Hierarchical Structure-aware Graph Prompting for Drug-Drug Interaction Prediction

Yuhan Ye[1,2,†], Jingbo Zhou[2,⋆], Shuangli Li[1,2], Congxi Xiao[1,2], Haochao Ying[3], and Hui Xiong[4,5,⋆]

[1]The University of Science and Technology of China
[2]Business Intelligence Lab, Baidu Research [3]Zhejiang University
[4] Thrust of Artifcial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), China [5] Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[1]xyqqj@mail.ustc.edu.cn, [2]{zhoujingbo, lishuangli, xiaocongxi}@baidu.com
[3]haochaoying@zju.edu.cn, [4,5]xionghui@ust.hk

Abstract. Drug-drug interaction (DDI) prediction holds crucial significance in biomedical applications such as polypharmacy and clinical decision-making. Considering the limited availability of labeled DDI relations, it is promising to effectively extract underlying knowledge from drug molecular graphs by self-supervised learning to enhance DDI prediction performance, owing to the recent successes in graph pre-training for molecular representation. Nonetheless, employing existing graph pre-training methods directly reveals significant disparities persisting between the pre-training tasks and the ultimate objective of DDI prediction. Addressing this, we propose HS-GPF, a novel hierarchical structure-aware graph prompting framework tailored for DDI prediction. Its key component is a specially designed graph prompt learning mechanism, which significantly integrates the pre-training and the final DDI task into a uniform task format. This is achieved through an adaptive dual-level prompting process featuring unique virtual tokens. Aligned with our hierarchical structure-aware pre-training, it effectively activates relevant knowledge for DDI prediction, fostering a more seamless integration between the pre-trained model and complex drug interactions. Extensive experiments across various scales of real-world datasets demonstrate that our method outperforms existing state-of-the-art baselines, even in challenging cold-start scenarios.

Keywords: Drug-Drug Interaction · Graph Self-supervised Learning · Graph Prompting.

1 Introduction

Drug-Drug Interactions (DDI) arise when multiple drugs are taken together, leading to pharmacological reactions that affect their absorption, distribution,

⋆ Corresponding authors. † This work was done when the first author was an intern at Baidu Research under the supervision of Jingbo Zhou.

metabolism, and excretion. Polypharmacy is vital for managing complex health conditions, notably in optimizing efficacy, as evidenced by the widespread use of combination chemotherapy in cancer treatment [7]. However, the associated risk of DDI poses a serious threat, potentially impacting drug efficacy and leading to adverse effects, increasing the risk of morbidity and mortality [17]. Therefore, it is crucial to identify potential drug interactions in practice, which is essential for minimizing side effects and maximizing synergistic benefits.

Extensive efforts have been invested in developing automated computational DDI prediction methods to address challenges associated with traditional wet chemical experiments, constrained by limitations in scale, cost, and duration [25]. From early machine learning methods relying on heterogeneous features [22, 30, 31] to recent graph learning-based approaches utilizing drug molecular graphs and drug-entities graphs for improved prediction [28, 1], both chemical structures [6, 15, 14] and drug-related relations [23, 10, 13] have demonstrated effectiveness.

Since the availability of labeled DDI relations is consistently constrained by the high cost associated with gathering such data, leveraging graph self-supervised learning methods is a promising solution to extracting preserved knowledge that should be beneficial for DDI prediction. For instance, the knowledge about local substructures within drugs and the external relationships among drug-related entities should be useful for enhancing DDI prediction accuracy, while existing methods for capturing such knowledge often rely heavily on supervision information [15, 14, 23, 10, 13]. Due to the scarcity of DDI labels [6], comprehensively capturing such knowledge to enhance DDI prediction performance remains challenging. Drawing inspiration from the achievements of self-supervised learning methods in molecular representation learning [32, 16], and recognizing the unexplored specialized approaches for DDI prediction, the adoption of a dedicated self-supervised learning approach emerges as a practical strategy for effectively capturing intricate drug relations and the significance of essential local substructures, thereby facilitating more accurate DDI predictions.

However, self-supervised tasks formulated based on the intrinsic structural features of molecules may exhibit certain disparities with respect to our overarching goal of DDI prediction. The established paradigms of "pre-train and fine-tune" with well-designed self-supervised graph learning tasks can aid in grasping the structural patterns of drug molecules. Nevertheless, since the objectives of these self-supervised learning tasks primarily revolve around capturing general graph patterns rather than DDI specifics, misalignments persist between the self-supervised tasks and the ultimate objective of DDI prediction. Hence, effectively transferring such preserved knowledge to downstream DDI tasks may incur knowledge loss and require additional efforts, warranting further investigation into optimal knowledge transfer mechanisms.

To overcome these limitations, we propose a novel Hierarchical Structure-aware Graph Prompting Framework (named HS-GPF) for DDI prediction that adopts the "pre-train, prompt, and fine-tune" strategy. Firstly, for comprehensive knowledge acquisition, we employ hierarchical self-supervised learning tasks, which consider drug molecules and their special substructures, known as motifs,

to organize complex drug interactive dependencies. Secondly, we define the DDI prediction problem as our final goal and introduce a dual-level graph prompting mechanism to narrow the gap in the fine-tuning stage. This strategy aligns with our multi-level pre-training objectives in a prompting-enhanced manner, facilitating the improved transfer and application of DDI pattern knowledge gained from pre-training.

In the pre-training phase, we implement a hierarchical self-supervised learning scheme tailored for learning DDI patterns at both drug and motif levels. At the drug level, each drug is represented as a molecular graph, enabling fine-grained structure modeling with atoms as nodes and chemical bonds as edges. We utilize a global similarity learning task to improve representation learning, integrating richer structural and semantic domain knowledge. At the motif level, drugs or motifs are modeled as nodes in a constructed drug-motif interaction graph for complex relationship modeling. We employ a mask edge prediction task to learn node connectivity, extracting intricate relation knowledge from the graph. Additionally, we include a contrastive learning component to aid in multi-level information alignment. In this way, our hierarchical self-supervised learning tasks capture structure knowledge and potential drug associations.

Moreover, to address the gap that arises in our graph pre-training when transitioning to the downstream DDI prediction task, we propose a novel prompt-based strategy tailored for DDI prediction. Inspired by successful prompt strategies in NLP [11], our adaptive prompting technique is specially crafted for DDI task, aiming to align the downstream task with the pretext task. Although some existing graph prompt methods explore the prompting idea on graph data, they often focus on a single pre-training task, such as link prediction [19, 12]. However, for our hierarchical self-supervised learning tasks, a unique drug- and motif-oriented prompting strategy is required to effectively handle these multi-level structures and objectives. To tackle this, we propose an adaptive dual-level prompting process aligned with our hierarchical perspectives. This prompt assigns different levels of attention to different perspectives, facilitating the effective integration of crucial motif information. It aims to better utilize internal structural information and external relational associations acquired during pre-training, enhancing the application of functional motifs and drug relationship knowledge for improved predictions. Specifically, we introduce motif and marker prompt functions tailored for different perspectives, further unifying the pre-training and fine-tuning models with consistent learning objectives. This approach enhances the activation of pre-trained knowledge, ultimately contributing to improved DDI prediction across various scenarios. The main contributions are summarized as follows:

– To the best of our knowledge, we are among the first to develop a novel hierarchical graph prompting learning architecture (HS-GPF) for DDI prediction. The incorporation of multi-level self-supervised learning tasks ensures a comprehensive thorough of structural knowledge and complex drug interactive dependencies. While the dual-level prompting strategy effectively leverages the knowledge acquired during pre-training.

- We introduce an innovative motif-based prompting strategy that seamlessly integrates representations from our hierarchical self-supervised tasks into DDI prediction. The novelty lies in dual-level prompt functions, actively guiding downstream tasks to maximize the utilization of learned knowledge.
- Extensive experiments on diverse real-world datasets demonstrate the effectiveness of the proposed HS-GPF over state-of-the-art approaches, even in challenging cold-start scenarios.

2 Related Work

DDI prediction. Early machine learning methods assume that structurally similar drugs may share similar DDIs. Fingerprint-based studies like [22, 17] utilized chemical fingerprints for structural similarity. Besides, [30] included side effects, and [31, 2] integrated heterogeneous drug features for enhanced similarity assessment. The focus has recently shifted towards exploiting graph neural networks for better predictions. Chen et al.[1] utilized graph convolution for molecular graph encoding, and Zitnik et al.[33] extracted information in a multimodal graph with drugs and proteins as nodes. Further advancements have exploited deeper insights. Huang et al.[6] mined chemical sequential patterns as functional substructures. Nyamabo et al.[15] extracted substructure information within the receptive field of GNN layers, identifying pairwise interactions between substructures. Wang et al.[23] treated DDI relations as an interaction graph to model important relationships. Lin et al.[10] and Lyu et al.[13] expanded the approach by constructing drug knowledge graphs to exploit topological data. Different from previous works, we take both external relations among drugs and internal meaningful substructures within drugs into consideration.

Self-supervised Learning of molecular graphs. Self-supervised learning on molecular graphs, employing diverse tasks, has been widely used for molecular representation learning [9]. For example, Hu et al. [5] have explored node-level tasks like context prediction and node masking, with graph-level tasks such as property prediction. Additionally, Zhang et al. [32] and Rong et al. [16] integrated graph-level motif label prediction task. Moreover, Sun et al. [18] introduced substructure substitution augmentation, while Fang et al. [3] applied the constructed chemical element knowledge graph to assist graph contrastive learning task. Primarily designed for learning molecular structural information, these methods may struggle to capture complex drug interactions. In contrast, we explore a novel graph prompt learning mechanism that independently adapts drug molecules and their motifs to the specific domain of the downstream task.

Prompt-based Learning. Most self-supervised learning methods focus on task-specific designs, overlooking the gap between pre-training and downstream tasks, which may restrict their prediction ability. Inspired by the success of prompt-based learning methods for various domains, some recent studies introduced learnable graph prompts to narrow this gap. GPPT [19] leveraged task and structure tokens for a graph-aware prompting function, while GraphPrompt [12] proposed a framework based on subgraph similarity, hinging on a learnable

prompt to actively guide downstream tasks to exploit pre-trained model in a task-specific manner. Different from existing efforts, our method introduces a unique prompt mechanism aligned with our hierarchical pre-training process.

3 Preliminaries

In this section, we present the notations used in this work and formulate the problem of DDI prediction.

Definition 1: Drug Molecular Graph. Given a drug set $D = \{D_1, ..., D_m\}$, each drug can be defined as a molecular graph $\mathcal{D}_i = (\mathcal{A}, \mathcal{B})$, where $\mathcal{A}$ denotes atom set, $\mathcal{B}$ denotes chemical bond set and m is the number of drugs.

Definition 2: Motif Graph. Given a molecular graph $\mathcal{D}_i$, we denote its meaningful motifs as: $M = \{M_1, ..., M_n\}$ where n is the number. Each motif graph is defined as a subgraph $\mathcal{M}_j = \{\mathcal{A}_{(j)}, \mathcal{B}_{(j)}\}$, with $\mathcal{A}_{(j)} \subseteq \mathcal{A}$ and $\mathcal{B}_{(j)} \subseteq \mathcal{B}$.

Definition 3: Drug-Motif Interaction Graph. We form a drug-motif interaction graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with drug nodes $\mathcal{V}^D$ and motif nodes $\mathcal{V}^M$. The edges $\mathcal{E}$ comprise three types: drug-drug, motif-motif, and drug-motif, encoding complex relationships. Construction details are provided in Appendix A.1.

Problem: Drug-Drug Interaction Prediction. Given a DDI dataset $\mathcal{T}$ consisting of tuples (D_i, D_j, r) and an interaction type set $\mathcal{I} = \{I_k\}_{k=1}^{C}$, where C is the total number of possible types, $\{D_i, D_j\} \in D$ represent a drug pair with interaction r of type I_k. We aim to learn a model $\mathcal{F}^{DDI} : D \times D \times \mathcal{I} \rightarrow y$ to predict the probability indicating the chance that the drug pair having interaction r.

4 Model Framework

As depicted in Figure 1, we introduce HS-GPF, a hierarchical structure-aware graph prompting framework for DDI prediction. It starts with hierarchical self-supervised tasks to capture structural and relational insights, enabling robust and transferable representations. After the pretraining phase, we implement a novel graph prompt learning mechanism to align optimization objectives between the pre-training and fine-tuning stages. Following the "pre-train, prompt, fine-tune" paradigm, we tailor the prompting function to accommodate the complexities of drug molecular data and DDI patterns, enhancing predictions.

4.1 Pre-train, Prompt, and Fine-tune

Before delving into the framework details, we present an overview of the "Pre-train, Prompt, and Fine-tune" pipeline. We begin by defining our prompting function as follows:

Definition 4: Dual-level Graph prompt function. We propose a novel dual-level graph prompt function containing a motif prompt function and a marker prompt function, which aids in reformulating the final DDI task to align with the template of the pre-training tasks. Firstly, for each drug molecular

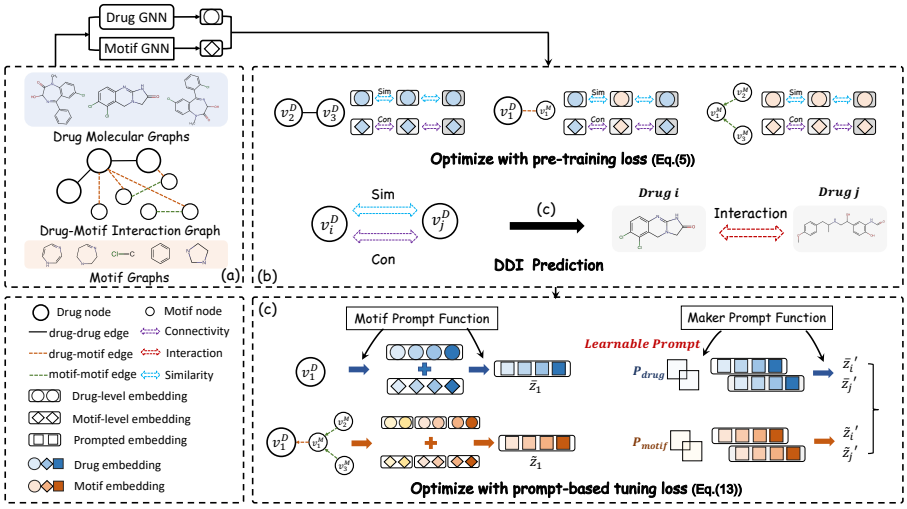


Fig. 1. Illustration of our proposed framework. (a): The inputs; (b): Pre-training and Prediction ; (c): Dual-level prompting process.

graph $\mathcal{D}i$ and the corresponding drug-motif interaction graph $\mathcal{G}$, we apply a motif prompt function f^m_{prompt} to reorganize each drug by aligning it with its relevant motifs:

$$\mathcal{M}^{(i)} = f^m_{prompt}(\mathcal{D}_i, \mathcal{G}), \tag{1}$$

where $\mathcal{M}^{(i)}$ denotes the motif prompt for drug i, redefining the drug to emphasize crucial motifs that influence its interaction dependencies.

Secondly, acknowledging the unique importance of drugs and associated motifs, we utilize a marker prompt function f^a_{prompt} that assigns different markers to various components, thereby delineating separate latent representation spaces for drugs and motifs:

$$\mathcal{D}^{'}_{ij} = f^a_{prompt}(\mathcal{D}_i, \mathcal{D}_j), \mathcal{M}^{(ij)'} = f^a_{prompt}(\mathcal{M}^{(i)}, \mathcal{M}^{(j)}), \tag{2}$$

where $\mathcal{D}^{'}_{ij}$ and $\mathcal{M}^{(ij)'}$ represent the marked prompted input drug pair and corresponding motifs, respectively. Our prompt strategy can be described by the following steps:

- **Hierarchical Pre-Training.** As introduced in Section 4.2, we employ hierarchical self-supervised tasks to train our backbone GNNs in the pre-training stage, acquiring drug and motif representations enriched with structural and relational information.
- **Prompt Addition.** We reformulate each drug pair into the prompted embedding by the dual-level prompting process as represented in Eq.(1-2) and Section 4.3.
- **Prompt-based Tuning.** In the fine-tuning stage, we replace the original pairs with specifically prompted ones for prediction. To enhance the integration of drugs and functional motifs, we predict the label from each prompted

pair separately and subsequently merge these predictions to obtain the final answer, as detailed in Section 4.3.

4.2 Hierarchical Pre-Training Phase

The hierarchical self-supervise tasks in the pre-training phase aim to capture hidden DDI patterns from a multi-level structural perspective. For drug-level perspective, we obtain drug-view representations from molecular graphs and employ a global similarity learning task to enhance learning. For motif-level perspective, we obtain motif-view representations from the drug-motif graph, utilizing a masked edge prediction task to explore local connectivity. Besides, we include a contrastive learning module for multi-level representation alignment.

Drug-level Global Similarity Learning At the drug-level, we represent each drug or motif instance as a molecular graph. For each graph pair $\{\widetilde{\mathcal{V}}_i, \widetilde{\mathcal{V}}_j\}$, where $\widetilde{\mathcal{V}}$ can be $\mathcal{D}$ or $\mathcal{M}$, we employ a drug GNN encoder g with parameters Φ_d to generate drug-view representations $\{h_i^D, h_j^D\}$. We employ a global similarity learning task guiding the encoder to learn pairwise similarity of representations, using their inherent chemical fingerprints as targets, which can be performed by:

$$\mathcal{L}_{drug} = \sum l_{drug}(\mathcal{F}^{sim}_{g_{(\Phi_d)}}(h_i^D, h_j^D)|y_{sim}), \tag{3}$$

where $\mathcal{F}^{sim}(\cdot)$ is the similarity knowledge learning model, which can use cosine similarity or a multi-layer perception. l_{drug} is the loss function like the mean squared error loss, and y_{sim} is the target global similarity intrinsic to the drugs.

Motif-level Local Connectivity Learning At the motif-level, we represent each drug or motif instance as a node in the drug-motif interaction graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. We randomly mask some edges of the positive edge set as $\mathcal{E}_p$ and generate the negative set as $\mathcal{E}_n$ through negative sampling strategy. For each node pair $\{v_i, v_j\}$ within $\mathcal{G}$, we obtain motif-view representations $\{h_i^M, h_j^M\}$ using a motif GNN encoder with parameters Φ_m. We introduce a masked edge prediction task to learn pairwise connective relationships between sampled pairs as:

$$\mathcal{L}_{motif} = \sum_{(v_i,v_j)\in\widetilde{\mathcal{E}}} l_{motif}(\mathcal{F}^{con}_{g_{(\Phi_m)}}(h_i^M, h_j^M)|y_{con}), \tag{4}$$

where $\widetilde{\mathcal{E}} = \mathcal{E}_p \cup \mathcal{E}_n$ represents the sampled edge set. $\mathcal{F}^{con}(\cdot)$ denotes the connectivity knowledge learning model, such as a multi-layer perceptron (MLP). l_{motif} is the loss function like binary cross-entropy loss, and y_{con} identifies the presence or absence of connective relationships between the node pairs.

Overall Pre-training Objective The motif-level connectivity and drug-level global similarity learning enrich the extraction of structural and relational insights from different perspectives. With the guidance of an extra contrastive

learning module, representations from multi-scale channels mutually supervise each other, ensuring a balanced hierarchical model training process. We integrate the self-supervised pre-training tasks with the contrastive learning component to formulate the following overall objective:

$$\begin{aligned}\mathcal{L}_{pre} = &\sum l_{drug}(\mathcal{F}^{sim}_{g(\Phi_d)}(h_i^D, h_j^D)|y_{sim}) + \\ &\sum l_{motif}(\mathcal{F}^{con}_{g(\Phi_m)}(h_i^M, h_j^M)|y_{con}) + \mathcal{L}_{cl}\end{aligned} \tag{5}$$

where $\mathcal{L}_{cl}$ denotes the contrastive learning loss (details in Appendix A.2).

4.3 Prompting for DDI Prediction

In this section, we introduce a dual-level prompting method in the fine-tuning process, to improve the performance of the downstream prediction. First, we explain that there is a discrepancy between the optimizing objectives in the pre-training stage and the fine-tuning process that targets DDI prediction. Then, we propose to prompt the pre-trained model from both motif- and drug-level perspectives, to narrow the gap between the two stages and more effectively extract the knowledge acquired by pre-training for final DDI prediction.

We elucidate the gap in optimizing objectives between the pre-training and fine-tuning stages from two aspects. On one hand, in our hierarchical pre-training phase, we delve into the knowledge of pairwise relationships among drug-drug, motif-motif, and drug-motif pairs. However, the fine-tuning phase focuses on predicting interactions specifically between the input drug-drug pairs, requiring a shift to distinct relational entities.

On the other hand, although we investigate the same drug-drug objects in motif-level learning tasks, the learned drug pairs share high-similarity representing different natures of relations compared to the input pairs in DDI prediction. Taking the dataset DeepDDI [17] as an example, only 0.97% of input pairs that have a DDI relation exhibit a chemical fingerprint similarity greater than 70%. Moreover, over 25% of drug-drug edge pairs share more than 75% of motifs in the constructed graph, while for the pairs with a DDI relation in DeepDDI, this percentage is merely 3.5%. Based on these statistics, it can be inferred that pairs with such high-similarity in pre-training display a tendency to share analogous meaningful functional groups. However, in the downstream DDI prediction task, where the emphasis lies on interaction relationships between input drug pairs, it becomes evident that these pairs involve a more diverse array of functional groups. The parameter space initialized by the pre-trained model does not entirely align with the final task, potentially leading to negative transfer.

To tackle this, we introduce a novel graph-based prompting learning mechanism outlined in Definition 4, tailored for our drug and motif perspectives. This mechanism effectively connects the pre-training objective with the fine-tuning task by employing a uniform task format, leveraging the relevant knowledge acquired during pre-training to enhance the final DDI prediction.

Prompting Function Design Aligned with our hierarchical pre-training phase, we propose a dual-level prompt function designed to activate and transform distinct knowledge from diverse perspectives.

In the motif-level perspective, a drug molecule is an entity composed of various components known as functional motifs, which significantly influence its properties, therapeutic effects, and overall interactions [4]. To exploit the relational knowledge between different nodes and identify key motifs for each drug, we propose a motif prompt function to leverage pre-training insights to capture deep associations. This function extracts connected motif embeddings of each drug via the drug-motif graph, reformulating the drug into significantly associated motifs in the representation space. It captures crucial motifs internal to the drug that influence various properties, preserving vital DDI knowledge. For each drug node v_i^D in the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the motif prompt function f_p^m work as:

$$\begin{aligned} \widetilde{z}_i &= f_p^m(h|h_i^D + h_i^M,\ v_i^D \in \mathcal{V}), \\ \overline{z}_i &= f_p^m(h|h_j^D + h_j^M,\ for\ v_j^M \in \mathcal{N}(v_i^D)), \end{aligned} \tag{6}$$

where $\widetilde{z}$ represents the prompted embedding of each drug, and $\overline{z}$ denotes the prompted embedding of significantly associated motifs, encompassing all its motif neighbors. The embedding h merges hierarchical representations, h^D and h^M, introduced in Section 4.2. After hierarchical pre-training, the obtained prompted drug embeddings aggregate information from multi-levels. Moreover, the prompted motif embeddings gather information from drug-motif and motif-motif pairs, serving as an alternative perspective of drug representations. They integrate knowledge acquired during pre-training from diverse pair objectives, providing essential functional motif insights into the drug, emphasizing target drug-drug relationships and effectively addressing the first identified gap.

In the drug-level perspective, the previously prompted drug and motif embeddings play distinct roles. From an entity view, motifs and drugs share similar learning processes but are distinct entities. In the fine-tuning stage for DDI prediction, motif and drug instances serve different purposes. Drug entities provide a comprehensive view for modeling relationships between drugs, while motif entities focus on specific structural characteristics, considering DDI problems in a more granular manner. From a relation view, prompted motif embeddings aggregate information from drug-motif and motif-motif pairs, emphasizing different aspects than drug embeddings. Although the focus shifts to drug-drug pairs, the acquired knowledge from the pre-training phase centers on topological and relational information, which needs to be effectively activated and transformed to align with the target interaction relationships in the downstream task.

To effectively activate and distinguish relevant knowledge learned during pre-training for the final DDI target, we introduce a marker prompt function tailored for motif and drug embeddings. To begin, we introduce learnable virtual tokens $[P_d, P_m]$ with hidden dimension d to denote the embedding space of drugs and motifs respectively. Then, we reparameterize the prompt tokens with reparameterization functions Θ which can be dense layers, such as MLP, acting as:

$$P'_d = \Theta_d(P_d), P'_m = \Theta_m(P_m), \tag{7}$$

Then, we present the prompt function as f_p^a, which adds different maker prompt tokens for different embeddings as follows, where $\oplus$ denotes concatenate:

$$f_p^a(\widetilde{z}) = [P'_d \oplus \widetilde{z}], f_p^a(\overline{z}) = [P'_m \oplus \overline{z}], \tag{8}$$

For a input drug pair i and j, we present the whole prompting process as:

$$(\widetilde{z}'_i, \widetilde{z}'_j) = f_p^a(\widetilde{z}_i, \widetilde{z}_j); (\overline{z}'_i, \overline{z}'_j) = f_p^a(\overline{z}_i, \overline{z}_j), \tag{9}$$

where $\widetilde{z}'$ and $\overline{z}'$ denote the final prompted embeddings for input drug pairs and associated motif pairs respectively. The function distinguishes the diverse knowledge about drugs and motifs gained from pre-training, guiding their distinct roles downstream. Furthermore, it facilitates the transfer of structure and correlation relationships acquired from pre-training into the downstream task domain, specifically targeting the interaction relationships among drug pairs. This effectively mitigates the second identified gap.

In summary, our dual-level prompt function considers and integrates pre-training and fine-tuning stages from different perspectives, bridging the existing gap and transferring prior knowledge.

Prompt-based Tuning To enhance the integration of drugs and their functional motifs, we leverage distinct prompted embeddings for improved prediction, considering their respective importance. Firstly, we separately make the predictions with a classifier p as:

$$\widetilde{y}_{pred} = p(\widetilde{z}'_i, \widetilde{z}'_j, r), \overline{y}_{pred} = p(\overline{z}'_i, \overline{z}'_j, r), \tag{10}$$

where $\widetilde{y}_{pred}, \overline{y}_{pred}$ represent the predicted probabilities respectively, r denotes the interaction embedding, p can be a linear projection head.

Secondly, we adaptively learn the weights of information from different scales, jointly influencing the final predictions. Formally, we employ a linear transformation followed by the Sigmoid function to calculate the importance scores:

$$\eta = Sigmoid(W_s(\widetilde{z}', \overline{z}')), \tag{11}$$

Finally, we allocate distinct importance to different components and integrate them to form the final prediction results as:

$$y_{pred} = \eta\ \widetilde{y}_{pred} + (1 - \eta)\ \overline{y}_{pred}, \tag{12}$$

Our prediction objective is to maximize the log-likelihood of correct output given our prompted embeddings:

$$\begin{aligned}
\mathcal{L}_{ddi} &= \sum_{(i,j,r)\in\mathcal{T}} l(y_{pred} \mid y_{true}) = \sum l(\mathcal{F}^{\mathcal{DDI}}(\mathcal{D}_i, \mathcal{D}_j, r) \mid y_{true}), \\
&= \sum l(p((\widetilde{z}'_i, \widetilde{z}'_j), (\overline{z}'_i, \overline{z}'_j)) \mid (r, y_{true})), \\
&= \sum l(p((f_p(h_i^M) + f_p(h_i^D)), (f_p(h_j^M) + f_p(h_j^D)))| \ y), \\
&= \sum l(p((f_p(h_i^M), f_p(h_j^M)) + (f_p(h_i^D), f_p(h_j^D)))| \ y), \\
&= \sum l(p(f_p(h_i^M, h_j^M))| \ y) + l(p(f_p(h_i^D, h_j^D))| \ y).
\end{aligned} \tag{13}$$

where l is the cross-entropy loss function, y denotes the true label indicating the interaction possibility. With our dual-level prompting mechanism, we effectively bridge the gap between pre-training and fine-tuning stages by reformulating the final DDI task (Eq.(13)) to the same format as pre-training task (Eq.(5)).

Table 1. Dataset Statistics

Datasets	Drugs	DDIs	Types
DeepDDI	1,704	191,511	86
DrugBankDDI	3,643	1,151,039	174
Twosides	645	4,576,287	963

5 Experiments

We evaluate our model on various real-world DDI datasets to confirm its scalability and robustness. The code of HS-GPF is available at `https://github.com/PaddlePaddle/PaddleHelix/tree/dev/apps/drug_drug_interaction/HSGPF`.

5.1 Experimental Settings

Datasets. Our experiments cover three datasets of different scales: the small-scale DeepDDI [17], the medium-scale collected DrugBankDDI and the large-scale Twosides [33]. The dataset statistics are summarized in Table 1 and follows.

- **DeepDDI**: We adopt the dataset introduced by [17] as our small-scale dataset, which includes 1,704 drugs and 191,511 DDI samples categorized into 86 unique interaction types, each defined by a standardized sentence.
- **DrugBankDDI**: We construct the medium-scale DrugBankDDI dataset from DrugBank (V5.1.9) [26], including 3,643 drugs and 1,151,039 DDI samples categorized into 174 unique interaction types through our process. Following [17], we categorize reaction types into different classes using standardized sentence structures and omit categories with fewer than 5 samples to ensure significant data.
- **Twosides**: We use this dataset released by [33] after filtering and preprocessing the original TWOSIDES dataset [20] as the large-scale dataset. This dataset contains 645 drugs with 963 interaction types and 4,576,287 DDI tuples. Unlike the first two datasets, this dataset focuses on interactions at the phenotypic level rather than the metabolic level to ensure data diversity.

Baselines. We assess our framework against state-of-the-art methods in three main categories as follows: (1) *Supervised methods*: GCN [8], GAT [21], GIN [27], SSP-MLP [17], SSI-DDI [15], MIRACLE [23], HM-GNN [29]. (2) *Graph pre-training methods*: MGSSL [32], MoCL [18], KCL [3]. (3) *Graph prompt methods*: GPPT [19] and GraphPrompt [12].

Table 2. Performance comparison on all datasets. The average and standard deviation (shown in brackets) results are reported across five folds. The best results are shown in bold and the second best results are shown with underlines.

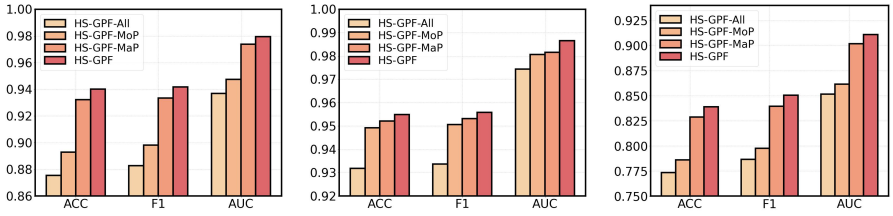
Method		DeepDDI			DrugBankDDI			Twosides		
		ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
Supervised Methods	GCN	0.782(.003)	0.800(.002)	0.862(.002)	0.769(.001)	0.792(.001)	0.862(.001)	0.735(.003)	0.742(.004)	0.807(.004)
	GAT	0.811(.002)	0.819(.002)	0.883(.002)	0.802(.002)	0.813(.002)	0.887(.001)	0.752(.005)	0.768(.005)	0.829(.005)
	GIN	0.840(.001)	0.848(.002)	0.909(.001)	0.866(.002)	0.872(.001)	0.939(.001)	0.762(.002)	0.784(.003)	0.842(.003)
	SSP-MLP	0.802(.013)	0.811(.009)	0.885(.009)	0.876(.006)	0.880(.005)	0.939(.004)	0.727(.076)	0.742(.035)	0.804(.102)
	SSI-DDI	0.844(.007)	0.851(.005)	0.920(.006)	0.882(.003)	0.885(.003)	0.945(.002)	0.781(.008)	0.796(.004)	0.854(.007)
	MIRACLE	0.859(.009)	0.885(.013)	0.930(.006)	0.923(.002)	0.926(.001)	0.968(.001)	0.800(.004)	0.816(.003)	0.874(.004)
	HM-GNN	0.872(.002)	0.878(.002)	0.934(.001)	0.927(.000)	0.930(.000)	0.966(.000)	0.789(.003)	0.801(.003)	0.863(.003)
Graph pre-training Methods	MGSSL	0.859(.001)	0.865(.001)	0.921(.002)	0.934(.000)	0.936(.000)	0.967(.000)	0.783(.007)	0.799(.003)	0.858(.004)
	MoCL	0.856(.002)	0.862(.003)	0.919(.002)	0.923(.001)	0.926(.001)	0.963(.000)	0.785(.003)	0.800(.002)	0.859(.003)
	KCL	0.878(.005)	0.880(.004)	0.941(.004)	0.866(.002)	0.870(.002)	0.930(.001)	0.785(.007)	0.787(.006)	0.860(.007)
Graph prompt Methods	GPPT	0.888(.006)	0.892(.006)	0.947(.004)	0.904(.004)	0.908(.003)	0.963(.001)	0.816(.009)	0.822(.009)	0.885(.012)
	GraphPrompt	0.894(.002)	0.900(.002)	0.948(.001)	0.932(.001)	0.934(.001)	0.974(.001)	0.790(.003)	0.799(.002)	0.864(.003)
	Ours	**0.941(.001)**	**0.942(.001)**	**0.980(.001)**	**0.955(.001)**	**0.956(.001)**	**0.987(.000)**	**0.839(.002)**	**0.851(.002)**	**0.911(.002)**

Experimental Configurations. In dataset preparation, we employed a stratified split based on DDI tuples, maintaining consistent proportions (60% training, 20% validation, 20% test) following[17]. This process are repeated five times with different random seeds to create five stratified randomized folds. We generated a negative sample for each DDI pair following[24] before training to maintain consistency. We reported the mean and standard deviation of the results across these folds, using *Acc, F1-score and AUC* as evaluation metrics. For the proposed HS-GPF, we employ 3-layer GINs with 64 hidden dimensions as our graph encoders. For pretraining, we set the running epoch as 300. For the graph prompting strategy, we set the reparameterization function layer to 1 for all datasets, with hidden dim as 16, 16, 32 for each dataset respectively. The data features are provided in Appendix B.

5.2 Performance Comparison

The performance comparison is presented in Table 2, and we have the following observations.

- Among supervised methods, MIRACLE and HM-GNN achieve better performance, underscoring the importance of capturing drug connections.
- For graph pre-training methods, MGSSL and MoCL outperform KCL on the DrugBankDDI dataset, while KCL excels on the DeepDDI dataset. This may be attributed to MGSSL and MoCL focusing on crucial substructures in self-supervised learning, highlighting the importance of learning the compositional information of motifs for molecular representation, especially beneficial for datasets with a large number of molecules. However, their performance on the Twosides dataset is less impressive, possibly due to the pretraining tasks' emphasis on the internal structural properties of molecules, causing notable negative transfer in downstream DDI prediction task.



(a) DeepDDI (b) DrugBankDDI (c) Twosides

Fig. 2. Ablation study for graph prompting.

– In some cases, supervised methods can achieve performance comparable or superior to graph pre-training methods, further highlighting the potential gap between pretraining and downstream tasks that can hinder effective knowledge transfer.
– For graph prompt methods, GPPT and GraphPrompt, demonstrate competitive performance on the DeepDDI and Twosides datasets. However, the performance on the DrugBankDDI dataset is only moderately effective, potentially due to a bias towards motif advantages when processing large-scale molecular data, as highlighted by the previously mentioned baselines.

Our proposed HS-GPF consistently outperforms the baselines across all datasets, thanks to its following advantages:

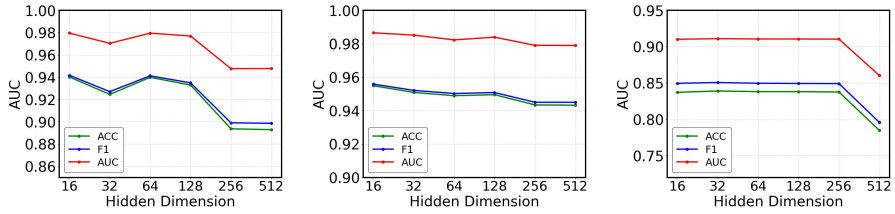
– We employ a dual-level prompt strategy for multi-level self-supervised tasks and DDI prediction, highlighting its effectiveness in bridging the gap between pre-training and fine-tuning.
– Our strategy integrates crucial information from both drugs and motifs, optimizing the incorporation of prompted embeddings for enhanced predictions.
– By prioritizing the connections among drugs and motifs, our strategy seamlessly unifies the learning of relationships between pre-training and downstream molecules, facilitating the transfer of latent interaction knowledge. This provides a relative advantage over existing methods and demonstrates improved robustness.

5.3 Ablation Study

We conduct ablation studies on all datasets to further validate the contributions of different designs in our framework, comparing our HS-GPF with its variants:

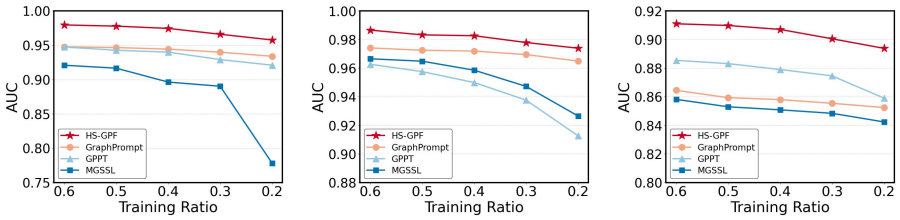
– **HS-GPF-MoP** removes the motif prompting strategy.
– **HS-GPF-MaP** removes the marker prompting strategy.
– **HS-GPF-All** drops pre-training and prompting process.

As observed from Figure 2, each component of our HS-GPF plays an indispensable role in DDI prediction. The specially designed dual-level prompt strategy tailored for multi-level self-supervised tasks can effectively leverage the advantages of pre-training knowledge, substantially assisting in prediction.



(a) DeepDDI (b) DrugBank (c) Twosides

Fig. 3. Parameter sensitivity analysis.



(a) DeepDDI (b) DrugBank (c) Twosides

Fig. 4. Sensitivity study to training sample size.

5.4 Sensitivity Analysis

We conduct experiments to analyze the robustness of our HS-GPF.

Effects of hyper-parameter variation. We assess the impact of critical hyper-parameters on our model's performance, focusing on the hidden dimension of our designed prompt tokens (Eq.(7)), which we vary from 16 to 512 across all datasets. The results in Figure 3 show that performance tends to decrease with increasing values, and excessively large values may cause a slight drop. This may result from the potential introduction of more noise to the embeddings with an overly complex prompt. Appropriate prompt learning can assist the model in aligning tasks between pre-training and fine-tuning, enhancing prediction.

Effects of training dataset size. We evaluated the robustness of our proposed method by analyzing its sensitivity to variations in the number of labeled training data. In our overall performance evaluation, we employed 5-fold cross-validation with a 60% training ratio. Here we extended our analysis to include comparative analyses with training ratios ranging from 20% to 60%, while consistently reserving 20% of the dataset as the test set for performance evaluation. We compared our method against leading DDI prediction baselines: MGSSL [32], GPPT[19] and GraphPrompt[12]. As shown in Figure 4, our HS-GPF consistently outperforms these baselines across various training ratios. In contrast, the compared baselines exhibit limitations to training data variations. Notably, the pre-training-based method MGSSL is more sensitive to data size changes than the prompt-based methods GPPT and GraphPrompt. As the training ratio decreases, our prompt mechanism minimizes the performance decline with limited

Table 3. Cold-start task performance comparison on all datasets. The average and standard deviation (shown in brackets) results are reported across five folds. The best results are shown in bold and the second best results are shown with underlines.

Methods		DeepDDI			DrugBankDDI			Twosides		
		Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC
Task A	GCN	0.735(.021)	0.726(.040)	0.809(.014)	0.777(.007)	0.753(.012)	0.850(.004)	0.646(.021)	0.603(.034)	0.702(.022)
	GAT	0.727(.005)	0.719(.010)	0.803(.006)	0.782(.006)	0.759(.015)	0.849(.004)	0.645(.015)	0.598(.028)	0.701(.018)
	GIN	0.754(.009)	0.734(.017)	0.835(.006)	0.785(.007)	0.758(.011)	0.856(.004)	0.644(.015)	0.586(.034)	0.716(.012)
	SSI-DDI	0.755(.008)	0.738(.014)	0.837(.007)	0.753(.004)	0.748(.005)	0.834(.003)	0.623(.013)	0.550(.032)	0.680(.012)
	MGSSL	0.746(.010)	0.725(.012)	0.816(.012)	0.788(.005)	0.761(.007)	0.851(.002)	0.647(.019)	0.603(.032)	0.706(.020)
	MoCL	0.761(.006)	0.738(.009)	0.839(.005)	0.791(.006)	0.763(.010)	0.862(.003)	0.651(.018)	0.604(.034)	0.721(.017)
	KCL	0.748(.009)	0.723(.011)	0.828(.009)	0.774(.004)	0.751(.007)	0.843(.006)	0.636(.015)	0.599(.034)	0.683(.013)
	HM-GNN	0.667(.071)	0.574(.146)	0.757(.063)	0.712(.052)	0.633(.101)	0.801(.039)	0.635(.030)	0.586(.047)	0.698(.034)
	GPPT	0.751(.016)	0.726(.026)	0.838(.013)	0.756(.005)	0.714(.009)	0.844(.005)	0.602(.048)	0.516(.100)	0.656(.052)
	GraphPrompt	0.764(.006)	0.743(.008)	0.848(.006)	0.785(.005)	0.751(.007)	0.862(.001)	0.648(.016)	0.604(.023)	0.715(.011)
	Ours	**0.794(.011)**	**0.772(.015)**	**0.871(.006)**	**0.817(.005)**	**0.796(.008)**	**0.889(.003)**	**0.658(.015)**	**0.622(.026)**	**0.727(.014)**
Task B	GCN	0.654(.017)	0.581(.029)	0.736(.020)	0.670(.011)	0.580(.024)	0.745(.008)	0.521(.018)	0.368(.043)	0.550(.038)
	GAT	0.654(.012)	0.583(.027)	0.737(.020)	0.672(.006)	0.585(.011)	0.743(.008)	0.522(.020)	0.375(.051)	0.551(.035)
	GIN	0.646(.017)	0.565(.037)	0.740(.014)	0.660(.007)	0.556(.012)	0.750(.004)	0.517(.025)	0.327(.086)	0.580(.028)
	SSI-DDI	0.643(.056)	0.575(.052)	0.713(.074)	0.674(.005)	0.585(.015)	0.756(.010)	0.504(.030)	0.291(.090)	0.531(.030)
	MGSSL	0.643(.020)	0.565(.032)	0.714(.026)	0.656(.007)	0.547(.017)	0.734(.004)	0.522(.027)	0.362(.070)	0.564(.031)
	MoCL	0.649(.012)	0.563(.021)	0.741(.010)	0.664(.010)	0.561(.022)	0.757(.008)	0.524(.028)	0.356(.077)	0.582(.033)
	KCL	0.663(.014)	0.591(.023)	0.740(.017)	0.675(.004)	0.598(.008)	0.746(.008)	0.520(.011)	0.378(.036)	0.560(.008)
	HM-GNN	0.600(.038)	0.454(.104)	0.683(.045)	0.613(.023)	0.447(.059)	0.705(.026)	0.524(.031)	0.376(.078)	0.574(.037)
	GPPT	0.647(.013)	0.560(.031)	0.744(.012)	0.623(.007)	0.477(.020)	0.714(.011)	0.522(.021)	0.370(.048)	0.573(.020)
	GraphPrompt	0.646(.010)	0.557(.018)	0.752(.011)	0.639(.005)	0.500(.011)	0.749(.003)	0.524(.038)	0.367(.089)	0.577(.031)
	Ours	**0.677(.017)**	**0.610(.027)**	**0.773(.019)**	**0.687(.006)**	**0.600(.015)**	**0.788(.006)**	**0.537(.024)**	**0.385(.044)**	**0.583(.029)**

labeled data, showcasing remarkable robustness. This underscores the benefits of prompt-based strategies in leveraging pre-trained knowledge for effective knowledge acquisition and transfer in the DDI domain under limited data availability.

5.5 Cold-start Task Analysis

To evaluate our model's performance when dealing with new drugs (i.e., cold-start problem), following [13], we split datasets based on drugs instead of DDI tuples and create two cold-start tasks under 5-fold cross-validation settings. These scenarios offer a more realistic and challenging evaluation scheme for the models. In *Task A*, the model is trained on DDI samples from training drugs, with predictions made for DDIs between training and testing drugs. For *Task B*, the only difference is that we predict DDIs between testing drugs. We compare HS-GPF exclusively with baselines applicable to cold-start scenarios.

It can be observed from Table 3 that our method maintains superior performance in both tasks, demonstrating its robustness in cold-start scenarios. We further have the following observations.

– Among GNN-based methods in supervised methods, GIN excels in warm-start scenarios but struggles in most cold-starts comparisons. *Task A* aligns more closely with warm-start scenarios than *Task B*, favoring stronger models for predicting possible interactions between new and known drugs. However, in *Task B*, the trend seems to be the opposite. This reversal may stem

from the trend of powerful models to overfit drug features in the training set, leading to a relatively inadequate generalization ability to new drugs.
– Graph pre-training methods and prompting methods appear to be the most competitive. MoCL and KCL perform well in most scenarios, likely due to the effectiveness of these enhancement techniques in learning crucial structural information of drug molecules, thereby enhancing generalization ability. In the *task B* scenario, KCL shows a slight advantage over MoCL, which may be due to the challenge of finding commonalities in substructures for drugs with substantial structural differences, where capturing elemental characteristics may be more versatile. GraphPrompt continues to demonstrate competitive performance, while GPPT does not perform well. This might be attributed to the cluster-based task token generation of GPPT, which may significantly impact its performance when dealing with new drugs due to substantial structural differences between new and known drugs.
– Our HS-GPF takes into consideration the complex drug interactive dependencies, offering robust support for research on DDI prediction, even in challenging cold-start scenarios. Through our innovative hierarchical self-supervised learning and prompting strategy, which leverages crucial DDI pattern knowledge, we not only enhance generalization in cold-start situations but also achieve a deeper comprehension of drug interactions.

6 Conclusion

In this paper, we introduce HS-GPF, a novel hierarchical structure-aware graph prompting framework for DDI prediction. The framework employs hierarchical self-supervised tasks capturing diverse drug relations. To enhance predictive capability and bridge the gap between pre-training and fine-tuning tasks, we propose a dual-level prompt strategy for effective knowledge transfer. Extensive experiments on various datasets demonstrated that HS-GPF outperforms state-of-the-art methods, including the challenging cold-start scenarios. In future work, we would like to explore how to apply this framework to other molecular graph applications.

Acknowledgment

This work was supported in part by Science and Technology Planning Project of Guangdong Province (Grant No. 2023A0505050111), Guangzhou-HKUST(GZ) Joint Funding ProgramGrant No.2023A03J0008Education Bureau of Guangzhou Municipality.

References

1. Chen, X., Liu, X., Wu, J.: Gcn-bmp: investigating graph representation learning for ddi prediction task. Methods **179**, 47–54 (2020)

2. Deng, Y., Xu, X., Qiu, Y., Xia, J., Zhang, W., Liu, S.: A multimodal deep learning framework for predicting drug–drug interaction events. Bioinformatics **36**(15), 4316–4322 (2020)
3. Fang, Y., Zhang, Q., Yang, H., Zhuang, X., Deng, S., Zhang, W., Qin, M., Chen, Z., Fan, X., Chen, H.: Molecular contrastive learning with chemical element knowledge graph. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 3968–3976 (2022)
4. Harrold, M.W., Zavod, R.M.: Basic concepts in medicinal chemistry (2014)
5. Hu, W., Liu, B., Gomes, J., Zitnik, M., Liang, P., Pande, V., Leskovec, J.: Strategies for pre-training graph neural networks. arXiv preprint arXiv:1905.12265 (2019)
6. Huang, K., Xiao, C., Hoang, T., Glass, L., Sun, J.: Caster: Predicting drug interactions with chemical substructure representation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 702–709 (2020)
7. Jia, J., Zhu, F., Ma, X., Cao, Z.W., Li, Y.X., Chen, Y.Z.: Mechanisms of drug combinations: interaction and network perspectives. Nature reviews Drug discovery **8**(2), 111–128 (2009)
8. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
9. Li, S., Zhou, J., Xu, T., Dou, D., Xiong, H.: Geomgcl: Geometric graph contrastive learning for molecular property prediction. In: Proceedings of the AAAI conference on artificial intelligence. vol. 36, pp. 4541–4549 (2022)
10. Lin, X., Quan, Z., Wang, Z.J., Ma, T., Zeng, X.: Kgnn: Knowledge graph neural network for drug-drug interaction prediction. In: IJCAI. vol. 380, pp. 2739–2745 (2020)
11. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G.: Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys **55**(9), 1–35 (2023)
12. Liu, Z., Yu, X., Fang, Y., Zhang, X.: Graphprompt: Unifying pre-training and downstream tasks for graph neural networks. In: Proceedings of the ACM Web Conference 2023. pp. 417–428 (2023)
13. Lyu, T., Gao, J., Tian, L., Li, Z., Zhang, P., Zhang, J.: Mdnn: A multimodal deep neural network for predicting drug-drug interaction events. In: IJCAI. pp. 3536–3542 (2021)
14. Nyamabo, A.K., Yu, H., Liu, Z., Shi, J.Y.: Drug–drug interaction prediction with learnable size-adaptive molecular substructures. Briefings in Bioinformatics **23**(1), bbab441 (2022)
15. Nyamabo, A.K., Yu, H., Shi, J.Y.: Ssi–ddi: substructure–substructure interactions for drug–drug interaction prediction. Briefings in Bioinformatics **22**(6), bbab133 (2021)
16. Rong, Y., Bian, Y., Xu, T., Xie, W., Wei, Y., Huang, W., Huang, J.: Self-supervised graph transformer on large-scale molecular data. Advances in Neural Information Processing Systems **33**, 12559–12571 (2020)
17. Ryu, J.Y., Kim, H.U., Lee, S.Y.: Deep learning improves prediction of drug–drug and drug–food interactions. Proceedings of the National Academy of Sciences **115**(18), E4304–E4311 (2018)
18. Sun, M., Xing, J., Wang, H., Chen, B., Zhou, J.: Mocl: Data-driven molecular fingerprint via knowledge-aware contrastive learning from molecular graph. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 3585–3594 (2021)

19. Sun, M., Zhou, K., He, X., Wang, Y., Wang, X.: Gppt: Graph pre-training and prompt tuning to generalize graph neural networks. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 1717–1727 (2022)
20. Tatonetti, N.P., Ye, P.P., Daneshjou, R., Altman, R.B.: Data-driven prediction of drug effects and interactions. Science translational medicine **4**(125), 125ra31–125ra31 (2012)
21. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al.: Graph attention networks. stat **1050**(20), 10–48550 (2017)
22. Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., Friedman, C.: Drug—drug interaction through molecular structure similarity analysis. Journal of the American Medical Informatics Association **19**(6), 1066–1074 (2012)
23. Wang, Y., Min, Y., Chen, X., Wu, J.: Multi-view graph contrastive representation learning for drug-drug interaction prediction. In: Proceedings of the Web Conference 2021. pp. 2921–2933 (2021)
24. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the AAAI conference on artificial intelligence. vol. 28 (2014)
25. Whitebread, S., Hamon, J., Bojanic, D., Urban, L.: Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. Drug discovery today **10**(21), 1421–1433 (2005)
26. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al.: Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic acids research **46**(D1), D1074–D1082 (2018)
27. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? arXiv preprint arXiv:1810.00826 (2018)
28. Xu, N., Wang, P., Chen, L., Tao, J., Zhao, J.: Mr-gnn: Multi-resolution and dual graph neural network for predicting structured entity interactions. arXiv preprint arXiv:1905.09558 (2019)
29. Yu, Z., Gao, H.: Molecular representation learning via heterogeneous motif graph neural networks. In: International Conference on Machine Learning. pp. 25581–25594. PMLR (2022)
30. Zhang, P., Wang, F., Hu, J., Sorrentino, R.: Label propagation prediction of drug-drug interactions based on clinical side effects. Scientific reports **5**(1), 12339 (2015)
31. Zhang, W., Chen, Y., Liu, F., Luo, F., Tian, G., Li, X.: Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. BMC bioinformatics **18**, 1–12 (2017)
32. Zhang, Z., Liu, Q., Wang, H., Lu, C., Lee, C.K.: Motif-based graph self-supervised learning for molecular property prediction. Advances in Neural Information Processing Systems **34**, 15870–15882 (2021)
33. Zitnik, M., Agrawal, M., Leskovec, J.: Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics **34**(13), i457–i466 (2018)