Motif-Oriented Representation Learning with Topology Refinement for Drug-Drug Interaction Prediction

Ran Zhang[1,2], Xuezhi Wang[1,2], Guannan Liu[3], Pengyang Wang[4], Yuanchun Zhou[1,2], Pengfei Wang[1,2*]

[1]Computer Network Information Center, Chinese Academy of Science
[2]University of Chinese Academy of Sciences
[3]Beihang University
[4]University of Macau
{zhangran, wxz}@cnic.cn, liugn@buaa.edu.cn, pywang@um.edu.mo, {zyc, pfwang}@cnic.cn

Abstract

Drug-Drug Interaction (DDI) prediction has attracted considerable attention in designing multi-drug combination strategies and avoiding adverse reactions. Notably, Artificial Intelligence (AI)-driven DDI prediction methods have emerged as a pivotal research paradigm. However, most AI-driven DDI prediction methods fall short in exploring intra-molecular motifs, and heavily rely on the overly idealized assumption of the complete inter-molecular topology, limiting their expressive capacities. To this end, we propose a **M**otif-**O**riented representation learning with **TO**pology **R**efinement for DDI prediction, namely **MOTOR**, to exploit both the multi-granularity motif information and the topological structure of DDI networks. Specifically, MOTOR effectively captures motif internal structures, motif local contexts, and motif global semantics. Furthermore, MOTOR employs an iterative learning strategy to continuously refine the DDI topology and optimize the corresponding drug representations. Extensive experimental results demonstrate that MOTOR exhibits superior performance with interpretable insights in DDI prediction tasks across three real-world datasets, thereby opening up new avenues in AI-driven DDI prediction.

Introduction

Recently, Drug-Drug Interaction (DDI) prediction has aroused significant attention in the pharmaceutical and healthcare industries. DDI, the concurrent or sequential administration of multiple drugs, can alter or diminish the effects of one or more medications. In the United States, adverse DDIs result in over 74,000 emergency room visits and more than 195,000 hospitalizations annually (Han et al. 2022a). With a global aging population, the prevalence of multiple chronic conditions in a single individual is expected to increase, leading to a heightened risk of DDIs. Effectively predicting DDIs can help identify potential risks in polypharmacy when managing multiple chronic conditions, thereby promoting the development of reliable and effective combination therapies. Due to the complexity of individual pharmacokinetics (*how the body processes drugs*) and pharmacodynamics (*how drugs affect the body*), tasks such as DDI detection are manually completed by professional pharmacologists. However, the vast body of potential DDIs renders the screening of these interactions via wet lab experiments laborious and time-consuming, spurring an urgent demand for efficient computational DDI prediction techniques.

The advances in Artificial Intelligence (AI) technology and the surge in drug-related data (Zhao et al. 2024) have led AI-driven DDI prediction methods to a prevalent trend (Luo et al. 2024). In the early years, text mining-based methods (Segura-Bedmar, Martínez Fernández, and Herrero Zazo 2013) leverage Natural Language Processing (NLP) techniques to extract relevant information from biomedical literature and predict potential DDIs. Similarity-based methods (Takeda et al. 2017; Zhang et al. 2018; Rohani, Eslahchi, and Katanforoush 2020) assume that if drug i interacts with drug j, then drugs similar to drug i are also likely to interact with drug j. Numerous studies (Vilar et al. 2012; Ferdousi, Safdari, and Omidi 2017) have emerged based on this assumption, aiming to identify DDIs by capturing drug features and measuring similarities between drugs. Recently, as both molecular structures and DDI networks exhibit typical graph structures, graph learning has demonstrated great potential and made impressive advancements in DDI prediction tasks. With the rise of Graph Neural Networks (GNNs), a range of graph-based methods have been successfully applied to identify potential DDIs (Cao, Fan, and Zeng 2020; Wang et al. 2022), which mainly depend on three strategies: intra-molecular approaches (Nyamabo, Yu, and Shi 2021), inter-molecular approaches (Zhao et al. 2021), and hierarchical approaches (Li et al. 2023).

While graph-based DDI prediction methods have achieved encouraging results, previous works still face limitations when tackling the following challenges: ***(1) Limited exploration of intra-molecular motifs:*** Atoms are the fundamental particles of the chemical elements, so many methods model DDI tasks from this atom perspective (Cao, Fan, and Zeng 2020; Wang et al. 2022; Zhang et al. 2023). However, motifs (unique structures or arrangements of atoms, bonds, or functional groups) commonly are structural units in biochemical reactions. While recent approaches have started to incorporate molecular substructures (Nyamabo, Yu, and Shi 2021; Li et al. 2023; Zhang et al. 2024b), they tend to neglect the inter-motif

*Corresponding author.

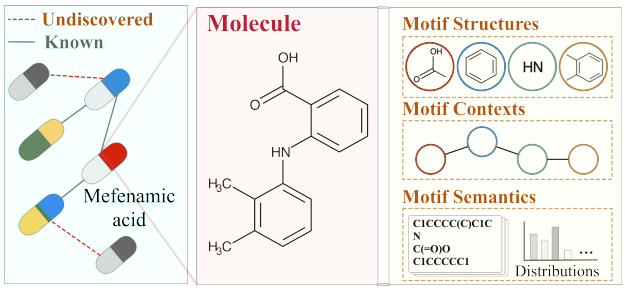


Figure 1: An example of DDIs with motif information.

interactions within molecules or the motif distribution across molecules. The internal structure of a motif reveals how atoms are organized and arranged within it. Since chemical reactions between molecules often involve the breaking and forming of inter-motif bonds, capturing motif local contexts is essential. Additionally, motif global semantics provide significant insights into motif distribution across different molecules, which can reveal potential pharmacological effects. Therefore, a comprehensive analysis of motifs (Zhang et al. 2024a), including their internal structures, local contexts, and global semantics, is crucial for accurate DDI prediction. ***(2) Overly idealized assumption of inter-molecular topology:*** a common issue in most DDI prediction studies is the tendency to overly idealize the observed DDI network as complete and assume it remains consistent with the properties of GNNs. Due to ongoing drug development and continuous experimental validation, the observed DDI dataset represents only a snapshot at a specific time. Furthermore, as the training set only captures a subset of real-world interactions, the assumption could always be violated. Relying on an incomplete and fixed interaction network topology can pose potential risks to the expressive power and prediction performance of GNNs (Franceschi et al. 2019; Pei et al. 2020).

To address these limitations, we aim to comprehensively investigate DDIs by exploring intra-molecular motifs and inter-molecular interactions. Figure 1 illustrates a real-world example. Mefenamic acid, a non-steroidal anti-inflammatory drug, interacts with various drugs, and some DDIs involving mefenamic acid remain undiscovered. From a motif perspective, mefenamic acid contains four interconnected motifs, each with a complex internal structure and a distinct distribution across different molecules. Motif structures indicate the atomic arrangement and combinations within motifs, while motif contexts capture the interactions among multiple motifs within a molecule. Motif semantics, derived from their distribution across molecules, reveal potential functional information. Along this line, we propose a **M**otif-**O**riented representation learning framework with iterative **TO**pology **R**efinement for DDI prediction, **MOTOR**. We employ multi-granularity motif-oriented learning to effectively capture internal structures, local contexts, and global semantics of motifs. Then, we continuously optimize the DDI topology to learn the corresponding drug representations. Furthermore, we facilitate mutual enhancement across motif-oriented learning and topology refinement modules.

Our major contributions can be summarized as follows:

- We propose a motif-oriented molecular representation learning framework to exploit motif information across internal structures, local contexts, and global semantics.
- We adopt an iterative learning strategy to continuously refine the DDI network topology and optimize the corresponding drug representations.
- MOTOR demonstrates state-of-the-art (SOTA) performance in three real-world DDI prediction tasks and shows potentially good interpretability.

Problem Statement

Notation. Given a ***DDI network*** $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{A})$, where $\mathcal{V} = \{v_1, v_2, v_3, ..., v_N\}$ is a set of N molecules and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the interaction set. Each edge $e_{ij} = (v_i, v_j)$ in $\mathcal{E}$ describes the interaction between two molecules v_i and v_j, where $v_i, v_j \in \mathcal{V}$. $\mathbf{X}$ is the attribute features of the DDI network. $\mathbf{A} \in \{0, 1\}^{N \times N}$ is the adjacency matrix of the DDI network, where 0 and 1 in $\mathbf{A}$ represent non-interaction and interaction between two corresponding drugs, respectively. A ***molecule*** mol can be defined as $G_{mol} = \{V_{mol}, E_{mol}\}$, where V_{mol} and E_{mol} denote the sets of atoms and bonds. Then we represent a intra-molecular ***motif*** m_s, a subgraph of G_{mol}, as $G_s = \{V_s, E_s\}$, where V_s and E_s denote the sets of atoms and bonds in motif m_s. Given a specific molecule, we define it as a ***motif graph*** $\mathcal{G}^m = (\mathcal{V}^m, \mathcal{E}^m)$, where $\mathcal{V}^m$ is the motif set and $\mathcal{E}^m$ is the corresponding bond set.

Objective. In our study, DDI prediction is a link prediction task defined on the DDI network. Given a specific DDI network and the corresponding motif information, our proposed model aims to establish a mapping function to generate the predicted adjacency matrix $\hat{\mathbf{A}}$, which indicates whether interactions occur between drug pairs.

Methodology

In this paper, we propose MOTOR, a motif-oriented learning and topology refinement framework for DDI prediction. As shown in Figure 2, we focus on learning the intra-molecular information through motifs, capturing internal structures, local contexts, and global semantics of motifs, achieving a comprehensive understanding from local to global perspectives. To effectively leverage the DDI network, we employ an iterative strategy to enhance both molecular representations and DDI network topology through graph representation learning and graph topology learning. Finally, we develop a DDI predictor to identify potential drug interactions.

Motif-Oriented Representation Learning

Motif Structure Learning. Given the intricate intertwining of molecular internal structures, we first adopt a micro perspective to analyze the fine-grained internal structures of motifs. We decompose Simplified Molecular Input Line Entry System (SMILES) descriptors into small and chemically meaningful motifs by Breaking of Recurrent Internal Chemical Structures (BRICS) (Degen et al. 2008).

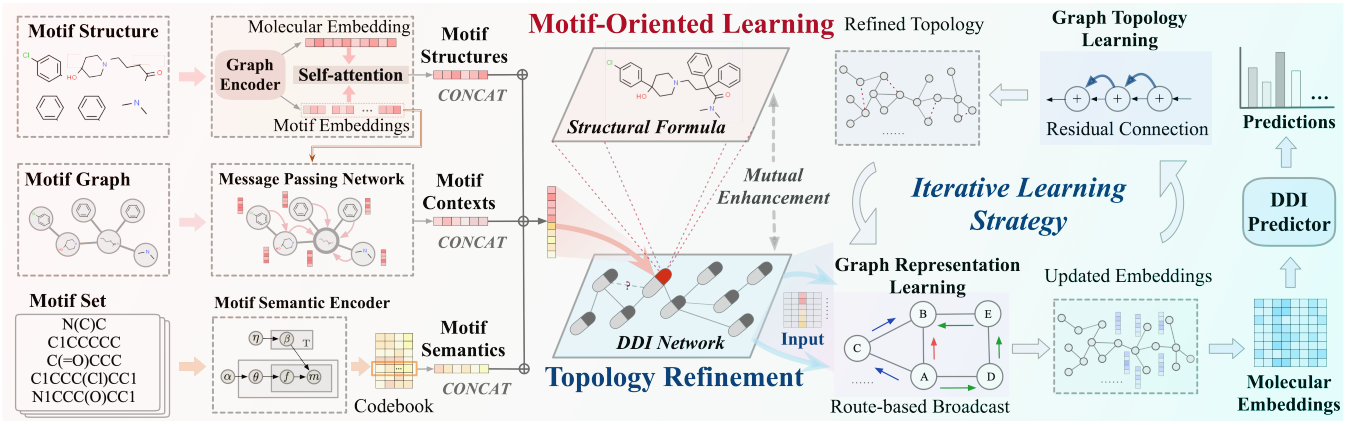


Figure 2: The overall framework of MOTOR. MOTOR contains two interdependent and mutually enhancing modules: (1) Motif-oriented learning: molecules are encoded into multi-granularity molecular representations by aggregating motif structures, motif contexts, and motif semantics. (2) Topology refinement: molecular representations are optimized along with topology refinement via an iterative learning strategy. Finally, DDI predictor is designed to generate the prediction results.

We derive atom features from nuclear charge numbers and bond features from various types of chemical bonds, such as single, double, triple, and aromatic. These attribute features are mapped to a hidden vector space $\mathbf{H}$ through linear transformations, with $\mathbf{h}_i^{(0)} = \mathbf{h}_i$ and $\mathbf{h}_{kl}^{(0)} = \mathbf{h}_{kl}$ as the initial states of atom a_i and bond b_{kl}. For each graph, $G = (V, E)$, we apply the attention mechanism to capture the inter-node dependencies, then propagate messages via edges and update node representations at the l-th layer:

$$\mathbf{h}_i^{(l)} = \sum_{j \in N_i} \text{softmax}\big(\text{ATT}(\mathbf{h}_i^{(l-1)}, \mathbf{h}_j^{(l-1)})\big) \cdot \mathbf{h}_j^{(l-1)} \quad (1)$$

where N_i denotes the set of neighborhood nodes of i and itself, and ATT is the attention calculation operator.

The edge representation is updated by combining its existing embedding with the hidden node embeddings at both ends. After L_1 iterations of message passing, all nodes and edges are aggregated to generate $\mathbf{h}_g$:

$$\mathbf{h}_g = \Big[\frac{1}{|V|} \sum_{v_i \in V} \mathbf{h}_i^{(\mathrm{L}_1)} \Big| \frac{1}{|E|} \sum_{e_{kl} \in E} \mathbf{h}_{kl}^{(\mathrm{L}_1)} \Big] \quad (2)$$

where $|\cdot|$ indicates the number of elements in the set, and $[\cdot|\cdot]$ denotes the vector concatenation operation.

We feed molecule G_{mol} and its corresponding motifs $\{G_s | \, m_s \in \mathcal{V}^m\}$ into the aforementioned mappings to obtain each molecular embedding $\mathbf{h}_{mol}$ and all motif embeddings $\{\mathbf{h}_{m_s} | \, m_s \in \mathcal{V}^m\}$. Given the varying significance of different motifs within a molecule, we apply a self-attention mechanism to investigate the local influences of distinct motifs on the overall molecule. Then we leverage the relevance between motifs and the molecule to learn the molecular representation with motif structures, denoted as $\mathbf{h}_{str}$:

$$\mathbf{h}_{str} = \sum_{m_s \in \mathcal{V}^m} \text{softmax}\Big(\frac{\mathbf{h}_{mol} \cdot \mathbf{h}_{m_s}^{\mathbf{T}}}{\sqrt{d_{m_s}}}\Big) \cdot \mathbf{h}_{m_s} \quad (3)$$

where d_{m_s} denotes the dimension of motif embedding $\mathbf{h}_{m_s}$, and $\mathbf{h}_{m_s}^{\mathbf{T}}$ denotes its transpose.

Motif Context Learning. Motifs within a molecule, organized through inter-motif bonds, are interdependent and can influence each other. For example, in phenol, the hydroxyl group facilitates the substitution of the phenyl group, while the phenyl group imparts acidity to the hydroxyl group. Therefore, we retain the bonds broken by BRICS, rebuild the inter-motif relationships to construct motif graphs, and investigate the dependencies between motif pairs within a specific intra-molecular context. Subsequently, we introduce a message-passing network to explore the mutual influence of spatially adjacent motif pairs. Given motif graph $\mathcal{G}^m = (\mathcal{V}^m, \mathcal{E}^m, \mathbf{X}^m, \mathbf{A}^m)$, $\mathbf{X}^m = \{\mathbf{h}_{m_s} | \, m_s \in \mathcal{V}^m\} \cup \{\mathbf{h}_{kl} | \, e_{kl} \in \mathcal{E}^m\}$ is the feature set derived from motif structure learning. Specifically, in motif context learning, the initial states $\mathbf{h}_{m_s}^{(0)} = \mathbf{h}_{m_s}$, $\mathbf{h}_{kl}^{(0)} = \mathbf{h}_{kl}^{(\mathrm{L}_1)}$. After L_2 layers of message passing as detailed in Equation 1, the molecular representation $\mathbf{h}_{con}$ is formed through the aggregation of motifs and bonds as follows:

$$\mathbf{h}_{con} = \Big[\frac{1}{|\mathcal{V}^m|} \sum_{m_s \in \mathcal{V}^m} \mathbf{h}_{m_s}^{(\mathrm{L}_2)} \Big| \frac{1}{|\mathcal{E}^m|} \sum_{b_{kl} \in \mathcal{E}^m} \mathbf{h}_{kl}^{(\mathrm{L}_2)} \Big] \quad (4)$$

Motif Semantic Learning. Motifs often display distinct patterns across different categories of molecules. That is, the global distribution of motifs can provide important insights into their functional semantics and pharmacological efficacy. Notably, topic models are widely applied to capture semantic information by analyzing the probability distribution of words or terms within a given corpus (Rosen-Zvi et al. 2012). Similarly, we develop a motif semantic encoder to encode the potential semantics between specific motifs and functions by exploring the motif distribution across diverse molecules. We treat each molecule as a composition of motifs, with motif functions analogous to word topics in documents. Given that a drug can have multiple pharmacological effects, we incorporate Latent Dirichlet Allocation (LDA) with Dirichlet-Multinomial Mixture (DMM) to flexibly capture the complex pharmacological semantics.

For each molecule mol: $\theta_{mol} \sim \text{Dirichlet}(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ denotes the prior distribution parameter of DMM, supporting the concurrent representation of multiple functions within each molecule; for a pharmacological function t: $\beta_t \sim \text{Dirichlet}(\boldsymbol{\eta})$, where $\boldsymbol{\eta}$ is the Dirichlet distribution parameter; for each motif m_s, motif function $f_s \sim \text{Multinomial}(\theta_{mol})$, motif $m_s \sim \text{Multinomial}(\beta_{f_s})$. The likelihood function can be formalized as:

$$P(m, f, \theta, \beta | \boldsymbol{\alpha}, \boldsymbol{\eta}) = \prod_{mol} P(\theta_{mol}|\boldsymbol{\alpha}) \prod_{t} P(\beta_t|\boldsymbol{\eta}) \prod_{s} P(f_s|\theta_{mol}) P(m_s|f_s, \beta_{f_s}) \tag{5}$$

During the training process, motifs are iteratively assigned to motif functions, and the probability distribution of motif functions for each molecule is updated continuously. Following the motif semantic encoder, we obtain a semantic codebook that contains the pharmacological functional distributions of all molecules. Each variable in this codebook indicates the relative importance of the corresponding motif function within the overall pharmacological profile. According to the codebook, the entry $\mathbf{h}_{sem} = [\theta_{mol,1}, \theta_{mol,2}, \ldots, \theta_{mol,\mathrm{T}}]$ is the motif semantic embedding of molecule mol. Here, T is the predefined number of motif functions, and $\theta_{mol,t}$ indicates the weight assigned to the t-th motif function.

By applying the three motif-oriented learning components described above, we explore the motif internal structures, motif local contexts, and motif global semantics. To achieve a comprehensive understanding of molecule, we concatenate the molecular embeddings as $\mathbf{g} = [\mathbf{h}_{str} \,|\, \mathbf{h}_{con} \,|\, \mathbf{h}_{sem}]$.

Iterative Topology Refinement

To capture high-order relationships between drugs, we position drugs as nodes within the DDI network. Given the inherent incompleteness of DDI networks, we employ an iterative learning strategy to refine the DDI network topology, and obtain the drug representations that incorporate the optimized neighborhood information.

Graph Representation Learning (GRL). GRL is designed to acquire informative molecular representations through message propagation on the DDI network. However, most graph-based DDI prediction methods aggregate neighborhood information based only on feature relevance, such as $q_{ij} = \text{ATT}(\mathbf{d}_i, \mathbf{d}_j)$, overlooking the structural details of drug pairs within the DDI network. To address the omission of structural details, we enhance GRL by equipping it with a route-based broadcast mechanism.

To be specific, we introduce the topological relevance term $r_{ij} = \text{Norm}\left(\sum_{p=1}^{\mathrm{P}} \tilde{\mathbf{I}}_{ij}^{p}\right)$, where $\tilde{\mathbf{I}}^p$ represents the adjacency matrix of the input DDI network raised to the power of p. The term r_{ij} quantifies the number of paths between drug i and drug j within the predefined number of hops P, with higher values indicating stronger topological relevance. For stable training, we apply a multi-head (C in total) multi-layer (L_3 in total) message-passing network to update drug i embedding $\mathbf{z}_i = \mathbf{d}_i^{(\mathrm{L}_3)}$ as follows:

$$\mathbf{d}_i^{(l+1)} = \Big\Vert_{c=1}^{\mathrm{C}} \sum_{j \in \tilde{\mathcal{N}}_i} \text{softmax}\big[w_q \cdot q_{ij}^{(l)} + w_r \cdot r_{ij}^{(l)}\big]_c \cdot \mathbf{d}_j^{(l)} + \mathbf{d}_i^{(l)} \tag{6}$$

where the initial states $\mathbf{d}_i^{(0)} = \mathbf{g}_i$, $\mathbf{d}_j^{(0)} = \mathbf{g}_j$ are obtained from motif-oriented learning, w_q and w_r are learnable parameters to balance $q_{ij}^{(l)}$ and $r_{ij}^{(l)}$, and $\tilde{\mathcal{N}}_i$ denotes the set of interacting drugs with drug i.

Graph Topology Learning (GTL). Ongoing advancements in novel drug development lead to delayed discovery of potential interactions in the DDI network due to limited research resources, time-consuming experiments, and evolving knowledge, resulting in an incomplete original DDI network. Despite its incompleteness, the original topology retains significant information about the optimal topology. The refined topology can be considered as a subtle adjustment from the original graph (Wang et al. 2021a). The edges between two nodes are influenced not only by the two node embeddings but also by previous observations. Therefore, we use residual connections to incorporate previous observations into the current predictions. Specifically, we train a Multi-Layer Perceptron (MLP) to estimate the interaction probability between drug pairs based on the embeddings obtained from GRL for the learned topology $\mathbf{L}$ as follows:

$$\mathbf{L}_{ij} = \sigma\big[\text{MLP}\big(\text{LayerNorm}(\mathbf{z}_i) \odot \text{LayerNorm}(\mathbf{z}_j)\big)\big] \tag{7}$$

Following that, we perform the addition and normalization operations between $\mathbf{L}$ and $\mathbf{A}$ to obtain the refined topology $\mathbf{I} = \text{Norm}(\mathbf{L} + \mathbf{A})$.

Iterative Learning Strategy. GRL and GTL represent two interconnected and mutually influencing aspects of DDI networks: features and structures. The informative drug representations from GRL facilitate the discovery of potential links, while the updated DDI network topology from GTL contributes to improved drug representations. Consequently, we employ an iterative learning strategy for joint learning of graph topology and node representations. During the training process, there is a reciprocal exchange of inputs and outputs between these two components. The learned topology $\mathbf{L}$ is derived from the output embeddings $\mathbf{z}$ of GRL. In the initial iteration, $\tilde{\mathcal{N}}_i$ and $\tilde{\mathbf{I}}$ in GRL rely on the original graph topology $\mathbf{A}$. In subsequent iterations, the refined topology $\mathbf{I}$ obtained from GTL replaces $\mathbf{A}$. That is, $\tilde{\mathcal{N}}_i$ and $\tilde{\mathbf{I}}$ are updated based on $\mathbf{I}$. The iterative process continuously aggregates the nuanced information from the refined topology into node representations, leading to more informative and comprehensive drug representations.

DDI Predictor. To detect the potential interaction between a specific drug pair (i, j), we apply a two-layer MLP to compress the drug embeddings into a link value p_{ij}, signifying the interaction probability between drug i and drug j. The DDI predictor produces two prediction scores, p_{ij}^{g} and p_{ij}^{z}, based on drug representations from motif-oriented learning and topology refinement, respectively. Additionally, GTL refines the DDI topology $\mathbf{I}$, which also yields the prediction $\mathbf{I}_{ij}$ for drug pair (i, j). The drug i embedding $\mathbf{z}_i$ relies on motif information and the optimized DDI network topology, so we use p_{ij}^{z} for the final DDI prediction during testing.

Model	ZhangDDI				ChCh-Miner				DeepDDI			
	AUROC	AP	F1	ACC	AUROC	AP	F1	ACC	AUROC	AP	F1	ACC
GCN	94.00	88.45	79.29	89.96	92.57	98.89	89.94	83.67	84.19	94.01	85.62	79.79
GAT	93.81	87.60	80.07	90.54	92.24	98.98	90.60	84.75	84.21	94.11	85.48	79.61
MR-GNN	96.18	92.63	82.93	91.90	93.11	95.95	88.13	85.03	93.35	94.56	90.07	87.54
EPGNN-DS	90.83	88.96	80.07	82.40	94.23	96.80	89.41	86.64	85.93	88.72	84.86	80.22
DeepDrug	93.35	92.33	82.89	85.67	98.38	99.16	94.67	93.18	91.74	92.99	89.39	86.28
MIRACLE	96.44	93.09	85.16	93.16	96.20	99.50	94.55	90.77	92.76	96.77	93.54	90.33
CSGNN	91.71	89.02	83.60	84.14	97.68	97.56	92.47	92.54	94.01	94.17	86.01	86.33
SSI-DDI	93.14	92.09	81.96	85.35	98.09	98.97	93.98	92.19	91.79	93.47	88.23	85.38
DeepDDS	93.20	92.08	82.79	85.63	97.10	98.51	92.21	90.38	94.38	95.68	91.27	88.87
DSN-DDI	91.13	86.42	87.68	86.65	96.69	96.34	88.12	88.89	93.22	92.87	85.60	85.41
MOTOR	**98.96**	**97.55**	**92.42**	**96.62**	**99.46**	**99.93**	**98.30**	**97.03**	**94.50**	**97.39**	**94.84**	**92.15**

Table 1: Overall performance of different DDI prediction methods. The best results are highlighted in **bold**, and the runner-up results are underlined. (Higher values indicate better performance.)

Loss Function. The loss comprises three components: supervised loss, inconsistency loss, and contrastive loss. It is worth noting that the latter two loss components facilitate mutual enhancement between modules.

Supervised loss $\mathcal{L}_S$ calculates Binary Cross-Entropy (BCE) between the predictions and the ground truth as follows:

$$\mathcal{L}_S = \sum_{(i,j)\in\Omega} \text{BCE}\big(p^g_{ij} \| y_{ij}\big) + \text{BCE}\big(p^z_{ij} \,\|\, y_{ij}\big) + \text{BCE}\big(\mathbf{I}_{ij} \| y_{ij}\big) \tag{8}$$

where Ω is the set of observed drug pairs in the training set, and y_{ij} denotes the true label of drug pair (i, j).

Inconsistency loss $\mathcal{L}_I$ measures the Kullback-Leibler (KL) divergence between the three predictions of unobserved samples to strengthen the consistency of different modules.

$$\mathcal{L}_I = \sum_{(i,j)\in\Omega^-} \text{KL}\big(p^z_{ij} \| p^g_{ij}\big) + \text{KL}\big(\mathbf{I}_{ij} \| p^g_{ij}\big) + \text{KL}\big(p^z_{ij} \| \mathbf{I}_{ij}\big) \tag{9}$$

where Ω^- (the complement of Ω) represents the set of unobserved drug pairs during training.

Contrastive loss $\mathcal{L}_C$ encourages MOTOR to learn more informative and distinguishable features from the different modules. Following (Nowozin, Cseke, and Tomioka 2016), we formulate the contrastive loss $\mathcal{L}_C$ by maximizing the similarity of positive samples and minimizing the similarity of negative samples as follows:

$$\mathcal{L}_C = \sum_{(i,j)\in\mathcal{E}} \log \frac{\exp(\text{JSD}(\mathbf{g}_i, \mathbf{z}_j))}{\exp(\text{JSD}(\mathbf{g}_i, \mathbf{z}_j)) + \sum\limits_{j'\notin\mathcal{N}_i} \exp(\text{JSD}(\mathbf{g}_i, \mathbf{z}_{j'}))} \tag{10}$$

where $\mathcal{E}$ is the set of drug pairs that can interact (positive samples) in the training set, JSD is Jensen–Shannon divergence, and exp is the exponential function.

Total loss $\mathfrak{L}$ is the weighted sum of supervised loss $\mathcal{L}_S$, inconsistency loss $\mathcal{L}_I$, and contrastive loss $\mathcal{L}_C$, formalized as:

$$\mathfrak{L} = \mathcal{L}_S + \lambda_1\mathcal{L}_I + \lambda_2\mathcal{L}_C \tag{11}$$

where λ_1 and λ_2 are hyper-parameters to balance supervised loss $\mathcal{L}_S$ with mutual learning loss components $\mathcal{L}_I$ and $\mathcal{L}_C$.

Experiment

Experimental Setup

Datasets. We evaluate MOTOR on three real-world DDI datasets: *ZhangDDI* (Zhang et al. 2017), *ChCh-Miner* (ZMarinka Zitnik and Leskovec 2017), and *DeepDDI* (Ryu, Kim, and Lee 2018). Following (Wang et al. 2021b), we remove unidentifiable SMILES during preprocessing. We then perform a stratified splitting to divide all drug pairs into a training set, a validation set, and a test set at a ratio of 6:2:2.

Evaluation Metrics. To verify the performance of DDI prediction, four widely used metrics are selected: Area Under the Receiver Operating Characteristic curve (AUROC), Average Precision (AP), F1-score (F1), and Accuracy (ACC). To reduce the impact of random errors, we report **mean** of the four metrics over five repetitions.

Baseline Methods. We compare with 10 representative baseline models as follows: GCN (Kipf and Welling 2016), GAT (Velickovic et al. 2017), MR-GNN (Xu et al. 2019), EPGNN-DS (Sun et al. 2020), DeepDrug (Cao, Fan, and Zeng 2020), MIRACLE (Wang et al. 2021b), CSGNN (Zhao et al. 2021), SSI-DDI (Nyamabo, Yu, and Shi 2021), DeepDDS (Wang et al. 2022), and DSN-DDI (Li et al. 2023).

Implementation Details. All experiments are conducted with EPYC 7742 CPU, and TESLA A100 GPU. We set $\text{L}_1 = \text{L}_2 = \text{L}_3 = 3$, $\text{P} = 3$, $\text{C} = 8$, $\lambda_1 = 0.8$, $\lambda_2 = 0.6$, $\text{T} = 15$, and fix the number of epochs to 100, the learning rate to 0.001. We initialize MOTOR by Xavier initialization and use Adam optimizer to update the parameters.

Overall Performance

Table 1 compares MOTOR's performance with the SOTA DDI prediction methods. Based on Table 1, we make the following observations:

(1) The experimental results demonstrate the effectiveness of MOTOR, as it consistently outperforms all baseline methods on three real-world datasets. Specifically, MOTOR exhibits the average DDI performance improvements

Model	ZhangDDI				ChCh-Miner				DeepDDI			
	AUROC	AP	F1	ACC	AUROC	AP	F1	ACC	AUROC	AP	F1	ACC
w/o str	98.82	97.21	90.32	95.51	99.40	99.92	97.75	96.08	94.13	97.15	94.56	91.79
w/o con	98.88	97.38	91.25	96.01	99.37	99.91	97.96	96.44	94.10	97.12	94.69	91.98
w/o sem	98.88	97.43	91.80	96.29	99.26	99.90	97.61	95.85	94.15	97.19	94.72	91.98
w/o grl	98.83	97.34	90.30	95.50	99.41	99.91	97.76	96.11	94.28	97.23	94.33	91.49
w/o gtl	98.92	97.44	90.62	95.67	99.43	99.92	97.87	96.30	94.29	97.25	94.41	91.59
w/o $\mathcal{L}_I$	98.91	97.45	91.39	96.08	99.34	99.91	98.16	96.79	94.27	97.23	94.49	91.69
w/o $\mathcal{L}_C$	98.52	96.73	90.26	95.57	99.00	99.86	97.62	95.86	92.97	96.39	94.77	92.04
MOTOR	**98.96**	**97.55**	**92.42**	**96.62**	**99.46**	**99.93**	**98.30**	**97.03**	**94.50**	**97.39**	**94.84**	**92.15**

Table 2: Component ablation study results. The best results are highlighted in **bold**.

of 3.80%, 2.25%, and 0.96% on three datasets. Notably, it shows significant increases of 4.74%, 3.63%, and 1.30% in F1, highlighting its superiority in DDI prediction tasks.

(2) MOTOR exceeds the baseline methods that rely solely on molecular structures or DDI networks, indicating the importance of integrating both multi-granularity motif information and topological structures.

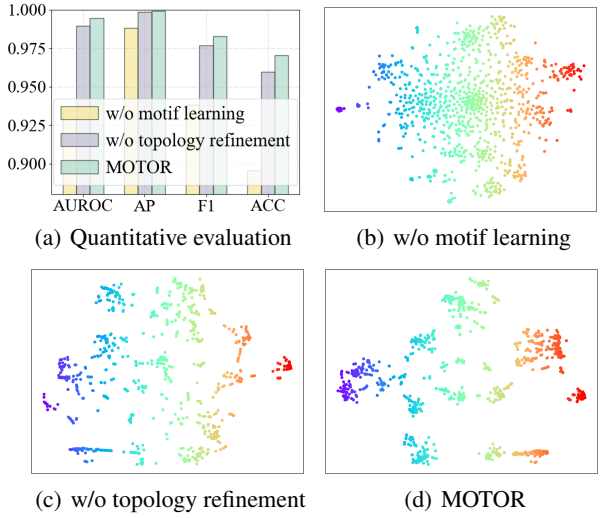
(3) MOTOR achieves the greatest performance improvement on ZhangDDI dataset with the highest edge density. This demonstrates MOTOR's proficiency in handling dense network topologies and its applicability to challenging real-world scenarios.

Ablation Study Analysis

We conduct ablation studies to explore the contributions of different components and modules within MOTOR. Specifically, we denote MOTOR without motif structure learning as *w/o str*, denote MOTOR without motif context learning as *w/o con*, denote MOTORwithout motif semantic learning as *w/o sem*, denote MOTOR without GRL as *w/o grl*, denote MOTOR without GTL as *w/o gtl*, denote MOTOR without $\mathcal{L}_I$ as *w/o* $\mathcal{L}_I$, and denote MOTOR without $\mathcal{L}_C$ as *w/o* $\mathcal{L}_C$. As listed in Table 2, each model component and loss component in MOTOR contributes to performance improvements. Furthermore, we conduct an ablation study on motif-oriented learning and topology refinement, respectively. Figure 3(a) presents the quantitative results, and for a more intuitive comparison, we use t-distributed Stochastic Neighbor Embedding (t-SNE) to project drug embeddings into a two-dimensional space, as shown in Figure 3(b)(c)(d). Consistent with the quantitative results, drug embeddings learned from MOTOR exhibit clearer boundaries and clustering tendencies than those from ablation models, showing the significant advantages of effectively integrating the two modules to enhance molecular understanding.

Discovery of Motif Semantics

Each molecule is represented as a mixture of motif functions describing its pharmacological effects. To uncover meaningful patterns within motifs, we analyze the parameter β_t^m and present the representative motifs for a specific motif function t in ZhangDDI dataset in Table 3. Specifically, motifs



(a) Quantitative evaluation (b) w/o motif learning

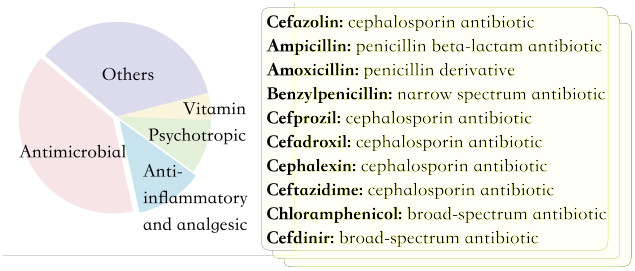
(c) w/o topology refinement (d) MOTOR

Figure 3: Module ablation study on ChCh-Miner dataset.

m_1 and m_2 are commonly found in various drugs to enhance metabolic stability. m_3 is frequently observed in antibiotics, while m_6 is common in Non-Steroidal Anti-Inflammatory Drugs (NSAIDs). m_4, m_5, and m_7 are substructures shared by some antibiotics and NSAIDs. In addition, m_8, m_9, and m_{10} resemble the core structures of analgesics. Referring to the semantic codebook, we analyze 43 molecules with the most significant motif function t and list the top 10 typical drugs in Figure 4. Among the 43 drugs, 17 are antibiotics and antiparasitics, five are anti-inflammatory and analgesic, with more than half showing antimicrobial and anti-inflammatory effects. These observations suggest that motif function t represents a hybrid pharmacological function associated with treating infections and inflammation. Notably, in clinical practice, antibiotics, NSAIDs, and analgesics are often co-administered to manage symptoms of inflammation accompanying infections. Experimental results demonstrate that MOTOR has the potential to capture latent motif semantics, providing interpretable insights into DDI mechanisms.

Rank#	Motif	Prob.
m_1	C(=O)O	0.241
m_2	C1CCCCC1	0.098
m_3	[C@@H]1N2C(=O)[C@@H][C@H]2SC1(C)C	0.042
m_4	C(=O)[C@@H]C	0.036
m_5	C(=O)[C@@H]CC	0.030
m_6	C1NOC(C)C1	0.018
m_7	C(=O)[C@@H]CCC	0.012
m_8	C1CCCC(F)C1	0.012
m_9	N1[C@H]C[C@@H]2CCC[C@@H]21	0.012
m_{10}	N1[C@H]C[C@H]2CCCC[C@@H]21	0.006

Table 3: The top 10 motifs for motif function t.



(a) Drug category overview (b) The top 10 typical drugs

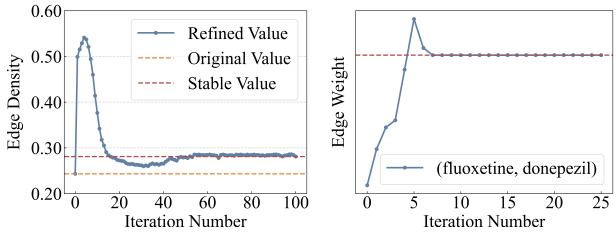
Figure 4: Pharmacological analysis of motif function t.

Evolution of Topology Refinement

To examine the evolution of DDI network topology during topology refinement, we first analyze the overall edge density of the DDI network. Figure 5(a) illustrates that topology refinement involves a comprehensive exploration of potential interactions and continuous adjustments of DDI weights. Then we visualize the evolution of the weight for one edge. Figure 5(b) details the normalized weight evolution of a positive sample, $(fluoxetine, donepezil)$, from ZhangDDI test set during the iterative learning process. Fluoxetine is a selective serotonin reuptake inhibitor for both acute and maintenance treatment of major depressive disorder, obsessive-compulsive disorder, and bulimia nervosa. Donepezil is an acetylcholinesterase inhibitor used to treat the behavioral and cognitive effects of Alzheimer's Disease. The therapeutic efficacy of fluoxetine can be decreased when used in combination with donepezil. Before topology refinement, MOTOR lacks visibility into this positive sample from the test set during training, resulting in an edge weight of 0. Through iterative topology refinement, the DDI weight gradually increases and stabilizes, indicating that the topology refinement module can effectively capture missing information and learn potential links.

Related Work

DDI Prediction. DDI detection through in vivo and in vitro experiments is the most accurate and reliable way. However, high experimental costs compel researchers to explore computational DDI prediction methods (Han et al. 2022b). *Text mining-based methods* extract DDIs from unstructured text sources (Segura-Bedmar, Martínez Fernández, and Herrero Zazo 2013; Zhang, Leng, and Liu 2020). Some *Similarity-based methods* (Zhang et al. 2018; Rohani, Eslahchi, and Katanforoush 2020) obtain latent feature matrices through matrix decomposition to discover potential DDIs. Recently, *graph-based methods* leveraging GNNs have demonstrated strong performance in DDI prediction (Guo et al. 2023; Huang, Ye, and Sakurai 2024).



(a) The trend of edge density (b) The trend of edge weight

Figure 5: The topology evolution over iterations.

Motif-based Molecular Representation Learning. Motifs are significant subgraph patterns that occur frequently in molecules (Milo et al. 2002). (Zhang et al. 2021) proposes a motif-based pre-training framework to capture the multi-scale information in molecular graphs. (Nyamabo, Yu, and Shi 2021) breaks the DDI prediction task down to identifying pairwise interactions between their respective motifs. (Yu and Gao 2022) builds a heterogeneous graph with motif nodes and molecular nodes to capture motif-level relationships. (Ye et al. 2024) presents a hierarchical cross-level graph contrastive learning framework to extract semantic motifs and learn complex connections from drug-motif interaction graphs. Motif-based molecular representation learning models investigate latent functional group information to enhance the understanding of molecules.

Graph Neural Networks. Recently, GNNs have attracted extensive research attention in handling graph-structured data (Ning et al. 2022; Corso et al. 2024). *Spectral-based GNNs* perform graph convolution operations in the spectral domain. (Bruna et al. 2013) first proposes a spectral graph-based extension of convolutional networks using the Fourier basis. (Xu et al. 2018) addresses graph isomorphism and related tasks. *Spatial-based GNNs* directly define graph convolution in the spatial domain by aggregating and transforming local information. (Hamilton, Ying, and Leskovec 2017) employs random sampling and aggregation with various pooling. (Velickovic et al. 2017) assigns different weights to different neighbors. At present, GNNs have emerged as a powerful tool for AI-driven DDI prediction methods.

Conclusion

In this paper, we propose MOTOR, integrating motif-oriented learning and topology refinement for DDI prediction. MOTOR captures the internal structures, local contexts, and global semantics of motifs. We then iteratively refine the DDI network topology. Furthermore, contrastive learning is introduced to facilitate mutual enhancement between modules. Extensive experiments validate the effectiveness and potential interpretability of MOTOR.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 92470204 and 62406306) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA0460104).

References

Bruna, J.; Zaremba, W.; Szlam, A.; and LeCun, Y. 2013. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*.

Cao, X.; Fan, R.; and Zeng, W. 2020. DeepDrug: a general graph-based deep learning framework for drug relation prediction. *biorxiv*.

Corso, G.; Stark, H.; Jegelka, S.; Jaakkola, T.; and Barzilay, R. 2024. Graph neural networks. *Nature Reviews Methods Primers*, 4(1): 17.

Degen, J.; Wegscheid-Gerlach, C.; Zaliani, A.; and Rarey, M. 2008. On the art of compiling and using'drug-like'chemical fragment spaces. *ChemMedChem*, 3(10): 1503.

Ferdousi, R.; Safdari, R.; and Omidi, Y. 2017. Computational prediction of drug-drug interactions based on drugs functional similarities. *Journal of biomedical informatics*, 70: 54–64.

Franceschi, L.; Niepert, M.; Pontil, M.; and He, X. 2019. Learning discrete structures for graph neural networks. In *International conference on machine learning*, 1972–1982. PMLR.

Guo, L.; Lei, X.; Chen, M.; and Pan, Y. 2023. MSResG: Using GAE and Residual GCN to Predict Drug–Drug Interactions Based on Multi-source Drug Features. *Interdisciplinary Sciences: Computational Life Sciences*, 1–18.

Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

Han, K.; Cao, P.; Wang, Y.; Xie, F.; Ma, J.; Yu, M.; Wang, J.; Xu, Y.; Zhang, Y.; and Wan, J. 2022a. A review of approaches for predicting drug–drug interactions based on machine learning. *Frontiers in pharmacology*, 12: 814858.

Han, K.; Cao, P.; Wang, Y.; Xie, F.; Ma, J.; Yu, M.; Wang, J.; Xu, Y.; Zhang, Y.; and Wan, J. 2022b. A review of approaches for predicting drug–drug interactions based on machine learning. *Frontiers in Pharmacology*, 12: 3966.

Huang, D.; Ye, X.; and Sakurai, T. 2024. Multi-party collaborative drug discovery via federated learning. *Computers in Biology and Medicine*, 108181.

Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Li, Z.; Zhu, S.; Shao, B.; Zeng, X.; Wang, T.; and Liu, T.-Y. 2023. DSN-DDI: an accurate and generalized framework for drug–drug interaction prediction by dual-view representation learning. *Briefings in Bioinformatics*, 24(1).

Luo, H.; Yin, W.; Wang, J.; Zhang, G.; Liang, W.; Luo, J.; and Yan, C. 2024. Drug-drug interactions prediction based on deep learning and knowledge graph: A review. *Iscience*.

Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; and Alon, U. 2002. Network motifs: simple building blocks of complex networks. *Science*, 298(5594): 824–827.

Ning, Z.; Wang, P.; Wang, P.; Qiao, Z.; Fan, W.; Zhang, D.; Du, Y.; and Zhou, Y. 2022. Graph soft-contrastive learning via neighborhood ranking. *arXiv preprint arXiv:2209.13964*.

Nowozin, S.; Cseke, B.; and Tomioka, R. 2016. f-gan: Training generative neural samplers using variational divergence minimization. *Advances in neural information processing systems*, 29.

Nyamabo, A. K.; Yu, H.; and Shi, J.-Y. 2021. SSI–DDI: substructure–substructure interactions for drug–drug interaction prediction. *Briefings in Bioinformatics*, 22(6): bbab133.

Pei, H.; Wei, B.; Chang, K. C.-C.; Lei, Y.; and Yang, B. 2020. Geom-gcn: Geometric graph convolutional networks. *arXiv preprint arXiv:2002.05287*.

Rohani, N.; Eslahchi, C.; and Katanforoush, A. 2020. ISCMF: Integrated similarity-constrained matrix factorization for drug–drug interaction prediction. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9: 1–8.

Rosen-Zvi, M.; Griffiths, T.; Steyvers, M.; and Smyth, P. 2012. The author-topic model for authors and documents. *arXiv preprint arXiv:1207.4169*.

Ryu, J. Y.; Kim, H. U.; and Lee, S. Y. 2018. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the National Academy of Sciences*, 115(18): E4304–E4311.

Segura-Bedmar, I.; Martínez Fernández, P.; and Herrero Zazo, M. 2013. Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). Association for Computational Linguistics.

Sun, M.; Wang, F.; Elemento, O.; and Zhou, J. 2020. Structure-based drug-drug interaction detection via expressive graph convolutional networks and deep sets (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 13927–13928.

Takeda, T.; Hao, M.; Cheng, T.; Bryant, S. H.; and Wang, Y. 2017. Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge. *Journal of cheminformatics*, 9(1): 1–9.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y.; et al. 2017. Graph attention networks. *stat*, 1050(20): 10–48550.

Vilar, S.; Harpaz, R.; Uriarte, E.; Santana, L.; Rabadan, R.; and Friedman, C. 2012. Drug—drug interaction through molecular structure similarity analysis. *Journal of the American Medical Informatics Association*, 19(6): 1066–1074.

Wang, J.; Liu, X.; Shen, S.; Deng, L.; and Liu, H. 2022. DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Briefings in Bioinformatics*, 23(1): bbab390.

Wang, R.; Mou, S.; Wang, X.; Xiao, W.; Ju, Q.; Shi, C.; and Xie, X. 2021a. Graph structure estimation neural networks. In *Proceedings of the Web Conference 2021*, 342–353.

Wang, Y.; Min, Y.; Chen, X.; and Wu, J. 2021b. Multi-view graph contrastive representation learning for drug-drug interaction prediction. In *Proceedings of the Web Conference 2021*, 2921–2933.

Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*.

Xu, N.; Wang, P.; Chen, L.; Tao, J.; and Zhao, J. 2019. Mr-gnn: Multi-resolution and dual graph neural network for predicting structured entity interactions. *arXiv preprint arXiv:1905.09558*.

Ye, Y.; Zhou, J.; Li, S.; Xiao, C.; Ying, H.; and Xiong, H. 2024. Hierarchical Cross-Level Graph Contrastive Learning for Drug-Drug Interaction Prediction. In *International Conference on Database Systems for Advanced Applications*, 232–242. Springer.

Yu, Z.; and Gao, H. 2022. Molecular representation learning via heterogeneous motif graph neural networks. In *International Conference on Machine Learning*, 25581–25594. PMLR.

Zhang, R.; Wang, X.; Liu, K.; Zhou, Y.; and Wang, P. 2024a. M2Mol: Multi-view Multi-granularity Molecular Representation Learning for Property Prediction. In *International Conference on Database Systems for Advanced Applications*, 264–274. Springer.

Zhang, R.; Wang, X.; Wang, P.; Meng, Z.; Cui, W.; and Zhou, Y. 2023. HTCL-DDI: a hierarchical triple-view contrastive learning framework for drug–drug interaction prediction. *Briefings in Bioinformatics*, 24(6): bbad324.

Zhang, R.; Wang, X.; Wang, S.; Liu, K.; Zhou, Y.; and Wang, P. 2024b. H2D: Hierarchical Heterogeneous Graph Learning Framework for Drug-Drug Interaction Prediction. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 4283–4287.

Zhang, T.; Leng, J.; and Liu, Y. 2020. Deep learning for drug–drug interaction extraction from the literature: a review. *Briefings in bioinformatics*, 21(5): 1609–1627.

Zhang, W.; Chen, Y.; Li, D.; and Yue, X. 2018. Manifold regularized matrix factorization for drug-drug interaction prediction. *Journal of biomedical informatics*, 88: 90–97.

Zhang, W.; Chen, Y.; Liu, F.; Luo, F.; Tian, G.; and Li, X. 2017. BioSNAP Datasets: Stanford Biomedical Network Dataset Collection. *BMC bioinformatics*, 18: 1–12.

Zhang, Z.; Liu, Q.; Wang, H.; Lu, C.; and Lee, C.-K. 2021. Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, 34: 15870–15882.

Zhao, C.; Liu, S.; Huang, F.; Liu, S.; and Zhang, W. 2021. CSGNN: Contrastive Self-Supervised Graph Neural Network for Molecular Interaction Prediction. In *IJCAI*, 3756–3763.

Zhao, Y.; Yin, J.; Zhang, L.; Zhang, Y.; and Chen, X. 2024. Drug–drug interaction prediction: databases, web servers and computational models. *Briefings in Bioinformatics*, 25(1): bbad445.

ZMarinka Zitnik, S. M., Rok Sosič; and Leskovec, J. 2017. BioSNAP Datasets: Stanford Biomedical Network Dataset Collection. http://snap.stanford.edu/biodata. Accessed: 2024-06-06.