. Jayatunga MKP, Xie W, Ruder L, *et al*. AI in small-molecule drug discovery: a coming wave. *Nat Rev Drug Discov* 2022;**21**:175–6.

25. Jumper J, Evans R, Pritzel A, *et al*. Highly accurate protein structure prediction with alphafold. *Nature* 2021;**596**(7873):583–9.

26. Lee I, Keum J, Nam H. DeepConv-DTI: prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Comput Biol* 2019;**15**(6): e1007129.

27. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model* 2010;**50**(5):742–54.

28. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* 2018;**34**(17): i821–9.

29. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM* 2017;**60**(6):84–90.

30. Nguyen T, Le H, Quinn TP, *et al*. GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics* 2021;**37**(8):1140–7.

31. 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): 573.

32. Zeng X, Zhu S, Liu X, *et al*. deepDR: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics* 2019;**35**(24):5191–8.

33. Zeng X, Xinqi T, Liu Y, *et al*. Toward better drug discovery with knowledge graph. *Curr Opin Struct Biol* 2022;**72**:114–26.

34. Wang Y, Zhai Y, Ding Y, Zou Q. SBSM-Pro: support bio-sequence machine for proteins. *arXiv preprint arXiv:2308.10275*. 2023. https://doi.org/10.48550/arXiv.2308.10275.

35. Liu Y, Shen X, Gong Y, *et al*. Sequence alignment/map format: a comprehensive review of approaches and applications. *Brief Bioinform* 2023;**24**:bbad320.

36. Yang X, Niu Z, Liu Y, *et al*. Modality-DTA: multimodality fusion strategy for drug–target affinity prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2022;**20**(2):1200–10.

37. Tang T, Zhang X, Liu Y, *et al*. Machine learning on protein–protein interaction prediction: models, challenges and trends. *Brief Bioinform* 2023;**24**(2): bbad076.

38. Song B, Luo X, Luo X, *et al*. Learning spatial structures of proteins improves protein–protein interaction prediction. *Brief Bioinform* 2022;**23**(2): bbab558.

39. Wang Y, Min Y, Chen X, Ji W. Multi-view graph contrastive representation learning for drug-drug interaction prediction. In: *Proceedings of the Web Conference*. ACM, Ljubljana, Slovenia, 2021:**2021**:2921–33.

40. He X, Deng K, Wang X, *et al*. LightGCN: Simplifying and powering graph convolution network for recommendation. In: Jimmy H, Yi C, Xueqi C, Jaap K, Vanessa M, Ji-Rong W, Yiqun L, editor, *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Virtual Event, China, 2020, pp. 639–48.

41. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: *International Conference on Learning Representations*. ICLR, Toulon, France, 2017.

42. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: *International Conference on Machine Learning*, pp. 1597–607. PMLR, Virtual Event, 2020.

43. He K, Fan H, Wu Y, Xie S, Girshick R. Momentum contrast for unsupervised visual representation learning. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. IEEE, Seattle, WA, USA, 2020, pp. 9729–38.

44. Oord Avd, Li Y, Vinyals O. Representation learning with contrastive predictive coding. *arXiv preprint arXiv: 1807.03748*. 2018. https://doi.org/10.48550/arXiv.1807.03748.

45. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Las Vegas, Nevada, USA, 2016, pp. 770–8.

46. 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.

47. Subramanian A, Narayan R, Corsello SM, *et al*. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**(6): 1437–1452.e17.

48. Candes E, Recht B. Exact matrix completion via convex optimization. *Commun ACM* 2012;**55**(6):111–9.

49. Rao N, Hsiang-Fu Y, Ravikumar PK, Dhillon IS. Collaborative filtering with graph information: consistency and scalable methods. *Adv Neural Inf Process Syst* 2015;**28**:2107–15.

50. Hartford J, Graham D, Leyton-Brown K, Ravanbakhsh S. Deep models of interactions across sets. In: *International Conference on Machine Learning*, Stockholm, Sweden, pp. 1909–18. PMLR, 2018.

51. Berg Rvd, Kipf TN, Welling M. Graph convolutional matrix completion. In: *KDD Workshop on Deep Learning Day*. ACM, London, UK, 2018.

52. Monti F, Bronstein M, Bresson X. Geometric matrix completion with recurrent multi-graph neural networks. *Adv Neural Inf Process Syst* 2017;**30**:3697–707.

53. Ying R, He R, Chen K, *et al*. Graph convolutional neural networks for web-scale recommender systems. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, London, UK, 2018, pp. 974–83.

54. Zhang M, Chen Y. Inductive matrix completion based on graph neural networks. In: *International Conference on Learning Representations*. ICLR, Addis Ababa, Ethiopia, 2020.

55. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: *International Conference on Learning Representations*, ICLR, San Diego, CA, USA, 2015.

56. Chen D, Lin Y, Li W, *et al*. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press, New York, New York, USA, 2020;**34**:3438–45.

57. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**(11):2579–605.

58. Sobolev V, Soboleva A, Denisova E, *et al*. Differential expression of estrogen-responsive genes in women with psoriasis. *J Pers Med* 2021;**11**(9):925.