REASONING OVER RELATIONS: MODELING INTER-RELATIONAL STRUCTURE FOR DRUG-DRUG INTERACTION PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Drug-drug interaction (DDI) prediction is essential for ensuring medication safety, and identifying the specific interaction type between a drug pair is of particular clinical value. Knowledge graphs (KGs) provide rich contextual information for this task. Existing methods typically model relations in KGs as fixed, type-specific representations, without encoding inter-relational structure. This limits the structural information available during entity representation learning. In this paper, we propose the Unconditioned Pair-anchored Relation Reasoning Network (UnPairRelNet), a multi-modal framework for DDI prediction. In particular, UnPairRelNet constructs a graph over relation types and generates relation representations that encode inter-relational structure, without dependence on any specific drug pair query. Subsequently, UnPairRelNet performs pair-conditioned bidirectional propagation anchored at the drug pair on the KG, with drug-pair attention aggregating local neighborhood information. The propagation reduces dependence on entity-specific parameters, naturally accommodating drugs unseen during training. In addition, UnPairRelNet encodes drug molecules at multiple granularities and fuses them with the KG representations for the final prediction. Extensive experiments on two benchmarks demonstrate that UnPairRelNet achieves the best performance on the majority of metrics and surpasses the strongest baseline by over 6% on macro F1 in scenarios involving novel drugs. The code is available at `https://anonymous.4open.science/r/DDI1-03F6`.

1 INTRODUCTION

Multidrug combination therapy has become an important treatment strategy (Webster, 2016; Mokhtari et al., 2017), for complex diseases such as cancer. The concurrent use of multiple drugs increases the probability of drug–drug interactions (DDIs) and raises the risk of harmful adverse reactions (Zhao et al., 2025; ur Rehman et al., 2026). Notably, such concurrent use can increase the risk of adverse drug events from approximately 13% to 82% (Prybys et al., 2002). Therefore, accurately predicting DDIs is crucial for reducing the risks of combination therapy and ensuring its safety and efficacy (Zhao et al., 2024; Luo et al., 2024). This task is further compounded by the continuous approval of new drugs (Handa et al., 2026), which requires prediction models to generalize to compounds unseen during training (Ma et al., 2026; Xu et al., 2026; Zhang et al., 2025a).

Early DDI records were primarily obtained through experimental approaches. However, experimental approaches are often constrained by cost, time, and scale, making it difficult to screen candidate combinations on a large scale (Whitebread et al., 2005). To overcome these limitations, research gradually shifted to computational models. Early computational models (Zhao et al., 2018; Liang et al., 2020) mostly adopted traditional machine learning methods, and their performance depended heavily on handcrafted features. Graph-based methods have since emerged as a widely adopted approach, as both molecular structures and biomedical knowledge naturally exhibit graph topology (Yuan et al., 2025; Zhang et al., 2025b; Li et al., 2026d; Chen et al., 2026; Zhang et al., 2026).

Existing graph-based methods can be broadly categorized into three groups (Su et al., 2024). Molecular graph-based methods (Yan et al., 2025; Wang et al., 2024b; Li et al., 2026b) encode drug molecules through message passing over atomic neighborhoods, capturing local chemical environ-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ments and global topological patterns. These methods derive representations from molecular structure itself, enabling natural generalization to unseen compounds, but operate on isolated molecules without inter-drug context. Knowledge graph-based methods (Abdullahi et al., 2025; Du et al., 2024; Wang et al., 2024a) leverage biomedical knowledge graphs (KGs), where drugs, proteins, diseases, and other entities are connected through heterogeneous typed relations, to enrich drug representations through neighborhood aggregation. Multi-modal methods (Ren et al., 2025; Lin et al., 2026; Li et al., 2026c) combine molecular graphs with KG information through parallel encoding branches and representation fusion, aiming to integrate intra-molecular structure with inter-molecular context.

A key challenge persists across the KG-based and multi-modal methods described above. These methods model relations by fixed, type-specific representations that modulate message passing on edges of each type. Such representations distinguish one relation type from another but do not encode inter-relational structure—how relations interact, co-occur, and organize around shared entities. Entity-level message passing, regardless of its depth, cannot recover such structure from entity-to-entity connections alone. Since relation representations directly modulate how information flows during propagation, the absence of inter-relational structure constrains the quality of the resulting entity representations.

A key observation supports the relevance of this challenge. On both benchmarks, approximately half of relation pairs share at least one entity (55.5% on Ryu's and 46.9% on DrugBank), with over 99% of co-occurrences exceeding random expectation, and specific DDI types exhibit distinct topological signatures. Given the inherent sparsity of knowledge graphs, the presence of such statistically significant structural patterns provides sufficient signal for relational modeling to yield meaningful gains.

To address this challenge, we propose the Unconditioned Pair-anchored Relation Reasoning Network (UnPairRelNet), a multi-modal framework for DDI type prediction. In particular, UnPairRelNet constructs a graph over relation types and generates relation representations that encode inter-relational structure. This process does not depend on any specific drug pair query. Subsequently, UnPairRelNet performs pair-conditioned bidirectional propagation on the KG anchored at the drug pair, with drug-pair attention aggregating local neighborhood information. The propagation pathway reduces dependence on entity-specific parameters, naturally accommodating drugs unseen during training. In addition, UnPairRelNet encodes drug molecules at multiple granularities and fuses them with the KG representations for the final prediction.

Our main contributions are summarized as follows:

- We propose a relational encoding paradigm for DDI prediction. By constructing a graph over relation types and generating representations that encode inter-relational structure, we provide structural context that is absent from fixed relation parameters.
- We introduce a pair-conditioned bidirectional propagation mechanism anchored at the drug pair, with drug-pair attention for local context aggregation. The mechanism reduces dependence on entity-specific parameters and naturally accommodates novel drugs.
- We integrate multi-granularity molecular representations with the KG branch within a unified framework.
- Extensive experiments on benchmarks and in scenarios involving novel drugs demonstrate the effectiveness of UnPairRelNet.

2 Related Work

Molecular graph-based methods. Molecular graph-based approaches represent drugs at the atomic level and learn representations through message passing over molecular topology (Deng et al., 2022; Xia et al., 2026). SSI-DDI (Nyamabo et al., 2021) decomposes DDI prediction into pairwise substructure interactions; GGI-DDI (Yu et al., 2024) identifies pivotal functional groups via granular computing; HLN-DDI (Luo et al., 2025) constructs an atom-motif-molecule hierarchy for multi-scale fusion; and AMIE-DDI (Niu et al., 2026) integrates molecule-based and motif-based views through adaptive fusion.

Knowledge graph-based methods. Biomedical knowledge graphs connect drugs, proteins, diseases, and other entities through heterogeneous typed relations, providing inter-molecular context

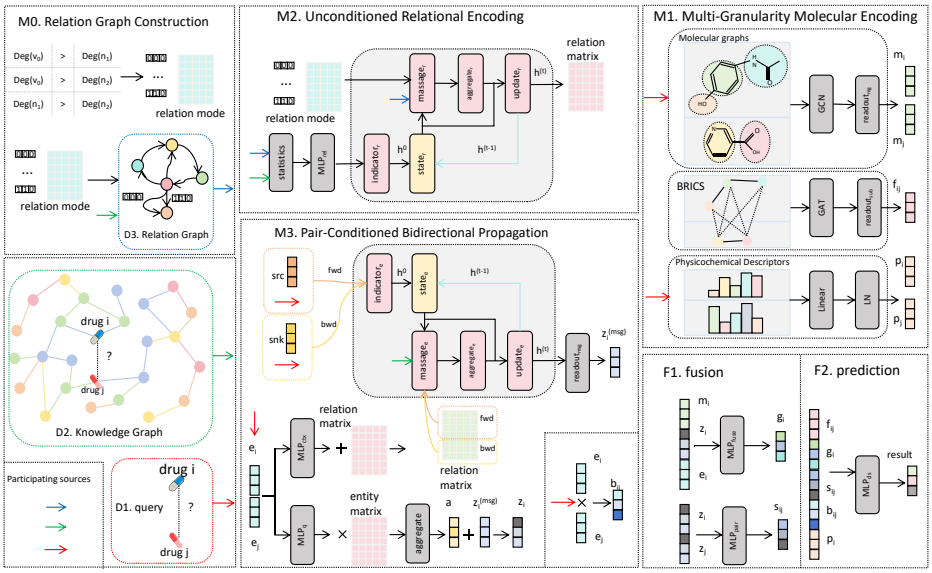


Figure 1: Overview of UnPairRelNet. Each drug is encoded at three granularities; relation types are encoded on a relation graph with degree-partial-order connectivity modes without any query signal; the resulting relation representations modulate pair-conditioned bidirectional propagation on the knowledge graph; all representations are fused for the final prediction.

beyond isolated molecules. KGNN (Lin et al., 2020) extends the receptive field to multiple hops to model high-order topology; LaGAT (Hong et al., 2022) generates pair-specific attention pathways for neighbor selection; and KnowDDI (Wang et al., 2024a) learns a knowledge subgraph for each drug pair.

Multi-modal methods. Recent studies jointly exploit molecular graphs and knowledge graphs to achieve complementarity between intra-molecular structure and inter-molecular context. SumGNN (Yu et al., 2021) anchors local subgraphs from biomedical KGs and integrates multi-channel knowledge and chemical structures for multi-typed DDI prediction; TIGER (Su et al., 2024) employs relation-aware self-attention in a dual-channel framework; MolecBioNet (Chen et al., 2025) integrates KG context with molecular structure through domain-specific pooling. In the broader context of KG reasoning, NBFNet (Zhu et al., 2021) and ULTRA (Galkin et al., 2024) have explored relation-level modeling; see Appendix G for a detailed discussion and comparison with our approach.

3 Methodology

3.1 Problem Formulation

We formulate DDI prediction as a multi-class classification problem. Let $\mathcal{D}$ denote the set of drugs and $\mathcal{R} = \{1, 2, \ldots, |\mathcal{R}|\}$ denote the set of possible interaction types. Given a dataset $\mathcal{S} = \{(d_i, r, d_j) \mid d_i, d_j \in \mathcal{D},\ r \in \mathcal{R}\}$, where each tuple (d_i, r, d_j) indicates that the drug pair (d_i, d_j) exhibits interaction type r, the goal is to learn a mapping $f : \mathcal{D} \times \mathcal{D} \to \mathcal{R}$ that predicts the interaction type for any given drug pair.

Our framework takes as input a biomedical knowledge graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{T})$, where $\mathcal{V}$ is the set of heterogeneous entities (drugs, proteins, diseases, pathways, etc.), $\mathcal{E}$ is the set of typed edges, and $\mathcal{T}$ is the set of relation types. The relations comprise DDI relations $\mathcal{T}_{\text{DDI}}$ connecting drug pairs, and background relations $\mathcal{T}_{\text{BKG}}$ connecting drugs to other entity types. In addition to the KG, each drug $d \in \mathcal{D}$ is associated with its molecular structure and a physicochemical descriptor vector.

We treat edges in $\mathcal{G}$ as directed. Each DDI relation is symmetrized so that its inverse coincides with the relation itself, while each BKG relation is augmented with a distinct inverse relation type; the relation type set $\mathcal{T}$ therefore includes these inverse types. For a directed edge (u, v), we denote its relation type by r_{uv}. The forward edge set comprises the edges of all original relation types, and the backward edge set comprises the augmented inverse BKG edges. Accordingly, $\mathcal{N}_{\text{fwd}}(v)$ and $\mathcal{N}_{\text{bwd}}(v)$ denote the neighbors of v reachable via forward and backward edges, respectively.

Figure 1 provides an overview of UnPairRelNet; the modules below correspond to the labeled panels.

3.2 Multi-Granularity Molecular Encoding

We encode each drug molecule at three granularities: atom-level, substructure-level, and physicochemical.

Atom-level Encoding. The molecular graph of drug d is constructed with atoms as nodes and chemical bonds as edges. Each atom and bond are encoded with a feature vector capturing its chemical properties. A graph neural network (GNN) is applied over the molecular graph, followed by a readout function that aggregates node representations into a graph-level embedding (Li et al., 2025), $\mathbf{m}_d = \text{Readout}(\text{GNN}(\mathcal{G}_d))$.

Substructure-level Encoding. To capture pharmacophore-level structural patterns, we decompose each molecule into chemically meaningful substructures using the Break of Retrosynthetically Interesting Chemical Substructures (BRICS) algorithm (Li et al., 2026a). Each substructure is represented by its Morgan fingerprint, and substructures are connected if they share a chemical bond in the original molecule. For a drug pair (d_i, d_j), we construct a heterogeneous interaction graph with two edge types: *intra-molecular* edges connecting substructures within the same molecule, and *inter-molecular* edges fully connecting substructures of d_i to those of d_j. A heterogeneous graph neural network is applied, where intra-molecular edges use standard aggregation and inter-molecular edges use multi-head attention, $\mathbf{f}_{ij} = \text{AttnReadout}(\text{HeteroGNN}(\mathcal{G}_{\text{sub}}^{(i,j)}))$, where the readout is a mean pooling weighted by the learned inter-molecular attention scores.

Physicochemical Descriptors. The descriptor vector $\mathbf{x}_d^{\text{desc}}$ is projected through a multi-layer perceptron (MLP) with layer normalization (LN), $\mathbf{p}_d = \text{LN}(\text{ReLU}(\mathbf{W}_p \mathbf{x}_d^{\text{desc}}))$. Physicochemical descriptors provide global molecular properties complementary to the local structural information captured by atom-level and substructure-level encodings.

3.3 Relation Graph Construction

Rather than treating relations as edge-type annotations, we construct a dedicated graph over relation types to generate their semantic representations. The relation graph $\mathcal{G}_{\text{rel}}$ has one node for each relation type $r \in \mathcal{T}$, and edges encode structural co-occurrence patterns between relations.

Two relations r_1 and r_2 are connected in $\mathcal{G}_{\text{rel}}$ if they share at least one entity in the KG. Let v_0 be a shared entity connecting r_1 and r_2, and let n_1 and n_2 denote the respective neighbors of v_0 under r_1 and r_2. The connection is assigned a degree-partial-order mode $m \in \{0, \ldots, 7\}$:

$$m = 4 \cdot \mathbf{1}[\deg(v_0) > \deg(n_1)] + 2 \cdot \mathbf{1}[\deg(v_0) > \deg(n_2)] + \mathbf{1}[\deg(n_1) > \deg(n_2)]. \quad (1)$$

Of the 8 possible modes, 6 are logically valid; the modes $m = 2$ and $m = 5$ (bit patterns $(0, 1, 0)$ and $(1, 0, 1)$) are impossible due to transitivity of degree ordering. Simple co-occurrence encoding records only whether two relations share an entity. The degree-partial-order encoding further captures the topological asymmetry of the connection: whether the shared entity occupies a structurally dominant or subordinate position relative to its neighbors. This distinction ensures that information flowing through different structural contexts is differentiated even when the same pair of relations is involved.

Each edge in $\mathcal{G}_{\text{rel}}$ is weighted by the inverse out-degree of its source node, $w(r_1, r_2) = 1/|\{r' : (r_1, r') \in \mathcal{E}_{\text{rel}}\}|$, to normalize the contribution of highly connected relations.

3.4 UNCONDITIONED RELATIONAL ENCODING

Given the relation graph $\mathcal{G}_{\text{rel}}$, all relation types are encoded simultaneously through propagation on $\mathcal{G}_{\text{rel}}$. Since the target relation in DDI type prediction is unknown at inference time, the encoding is performed without any query-specific signal.

All relation nodes are initialized with a structural prior derived from relation-level statistics: $\mathbf{h}_r^{(0)} = \text{MLP}_{\text{init}}(\mathbf{s}_r)$, where $\mathbf{s}_r$ is a vector of structural statistics for relation r. The encoding proceeds through L layers of message passing on $\mathcal{G}_{\text{rel}}$. Each edge is associated with a learnable type embedding $\boldsymbol{\mu}_{r,r'}$ corresponding to its degree-partial-order mode:

$$\mathbf{h}_r^{(l+1)} = \text{Update}_{\text{rel}}\Big(\mathbf{h}_r^{(l)},\ \text{Agg}_{\text{rel},\, r' \in \mathcal{N}(r)}\big(w(r', r) \cdot (\mathbf{h}_{r'}^{(l)} \odot \boldsymbol{\mu}_{r',r})\big)\Big), \tag{2}$$

where $\odot$ denotes element-wise multiplication, $w(r', r)$ is the edge weight defined above, and Agg_{rel} is mean aggregation. After L layers, the output is projected to obtain the relation representation matrix $\mathbf{R} = \text{Proj}_{\text{rel}}(\mathbf{H}^{(L)}) \in \mathbb{R}^{|\mathcal{T}| \times d_r}$, where $\mathbf{H}^{(L)}$ stacks the final-layer states $\mathbf{h}_r^{(L)}$ over all $r \in \mathcal{T}$, and $\mathbf{R}[r]$ denotes the representation of relation r. Since the encoding does not depend on any specific drug pair, the resulting representations are shared across all drug pairs.

3.5 PAIR-CONDITIONED BIDIRECTIONAL PROPAGATION

Given the relation representations $\mathbf{R}$ obtained from the relation graph and a drug pair (d_i, d_j), context-aware drug representations are generated through bidirectional propagation on the KG.

The representations $\mathbf{R}$ are derived from the global topological structure of the relation graph and are not anchored to any specific drug pair. To anchor the propagation to the drug pair under consideration, we inject a pair context derived from drug embeddings, $\mathbf{c}_{ij} = \text{MLP}_{\text{ctx}}([\mathbf{e}_i \| \mathbf{e}_j])$, where $\mathbf{e}_i, \mathbf{e}_j$ are learned drug embeddings used for conditioning, pair attention, and fusion, but not as propagation states. For drugs unseen during training, $\mathbf{e}$ is replaced by a projection of the physicochemical descriptor vector, ensuring that the conditioning remains available for novel drugs. The pair-conditioned relation representations are $\mathbf{R}^{(ij)} = \mathbf{R} + \mathbf{c}_{ij}$.

Propagation proceeds on the KG in two directions. Entity nodes are initialized to zero except the node d_i, which receives the shared source signal, i.e., $\mathbf{h}_{d_i}^{(0)} = \mathbf{g}_{\text{src}}$ and $\mathbf{h}_v^{(0)} = \mathbf{0}$ for $v \neq d_i$. In the forward direction, messages propagate along forward edges:

$$\mathbf{h}_v^{(l+1)} = \text{Update}_{\text{ent}}\Big(\mathbf{h}_v^{(l)},\ \text{Agg}_{\text{ent},\, u \in \mathcal{N}_{\text{fwd}}(v)}\big(\mathbf{h}_u^{(l)} \odot \mathbf{R}^{(ij)}[r_{uv}]\big)\Big), \tag{3}$$

where $\mathbf{R}^{(ij)}[r_{uv}]$ retrieves the pair-conditioned representation of the relation type of the traversed edge. In the backward direction, the shared sink signal $\mathbf{g}_{\text{snk}}$ is injected at node d_j, and messages propagate along backward edges, modulated by the corresponding relation representations.

The drug representations are obtained by extracting and concatenating features at both anchor positions from both directions, $\mathbf{z}_i = [\mathbf{H}_{\text{fwd}}[d_i] \| \mathbf{H}_{\text{bwd}}[d_i]]$ and $\mathbf{z}_j = [\mathbf{H}_{\text{fwd}}[d_j] \| \mathbf{H}_{\text{bwd}}[d_j]]$. This bidirectional design ensures that each drug's representation captures both outgoing context and incoming context, naturally respecting the symmetric nature of DDI relations.

Note that the source/sink signals, relation representations, and GNN weights are shared across all entities; no entity-specific parameters participate in the propagation itself.

To further enrich the drug representations with local neighborhood information, we apply a drug-pair attention mechanism. For each drug d_i, its one-hop drug neighbors are aggregated using an attention query derived from the drug pair:

$$\mathbf{q}_{ij} = \text{MLP}_q([\mathbf{e}_i \| \mathbf{e}_j]), \quad \alpha_k = \text{softmax}\Big(\frac{\mathbf{q}_{ij}^\top \mathbf{W}_k \mathbf{e}_k}{\sqrt{d_e}}\Big), \tag{4}$$

$$\mathbf{a}_i = \sum_{k \in \mathcal{N}(d_i)} \alpha_k \mathbf{e}_k. \tag{5}$$

The aggregated neighbor information is projected and added to the propagation output via a residual connection:

$$\mathbf{z}_i \leftarrow \mathbf{z}_i + \text{Proj}_{\text{attn}}(\mathbf{a}_i), \quad \mathbf{z}_j \leftarrow \mathbf{z}_j + \text{Proj}_{\text{attn}}(\mathbf{a}_j). \tag{6}$$

3.6 FUSION AND PREDICTION

The KG propagation output captures inter-drug structural context, while the molecular encodings capture intra-drug chemical structure. We fuse these information sources for the final prediction.

For each drug in the pair, we fuse molecular and KG information, $\mathbf{g}_i = \text{MLP}_{\text{fuse}}([\mathbf{m}_i \| \mathbf{z}_i \| \mathbf{e}_i])$ and likewise for $\mathbf{g}_j$. A bilinear interaction between drug embeddings is additionally computed to capture pair-level compatibility, $\mathbf{b}_{ij} = \mathbf{e}_i^\top \mathbf{W}_{\text{bil}} \mathbf{e}_j$.

An auxiliary classification head predicts the relation type directly from the KG propagation output, $\hat{r}_{\text{aux}} = \mathbf{W}_{\text{aux}} \cdot \text{MLP}_{\text{pair}}([\mathbf{z}_i \| \mathbf{z}_j])$. The auxiliary prediction provides an additional supervisory signal that encourages the propagation to produce relation-discriminative features.

All available representations are concatenated for the final classification:

$$\mathbf{o} = [\mathbf{f}_{ij} \| \mathbf{s}_{ij} \| \mathbf{g}_i \| \mathbf{g}_j \| \mathbf{b}_{ij} \| \mathbf{p}_i \| \mathbf{p}_j], \tag{7}$$

where $\mathbf{f}_{ij}$ is the substructure-level representation, $\mathbf{s}_{ij} = \text{MLP}_{\text{pair}}([\mathbf{z}_i \| \mathbf{z}_j])$ is the pair-level KG summary, and $\mathbf{p}_i, \mathbf{p}_j$ are the physicochemical descriptor projections. The prediction is:

$$\hat{r} = \text{MLP}_{\text{cls}}(\mathbf{o}). \tag{8}$$

The focal loss addresses the extreme class imbalance in DDI type distributions:

$$\mathcal{L}_{\text{focal}} = -\frac{1}{B} \sum_{(d_i, r, d_j) \in \mathcal{B}} (1 - \pi_r)^\gamma \log(\pi_r), \tag{9}$$

where $\mathcal{B}$ is a mini-batch of drug pairs and r is the true interaction type of each pair. The auxiliary loss employs class weights:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{B} \sum_{(d_i, r, d_j) \in \mathcal{B}} \omega_r \log(\hat{r}_{\text{aux},r}), \quad \omega_r \propto \frac{1}{\sqrt{n_r}}, \tag{10}$$

where n_r is the number of training samples of relation r. The total loss combines a focal loss for the main classification with a cross-entropy loss for the auxiliary prediction:

$$\mathcal{L} = \mathcal{L}_{\text{focal}}(\hat{r}, r) + \lambda \cdot \mathcal{L}_{\text{CE}}(\hat{r}_{\text{aux}}, r). \tag{11}$$

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets. We evaluate UnPairRelNet on two benchmarks for multi-type DDI prediction: the Ryu's dataset (Ryu et al., 2018) and the DrugBank dataset (Chen et al., 2025), the latter being larger and more complex. All experiments, on both benchmarks and under the two novel-drug settings, are conducted under 5-fold cross-validation.

Evaluation Metrics. We report four metrics: accuracy (ACC), macro F1, macro PR-AUC, and Cohen's κ. For brevity, "macro" is omitted in table headers hereafter.

Baselines. We compare UnPairRelNet against nine representative baselines spanning three categories: molecular graph-based methods (DeepDDI(Ryu et al., 2018), SSI-DDI(Nyamabo et al., 2021), GGI-DDI(Yu et al., 2024)), KG-based methods (KGNN(Lin et al., 2020), LaGAT(Hong et al., 2022), KnowDDI(Wang et al., 2024a)), and multi-modal methods (SumGNN(Yu et al., 2021), TIGER(Su et al., 2024), MolecBioNet(Chen et al., 2025)).

Novel-Drug Settings. To evaluate generalization to drugs unseen during training, we adopt two settings derived from the Ryu's dataset: *S1*, where each test triplet contains one drug that appears in the training set and one that does not; and *S2*, where neither drug in the test triplet appears in the training set. Under both settings, the KG topology at inference differs from that at training.

Dataset statistics, implementation details, etc. are provided in Appendices A and E.

Table 1: Performance comparison on Ryu's and DrugBank datasets. Best results in **bold**, second best underlined.

Methods	Ryu's Dataset				DrugBank Dataset			
	ACC	F1	PR-AUC	Cohen's κ	ACC	F1	PR-AUC	Cohen's κ
DeepDDI	0.868	0.703	0.811	0.842	0.862	0.676	0.810	0.848
SSI-DDI	0.848	0.618	0.704	0.818	0.743	0.327	0.438	0.717
GGI-DDI	0.889	0.811	0.880	0.869	0.859	0.762	0.836	0.846
KGNN	0.916	0.814	0.872	0.900	0.907	0.764	0.832	0.898
SumGNN	0.904	0.842	0.912	0.886	0.864	0.732	0.855	0.851
LaGAT	0.935	0.843	0.899	0.922	0.900	0.741	0.829	0.891
KnowDDI	0.923	0.880	0.934	0.909	0.894	0.772	0.868	0.885
TIGER	0.938	0.880	0.894	0.926	0.917	0.791	0.763	0.910
MolecBioNet	0.952	**0.912**	0.942	0.943	0.935	**0.846**	0.910	0.929
Ours	**0.956**	0.883	**0.951**	**0.948**	**0.954**	0.830	**0.937**	**0.949**

4.2 Performance Comparison

To evaluate the effectiveness of UnPairRelNet on DDI type prediction, we compare it against representative baselines on both benchmarks. The results are summarized in Table 1. Results with standard deviations over 5 folds are provided in Appendix H.

On the Ryu's dataset, UnPairRelNet achieves the best performance on accuracy, macro PR-AUC, and Cohen's κ, and ranks second on macro F1. On the DrugBank dataset, UnPairRelNet achieves the best accuracy, macro PR-AUC, and Cohen's κ, surpassing the strongest baseline by 1.9%, 2.7%, and 2.0%, respectively, while ranking second on macro F1.

Two trends emerge from the comparison. *First*, the performance advantage of UnPairRelNet over baselines is more pronounced on DrugBank than on Ryu's: the margins on accuracy, macro PR-AUC, and Cohen's κ are all larger, and the gap to the best macro F1 is smaller. This suggests that the relational modeling capability of UnPairRelNet yields greater benefits as the number of interaction types and the complexity of the KG increase. *Second*, single-modality baselines are generally outperformed by multi-modal baselines, and UnPairRelNet further improves upon existing multi-modal methods.

These results indicate that the inter-relational structure modeling employed by UnPairRelNet provides structural information absent from the fixed, type-specific relation parameters used by existing baselines. This information guides propagation to produce more accurate entity representations, leading to the observed performance gains.

4.3 Generalization to Novel Drugs

Table 2: Performance comparison on S1 and S2 settings. Best results in **bold**.

Methods	S1 setting				S2 setting			
	ACC	F1	PR-AUC	Cohen's κ	ACC	F1	PR-AUC	Cohen's κ
DeepDDI	0.652	0.536	0.592	0.580	0.394	0.239	0.202	0.269
SSI-DDI	0.597	0.538	0.580	0.514	0.389	0.245	0.204	0.251
GGI-DDI	0.642	0.531	0.573	0.566	0.394	0.230	0.187	0.266
KGNN	0.392	0.119	0.115	0.267	0.151	0.018	0.016	0.014
SumGNN	0.571	0.420	0.498	0.476	0.396	0.140	0.178	0.204
LaGAT	0.511	0.289	0.314	0.399	0.201	0.015	0.015	0.014
KnowDDI	0.518	0.296	0.351	0.415	0.300	0.016	0.061	0.053
TIGER	0.311	0.110	0.116	0.159	0.268	0.068	0.061	0.072
MolecBioNet	0.625	0.563	0.591	0.544	0.395	0.233	0.199	0.264
Ours	**0.673**	**0.625**	**0.674**	**0.605**	**0.415**	**0.313**	**0.251**	**0.280**

In practical clinical settings, new drugs are continuously approved and enter the market. This requires DDI prediction models to generalize to drugs unseen during training. From the perspective

of the KG, the introduction of novel drugs implies that the graph topology at inference differs significantly from that at training, posing a fundamental challenge for methods that rely on the KG. To evaluate the generalization capability of UnPairRelNet in this scenario, we conduct experiments under the S1 and S2 settings. The results are summarized in Table 2. Results with standard deviations over 5 folds are provided in Appendix H.

In the S1 setting, UnPairRelNet achieves the best performance on all four metrics, surpassing the strongest baseline by 2.1% on accuracy, 6.2% on macro F1, 8.2% on macro PR-AUC, and 2.5% on Cohen's κ. In the S2 setting, UnPairRelNet likewise achieves the best performance on all four metrics, surpassing the strongest baseline by 1.9% on accuracy, 6.8% on macro F1, 4.7% on macro PR-AUC, and 1.1% on Cohen's κ.

A striking observation from Table 2 is that molecular graph-based baselines substantially outperform KG-based baselines on most metrics. This reveals a fundamental limitation of existing KG-based methods when facing novel drugs. These methods rely on entity-specific learnable parameters, which remain untrained for drugs unseen during training and cannot provide meaningful representations. Simultaneously, the altered graph topology shifts the message passing context for all entities, further degrading representation quality.

UnPairRelNet overcomes this limitation through two properties of its KG branch. First, the relation representations are derived from the global topological structure of the relation graph rather than from any specific entity: the same unconditioned encoding function is applied to the inference-time graph, so the representations track the current topology by construction instead of relying on parameters frozen at training time. Second, the propagation pathway involves no entity-specific parameters and therefore applies to unseen drugs without retraining; where the conditioning and attention inputs require drug embeddings, unseen drugs fall back to descriptor projections as described in Section 3.5. The molecular graph branch adds structure-based representations available for every drug. Together, these designs allow the model to retain both information sources when the graph topology changes significantly.

Notably, UnPairRelNet not only outperforms KG-based baselines but also substantially surpasses molecular graph-based and multi-modal baselines. On macro F1, UnPairRelNet exceeds the strongest baseline by 6.2% (S1) and 6.8% (S2), suggesting that the structural context supplied by the KG branch complements molecular structure rather than being redundant with it, even in scenarios involving novel drugs.

4.4 Ablation Study

Table 3: Ablation study on the Ryu's dataset. ↓ indicates the decrease in macro F1 relative to the full model. Best results in **bold**. MG: molecular graph; Uncond.: unconditioned. Rows above the rule add components to the molecular-graph-only base; rows below remove one component from the full model.

Variant	ACC	Macro-F1	PR-AUC	Cohen's κ
MG Only	0.905	0.815	0.907	0.888
+ Uncond. Propagation	0.944	0.833	0.927	0.934
UnPairRelNet (Full)	**0.953**	**0.878**	**0.942**	**0.944**
w/o Descriptors	0.951	0.843 (↓0.035)	0.940	0.942
w/o Pair Attn & Context	0.948	0.853 (↓0.025)	0.938	0.938
w/o KG Propagation	0.946	0.847 (↓0.031)	0.937	0.936

To understand the contribution of each component, we conduct an ablation study on the Ryu's dataset with macro F1 as the primary metric, since it is most sensitive to minority interaction types. We evaluate two types of configurations: variants that add components to a molecular-graph-only base (MG Only), and variants that remove a single component from the full model, denoted by "w/o". All variants share the same hyperparameters and the same molecular graph encoder architecture. The results are summarized in Table 3.

The full model achieves a macro F1 of 0.878, surpassing the molecular-graph-only variant (0.815) by 0.063, showing that the KG branch provides complementary information beyond molecular structure alone. Among the components removed from the full model, physicochemical descriptors contribute

the most (-0.035), followed by KG propagation (-0.031) and pair attention with context (-0.025); per-component discussion is provided in Appendix I. The "w/o KG Propagation" variant disables only multi-hop message passing, retaining entity embeddings and their bilinear interaction as a minimal pairwise signal.

4.5 Representation Diagnostic via Embedding Norm Analysis

Table 4: Representative outlier drugs identified by UnPairRelNet (degree < 10, embedding norm > 0.6). All identities confirmed via PubChem. PB = protein binding. HSDB = Hazardous Substances Data Bank.

Drug Name	Deg.	Norm	Chemical Class	Key Evidence
Sulfamethizole	1	0.908	Sulfonamide	PB 98–99%; 12 HSDB
Dexrazoxane	5	0.764	Bis(2,5-dioxopiperazine)[†]	DDI records; PB <2%
Propylthiouracil	6	0.688	Antithyroid (thiourea)	Inhibits T3/T4; PB 82%
Carbimazole	6	0.683	Antithyroid (imidazole)	Same use as PTU; PB 85%
Chlorambucil	9	0.616	Bis-chloroethylamine	PB 99%; 8 HSDB

[†]Identified from molecular structure.

Beyond predictive accuracy, we investigate whether the learned representations encode drug-level signals that cannot be explained by graph topology alone. We compute the ℓ_2 norm of each drug's learned embedding and examine its relationship with graph degree.

Embedding norm and degree exhibit a moderate positive correlation (Spearman $\rho = 0.42$), indicating that the model encodes topological information. However, 12 drugs with minimal interaction records (average degree 3.8) exhibit anomalously high norms (> 0.6), deviating markedly from the overall degree-to-norm trend. Through exact chemical identification via PubChem, five of these outliers correspond to drugs with documented pharmacological annotations, spanning diverse chemical scaffolds (Table 4).

Three observations support the interpretation that these elevated norms reflect genuine representational content rather than artifacts. *First*, the outliers form a coherent cluster in embedding space (mean pairwise cosine 0.209 vs. random baseline 0.010, a 20-fold enrichment), yet do not align with known high-degree hubs (cosine ≈ -0.1, indistinguishable from controls), indicating a pattern distinct from topological degree. *Second*, the elevated norms are not attributable to background KG connectivity (outlier mean 580 edges vs. control 926), ruling out background KG topology as the source. *Third*, four of the five outliers listed in Table 4 exhibit elevated protein binding rates (82–99%; dexrazoxane being the exception at <2%), suggesting that physicochemical properties may contribute to the elevated norms.

To attribute this beyond-degree content to a specific branch, we repeat the diagnostic on the checkpoint of the variant without KG propagation (Section 4.4), which retains the molecular encoder, physicochemical descriptors, and pair attention while removing multi-hop propagation. In this variant, the degree–norm correlation strengthens (Spearman $\rho = 0.68$ vs. 0.42), and none of the 12 outliers remains elevated (mean norm percentile 13 vs. 97), indicating that the elevated norms require the propagation pathway and do not arise from the molecular or descriptor inputs alone.

These results indicate that the learned representations are not reducible to topological degree, and that the beyond-degree component is supplied by the KG propagation pathway rather than by topological memorization or by the molecular inputs alone. Detailed identification results, pharmacological annotations, and the per-drug control comparison are provided in Appendix F.

5 Conclusion

We presented UnPairRelNet, a multi-modal framework for drug-drug interaction type prediction that encodes relation types through unconditioned propagation on a relation graph with degree-partial-order connectivity patterns, and modulates bidirectional knowledge-graph propagation with the resulting relation representations. Experiments on two benchmarks show the best performance on the majority of metrics, with larger margins on the more complex benchmark, and the best performance on all metrics when generalizing to novel drugs.

AI USE STATEMENT

In this work, we used generative AI tools to assist with and polish the writing, to support literature retrieval and the discovery of related work, and to aid research ideation and execution. All AI-assisted text, literature suggestions, research ideas, and outputs were critically reviewed and verified by the authors. The authors independently checked the accuracy and relevance of suggested references, validated AI-assisted research outputs, and revised the resulting content as necessary. We take full responsibility for the final content of this work, including all text, claims, citations, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Our source code and the data used in our experiments are available at `https://anonymous.4open.science/r/DDI1-03F6`.

REFERENCES

Tassallah Abdullahi, Ioanna Gemou, Nihal V Nayak, Ghulam Murtaza, Stephen H Bach, Carsten Eickhoff, and Ritambhara Singh. K-paths: Reasoning over graph paths for drug repurposing and drug interaction prediction. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 5–16, 2025.

Jing Chen, Qiang Deng, Peimeng Zhen, Jialu Hu, Yongtian Wang, Jiajie Peng, Zhuhong You, Xuequn Shang, Xu Zhang, and Tao Wang. Knowledge-graph-enhanced multi-scale modeling for drug-drug interaction prediction. *Molecular Therapy Nucleic Acids*, 37(1), 2026.

Mengjie Chen, Ming Zhang, and Cunquan Qu. Towards interpretable drug-drug interaction prediction: A graph-based approach with molecular and network-level explanations. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 203–214, 2025.

Yifan Deng, Yang Qiu, Xinran Xu, Shichao Liu, Zhongfei Zhang, Shanfeng Zhu, and Wen Zhang. META-DDIE: predicting drug–drug interaction events with few-shot learning. *Briefings in Bioinformatics*, 23(1):bbab514, 2022.

Haotong Du, Quanming Yao, Juzheng Zhang, Yang Liu, and Zhen Wang. Customized subgraph selection and encoding for drug-drug interaction prediction. In *Advances in Neural Information Processing Systems*, volume 37, pp. 109582–109608, 2024.

Mikhail Galkin, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu. Towards foundation models for knowledge graph reasoning. In *International Conference on Learning Representations*, volume 2024, pp. 31598–31619, 2024.

Koichi Handa, Hideaki Mamada, Shinji Nakayama, Tadahaya Mizuno, Hiroaki Iwata, Tsuyoshi Esaki, and Yohei Kosugi. Advances in in silico predictive models for DDI prediction: implications and practical applications in drug discovery. *Drug Metabolism and Pharmacokinetics*, pp. 101532, 2026.

Yue Hong, Pengyu Luo, Shuting Jin, and Xiangrong Liu. LaGAT: link-aware graph attention network for drug–drug interaction prediction. *Bioinformatics*, 38(24):5406–5412, 2022.

Kefei Li, Jianbo Qiao, Yuntao Yang, Hanjun Zhao, Ding Wang, Junru Jin, Zhongmin Yan, and Leyi Wei. FGSGT-DDI: An LLM-enhanced functional group semantic graph transformer for drug–drug interaction prediction. *Journal of Chemical Information and Modeling*, 2026a.

Mengli Li, Chao Cao, Quan Zou, Leyi Wei, and Yansu Wang. MHAFR-DDI: a multimodal hierarchical attention fusion and relation-aware architecture for drug–drug interaction event prediction. *Briefings in Bioinformatics*, 27(1):bbag077, 2026b.

Wenjun Li, Yiting Zhou, Wanjun Ma, Weijun Liang, and Xiwei Tang. MDG-DDI: multi-feature drug graph for drug-drug interaction prediction. *BMC Bioinformatics*, 26(1):268, 2025.

Yu Li, Jia-Ming Liu, Yue-Chao Li, Hai-Ru You, Zhu-Hong You, Yang Yuan, and Cheng-Gang Mi. KA-DDI: A knowledge-adaptive contrastive learning framework for drug-drug interaction prediction. *Expert Systems with Applications*, pp. 131076, 2026c.

Yuanxian Li, Yuan Du, Hong Peng, Zhenli He, Xin Jin, and Cheng Xie. GIL-DDI: multi-view graph invariant learning for unknown drug–drug interaction prediction. *Knowledge and Information Systems*, 68(1):48, 2026d.

Haiyan Liang, Lei Chen, Xian Zhao, and Xiaolin Zhang. Prediction of drug side effects with a refined negative sample selection strategy. *Computational and Mathematical Methods in Medicine*, 2020(1):1573543, 2020.

Xiaoli Lin, Siyuan Zhang, Bo Li, Ping Liang, and Jun Pang. HSAF-DDI: Heterogeneous semantic-aware drug–drug interaction prediction with hierarchical adaptive fusion. *Journal of Chemical Information and Modeling*, 66(16):10379–10395, 2026.

Xuan Lin, Zhe Quan, Zhi-Jie Wang, Tengfei Ma, Xiangxiang Zeng, et al. KGNN: Knowledge graph neural network for drug-drug interaction prediction. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, volume 380, pp. 2739–2745, 2020.

Huimin Luo, Weijie Yin, Jianlin Wang, Ge Zhang, Wenjuan Liang, Junwei Luo, and Chaokun Yan. Drug-drug interactions prediction based on deep learning and knowledge graph: A review. *iScience*, 27(3), 2024.

Yue Luo, Lei Deng, and Zhijian Huang. HLN-DDI: hierarchical molecular representation learning with co-attention mechanism for drug-drug interaction prediction. *BMC Bioinformatics*, 26(1): 152, 2025.

Tengfei Ma, Kun Chen, Yongsheng Zang, Yujie Chen, Xuanbai Ren, Bosheng Song, Hongxin Xiang, Yiping Liu, et al. Self-supervised blending structural context of visual molecules for robust drug interaction prediction. In *Advances in Neural Information Processing Systems*, volume 38, pp. 133133–133161, 2026.

Reza Bayat Mokhtari, Tina S Homayouni, Narges Baluch, Evgeniya Morgatskaya, Sushil Kumar, Bikul Das, and Herman Yeger. Combination therapy in combating cancer. *Oncotarget*, 8(23): 38022, 2017.

Dongjiang Niu, Xiaofeng Wang, Zengqian Deng, Bowen Tang, and Zhen Li. Interaction-aware adaptive network for drug-drug interaction prediction. *Pattern Recognition*, pp. 113172, 2026.

Arnold K Nyamabo, Hui Yu, and Jian-Yu Shi. SSI–DDI: substructure–substructure interactions for drug–drug interaction prediction. *Briefings in Bioinformatics*, 22(6):bbab133, 2021.

K Prybys, A Gee, et al. Polypharmacy in the elderly: clinical challenges in emergency practice. *Part 1 Overview, Etiology, and Drug Interactions, Emergency Medicine Reports*, 23(11):145–151, 2002.

Zhonghao Ren, Xiangxiang Zeng, Yizhen Lao, Zhuhong You, Yifan Shang, Quan Zou, and Chen Lin. Predicting rare drug-drug interaction events with dual-granular structure-adaptive and pair variational representation. *Nature Communications*, 16(1):3997, 2025.

Jae Yong Ryu, Hyun Uk Kim, and Sang Yup Lee. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the National Academy of Sciences*, 115(18):E4304–E4311, 2018.

Xiaorui Su, Pengwei Hu, Zhu-Hong You, Philip S Yu, and Lun Hu. Dual-channel learning framework for drug-drug interaction prediction via relation-aware heterogeneous graph transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 249–256, 2024.

Faisal Asad ur Rehman, Abdulkarim Kanaan Jebna, Touqeer Ahmad, Arif Ur Rahman, and Fasee Ullah. Prediction of drug–drug interactions based on artificial intelligence: A systematic literature review. *Artificial Intelligence and Applications*, 4(3):280–296, 2026.

Yaqing Wang, Zaifei Yang, and Quanming Yao. Accurate and interpretable drug-drug interaction prediction enabled by knowledge subgraph learning. *Communications Medicine*, 4(1):59, 2024a.

Ziyan Wang, Zhankun Xiong, Feng Huang, Xuan Liu, and Wen Zhang. ZeroDDI: a zero-shot drug-drug interaction event prediction method with semantic enhanced learning and dual-modal uniform alignment. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI '24*, pp. 6071–6079, 2024b. doi: 10.24963/ijcai.2024/671.

Rachel M Webster. Combination therapies in oncology. *Nature Reviews Drug Discovery*, 15(2): 81–83, 2016.

Steven Whitebread, Jacques Hamon, Dejan Bojanic, and Laszlo Urban. Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. *Drug Discovery Today*, 10(21):1421–1433, 2005.

Yan Xia, Yazi Li, Shuting Jin, Junlin Xu, Yajie Meng, Leyi Wei, Zilong Zhang, Xin Gao, and Feifei Cui. MFDL-DDI: An effective deep learning-based framework for predicting drug-drug interactions through multimodal information fusion. *Pattern Recognition*, pp. 113431, 2026.

Dong Xu, Jiantao Wu, Qihua Pan, Sisi Yuan, Zexuan Zhu, and Junkai Ji. Rethinking drug-drug interaction modeling as generalizable relation learning. arXiv preprint arXiv:2601.15771, 2026.

Zimo Yan, Jie Zhang, Zheng Xie, Yiping Song, and Hao Li. A multi-scale graph neural process with cross-drug co-attention for drug-drug interactions prediction. arXiv preprint arXiv:2509.15256, 2025.

Hui Yu, Jing Wang, Shi-Yu Zhao, Omayo Silver, Zun Liu, JingTao Yao, and Jian-Yu Shi. GGI-DDI: Identification for key molecular substructures by granule learning to interpret predicted drug–drug interactions. *Expert Systems with Applications*, 240:122500, 2024.

Yue Yu, Kexin Huang, Chao Zhang, Lucas M Glass, Jimeng Sun, and Cao Xiao. SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics*, 37(18):2988–2995, 2021.

Yongna Yuan, Jiaqi Yue, Ruisheng Zhang, and Wei Su. PHGL-DDI: A pre-training based hierarchical graph learning framework for drug-drug interaction prediction. *Expert Systems with Applications*, 270:126408, 2025.

Jiahui Zhang, Shuai Zhang, Xuqiang Li, Di Wu, Sihan Wang, Limin Li, Wenjie Du, and Yang Wang. IIB-DDI: Invariant information bottleneck theory for out-of-distribution drug-drug interaction prediction. *IEEE Transactions on Computational Biology and Bioinformatics*, 22(3): 1009–1022, 2025a.

Ran Zhang, Xuezhi Wang, Guannan Liu, Pengyang Wang, Yuanchun Zhou, and Pengfei Wang. Motif-oriented representation learning with topology refinement for drug-drug interaction prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 1102–1110, 2025b.

Yongqing Zhang, Yugui Xu, Zhigan Zhou, Hao Yuan, Jin Wu, Quan Zou, and Lei Xu. MCFusion-DDI: Multimodal cross-attention fusion of local-global features and latent drug associations for explainable DDI prediction. *Neural Networks*, pp. 109278, 2026.

Haochen Zhao, Jian Zhong, Xiao Liang, Chenliang Xie, and Shaokai Wang. Application of machine learning in drug side effect prediction: databases, methods, and challenges. *Frontiers of Computer Science*, 19(5):195902, 2025.

Xian Zhao, Lei Chen, and Jing Lu. A similarity-based method for prediction of drug side effects with heterogeneous information. *Mathematical Biosciences*, 306:136–144, 2018.

Yan Zhao, Jun Yin, Li Zhang, Yong Zhang, and Xing Chen. Drug–drug interaction prediction: databases, web servers and computational models. *Briefings in Bioinformatics*, 25(1):bbad445, 2024.

Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal Xhonneux, and Jian Tang. Neural Bellman-Ford networks: A general graph neural network framework for link prediction. In *Advances in Neural Information Processing Systems*, volume 34, pp. 29476–29490, 2021.

A Dataset Statistics

A.1 Notation

Table 5 summarizes the main symbols used throughout the paper.

A.2 Overview

Table 6 summarizes the key statistics of the two benchmarks. Both datasets share the same background knowledge graph (BKG) schema but differ in the number of drugs, DDI relation types, and triplet counts. DrugBank contains more DDI relation types (221 vs. 86) and 3.2 times as many DDI triplets as Ryu's.

A.3 Entity Type Distribution

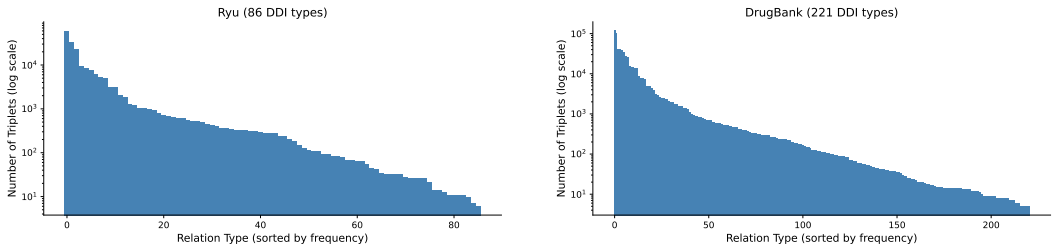
Table 7 reports the entity type distribution over all nodes of the shared schema. The two datasets differ only in the number of drug nodes (178,418 vs. 178,415).

A.4 BKG Relation Types

Table 8 lists the relation types stored in the BKG edge file. Relation ID 0 corresponds to the DDI target relation, which is stored together with the 11 background relation types (IDs 1–11); the background relation set $\mathcal{T}_{\text{BKG}}$ therefore comprises 11 types, and the relation-type count of 12 in Table 6 includes the DDI edges.

A.5 DDI Type Distribution

Both datasets exhibit a pronounced long-tail distribution over DDI relation types, as shown in Figure 2. On the Ryu's dataset, the most frequent type accounts for 60,936 triplets while the least frequent has only 6. On DrugBank, the disparity is even larger (119,017 vs. 5). This imbalance motivates the use of macro-averaged metrics and focal loss in our framework.



(a) Ryu's (86 DDI types) (b) DrugBank (221 DDI types)

Figure 2: Long-tail distribution of DDI relation types. The x-axis represents relation types sorted by frequency in descending order; the y-axis shows the number of triplets per type (log scale).

B Metric Definitions

Accuracy (ACC). The fraction of correctly predicted interaction types over all test samples.

Macro F1. The unweighted average of per-class F1 scores: Macro F1 $= \frac{1}{|\mathcal{R}|}\sum_{r\in\mathcal{R}} F1_r$, where $F1_r$ is the F1 score for relation type r. This metric treats all classes equally regardless of their frequency, making it sensitive to minority class performance.

Macro PR-AUC. The unweighted average of per-class precision-recall area under the curve. Like macro F1, this metric is sensitive to minority classes and is more informative than accuracy under class imbalance.

Cohen's κ. A chance-corrected measure of agreement between predictions and ground truth: $\kappa = \frac{p_o - p_e}{1 - p_e}$, where p_o is the observed agreement (equal to accuracy for single-label prediction) and p_e is the agreement expected by chance, computed from the marginal distributions of predictions and ground-truth labels. Values range from -1 to 1, with higher values indicating better agreement beyond chance.

Macro F1 and macro PR-AUC average per-class scores without frequency weighting and therefore account for the extreme class imbalance in DDI type distributions (see Figure 2); ACC and Cohen's κ complement them as global agreement measures.

C Baseline Descriptions

We compare UnPairRelNet against nine representative baselines spanning three categories. All baselines are evaluated under the same data splits and evaluation protocol.

C.1 Molecular Graph-based Methods

DeepDDI (Ryu et al., 2018) predicts DDI types using structural similarity profiles derived from molecular substructures, combined with a deep neural network classifier. It is the original method associated with the Ryu's dataset.

SSI-DDI (Nyamabo et al., 2021) decomposes DDI prediction into pairwise interactions between drug substructures, operating directly on raw molecular graph representations for richer feature extraction.

GGI-DDI (Yu et al., 2024) employs granular computing to identify pivotal functional groups of drugs and detects DDIs through interactions among these granules.

C.2 Knowledge Graph-based Methods

KGNN (Lin et al., 2020) applies graph neural networks on biomedical knowledge graphs, extending the receptive field to multiple hops to model high-order topological information and obtain long-distance drug correlations.

LaGAT (Hong et al., 2022) generates pair-specific attention pathways for neighbor selection, using one drug's embedding as a query vector to calculate attention weights for selecting topological neighbors of the other drug.

KnowDDI (Wang et al., 2024a) enhances drug representations by adaptively leveraging neighborhood information from biomedical knowledge graphs and learns a knowledge subgraph for each drug pair to interpret the predicted DDI.

C.3 Multi-modal Methods

SumGNN (Yu et al., 2021) anchors local subgraphs from biomedical knowledge graphs and integrates multi-channel knowledge, including chemical structures, for multi-typed DDI prediction.

TIGER (Su et al., 2024) employs a Transformer-based relation-aware self-attention mechanism in a dual-channel framework, where drug molecular graphs and biomedical knowledge graphs are fed into two respective channels.

MolecBioNet (Chen et al., 2025) integrates molecular and biomedical knowledge through domain-specific pooling strategies, including context-aware subgraph pooling and attention-guided influence pooling, for robust and interpretable DDI prediction.

C.4 Implementation Sources

For all baselines, we use the official implementations released by the original authors when available; for those without released code, we re-implement them following the original papers. Hyperparameters for each baseline follow the settings reported in the original papers unless otherwise noted.

D Data Splitting Protocols

D.1 Transductive Data Splitting

We adopt a 5-fold cross-validation protocol. The complete set of DDI triplets is randomly partitioned into 5 folds at the **triplet level** (not drug-pair level), ensuring that the same drug may appear in both training and test sets within a fold. For each fold, the split ratio is 60% training, 20% validation, and 20% testing. The partition is performed once and fixed across all experiments to ensure reproducibility. Table 9 summarizes the sample sizes.

D.2 Inductive Data Splitting

We define two inductive evaluation settings that test the model's ability to generalize to **unseen drugs**:

- **S1 (One-side unseen):** Each test triplet contains one drug that appears in the training graph (seen) and one drug that does not (unseen). This evaluates the model's ability to leverage partial topological context.
- **S2 (Both-side unseen):** Each test triplet contains two drugs that are both absent from the training graph. This evaluates the model's ability to generalize when topological context is substantially reduced.

Construction procedure:

1. From the full set of 1,704 drugs, a subset of drugs (306 on average for S1, 340 for S2; Table 10) is designated as *unseen* and removed from the training graph.
2. The training set consists of all DDI triplets where **both** drugs are seen.
3. For S1, the test set consists of triplets where exactly one drug is unseen. For S2, the test set consists of triplets where both drugs are unseen.
4. A *support set* of held-out DDI edges incident to unseen drugs is provided at inference time to construct the inference graph. For S1, it contains $\sim$300 edges, most of which connect unseen drugs to the seen graph. For S2, it contains $\sim$61,000 edges, predominantly connecting unseen drugs to the seen graph (Table 10).
5. The background knowledge graph (BKG) is fully retained in both training and inference graphs.

The inductive splits are constructed once per fold and fixed across all experiments. Table 10 summarizes the inductive splits.

Key design choices:

- The split is performed at the **triplet level**, consistent with the transductive setting.
- Unseen drugs retain their molecular graphs (SMILES) and physicochemical descriptors, ensuring that molecular-level information remains available.
- The support set is substantially larger in S2 than in S1: in S2, edges with exactly one unseen endpoint are excluded from both training and testing by construction and are therefore retained as support, whereas in S1 most such edges constitute the test set, leaving only a small reserved sample for support.
- No relation types are held out in either setting; all 86 DDI relation types remain visible during training.

E Implementation Details

Our implementation is available at `https://anonymous.4open.science/r/DDI1-03F6`.

E.1 Model Architecture

Table 11 lists the architecture hyperparameters. All dimensions refer to the Ryu's configuration; the DrugBank runs use the same values unless noted in the released configuration files.

E.2 Training Configuration

Table 12 reports the training configuration shared by all main experiments; ablation variants differ only in the components removed (Appendix I).

E.3 Physicochemical Descriptors

Table 6 reports the descriptor dimensionality; here we detail its composition and usage. Each drug is represented by a 187-dimensional descriptor vector composed of two parts:

- **20 physicochemical properties:** molecular weight, LogP, topological polar surface area (TPSA), number of hydrogen-bond donors/acceptors, number of rotatable bonds, number of aromatic rings, fractional CSP3, and other ADME-relevant descriptors computed via RDKit.
- **167 MACCS keys:** the standard MDL MACCS structural key fingerprint (166 bits + 1 padding bit), encoding the presence or absence of specific substructural patterns.

The descriptor vector is precomputed for all drugs and stored offline. During training and inference, descriptors are looked up by drug ID and projected via the MLP in Section 3.2. For unseen drugs in inductive settings, a projection of the descriptor vector substitutes for the learned drug embedding $\mathbf{e}$ in conditioning, pair attention, and fusion (Section 3.5); the molecular graph branch remains available for these drugs since their SMILES are retained (Appendix D).

E.4 Molecular Graph Encoding

The two molecular encoders correspond to the atom-level and substructure-level branches of Section 3.2.

Molecular graph (MolCov): Each drug's SMILES string is converted to a molecular graph where atoms are nodes and chemical bonds are edges. Node features include atom type, degree, hybridization, aromaticity, and formal charge. A 2-layer Graph Convolutional Network with hidden dimension 256 produces atom-level embeddings, which are aggregated via mean pooling to obtain a fixed-size molecular representation.

Fragment graph (MG encoder): Each drug is decomposed into fragments using the BRICS algorithm. The resulting fragment graph is a heterogeneous graph where nodes represent fragments and intra-molecular edges represent BRICS bond types; for a drug pair, inter-molecular edges fully connect the fragments of the two drugs. A heterogeneous GNN is applied, in which intra-molecular edges use standard (GCN) aggregation and inter-molecular edges use multi-head attention (HeteroGAT), followed by an attention-weighted mean readout, consistent with Section 3.2.

F Representation Diagnostic: Details

F.1 Hub Drug Validation

As a sanity check, we verify that the interaction topology is pharmacologically meaningful. The ten highest-degree drugs in the DDI graph all have pharmacological annotations retrievable in PubChem (Table 13).

F.2 Outlier Drug Identification

Table 14 provides the complete list of the 12 identified outlier drugs. Identities are established via exact PubChem CID matches, except tigecycline, which is identified by molecular formula and characteristic substructures (table footnotes).

F.3 Degree–Norm Scatter Plot

Figure 3 visualizes the relationship between embedding norm and graph degree. The shaded region marks the outlier criterion (degree < 10 and norm > 0.6), within which 12 drugs fall while the remaining low-degree drugs concentrate at norms below 0.2.

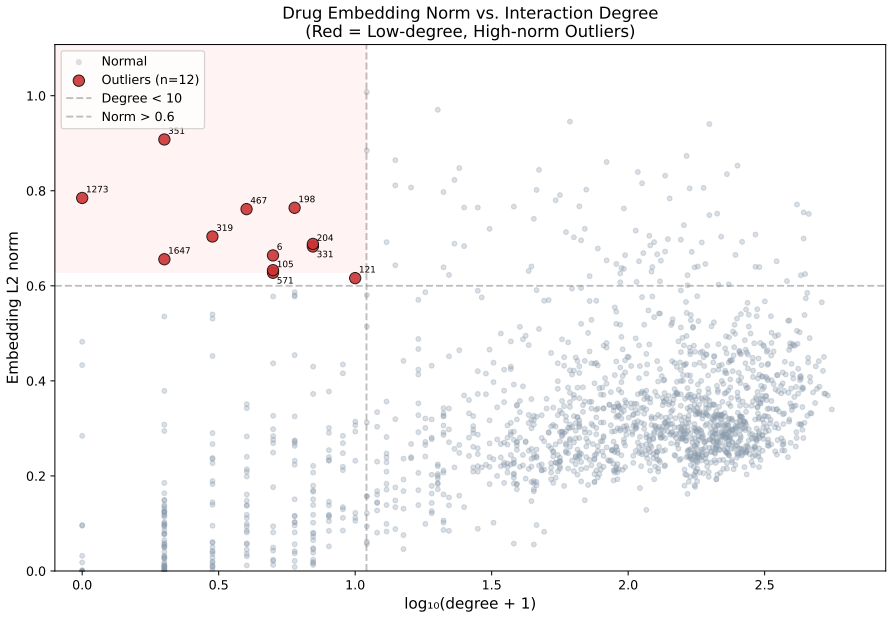


Figure 3: Embedding L2 norm vs. interaction degree on the Ryu's dataset (Fold 1). Red points mark the 12 outlier drugs (degree < 10, norm > 0.6); annotated numbers are entity IDs. Dashed lines indicate the outlier thresholds.

F.4 Degree–Norm Correlation

Table 15 reports the relationship between graph degree and embedding norm.

F.5 Attribution of the Beyond-Degree Component

To identify which branch supplies the beyond-degree component observed in Section 4.5, we recompute the embedding-norm diagnostic on the checkpoint of the variant without KG propagation (Table 3), which retains the molecular encoder, physicochemical descriptors, and pair attention while removing multi-hop propagation and its conditioning of relation representations.

Table 16 and Figure 4 summarize the comparison. Without propagation, the degree–norm correlation strengthens from 0.42 to 0.68, the 12 outlier drugs fall from the 97th to the 13th mean norm percentile, the outlier-to-control norm ratio drops from 5.4 to 0.2, and no low-degree drug reproduces the outlier pattern under the same relative criterion (overlap 0/12). Since the molecular and descriptor inputs remain active in this variant, the elevated norms cannot be attributed to them; they originate from the propagation pathway. Because the variant removes propagation together with its conditioning of relation representations, the attribution is to the propagation mechanism as a whole.

F.6 Representative Cases

Among the 12 outliers, six have documented pharmacological annotations in PubChem; the supporting evidence for all 12 drugs is recorded in Table 14, and five representative cases are listed in Table 4 (Section 4.5) and discussed below.

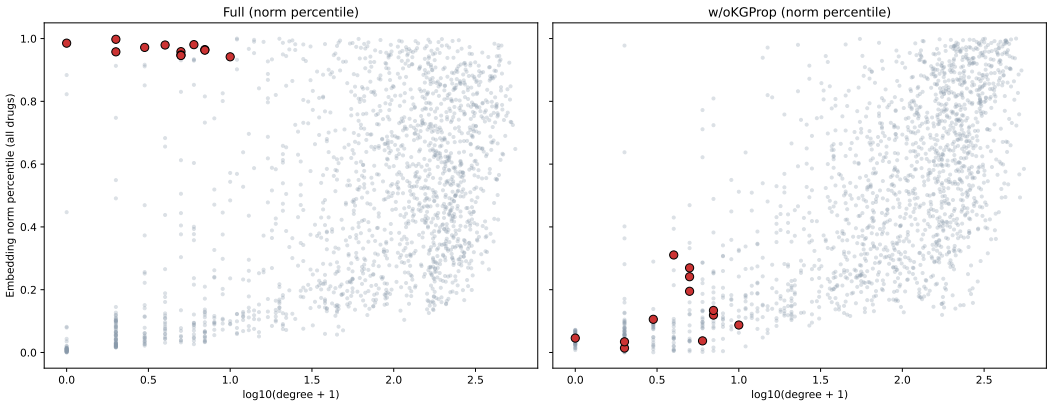


Figure 4: Embedding norm percentile vs. degree for the full model (left) and the variant without KG propagation (right). Red points mark the 12 outlier drugs. Without propagation, the outliers fall back onto the degree trend and the overall degree–norm relationship tightens.

Sulfamethizole (norm 0.908, degree 1) exhibits the highest embedding norm among all drugs despite having only a single recorded interaction in the training set. PubChem records a protein binding rate of 98–99% and 12 HSDB records.

Propylthiouracil (norm 0.688, degree 6) is a thiourea derivative with antithyroid activity that inhibits synthesis of thyroid hormones T3 and T4. *Carbimazole* (norm 0.683, degree 6) is an antithyroid agent sharing the same therapeutic indication. Despite belonging to the same therapeutic category, they are flagged independently, consistent with the interpretation that the representation captures category-level interaction propensity.

Dexrazoxane (norm 0.764, degree 5) has interaction-related records in PubChem, with a protein binding rate below 2%.

Chlorambucil (norm 0.616, degree 9) contains a bis(2-chloroethyl)amino group, the defining structural motif of nitrogen mustards, and has a protein binding rate of 99% with 8 HSDB records.

The remaining seven outliers include non-systemic agents and compounds with minimal documented interaction profiles, indicating that an elevated embedding norm alone does not imply documented interaction risk.

F.7 Embedding Space Analysis

Table 17 summarizes the BKG connectivity and embedding space clustering analysis.

G Additional Related Work

Relation reasoning on knowledge graphs. NBFNet (Zhu et al., 2021) formulates entity representation learning as path-based reasoning through generalized message passing, but represents relations by fixed learnable embeddings that do not capture inter-relational structure. ULTRA (Galkin et al., 2024) addresses this limitation by constructing a graph over relation types with co-occurrence edges and generating relation representations through message passing on this relation graph. However, ULTRA conditions the relation encoding on each query by injecting the query relation as a prior signal, which assumes that the query relation is known at encoding time. In DDI type prediction, the target relation is the prediction target itself and is therefore unknown at inference time, rendering such conditional paradigms inapplicable. Moreover, ULTRA's co-occurrence encoding records only whether two relations share an entity, without capturing the topological asymmetry of the connection. On the entity modeling side, ULTRA performs unidirectional propagation anchored at a single head entity with the query relation as prior, which is designed for link prediction where one endpoint is given. In contrast, DDI prediction involves a drug pair where both endpoints are given and the interaction type is to be determined, calling for a propagation paradigm anchored at both

endpoints simultaneously. Our work proposes an unconditioned relational encoding with degree-partial-order connectivity patterns that captures topological asymmetry without dependence on any specific query, together with a bidirectional propagation scheme anchored at the drug pair, where structurally-informed relation representations modulate information flow in both directions.

H Full Results with Standard Deviations

Tables 18 and 19 report the mean and standard deviation over 5 folds for the main comparison on the two benchmarks, and Tables 20 and 21 report the two novel-drug settings on the Ryu's dataset.

H.1 Fold Consistency Analysis

To assess the stability of per-class predictions across folds, we compare the per-class F1 scores between Fold 1 and Fold 2 on the Ryu's dataset. Among 86 relation types, 54 (62.8%) exhibit a difference below 0.02, indicating high stability for the majority of classes. The 10 classes with differences exceeding 0.10 are predominantly rare types with limited training samples, where predictions are sensitive to the specific fold partition. Three classes (IDs 49, 60, 73) receive zero F1 in both folds, representing relation types with insufficient training signal regardless of the model architecture.

I Component Analysis

I.1 Ablation Protocol

All ablation variants share the hyperparameters in Table 12 and the identical molecular graph encoder. When a component is removed, its associated modules and loss terms are removed accordingly; all remaining components and loss terms are unchanged. The "w/o KG Propagation" variant disables only multi-hop message passing, retaining entity embeddings and their bilinear interaction as a minimal pairwise signal.

I.2 Ablation Discussion

Adding unconditioned propagation to the molecular-graph-only base yields an improvement of 0.018 in macro F1 (0.815 $\rightarrow$ 0.833), showing that the global topological profile of the KG is itself a positive signal even without pair-specific conditioning or auxiliary components. The larger decrease when removing propagation from the full model (-0.031) compared with adding it to the molecular-graph-only base ($+0.018$) indicates a positive interaction effect: propagation provides the most benefit when combined with pair attention and descriptors, which leverage the propagated representations to produce more discriminative features.

Among the three components removed from the full model, physicochemical descriptors contribute the most (-0.035 in macro F1), followed by KG propagation (-0.031) and pair attention with context (-0.025). The contribution of physicochemical descriptors suggests that global molecular properties provide chemical priors complementary to the topological information captured by graph-based encodings; the per-dimension analysis in Figure 5 supports this view. The contribution of KG propagation indicates that structurally informed relation representations guide propagation to produce more discriminative entity representations. The contribution of pair attention and context indicates that anchoring the global propagation to the specific drug pair and aggregating local neighborhood information further enhances discriminative power.

I.3 Metric Behavior in Ablations

Among the full model and its ablation variants, accuracy, macro PR-AUC, and Cohen's κ exhibit small variations (<0.01), as these metrics are dominated by the majority class. The larger gap between the molecular-graph-only variant and the full model reflects the fundamental difference in information sources rather than metric sensitivity. Macro F1, which weights all classes equally, is therefore the most informative metric for ablation analysis.

I.4 Descriptor Importance Analysis

To understand which components of the 187-dimensional descriptor vector contribute most to relation discrimination, we compute the sample-weighted between-class variance of each dimension, restricted to relation types with at least 50 training triplets to mitigate noise from rare classes. (Without this restriction, a few rare classes with extreme descriptor means produce spurious outliers in the variance estimates.)

Figure 5 shows two complementary patterns. At the individual-dimension level, a small number of physicochemical properties (notably dims 15 and 17) exhibit the highest between-class variance, indicating that certain global molecular properties carry substantial relation-specific information. At the group level, however, the 167 MACCS keys collectively account for a larger share of the total discriminative signal: although each key contributes moderately, their broad coverage of substructural patterns (functional groups, ring systems) provides information that global properties alone cannot capture. This complementarity is consistent with the ablation result that removing descriptors causes the largest performance drop (-0.035 in macro F1, Table 3).

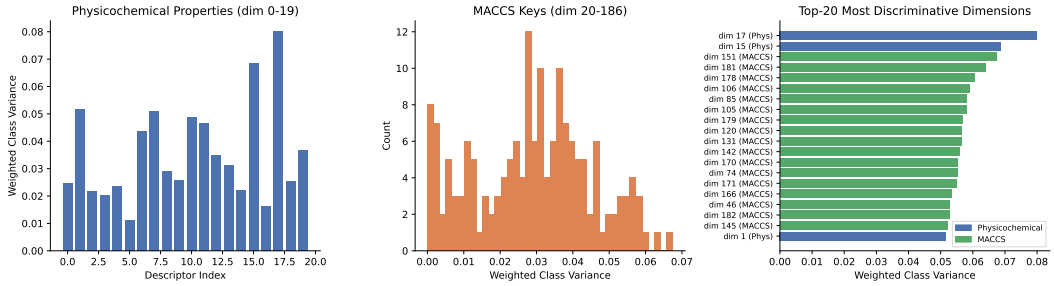


Figure 5: Weighted between-class variance of the 187 descriptor dimensions (classes with $\geq$50 training triplets). Left: physicochemical properties (dim 0–19). Middle: distribution of MACCS key importance (dim 20–186). Right: top-20 most discriminative dimensions, color-coded by group.

Table 5: Summary of notation.

Symbol	Meaning
Sets and graphs	
$\mathcal{D}$	Set of drugs
$\mathcal{R}$	Set of interaction types
$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{T})$	Biomedical knowledge graph
$\mathcal{T}_{\text{DDI}}/\mathcal{T}_{\text{BKG}}$	DDI / background relation types
$\mathcal{G}_{\text{rel}} = (\mathcal{T}, \mathcal{E}_{\text{rel}})$	Relation graph over relation types
$\mathcal{G}_d$	Molecular graph of drug d
$\mathcal{G}_{\text{sub}}^{(i,j)}$	Substructure interaction graph of pair (d_i, d_j)
$\mathcal{B}$	Mini-batch of drug pairs
Indices and neighborhoods	
d_i, d_j	Drug pair under consideration
r, r_1, r_2	Relation (interaction) types
u, v	Entity nodes (dummy indices of directed edges)
v_0, n_1, n_2	Shared entity and its two neighbors
m	Degree-partial-order mode (0–7)
r_{uv}	Relation type of directed edge (u, v)
$\mathcal{N}(r)$	Relation neighbors of r in $\mathcal{G}_{\text{rel}}$
$\mathcal{N}(d_i)$	One-hop drug neighbors of d_i
$\mathcal{N}_{\text{fwd}}(v), \mathcal{N}_{\text{bwd}}(v)$	Neighbors of v via forward / backward edges
Representations	
$\mathbf{e}_i$	Learned drug embedding (used for conditioning, pair attention, and fusion, but not as propagation states)
$\mathbf{m}_d$	Atom-level molecular embedding
$\mathbf{f}_{ij}$	Substructure-level pair representation
$\mathbf{x}_d^{\text{desc}}, \mathbf{p}_d$	Raw descriptor vector; its projection
$\mathbf{s}_r$	Structural statistics of relation r
$\mathbf{h}_r^{(l)}, \mathbf{H}^{(L)}$	Relation node state; stacked relation states
$\boldsymbol{\mu}_{r,r'}$	Learnable mode type embedding
$\mathbf{R}, \mathbf{R}^{(ij)}$	Relation representations (global / pair-conditioned)
$\mathbf{c}_{ij}$	Pair context
$\mathbf{h}_v^{(l)}, \mathbf{h}_v^{(0)}$	Entity node state; initialization (anchor receives signal)
$\mathbf{g}_{\text{src}}, \mathbf{g}_{\text{snk}}$	Shared source / sink signals
$\mathbf{H}_{\text{fwd}}, \mathbf{H}_{\text{bwd}}$	Entity state matrices of the two directions
$\mathbf{z}_i$	Propagation output for drug d_i
$\mathbf{q}_{ij}, \alpha_k, \mathbf{a}_i$	Attention query, weights, aggregated neighbors
$\mathbf{g}_i$	Fused representation of drug d_i
$\mathbf{b}_{ij}$	Bilinear pair interaction
$\mathbf{s}_{ij}$	Pair-level KG summary
$\mathbf{o}$	Final concatenated representation
Operators	
$\text{Update}_{\text{rel}}, \text{Agg}_{\text{rel}}$	Relational encoding update / aggregation
$\text{Update}_{\text{ent}}, \text{Agg}_{\text{ent}}$	Entity propagation update / aggregation
$\text{Proj}_{\text{rel}}, \text{Proj}_{\text{attn}}$	Relation output projection; attention residual projection
Scalars, dimensions, losses	
$w(r_1, r_2)$	Relation graph edge weight
$\deg(\cdot)$	Entity degree in the KG
L	Number of propagation layers
d_r, d_e	Relation embedding dim; drug embedding dim
B	Batch size
π_r	Predicted probability of the true type r
ω_r, n_r	Class weight; training samples of relation r
γ, λ	Focal focusing parameter; loss weight
$\mathcal{L}_{\text{focal}}, \mathcal{L}_{\text{CE}}$	Focal loss; auxiliary cross-entropy

Table 6: Dataset statistics for Ryu's and DrugBank. Triplet counts refer to the transductive split of Fold 1; the BKG counts refer to the BKG edge file, which stores the DDI edges (relation ID 0) together with the background relation edges.

Statistic	**Ryu's**	**DrugBank**
DDI Network		
Total DDI Triplets	191,388	617,277
Train	114,832	370,365
Valid	38,278	123,456
Test	38,278	123,456
Drugs	1,704	2,152
DDI Relation Types	86	221
Rel. Freq. (Min / Median / Max)	6 / 277 / 60,936	5 / 112 / 119,017
Knowledge Graph (BKG file)		
Triplets in BKG file (incl. DDI)	2,152,993	1,691,368
Entities in BKG triplets	241,836	241,837
Relation types in BKG file (incl. DDI)	12	12
Molecular Features		
Physicochemical Descriptors (dim)	187	187

Table 7: Entity type distribution over all nodes of the background knowledge graph schema. Both datasets share the same schema. Counts cover all nodes in the schema file; a small number of nodes (5 in Ryu's, 1 in DrugBank) do not appear in any BKG triplet and are therefore excluded from the entity counts in Table 6.

Entity Type	**Ryu's**	**DrugBank**
Drug	178,418	178,415
Chemical Entity	26,658	26,658
Protein	18,607	18,607
Disease	5,962	5,962
Genetic Disorder	4,409	4,409
Side Effect	4,303	4,303
Pathway	3,484	3,484
Total	**241,841**	**241,838**

Table 8: Relation types stored in the BKG edge file. ID 0 is the DDI target relation; IDs 1–11 are the background relation types.

Relation	**ID**
DDI (target relation)	0
Disease–Genetic Disorder	1
Disease–Pathway Association	2
Drug–Protein Interaction	3
Drug–Disease Association	4
Drug–Chemical Entity	5
Drug–Pathway Association	6
Drug–Side Effect Association	7
Protein–Protein Interaction	8
Protein–Disease Association	9
Protein–Pathway Association	10
Related Genetic Disorder	11

Table 9: Per-fold sample sizes on the Ryu's dataset (transductive). All folds share the same total of 191,388 triplets.

Fold	Train	Valid	Test
1	114,832	38,278	38,278
2	114,832	38,278	38,278
3	114,832	38,278	38,278
4	114,833	38,278	38,277
5	114,833	38,278	38,277

Table 10: Inductive split statistics on the Ryu's dataset (averaged over 5 folds). Support edges are held-out DDI edges incident to unseen drugs that are retained in the inference graph to provide structural context.

Statistic	S1	S2
Unseen Drugs	306.0 ± 41.6	340.0 ± 0.0
Train Triplets	$130{,}130 \pm 9{,}225$	$123{,}363 \pm 2{,}247$
Valid Triplets	$30{,}479 \pm 4{,}593$	$3{,}566 \pm 242$
Test Triplets	$30{,}479 \pm 4{,}593$	$3{,}567 \pm 242$
Support Edges	301 ± 39	$60{,}893 \pm 1{,}764$

Table 11: Model architecture hyperparameters.

Hyperparameter	Value
Relation Encoding	
Degree-partial-order modes	6 valid (of 8 patterns)
Relation graph edge threshold	0.005
Relation NBFNet layers	2
Relation embedding dim (d_r)	64
Relation statistics dim	15
Entity Encoding	
Drug embedding dim (d_e)	64
Entity NBFNet layers	2
Entity embedding dim (per direction)	128
Bidirectional concat dim	256
Pair attention max neighbors (K)	64
Pair context projection dim	128
Molecular Encoding	
MolCov layers	2
MolCov hidden dim	256
MG encoder (HeteroGAT + GCN) layers	2
Fragment decomposition	BRICS
GCN dropout	0.2
Fusion & Classification	
Joint feature dim	256
Projection head	Linear(256→256)→ReLU→Linear(256→ $\lvert\mathcal{R}\rvert$)

Table 12: Training configuration.

Setting	Value
Optimization	
Optimizer	Adam
Learning rate	5×10^{-4}
Weight decay	5×10^{-5}
LR schedule	Multiplicative decay, rate 0.93
Gradient clipping (max norm)	2.0
Batching	
Batch size (per GPU)	4
Gradient accumulation steps	4
Number of GPUs	8
Effective batch size	$4 \times 4 \times 8 = 128$
Training Schedule	
Max epochs	15
Early stopping	Patience 10 epochs
Evaluation frequency	Every 4,000 iterations
Random seed	42
Loss Function	
Focal loss γ	2.5
Label smoothing ϵ	0.03
Class weight scheme	$\omega_r \propto 1/\sqrt{n_r}$, normalized to mean 1, clipped at 5.0
Auxiliary loss weight (λ)	0.08
Regularization	
Propagation dropout (inductive)	0.3
Propagation dropout (transductive)	0.0
Relation init noise std	0.01
Hardware & Time	
GPUs	$8 \times$ NVIDIA A800 (80 GB)
Parallelism	PyTorch DDP
Training time (per fold)	$\sim$8 hours

Table 13: Top-10 highest-degree drugs in the DDI graph with pharmacological annotations retrievable in PubChem.

Drug Name	Deg.	Class	Documented Role
Phenobarbital	553	Barbiturate	CYP2B6/CYP3A inducer; 40 HSDB
Phenytoin	534	Hydantoin	Narrow TI; hepatic oxidation
Fosphenytoin	523	Phenytoin prodrug	Hydrolyzed to phenytoin
Primidone	518	Deoxybarbiturate	Metabolized to phenobarbital
Fluvoxamine	517	SSRI	Serotonin reuptake inhibition
Carbamazepine	513	Dibenzazepine	CYP3A inducer; CYP3A4 substrate
Pentobarbital	508	Barbiturate, short-acting	Prolongs Cl^- channel open time
Amiodarone	497	Class III antiarrhythmic	CYP450 inhibitor; reduces warfarin clearance
Venlafaxine	495	SNRI	Dual reuptake inhibition
Nefazodone	487	Phenylpiperazine	5-HT_2 receptor antagonist

Table 14: All 12 identified outlier drugs. PB = protein binding rate. HSDB = Hazardous Substances Data Bank.

Drug Name	Deg.	Norm	CID	Class	Evidence
Sulfamethizole	1	0.908	5328	Sulfonamide	PB 98–99%; 12 HSDB
Nonoxynol-9	0	0.785	72385	Polyethoxylated	Non-systemic agent
Dexrazoxane	5	0.764	71384	Bis-dioxopiperazine†	DDI records; PB <2%
Diethylcarbamazine	3	0.761	3052	Piperazine derivative	CID confirmed
Cefdinir	2	0.704	6803597	β-Lactam†	CID confirmed
Propylthiouracil	6	0.688	657298	Antithyroid (thiourea)	Inhibits T3/T4; PB 82%
Carbimazole	6	0.683	31072	Antithyroid (imidazole)	Same use as PTU; PB 85%
L-Phenylalanine	4	0.664	6140	Amino acid	Metabolism confirmed
Tigecycline	1	0.656	—	Formula-identified‡	Chemical-target interactions
Progabide	4	0.633	44115	CID confirmed	—
Cefmetazole-like	4	0.627	2626	CID confirmed	CID confirmed
Chlorambucil	9	0.616	2708	Bis-chloroethylamine	PB 99%; 8 HSDB

†Identified from molecular structure.
‡Molecular formula $C_{29}H_{39}N_5O_8$.

Table 15: Embedding norm statistics by interaction level on the Ryu's dataset (Fold 1).

Group	Mean Norm	Std
High-interaction (top 20%)	0.369	0.108
Low-interaction (bottom 20%)	0.192	0.199
Spearman ρ (degree vs. norm)	0.421 ($p < 10^{-70}$)	

Table 16: Embedding-norm diagnostic with and without KG propagation (Ryu's, Fold 1). Percentiles are computed over all 1,704 drugs; the ratio compares outlier mean norm to the mean norm of other low-degree (degree < 10) drugs.

Model	Spearman ρ	Outlier mean pct.	Outlier/control ratio
Full	0.42	0.97	5.4
w/o KG propagation	0.68	0.13	0.2

Table 17: BKG connectivity and embedding space clustering analysis of outlier drugs.

Metric	Value	Interpretation
Mean BKG edges (outliers)	580	Lower than controls
Mean BKG edges (controls)	926	—
Mean pairwise cosine (outliers)	0.209	Internal cohesion
Mean pairwise cosine (random)	0.010	Baseline
Enrichment ratio	20×	Non-random clustering
Cosine to high-degree hubs	≈ -0.1	No alignment with hubs
Cosine to hubs (controls)	≈ -0.1	No difference from controls

Table 18: Performance comparison on Ryu's Dataset (mean ± std over 5 folds). Best results in **bold**.

Methods	Ryu's Dataset			
	ACC	F1	PR-AUC	Cohen's κ
DeepDDI	0.868±0.005	0.703±0.006	0.811±0.010	0.842±0.006
SSI-DDI	0.848±0.006	0.618±0.017	0.704±0.014	0.818±0.007
GGI-DDI	0.889±0.006	0.811±0.019	0.880±0.012	0.869±0.007
KGNN	0.916±0.003	0.814±0.021	0.872±0.014	0.900±0.004
SumGNN	0.904±0.002	0.842±0.005	0.912±0.013	0.886±0.002
LaGAT	0.935±0.001	0.843±0.014	0.899±0.012	0.922±0.001
KnowDDI	0.923±0.002	0.880±0.010	0.934±0.011	0.909±0.002
TIGER	0.938±0.003	0.880±0.007	0.894±0.011	0.926±0.004
MolecBioNet	0.952±0.002	**0.912±0.008**	0.942±0.006	0.943±0.003
Ours	**0.956±0.002**	0.883±0.005	**0.951±0.010**	**0.948±0.002**

Table 19: Performance comparison on DrugBank Dataset (mean ± std over 5 folds). Best results in **bold**.

Methods	DrugBank Dataset			
	ACC	F1	PR-AUC	Cohen's κ
DeepDDI	0.862±0.003	0.676±0.026	0.810±0.016	0.848±0.004
SSI-DDI	0.743±0.009	0.327±0.005	0.438±0.009	0.717±0.010
GGI-DDI	0.859±0.008	0.762±0.015	0.836±0.016	0.846±0.009
KGNN	0.907±0.002	0.764±0.011	0.832±0.007	0.898±0.002
SumGNN	0.864±0.002	0.732±0.019	0.855±0.011	0.851±0.002
LaGAT	0.900±0.001	0.741±0.010	0.829±0.008	0.891±0.001
KnowDDI	0.894±0.001	0.772±0.015	0.868±0.008	0.885±0.001
TIGER	0.917±0.001	0.791±0.008	0.763±0.011	0.910±0.001
MolecBioNet	0.935±0.002	**0.846±0.010**	0.910±0.008	0.929±0.002
Ours	**0.954±0.000**	0.830±0.006	**0.937±0.006**	**0.949±0.000**

Table 20: Performance comparison on S1 setting (mean ± std over 5 folds). Best results in **bold**.

Methods	S1 setting			
	ACC	F1	PR-AUC	Cohen's κ
DeepDDI	0.652±0.019	0.536±0.027	0.592±0.043	0.580±0.016
SSI-DDI	0.597±0.025	0.538±0.040	0.580±0.044	0.514±0.029
GGI-DDI	0.642±0.018	0.531±0.033	0.573±0.036	0.566±0.022
KGNN	0.392±0.018	0.119±0.025	0.115±0.023	0.267±0.019
SumGNN	0.571±0.016	0.420±0.030	0.498±0.035	0.476±0.019
LaGAT	0.511±0.022	0.289±0.028	0.314±0.026	0.399±0.023
KnowDDI	0.518±0.013	0.296±0.031	0.351±0.046	0.415±0.017
TIGER	0.311±0.010	0.110±0.015	0.116±0.023	0.159±0.013
MolecBioNet	0.625±0.011	0.563±0.040	0.591±0.047	0.544±0.011
Ours	**0.673±0.017**	**0.625±0.057**	**0.674±0.046**	**0.605±0.018**

Table 21: Performance comparison on S2 setting (mean ± std over 5 folds). Best results in **bold**.

Methods	S2 setting			
	ACC	F1	PR-AUC	Cohen's κ
DeepDDI	0.394±0.018	0.239±0.021	0.202±0.021	0.269±0.023
SSI-DDI	0.389±0.025	0.245±0.023	0.204±0.016	0.251±0.028
GGI-DDI	0.394±0.047	0.230±0.016	0.187±0.021	0.266±0.041
KGNN	0.151±0.029	0.018±0.003	0.016±0.001	0.014±0.020
SumGNN	0.396±0.032	0.140±0.021	0.178±0.012	0.204±0.039
LaGAT	0.201±0.042	0.015±0.002	0.015±0.000	0.014±0.008
KnowDDI	0.300±0.054	0.016±0.004	0.061±0.005	0.053±0.033
TIGER	0.268±0.030	0.068±0.036	0.061±0.028	0.072±0.041
MolecBioNet	0.395±0.020	0.233±0.026	0.199±0.021	0.264±0.020
Ours	**0.415±0.030**	**0.313±0.033**	**0.251±0.029**	**0.280±0.037**

Table 22: Distribution of per-class F1 differences between Fold 1 and Fold 2 on the Ryu's dataset.

Difference Range	**Number of Classes**
< 0.02	54
0.02 – 0.05	14
0.05 – 0.10	8
> 0.10	10