

Learning to Denoise Biomedical Knowledge Graph for Robust Molecular Interaction Prediction

Tengfei Ma, Yujie Chen, Wen Tao, Dashun Zheng, Xuan Lin, Patrick Cheong-Iao Pang, Yiping Liu, Yijun Wang, Longyue Wang, Bosheng Song, Xiangxiang Zeng*, *Senior Member, IEEE*, and Philip S. Yu, *Fellow, IEEE*

Abstract—Molecular interaction prediction plays a crucial role in forecasting unknown interactions between molecules, such as drug-target interaction (DTI) and drug-drug interaction (DDI), which are essential in the field of drug discovery and therapeutics. Although previous prediction methods have yielded promising results by leveraging the rich semantics and topological structure of biomedical knowledge graphs (KGs), they have primarily focused on enhancing predictive performance without addressing the presence of inevitable noise and inconsistent semantics. This limitation has hindered the advancement of KG-based prediction methods. To address this limitation, we propose BioKDN (Biomedical Knowledge Graph Denoising Network) for robust molecular interaction prediction. BioKDN refines the reliable structure of local subgraphs by denoising noisy links in a learnable manner, providing a general module for extracting task-relevant interactions. To enhance the reliability of the refined structure, BioKDN maintains consistent and robust semantics by smoothing relations around the target interaction. By maximizing the mutual information between reliable structure and smoothed relations, BioKDN emphasizes informative semantics to enable precise predictions. Experimental results on real-world datasets show that BioKDN surpasses state-of-the-art models in DTI and DDI prediction tasks, confirming the effectiveness and robustness of BioKDN in denoising unreliable interactions within contaminated KGs. Code is available at <https://github.com/xiaomingaaa/BioKDN>.

Index Terms—Molecular Interaction Prediction, Knowledge Graph Reasoning, Knowledge-enhanced Network

I. INTRODUCTION

The prediction of molecular interactions, including drug-target interaction (DTI) prediction [1] and drug-drug interaction (DDI) prediction [2]–[4], is pivotal to drug discovery and therapeutics. The success of interaction prediction based on knowledge graphs (KGs) in social networks [5], [6] and recommendations [7]–[9] encourages researchers to develop various KG-based computational methods to accelerate drug development [10]–[13]. The biomedical KGs contain a large number of entities with specific patterns and semantics (e.g.,

Tengfei Ma, Yujie Chen, Wen Tao, Yiping Liu, Yijun Wang, Bosheng Song, and Xiangxiang Zeng are with the College of Computer Science and Electronic Engineering, Hunan University, China. (E-mail: {tfma, yjchen, taowen, ypliu, wyjun, boshengsong, xzeng}@hnu.edu.cn).

Dashun Zheng and Patrick Cheong-Iao Pang are with the Faculty of Applied Sciences, Macao Polytechnic University, Macao, China. (E-mail: p2212871@mpu.edu.mo; mail@patrickpang.net).

Xuan Lin is with the College of Computer Science, Xiangtan University, China. (E-mail: jack_lin@xtu.edu.cn).

Longyue Wang is with the Tencent AI Lab, Shenzhen, China. (E-mail: vinnilywang@tencent.com).

Philip S. Yu is with the Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7053 USA. (E-mail: psyu@uic.edu).

*Corresponding author: Xiangxiang Zeng.

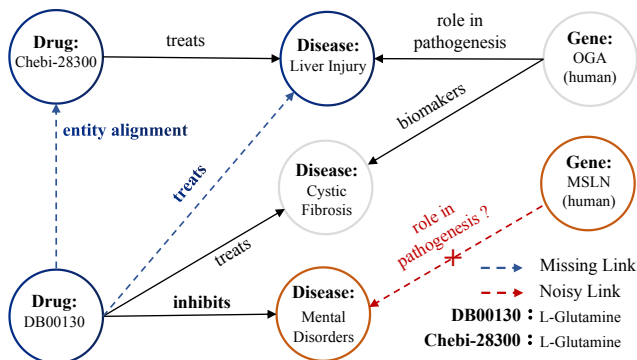


Fig. 1. The explanatory case of noise within DRKG. *Chebi:28300* and *DB00130* indicate the same drug *L-Glutamine*, but they are treated as different entities, which results in entity unalignment and facts missing. Meanwhile, the source document represents the gene *MSLN* as a biomarker for cancer patients, and no confidence indicates it has a role in the disease *Mental Disorders*, which introduces noisy interactions into the KG.

the associated pathways between drugs and targets). The interaction patterns between biomedical entities can enhance the prediction of biological connections among various entities [10]. However, accurately recognizing the unknown interactions between various molecular entities with computational models remains challenging.

Previous methods utilized the topological properties of the integrated association networks (e.g., *drug-disease-association* networks) to learn low-dimensional vector representations for predicting unknown interactions [14]–[16]. These methods adopted network-based models, which ignore the semantic relations between various entities (e.g., drug, pathway, disease). Subsequently, a line of works applied the knowledge graph embedding methods to learn the semantic relations with multi-dimensional embeddings for predicting DTI [17] and DDI [18], [19]. However, these approaches often struggle to learn the topological structure of complex biomedical knowledge graphs (KGs) efficiently. More recently, various models [2], [3], [10], [20] utilizing heterogeneous graph neural networks have achieved promising results. These methods focus on learning the semantic knowledge and local structure of the neighboring relational paths, which enables them to capture the semantic relations and tractable pathways surrounding the predicted interaction.

Despite their effectiveness, existing KG-based models suffer from noisy interactions and relations within biomedical KGs [21], [22]. The majority of biomedical knowledge graphs are generated from unstructured text and multi-source

databases using natural language processing technology [23], [24], which can lead to the presence of misaligned entities and counterfactual links within the KGs. For example, as shown in Fig. 1, the different entities *Chebi:28300* and *DB00130* within DRKG represent the same drug *L-Glutamine*¹, which is misaligned. The source document represents the gene *MSLN* is a biomarker of cancer patients, and no evidence indicates it has a role in the disease *Mental Disorders*, introducing noise into the DRKG [24]. In addition, the different relations *drug_treats_disease* and *drug_inhibits_disease* represent the same meaning “one drug can therapy one disease” in the contexts of Fig. 1, which brings inconsistent semantics.

In this case, current KG-based methods are ineffective due to the noisy interactions and inconsistent semantics. Based on the above observations, we propose a novel knowledge-enhanced denoising network, called BioKDN. We design a structure reliability learning module for the local subgraph guided by downstream tasks, to contain reliable interactions. Inspired by the successful application of smoothing techniques for image denoising [25], [26] by blurring noisy pixels, we develop a smooth semantic preservation module that blurs the similar relations to keep consistent semantics and ignore task-irrelevant edges. This reduces the negative impact of noisy interactions and inconsistent semantics existing in the KG. To further focus on the knowledge-enhanced informative interactions, we maximize the mutual information between the representations of reliable structure and smoothed semantics. BioKDN improves the AUC-ROC and Micro-Recall by 2.19% and 2.76% respectively on the DrugBank dataset in both DTI and DDI prediction tasks.

In summary, the contributions of BioKDN include:

- We approach knowledge-enhanced molecular interaction prediction from a new perspective by adaptively reducing the negative impact of noisy interactions.
- We innovatively propose learning the reliable structure and smoothing semantics by blurring similar relations, which reduces the negative influence of noisy interactions and maintains consistent semantics.
- We emphasize knowledge-enhanced reliable interactions by maximizing the mutual information between the learned structure and smoothed semantics to efficiently drop information unrelated to downstream tasks.
- Extensive experiments of the DTI and DDI prediction on benchmark datasets and contaminated KG demonstrate that BioKDN outperforms the state-of-the-art baselines.

II. RELATED WORK

A. KG-based Molecular Interaction Prediction

Molecular interaction prediction is increasingly adopted in biomedical knowledge graphs to identify unknown biological relations and interactions between various molecular entities [27]. The line of work mainly focuses on the completion of DTI and DDI relations on KGs. TriModel [17] and KG-DDI [18] proposed novel knowledge graph embedding models to learn the informative global structure and semantic

knowledge for completing the relations of DTI and DDI, respectively. To obtain rich neighborhood information and semantic relations of KG, KGNN [2] proposed a graph neural network to learn the structural relations, which enhances the prediction of the DDI relations. Subsequently, KGE-NFM [1] developed a unified knowledge graph embedding framework to predict missing DTI links by combining the knowledge graph and recommendation system. However, these methods only consider the structure of the biomedical KGs. Recent methods proposed various fusion models to integrate the features of molecular graphs and KG embeddings for enhancing DTI [10] and DDI [28] prediction. To focus on the local structure of the predicted entity pairs, SumGNN [3] designed a new method to efficiently emphasize the subgraph structure of the biomedical KG, which aids the drug interaction prediction. GraIL [29] and SNRI [30] proposed to model the enclosing subgraph structure and neighboring relational paths around the target triple to effectively predict unknown links. Besides, MINES [31] further introduces intercommunication mechanism and performs prediction on neighbor-enhanced subgraph, which also achieves promising performance. To filter out irrelevant entities, AdaProp [32] designed an incremental sampling mechanism to preserve promising targets. However, the presence of noise such as entity misalignment, and false positive triples in the KGs greatly degrades the performance of these methods. To address the above limitations, we develop reliable structure learning and smooth semantic preservation modules to denoise unreasonable interactions and maintain consistent semantics.

B. Denoising Methods on Graphs

Denoising on graphs has been successfully applied to the recommendation [33] and social networks [34]. RGCF [35] proposed a self-supervised robust graph collaborative filtering model to denoise unreliable interactions and preserve the diversity in a contrastive way for the recommendation. Similarly, SGDL [36] provided a universal solution using self-guided learning to denoise implicit noisy feedback that can be generalized to various recommendation tasks. However, these methods are limited in their ability to denoise noisy interactions with positive and negative feedback in domain-specific networks, and difficult to work on the complex biomedical KGs. To tackle these limitations, inspired by the smoothing insight in image denoising [25], [26], we blur the complex relations to contain consistent semantics and learn the reliable interactions from the local subgraph.

III. METHODOLOGY

A. Preliminaries

1) *Biomedical Knowledge Graph*: We define a biomedical knowledge graph DRKG [24] as $\mathcal{G}_{kg} = \{(h, r, t) | h, t \in \mathcal{E}, r \in \mathcal{R}\}$ where each triple (h, r, t) describes a relation r (e.g., DTI and DDI) between the biomedical entities h and t .

2) *Local Subgraph*: Based on GraIL [29], when given a KG \mathcal{G}_{kg} and a molecular pair (u, v) , we extract a local subgraph surrounding the target link. Initially, we obtain the k -hop neighboring nodes $\mathcal{N}_k(u) = \{s | d(u, s) \leq k\}$ and $\mathcal{N}_k(v) =$

¹<https://go.drugbank.com/drugs/DB00130>

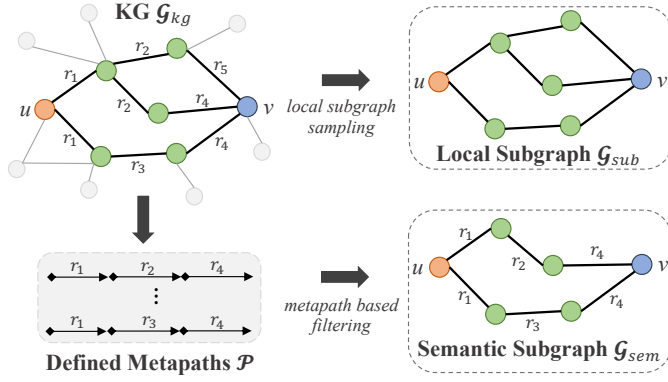


Fig. 2. An example of extracting local (structural reliability) and semantic subgraphs (semantic consistency) surrounding the target molecular pair (u, v) .

$\{s | d(v, s) \leq k\}$ for both u and v , where $d(\cdot, \cdot)$ represents the shortest path distance between target pair on \mathcal{G}_{kg} . We then obtain the set of nodes $V = \{s | s \in \mathcal{N}_k(u) \cap \mathcal{N}_k(v)\}$ as vertices of the local subgraph. Finally, we extract the edges E linked by the set of nodes V from \mathcal{G}_{kg} as the local subgraph $\mathcal{G}_{sub} = (V, E)$. We show an example in Fig. 2.

3) *Semantic Subgraph:* Given a molecular pair (u, v) , we define a set of metapaths $\mathcal{P} = \{P_1, \dots, P_i, \dots, P_n\}$ between nodes u and v , where n is the number of defined metapaths and $P_i = (r_1 \rightarrow \dots \rightarrow r_k)$. According to the defined set of metapaths, we can extract various relational paths around the given link from original KG \mathcal{G}_{kg} to construct a semantic subgraph \mathcal{G}_{sem} as shown in Fig. 2.

4) *Problem Definition:* In this paper, we focus on predicting the missing molecular interactions based on biomedical KG \mathcal{G}_{kg} by adaptively extracting high-quality facts and offering additional knowledge. We treat the molecular interaction prediction as a classification task, aiming to estimate the interaction probability of various relations (e.g., DTI and DDI). For a given unknown molecular pair (u, v) with relation r , we propose a model to predict the interaction probability denoted as $p_{(u,r,v)} = \mathcal{F}((u, r, v) | \Theta, \mathcal{G}_{kg}, \mathcal{G}_{sub}, \mathcal{G}_{sem})$ by maximizing the mutual information between the local and semantic subgraphs, where r represents the DTI or DDI.

B. Initialization of Knowledge Graph

In this paper, we utilize DRKG [24] as our external biomedical KG. DRKG contains complex relationships between biological entities (e.g., symmetric and inverse interactions among genes). To effectively learn the semantic knowledge within the DRKG, we use the knowledge graph embedding method with relation rotation following in [37]. Given a triple (h, r, t) , we expect that $\mathbf{x}_t = \mathbf{x}_h \odot \mathbf{e}_r$, where the \odot represents the element-wise product. $\mathbf{x}_h, \mathbf{x}_t$ and \mathbf{e}_r represent the embeddings of entities h, r and the relation r , respectively. The score function is defined as follows:

$$s(h, r, t) = \|\mathbf{x}_h \odot \mathbf{e}_r - \mathbf{x}_t\|, \quad (1)$$

By minimizing the score of positive triples and maximizing the score of negative ones, we obtain the entity and relation embeddings \mathbf{X} and \mathbf{E} as the initial features of the KG.

C. Subgraph Denoising

1) *Structure Reliability Learning:* To enable robust estimation of noisy interactions in KGs, we propose a structural reliability learning module for the local subgraph. This module can dynamically adjust the reliable subgraph structure by using the pre-trained node features and the feedback of the downstream prediction tasks. The underlying assumption is that nodes with similar features or structures are more likely to interact with each other than those with irrelevant features or structures [38], [39]. Our objective is to assign weights to all edges between the set of nodes using a reliability estimation function denoted as $F(\cdot, \cdot)$, which relies on pre-trained node features. Then, the refined local subgraph can be generated by filtering out noisy edges with low weight and retaining the reliable links with larger ones, as shown in Fig. 3b. Specifically, given an extracted local subgraph $\mathcal{G}_{sub} = (V, E)$ around the molecular pair (u, v) , we model all possible edges between the nodes as a set of mutually independent Bernoulli random variables parameterized by the learned attention weights π .

$$\mathcal{G}'_{sub} = \bigcup_{i,j \in V} \{(i, j) \sim \text{Ber}(\pi_{i,j})\}. \quad (2)$$

Here, V represents the set of nodes within the local subgraph and $(i, j) \in E$ denotes the edge between nodes i and j . We optimize the reliability probability π jointly with the downstream molecular interaction prediction tasks. The value of $\pi_{i,j}$ describes the task-specific reliability of edge (i, j) where smaller $\pi_{i,j}$ indicates that the edge (i, j) is more likely to be noisy and should be assigned a lower weight or be removed. For each edge between node pair (i, j) , the reliable probability $\pi_{i,j} = F(i, j)$ can be calculated as follows:

$$\begin{aligned} \pi_{i,j} &= \text{sigmoid}(Z(i)Z(j)^T), \\ Z(i) &= \text{MLP}(\mathbf{X}(i)), \end{aligned} \quad (3)$$

where $\mathbf{X}(i)$ represents the pre-trained feature of node i derived from Section III-B, $Z(i)$ is the learned embedding of node feature $\mathbf{X}(i)$, and $\text{MLP}(\cdot)$ denotes a two-layer perceptron in this work. Since the extracted local subgraph \mathcal{G}_{sub} is not differentiable with the probability π as Bernoulli distribution, we use the reparameterization trick and relax the distribution [33], [40] $\text{Ber}(\pi_{i,j})$ in Equation (2) as follows:

$$\text{Ber}(\pi_{i,j}) \approx \text{sigmoid}\left(\frac{1}{t} \left(\log \frac{\pi_{i,j}}{1 - \pi_{i,j}} + \log \frac{\epsilon}{1 - \epsilon} \right)\right), \quad (4)$$

where $\epsilon \sim \text{Uniform}(0, 1)$, $t \in \mathbb{R}^+$ indicates the temperature for the concrete distribution. After relaxation, the binary entries (i, j) sampled from a Bernoulli distribution are converted into a deterministic function of $\pi_{i,j}$ and ϵ . With $t > 0$, the function is smoothed with a well-defined gradient $\frac{\partial \text{Ber}(\pi_{i,j})}{\partial \pi_{i,j}}$ that enables the optimization of learnable subgraph structure during the training process. The subgraph structure after the concrete relaxation is a weighted fully connected graph, which is computationally expensive. We hence drop the edges of the subgraph with a probability of less than 0.5 and get the refined graph $\mathcal{G}'_{sub} = (V, E')$. Subsequently, we perform the L -layer

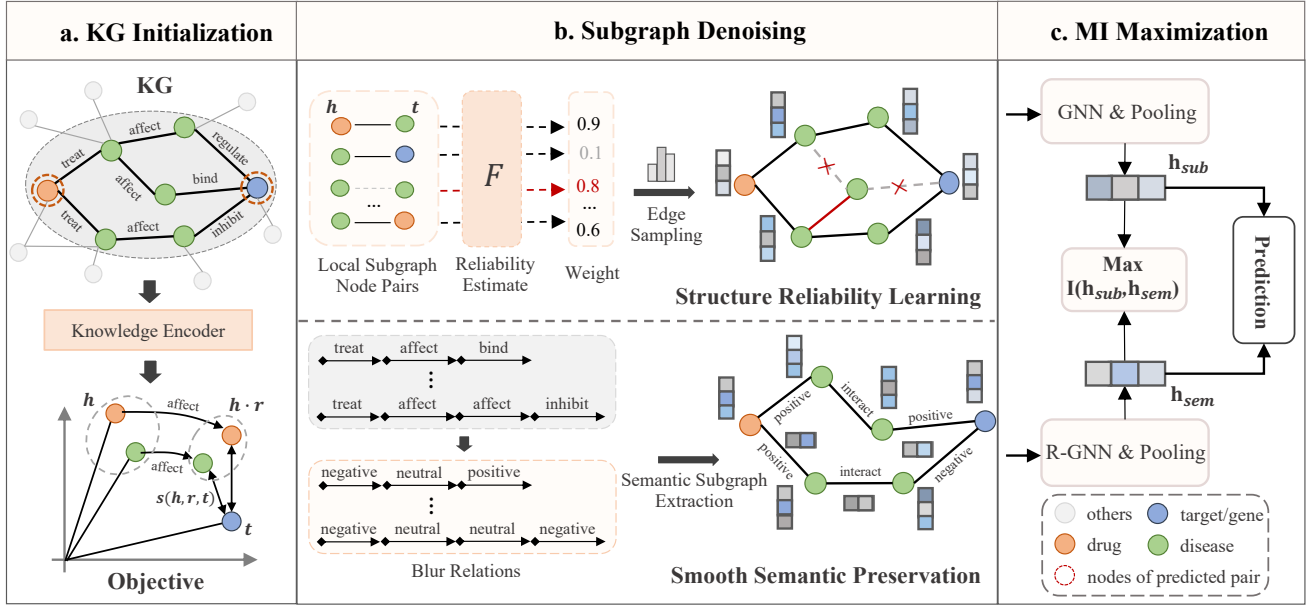


Fig. 3. The BioKDN framework comprises three modules for predicting links in a given KG: (a) Initializing the entity and relation embeddings of the KG using RotatE; (b) Denoising the noisy interactions around the predicted link by learning reliable structure and preserving consistent semantics; (c) Maximizing the mutual information between refined and semantic subgraphs to focus on informative interactions.

GCNs [41] on the refined subgraph with pre-trained node features to obtain its global representation \mathbf{h}_{sub} as follows:

$$\begin{aligned} h^l &= \text{GCN} \left(h^{l-1}, \mathcal{G}'_{sub} \right), \\ \mathbf{h}_{sub} &= \frac{1}{|V|} \sum_{i \in V} \sigma(f(h^L(i))), \end{aligned} \quad (5)$$

where the initial $h^0 = \mathbf{X}$ and $\sigma(\cdot)$ represents the activation function. $f(\cdot)$ denotes the feature transformation operation.

2) *Smooth Semantic Preservation*: Biomedical KGs often contain noise and inconsistent relations (e.g., the different relations *drug_treats_disease* and *drug_inhibits_disease* represent the same semantic “drug can therapy disease”) that bring negative impact into the downstream molecular interaction prediction tasks [23]. Inspired by the smoothing insight of image denoising [25], [26], we design a knowledge-enhanced smooth semantic preservation module, which blurs similar relations and preserves the smoothed relational semantics to mitigate the adverse impact of potential inconsistency. Specifically, we utilize prior knowledge to generalize the interactions between biological entities into positive, neutral, and negative semantics based on the consistent meaning of different relations, thereby smoothing the KG. Subsequently, we develop a semantic subgraph extraction module to explore the neighboring relations by extracting the paths with pre-defined metapaths (i.e., patterns among drug, disease, and gene entities). Given the molecular pair (u, v) , we use the defined metapaths to extract relational paths and construct a semantic subgraph \mathcal{G}_{sem} as shown in Fig. 2. After obtaining the semantic subgraph, we design a L -layer relational graph neural network (R-GNN) inspired by [30], [42] to obtain the global semantic representation of \mathcal{G}_{sem} . We define the

updating function of the nodes in l -th layer as:

$$\begin{aligned} \mathbf{x}_i^l &= \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_r(i)} \alpha_{i,r} \mathbf{W}_r^l \phi(\mathbf{e}_r^{l-1}, \mathbf{x}_j^{l-1}), \\ \alpha_{i,r} &= \text{sigmoid}(\mathbf{W}_1 [\mathbf{x}_i^{l-1} \oplus \mathbf{x}_j^{l-1} \oplus \mathbf{e}_r^{l-1}]), \end{aligned} \quad (6)$$

where $\mathcal{N}_r(i)$ and $\alpha_{i,r}$ denote the neighbors and the weight of node i under the relation r , respectively. \oplus indicates the concatenation operation. \mathbf{W}_r^l represents the transformation matrix of relation r , ϕ is the aggregation operation $\phi(\mathbf{x}, \mathbf{e}) = \mathbf{x} - \mathbf{e}$ to fuse the hidden features of nodes and relations, and \mathbf{e}_r^{l-1} is optimized from \mathbf{e}_r^0 layer by layer. In addition, we initialize the node feature \mathbf{x}_i^0 and relation representation \mathbf{e}_r^0 using the pretrained embeddings \mathbf{X} and \mathbf{E} . Finally, we obtain the global representation \mathbf{h}_{sem} of the semantic subgraph \mathcal{G}_{sem} as follows:

$$\mathbf{h}_{sem} = \frac{1}{|V_{sem}|} \sum_{i \in V_{sem}} \sigma(f(\mathbf{x}_i^L)), \quad (7)$$

where V_{sem} is the node set of semantic subgraph \mathcal{G}_{sem} . For more information about the smoothed KG and pre-defined metapaths, please refer to Section IV-B3.

D. MI Maximization in Subgraphs

To preserve the knowledge-enhanced reliable interactions, we incorporate fusion information from diverse perspectives of the reliable structure and smoothed semantic relations through the maximization of mutual information (MI). Specifically, we adopt InfoNCE [43] to estimate mutual information between the representations of local structure and semantic subgraphs globally. In a formal context, when discussing the concept of subgraph mutual information, we treat the representations originating from both the reliable structure and the semantic subgraph, each extracted from a common link, as positive

pairs. Conversely, the representations stemming from two distinct links within the refined local structure and the smoothed semantic are treated as adversarial pairs:

$$I(\mathbf{h}_{sub}; \mathbf{h}_{sem}) = -\log \frac{\exp(d(\mathbf{h}_{sub}, \mathbf{h}_{sem})/\tau)}{\sum_{m \in P} \exp(d(\mathbf{h}_{sub}, \mathbf{h}_{sem}^m)/\tau)}, \quad (8)$$

where $d(\cdot, \cdot)$ is set as a cosine similarity function to measure the similarity between two representation vectors and τ is a hyper-parameter indicating the temperature; P represents all link pairs to be predicted and \mathbf{h}_{sem}^m denotes the representation of global semantic subgraph for the link m .

E. Prediction and Optimization

We consider the molecular interaction prediction a classification task. For a given molecular pair (u, v) with relation r , we model the interaction probability $p_{(u,r,v)}$ of the pair by adopting the learned representations from the structure and knowledge-enhanced semantic views as follows:

$$p_{(u,r,v)} = \sigma(f([\mathbf{h}_{sub} \oplus \mathbf{h}_{sem}])), \quad (9)$$

where the \oplus indicates the concatenate operation. We then adopt the cross-entropy loss:

$$\ell(u, v) = -\sum_{r \in \mathcal{R}} \log(p_{(u,r,v)})y_{(u,r,v)}, \quad (10)$$

where $y_{(u,r,v)}$ is the real label of the given link. To denoise unreliable interactions from the structure and semantic subgraphs, we jointly optimize the link prediction task and the self-supervised MI maximization contrastive learning:

$$\ell_{total}(u, v) = \ell(u, v) + \lambda I(\mathbf{h}_{sub}; \mathbf{h}_{sem}), \quad (11)$$

where λ is a hyper-parameter that weighs the contribution of the self-supervised MI mechanism.

F. Computational Complexity of BioKDN

BioKDN consists of three main modules: structural reliability learning (SRL), smooth semantic preservation (SSP), and mutual information maximization (MI). The SRL contains a graph structure learning model whose computational complexity is $kD \cdot d + (kD)^2 \cdot d$, where k is the subgraph size, D denotes the average degree of DRKG, d is the embedding dimension, and kD represents the average number of nodes within the subgraph. The SSP contains an RGNN model over subgraphs whose computational complexity is $kD + \frac{kD^2}{2} + (kD) \cdot d \cdot L$, where L is the number of layers. The MI module contains an infoNCE loss computation whose computational complexity is $(N \cdot B + N)d$, where N is the number of samples and B denotes the batch size. The overall computational complexity of BioKDN is $kD((L+2)d^2 + \frac{D}{2} + 1) + (N \cdot B + N)d$. This shows the complexity of the BioKDN approximation polynomial, and its efficiency depends mainly on the sample size, the batch size, the average degree of KG, and the embedding dimension. Therefore, BioKDN can be scaled to large-scale sparse KGs (low average node degree), and the training efficiency can be balanced for large-scale datasets by reducing the embedding dimension and batch size.

G. Discussion of Smoothing Operation

Biomedical knowledge-enhanced molecular interaction prediction has several key applications such as drug-target interaction (DTI) prediction and drug-drug interaction (DDI) prediction, which is critical for drug discovery and clinical research. However, most biomedical KGs are constructed from publication text and multi-source databases that often contain unreliable interactions and inconsistent semantics of similar relations. For example, the different relations *drug_treat_disease* and *drug_inhibit_disease* represent the same semantic “a drug can therapy a disease” actually. However, the KG embedding models will adopt different multi-dimensional vectors to represent their semantics. This will introduce the inconsistency of different semantic relations, which degrades the robustness of representation models. Inspired by the smooth insights of image denoising [25], [26] by blurring noisy pixels, we smooth DRKG [24] by blurring inconsistent and sparse relations. The results in Section IV-F indicate this operation can effectively reduce the negative impact of inconsistent relations.

IV. EXPERIMENTS

In this section, BioKDN² performs the molecular interaction prediction task for two key relations (i.e., *drug-target interaction* and *drug-drug interaction*) on biomedical KG.

A. Experimental Setups

1) *Datasets*: For link prediction of the relation DTI, we empirically perform experiments on two real-world datasets: (1) **DrugBank** [44] collects the unique bioinformatics and cheminformatics resources that contain 12,063 drug-target pairs with 2,515 drugs, and 2,972 targets. (2) **DrugCentral** [45] is a drug database built from multiple sources, which contains 9,317 interactions between 1,061 drugs and 1,388 targets. For DDI prediction, we evaluate BioKDN on two wide-used datasets: (3) **DrugBank** [44] contains 191,984 drug pairs with 86 types associated pharmacological relations for 1,703 drugs (e.g., *increase of cardiotoxic activity*). (4) **TWOSIDES** [46] dataset contains 335 drugs with 26,443 drug pairs for 200 various side effect types. Following [47], we ensure each DDI type has at least 900 drug pairs by keeping 200 commonly occurring types. We adopt the comprehensive DRKG [24] as the biomedical knowledge graph, which contains 97,238 entities and 5,874,261 triples. To smooth the semantic relations and filter out task-irrelevant edges of the DRKG, we blur the interactions between drugs, genes, and diseases into three types according to the semantic similarity of various relations.

2) *Evaluation Metrics*: In this paper, we consider the molecular interaction predictions as classification tasks (i.e., Binary classification for DTI prediction [10] and Multi-class/type for DDI prediction [3]). To quantify the prediction performance of the two molecular interaction prediction tasks, we adopted the *area under the receiver operating characteristic curve* (AUC-ROC), the *area under the precision-recall curve* (AUC-PR) for binary and multi-type classification, and

²Code: <https://github.com/xiaomingaaa/BioKDN>

TABLE I
THE BLURRED AND ORIGINAL RELATIONS OF DRKG.

Blurred Relations	Original Relations
Gene_neutral_Gene	HumGenHumGen, VirGenHumGen, Q (production), E (affect) Rg (regulation), I (signaling pathway), H (same protein)
Gene_positive_Gene	V+ (activates), B (bind), E+ (increases expression), W (enhance)
Drug_neutral_Gene	DrugVirGen, DrugHumGen, E (affect), K (metabolism)
Drug_positive_Gene	target, enzyme, carrier, A+ (agonism), E+ (increase), B (bind), O (transport), Z (enzyme activity)
Drug_negative_Gene	N (inhibit), A- (antagonism), E- (decrease)
Gene_negative_Drug	ANTAGONIST, INHIBITOR, BLOCKER, CHANNEL BLOCKER, ANTIBODY
Gene_neutral_Drug	OTHER, ALLOSTERIC MODULATOR
Gene_positive_Drug	BINDER, MODULATOR, AGONIST, POSITIVE ALLOSTERIC MODULATOR, ACTIVATOR, PARTIAL AGONIST
Drug_positive_Disease	Sa (side effect), J (role in disease pathogenesis)
Drug_negative_Disease	treats, T (therapy), C (inhibits cell growth), Pa (alleviate), Pr (prevent)
Drug_neutral_Disease	Mp (biomarkers)
Gene_positive_Disease	L (improper), U (causal), Y (polymorphisms), J (role in pathogenesis), G (promote), Ud (mutations)
Gene_negative_Disease	Te (possible therapeutic effect)
Gene_neutral_Disease	D (targets), Md (diagnostic), X (overexpression in disease)
Disease_positive_Gene	DuG (upregulate)
Disease_negative_Gene	DdG (downregulate)
Disease_neutral_Gene	DaG (associate)
Disease_neutral_Disease	DrD (resemble)

TABLE II
THE EXAMPLES OF PRE-DEFINED METAPATHS.

Prediction Task	Metapaths
drug-target interaction	(drug_neutral_disease, disease_neutral_gene)
	(drug_neutral_gene, gene_neutral_gene),
	(drug_positive_disease, disease_neutral_gene),
	(drug_negative_disease, disease_positive_gene),
...	
drug-drug interaction	(drug_neutral_disease, disease_neutral_drug),
	(drug_neutral_gene, gene_positive_drug),
	(drug_positive_disease, disease_positive_drug),
	(drug_negative_disease, disease_positive_drug),
...	

TABLE III
OPTIMAL HYPER-PARAMETER OF BIOKDN FOR EACH TASK.

	DTI	DDI
learning rate	10^{-2}	10^{-2}
batch size	64	64
embed dim	128	64
λ	0.5	0.1
π	0.5	0.3
τ	0.5	1
k -hop	2	2
KG initializer	RotatE	RotatE

F1 score, Recall score for multi-class classification task. The details of these metrics are as follows:

- **AUC-ROC** is calculated by $\sum_{k=1}^n TP_k \Delta FP_k$, where k is k -th true-positive and false-positive operating point TP_k, FP_k , and n represents the number of samples.
- **AUC-PR** is calculated by $\sum_{k=1}^n Prec_k \Delta Rec_k$, where k is k -th true-positive and false-positive operating point $Prec_k, Rec_k$, n indicates the number of samples.
- **F1 score**: average F1 score over different classes as $\frac{1}{N} \sum_{k=1}^N \frac{2P_k \cdot R_k}{P_k + R_k}$, where N is the number of classes and P_k, R_k indicate the precision and recall for k -th class. This metric is more sensitive to the results for classes where samples are fewer.
- **Recall score**: average Recall score over different classes as $\frac{1}{N} \sum_{k=1}^N \frac{TP_k}{TP_k + FN_k}$.

We perform 10-fold cross-validation on all datasets, and for each iteration, two blocks will be selected as the validation set and the test set respectively. For binary and multi-type classification tasks, we select the best model based on the AUC-ROC of the validation set. In addition, we select the optimal model based on the F1 score of the validation set for the multi-class classification task.

3) *Smoothed Relations*: To effectively capture the consistent semantic information in DRKG and reduce the negative impact of noisy interactions and irrelevant edges on downstream tasks, we blur the relations according to the semantic similarity of the original relationship. As shown in Table I, we filter out all entities except drug, gene (i.e., target), and disease, and classify the relationships between them into three categories (i.e., *positive*, *neutral*, and *negative*). Take the original relations E+ and A+ as an example, they are similar as the semantic “*activation*” and as the blurred association

Drug_positive_Gene, which represent drugs have a role in increasing the expression of genes. In Fig. 4, the distribution of blurred relations of the smoothed KG is shown. We observe that the sparse links are blurred, which reduces the negative influence of noisy interactions.

B. Data Preprocessing

1) *Sample Details*: For the DrugBank and DrugCentral datasets in predicting DTIs, we ensure the positive and negative samples for each drug are balanced by generating them randomly. Following SumGNN [3], we keep the same settings by stratified sampling DDI from DrugBank for cross-validation. For the TWOSIDES dataset, we follow the method in [47] to generate negative counterparts for every positive sample from the unknown set of drug pairs. We perform 10-fold cross-validation and select the best model based on the AUC-ROC of the validation set. The average performance and standard deviation evaluated on the test set are reported. **To avoid data leakage**, we removed observable DTI and DDI facts from the external KG in advance.

2) *Entities Mapping*: DRKG provides the IDs for each entity that can link to widely used external databases (e.g., DrugBank IDs for drugs, Gene IDs for targets/proteins). Thus we link drugs and targets from the four datasets to corresponding entities in DRKG by using their IDs.

3) *Pre-defined Metapaths of DRKG*: To model relational paths and semantic information of DRKG efficiently, we define a set of reasonable metapaths for constructing semantic subgraphs. Specifically, we focus on the prediction of relations *drug-target interaction* (DTI) and *drug-drug interaction* (DDI). For DTI prediction, we extract semantic subgraphs

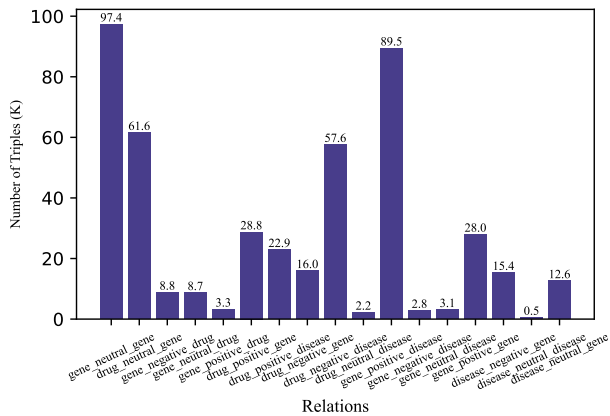


Fig. 4. The relation distribution of smoothed KG.

TABLE IV

THE PERFORMANCE ON DRUGBANK AND DRUGCENTRAL FOR DTI PREDICTION. THE **BOLDFACE** DENOTES THE HIGHEST SCORE AND UNDERLINE INDICATES THE SECOND HIGHEST SCORE.

Methods	DrugBank		DrugCentral	
	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR
GCN-KG	80.41 ± 0.15	79.33 ± 0.21	84.66 ± 0.32	83.88 ± 0.15
RotatE	77.65 ± 0.31	75.99 ± 0.13	81.18 ± 0.62	81.03 ± 0.11
GraIL	80.54 ± 0.17	81.37 ± 0.37	82.74 ± 0.45	82.89 ± 0.57
TriModel	81.23 ± 0.13	81.85 ± 0.22	80.91 ± 0.21	81.59 ± 0.55
SNRI	81.33 ± 0.39	82.09 ± 0.33	82.74 ± 0.45	82.89 ± 0.57
AdaProp	81.45 ± 0.15	82.01 ± 0.24	83.65 ± 0.22	83.27 ± 0.36
KG-MTL	82.55 ± 0.31	81.79 ± 0.52	84.39 ± 0.55	83.13 ± 0.74
KGE-NFM	<u>82.71 ± 0.22</u>	<u>82.09 ± 0.52</u>	86.34 ± 0.16	84.65 ± 0.14
BioKDN	84.90 ± 0.35	84.52 ± 0.44	88.79 ± 0.23	87.36 ± 0.19
ours w/o SRL	82.32 ± 0.13	82.55 ± 0.11	86.26 ± 0.07	85.24 ± 0.12
ours w/o SSP	81.78 ± 0.23	83.19 ± 0.21	86.54 ± 0.14	86.28 ± 0.15
ours w/o MI	82.05 ± 0.09	82.18 ± 0.18	86.34 ± 0.09	86.98 ± 0.17

using all 1-hop metapaths between drug and target (i.e., gene). Similarly, in DDI prediction, we adopt 1-hop metapaths between drug and drug as the relational schemes to extract semantic subgraphs. We illustrate some pre-defined metapaths in Table II. DRKG is a large-scale KG that integrates other widely-used biomedical KGs (e.g., Hetionet and GNBR). Thus DRKG contains 5,874,261 triples with 107 relations, which covers the relations from other KGs. We have categorized the biomedical relations from DRKG into *positive*, *neutral*, and *negative* in Table I. Thus we can map the relations of other biomedical KGs (e.g., Hetionet) directly using this mapping table, without re-manual processing.

C. Implementation Details of BioKDN

All the experiments in this work were conducted on a Linux server with an Intel Xeon(R) Platinum 8255C Processor (12 vCPU @2.5GHz), 43GB of RAM, and 2 RTX 2080Ti cards (11GB of RAM each). We implement BioKDN in Pytorch with Python 3.8.5.

To avoid data leakage, we removed observable DTI and DDI facts from the KG in advance. The average performance and standard deviation evaluated on the test set are reported on Table IV and Table V. We adopt the Xavier initializer [48] to initialize the model parameters and optimize BioKDN with Adam [49]. The grid search is applied

TABLE V
THE RESULTS COMPARISON ON DRUGBANK AND TWOSIDES FOR DDI PREDICTION. THE BEST IS MARKED WITH **BOLDFACE** AND THE SECOND BEST IS WITH UNDERLINE.

Methods	DrugBank (Multi-class)		TWO SIDES (Multi-label)	
	Micro-F1	Micro-Rec	AUC-ROC	AUC-PR
GCN-KG	79.34 ± 0.16	82.56 ± 0.23	85.22 ± 0.32	82.57 ± 0.12
RotatE	76.41 ± 0.11	80.71 ± 0.15	85.92 ± 0.18	82.69 ± 0.21
GraIL	83.39 ± 0.35	76.11 ± 0.46	83.72 ± 0.18	80.73 ± 0.09
KGNN	76.13 ± 0.32	74.62 ± 0.42	86.97 ± 0.23	82.71 ± 0.41
SNRI	84.57 ± 0.13	82.13 ± 0.19	85.24 ± 0.54	81.75 ± 0.27
AdaProp	83.26 ± 0.10	82.01 ± 0.12	86.41 ± 0.13	82.53 ± 0.21
SumGNN	<u>85.58 ± 0.10</u>	<u>82.79 ± 0.19</u>	<u>87.42 ± 0.16</u>	82.65 ± 0.07
BioKDN	87.51 ± 0.11	85.55 ± 0.13	88.62 ± 0.09	84.73 ± 0.12
ours w/o SRL	85.67 ± 0.19	82.31 ± 0.13	86.21 ± 0.21	82.36 ± 0.25
ours w/o SSP	84.59 ± 0.15	83.01 ± 0.19	85.38 ± 0.06	83.12 ± 0.22
ours w/o MI	85.07 ± 0.31	82.87 ± 0.24	85.96 ± 0.05	82.81 ± 0.32

to retrieve optimal hyper-parameters. We tune the learning rate among $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, the embedding size in $\{32, 64, 128, 256, 512\}$, the size of subgraphs in $\{1, 2, 3, 4\}$ -hop and the pruning threshold π in $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. Besides, we set weight $\lambda = 0.5$ of the loss function. In addition, we search the weight λ of the contrastive loss among $\{0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ and the temperature τ among $\{0.5, 1, 2\}$. The final used hyperparameters of BioKDN for each prediction task are shown in Table III.

D. Baselines

To verify the performance of BioKDN, we compare it against various baselines as follows:

- **GCN-KG** and **RotatE** [37] adopted the graph neural network [41] and relational rotation in complex space to learn the embedding of entities and relations from the DRKG, and then predicted the links for DTI or DDI using the embeddings.
- **GraIL** [29] utilized a local subgraph for inductive relation prediction on KGs. To model neighboring relations effectively, **SNRI** [30] adopted the semantic subgraphs by extracting semantic relational paths to learn informative embedding. To filter out irrelevant entities, **AdaProp** [32] designed an incremental sampling mechanism to preserve the nearby targets.
- **TriModel** [17] and **KGE-NFM** [1] developed new methods to learn the relational representation of entities and relations, then predict the unknown links for DTI. Additionally, **KG-MTL** [10] proposed a global knowledge enhanced multi-task method to predict unknown DTI.
- **KGNN** [2] aggregated neighborhood information for each node from their local receptive via GNN on the biomedical knowledge graph for link prediction of relation DDI. **SumGNN** [3] focused on extracting information from the local subgraph in a learnable way and considered the DDI prediction as multi-type and multi-class classification tasks.

Our approach focuses on subgraph-based reasoning on the biomedical knowledge graph to robustly predict molecular interactions. From the technical perspective, we compare the SOTA models based on knowledge graph representation (e.g.,

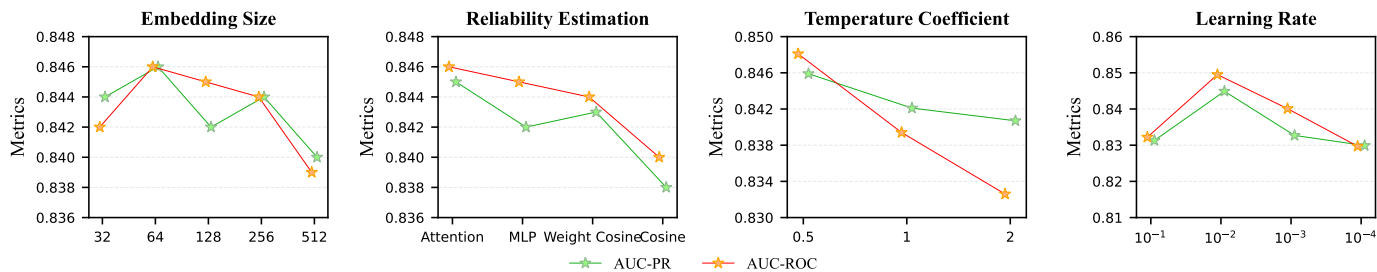


Fig. 5. Hyper-parameter sensitivity analysis of DTI prediction based on the DrugBank dataset.

RotatE, 2018) and subgraph-based robust GNN (e.g., AdaProp, 2023 and SNRI, 2022). Meanwhile, we compare the optimal models (e.g., SumGNN, 2021, KGE-NFM, 2021, and KG-MTL, 2022) for KG-based molecular interaction prediction from the application perspective. This study provides a detailed and systematic comparison of notable methods in subgraph reasoning alongside KG-enhanced molecular interaction prediction models.

E. Implementation Details of Baselines

For baselines, we adopt official code packages from the authors for RotatE³ [37], GraIL⁴ [29], SNRI⁵ [30], AdaProp⁶ [32], TriModel⁷ [17], KGE-NFM⁸ [1], KGNN⁹ [2] and SumGNN¹⁰ [3]. We use the GCN module of the DGL¹¹ library on external DRKG [24] to implement GCN-KG and capture the embeddings of all molecules. For fair comparisons, we tune them for optimal performance.

F. Comparison with Baselines

We report the performance of our model and baselines for predicting molecular interactions of the relations DTI and DDI in Table IV and Table V, respectively. As shown in Table IV, we observe BioKDN achieves the best prediction results in the DTI task on both DrugBank and DrugCentral datasets. Specifically, BioKDN improves the AUC-ROC and AUC-PR by at least 2.19% and 2.43% respectively on the DrugBank dataset and achieves the 2.45% and 2.71% absolute increase over the best baseline on DrugCentral data. For the prediction of DDI, we find that the boosts of BioKDN on DrugBank for the multi-class task in Micro-F1 and Micro-Recall score up to 1.93% and 2.76% respectively. The performance of BioKDN on the TWOSIDES dataset has achieved 1.2%, and 2.02% improvement in AUC-ROC and AUC-PR compared with the best baseline.

Furthermore, we have the following observations: (1) Compared with RotatE, the GCN-KG utilizing the neighboring

TABLE VI
THE PERFORMANCE OF BIOKDN ON DRUGBANK FOR DTI PREDICTION WITH VARIOUS KGE METHODS.

Metrics	TransE	RotatE	DistMult
AUC-ROC	84.01	84.88	84.23
AUC-PR	83.87	84.37	84.11

information and structures achieves better performance on DTI and DDI prediction, which indicates that the neighboring structure benefits the downstream prediction tasks. (2) Compared with TriModel and KGNN, the SNRI using local semantic relations performs better than them on DrugBank (3) Among the subgraph-based methods (i.e., GraIL, SNRI, and SumGNN), BioKDN can achieve superior improvement. This is because the noisy interactions in the local subgraph may make it hard for models to learn reliable neighborhood information effectively, degrading their performance. (4) BioKDN outperforms AdaProp in pruning task-irrelevant facts, indicating that knowledge-enhanced smoothing semantics can effectively improve the performance of denoising noisy interactions. In short, the BioKDN reduces the negative influence of noise by denoising unreliable interactions in a learnable way and preserving knowledge-enhanced smoothing semantics, resulting in superior results.

G. Ablation Study

To investigate the impact of each module in BioKDN, we perform an ablation study on all datasets for DTI and DDI prediction by removing: (i) structure reliability learning (called **ours w/o SRL**), (ii) smooth semantic preservation (called **ours w/o SSP**), (iii) Mutual information (MI) maximization of dual-view subgraphs (called **ours w/o MI**), respectively. We can observe that all variants of BioKDN perform worse than the original model in Table IV and Table V, which verifies the effectiveness of each component.

ours w/o SRL. We observe a significant reduction in performance across all datasets for DTI and DDI prediction after removing the structure reliability learning module. This suggests that the unreliable subgraph structure is less expressive for downstream tasks and cannot effectively eliminate the negative influence of noisy interactions. In contrast, a complete

³github.com/DeepGraphLearning/KnowledgeGraphEmbedding

⁴<https://github.com/kkteru/grail>

⁵<https://github.com/Tebmer/SNRI>

⁶<https://github.com/LARS-research/AdaProp>

⁷The official source is not accessed, we requested the code from the author at: <https://github.com/xiaomingaaa/BioKDN/baseline/trimodel.py>

⁸<https://zenodo.org/records/5500305>

⁹<https://github.com/xzenglab/KGNN>

¹⁰<https://github.com/yueyu1030/SumGNN>

¹¹<https://github.com/dmlc/dgl>

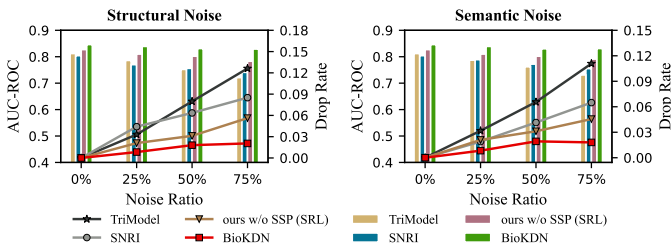


Fig. 6. Performance over various noisy types with different ratios for DTI prediction. The bar represents the AUC-ROC and the line indicates the degradation ratio on the value. The smaller the drop rate, the better.

reliable structure can improve the performance of the original model by removing possible noise and retaining trustworthy interactions.

ours w/o SSP. From the results reported in Table IV and Table V, we notice a significant degradation in performance on all datasets when omitting the semantic subgraph of the smooth semantic preservation module. This observation demonstrates the effectiveness of the semantic subgraph, extracted by pre-defined metapaths, in keeping the consistent semantics of task-relevant relations. Intuitively, **ours w/o SRL** together with **ours w/o SSP** demonstrate the effectiveness of denoising unreliable interaction from the local structure and smooth semantic views.

ours w/o MI. Additionally, removing the MI maximization from BioKDN results in a reduction of performance on all datasets for both tasks. The results demonstrate that the learned reliable substructure, guided by knowledge-enhanced smoothing semantics and the removal of task-irrelevant relations, is effective in enhancing the performance of KG-based methods. These findings of the various variants show that the original model BioKDN can effectively enhance the superior performance of the link prediction tasks.

H. Performance on Different KG initializers

We used knowledge graph embedding (KGE) methods to initialize the vector representation of entities and relations. To find the best KGE methods, we performed BioKDN on TransE [50], RotatE [37], and DistMult [51] for the DTI prediction task. The performance of them is reported in Table VI. We can observe that RotatE achieves the best result, which shows that modeling complex relations in advance is beneficial to prediction performance. Therefore, we adopt RotatE¹² to learn the vector representation of KG elements (e.g., drug/target molecules).

I. Hyper-parameter Sensitivity

1) *Impact of embedding size:* We explore the effect of hidden embedding size by varying it from 32 to 516. The left of Fig. 5 depicts the changing trend of the AUC-ROC and AUC-PR values on the DrugBank dataset evaluated on BioKDN. Based on the results, we observe that the AUC-ROC values of BioKDN variation across different embedding sizes collapsed into a hunchback shape. The reason could be

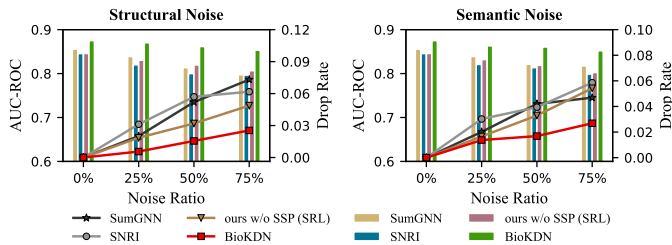


Fig. 7. Performance over various noisy types with different ratios for DDI prediction. The bar represents the Micro-F1 and the line indicates the degradation ratio on the value. The smaller the drop rate, the better.

that enough embedding size can represent more information, while the larger one will progressively introduce a lot of noise with the degradation of BioKDN.

2) *Impact of reliability estimation:* To investigate the impact of various reliability estimation functions $F(\cdot, \cdot)$ defined in the Section **Structure Reliability Learning**, we conduct experiments by varying the estimation types to the *Attention*, *MLP*, *Weighted_Cosine*, and *Cosine*. As illustrated in Fig. 5, the *Attention* with linear attention modeling the reliability weight between nodes set has the best performance. The *MLP* adopts a 2-layer preceptor, which achieves a secondary best result by learning the weight of the node pairs from the extracted local subgraph. This is because the attention mechanism can better model the importance of node pairs compared with the *MLP*, which improves the effectiveness of estimating the reliable edges. Additionally, the parametric *Weighted_Cosine* is better than the non-parametric *Cosine* indicating the learned weight guided by downstream tasks is more efficient.

3) *Impact of temperature coefficient τ :* We evaluate the influence of different temperatures τ for BioKDN in Fig. 5. We can see that as τ becomes larger, the effect of BioKDN gradually decreases. This indicates that a large τ will cause BioKDN to be unstable, thus making the effect worse. Therefore, we set $\tau = 0.5$ for the DTI task.

4) *Impact of learning rate:* We conduct experiments to study the influence of the learning rate by varying it to be 10^{-1} , 10^{-2} , 10^{-3} , and 10^{-4} . The results show a hunchback shape in Fig. 5 and we observe that BioKDN achieves the best result when 10^{-2} is adopted. Finally, we set the learning rate as 10^{-2} for downstream tasks.

5) *Impact of λ and π :* We conduct experiments to study the impact of the pruning threshold π and the loss weight λ of MI module shown in Table VII. λ is tested on various scales when $\pi = 0.1$ and π is evaluated on various values when $\lambda = 0.5$. We can observe that the AUC-ROC values of BioKDN across different λ and π collapsed into a hunchback shape, which shows a balanced coordination is beneficial. Thus the performance of BioKDN was reported under $\lambda = 0.5$ and $\pi = 0.5$.

J. Robustness of Interaction Noises

1) *Performance on Contaminated KGs:* To verify the effectiveness of structure reliability learning and smooth semantic preservation modules in denoising interaction noise,

¹²We have provided details in Section III-B

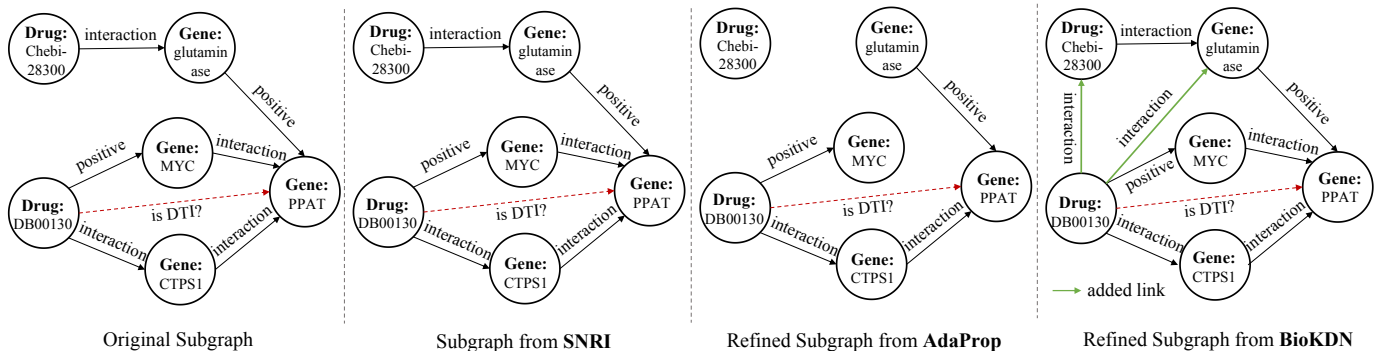


Fig. 8. The original noisy subgraph and the refined subgraphs constructed from SNRI, AdaProp, and BioKDN.

TABLE VII
THE AUC-ROC VALUES OF BIOKDN ON DRUGBANK FOR DTI PREDICTION ACROSS DIFFERENT λ AND π .

Parameter	0.1	0.3	0.5	0.7	0.9	1
λ	83.11	83.97	84.78	84.26	84.16	83.96
π	84.13	84.01	84.83	83.67	82.21	-

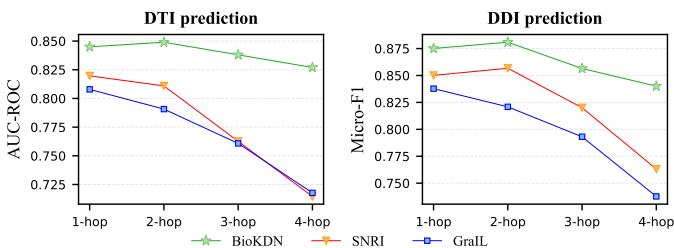


Fig. 9. The performance of BioKDN on Drugbank compared with SNRI and GraIL on various sized subgraphs.

we generate different proportions of *structural* and *semantic* negative interactions (i.e., 25%, 50%, and 75%) to contaminate the training knowledge graph. The reported performance of various models is evaluated on the unchanged test set shown in Fig. 6. Structural noises are generated by sampling unknown triples from all possible *entity-relation-entity* combinations, while semantic noises are sampled from all missing triples with reasonable *entity-relation-entity* schemes (e.g., the scheme (drug, *drug_disease_treat*, disease)). We then perform BioKDN and its variants (i.e., **ours w/o SSP** for structural noise and **ours w/o SRL** for semantic noise) on the noisy KG and compare their performance with semantic subgraph-based SNRI and global KG-based TriModel. As shown in Fig. 6 and Fig. 7, the AUC-ROC and Micro-F1 values on DrugBank for DTI and DDI prediction tasks are reported. Additionally, the corresponding performance degradation ratios are presented.

We observe that as more noise is added, the performance of all models deteriorates for both structural and semantic experiments. This is because the introduced noise weakens the expressive power of the aggregated neighbor information. However, BioKDN and its variants exhibit smaller degradation than other methods for both types of noise. The variants **ours w/o SSP** and **ours w/o SRL** show lighter changes as

the noise increases than SNRI and TriModel on structural and semantic noise respectively, which indicates the structure reliability learning and smooth semantic subgraph can effectively denoise noisy interactions and ignore task-irrelevant relations. Furthermore, the gaps between BioKDN and SNRI or SumGNN grew larger with increasing noise. This is because BioKDN pays more attention to the informative interactions by maximizing the mutual information between the refined and semantic subgraphs. This phenomenon shows that BioKDN can effectively mitigate noises using reliable structure and smoothed semantics.

2) *Impact of k-hop Subgraph*: To evaluate the global effectiveness of BioKDN, we conduct experiments to compare the impact of *k-hop* subgraphs with baselines SNRI [30] and GraIL [29]. As shown in Fig. 9, BioKDN experiences less performance degradation than SNRI and GraIL in both DTI and DDI predictions. This occurs because as the subgraph size increases, more noise is introduced, which can negatively impact model performance. However, the automatic denoising mechanism of BioKDN helps minimize the adverse effects of noisy interactions. This finding demonstrates that BioKDN can effectively handle noise by leveraging a reliable structure and smooth semantics.

K. Case Study

We conduct a case study for predicting DTI relation between the drug *DB00130* and the gene *PPAT* to demonstrate the effectiveness of BioKDN, shown in Fig. 8. In the original subgraph, the different entities *DB00130* and *Chebi:28300* represent the same drug *L-Glutamine*, but they are unalignment, resulting in the absence of an interactive edge (i.e., *DB00130, same_as, glutaminase*). By learning the reliable structure of the original subgraph, BioKDN effectively establishes a connection between the drug *DB00130* and the gene *glutaminase*, which brings favorable information for predicting the DTI relation between *DB00130* and *PPAT*. However, SNRI directly reasoning on the original subgraph without considering the noises. In addition, AdaProp uses a novel sample strategy to filter out task-irrelevant edges within the local subgraph and prune them, potentially ignoring reliable information and failing to build missing edges to bring information interaction. This case shows that BioKDN can effectively learn a reliable structure, enhancing the performance of link prediction.

V. CONCLUSION

In this paper, we proposed a biomedical knowledge-enhanced denoising network, called BioKDN, to mitigate the negative influence caused by the noise and inconsistent semantics hiding in the KG. BioKDN designed both structure reliability learning and semantic smoothing modules to extract reliable structure and keep semantic relations consistent for robust DTI and DDI predictions. Our experiments on four datasets and contaminated KGs for DTI and DDI predictions demonstrate that BioKDN significantly outperforms several existing state-of-the-art methods. These results verify the robustness of our model against interaction noises. However, the semantic smoothing strategy of BioKDN is limited when applied to other domains for denoising. In the future, we will design a generalized denoising model based on KGs to remove noisy facts in the real world rather than just being irrelevant to downstream tasks. In addition, the effectiveness of reliable structure learning motivates us to design a robust relation discovery method for broader applications. Based on this, we can construct a high-quality and reliable biomedical KG as the foundation for accelerating drug discovery.

ACKNOWLEDGEMENT

The work was supported by National Natural Science Foundation of China (Grant Nos. 62425204, 62122025, U22A2037, 62250028, 62272151, 61972138, 62106073, 62206089, 62202413, 62450002, 62432011), Hunan Provincial Natural Science Foundation of China (Grant Nos. 2024JJ4015, 2023JJ40178, 2022JJ20016), the science and technology innovation Program of Hunan Province (Grant Nos. 2022RC1099), Excellent Youth Funding of Hunan (Grant Nos. 23B0129), NSF under grants III-2106758, and POSE-2346158.

REFERENCES

- [1] Q. Ye, C.-Y. Hsieh, Z. Yang, Y. Kang, J. Chen, D. Cao, S. He, and T. Hou, "A unified drug-target interaction prediction framework based on knowledge graph and recommendation system," *Nature Communications*, vol. 12, no. 1, p. 6775, 2021.
- [2] X. Lin, Z. Quan, Z.-J. Wang, T. Ma, and X. Zeng, "Kgnn: Knowledge graph neural network for drug-drug interaction prediction." in *IJCAI*, 2020, pp. 2739–2745.
- [3] Y. Yu, K. Huang, C. Zhang, L. M. Glass, J. Sun, and C. Xiao, "Sumgnn: multi-typed drug interaction prediction via efficient knowledge graph summarization," *Bioinformatics*, vol. 37, no. 18, pp. 2988–2995, 2021.
- [4] X. Lin, L. Dai, Y. Zhou, Z.-G. Yu, W. Zhang, J.-Y. Shi, D.-S. Cao, L. Zeng, H. Chen, B. Song *et al.*, "Comprehensive evaluation of deep and graph learning on drug–drug interactions prediction," *Briefings in Bioinformatics*, vol. 24, no. 4, p. bbad235, 2023.
- [5] Q. He, J. Yang, and B. Shi, "Constructing knowledge graph for social networks in a deep and holistic way," in *WWW*, 2020, pp. 307–308.
- [6] K. Liang, L. Meng, M. Liu, Y. Liu, W. Tu, S. Wang, S. Zhou, X. Liu, F. Sun, and K. He, "A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–20, 2024.
- [7] X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T.-S. Chua, "Explainable reasoning over knowledge graphs for recommendation," in *AAAI*, no. 01, 2019, pp. 5329–5336.
- [8] Y. Yang, C. Huang, L. Xia, and C. Li, "Knowledge graph contrastive learning for recommendation," in *SIGIR*, 2022, pp. 1434–1443.
- [9] J. Guo, Y. Zhou, P. Zhang, B. Song, and C. Chen, "Trust-aware recommendation based on heterogeneous multi-relational graphs fusion," *Information Fusion*, vol. 74, pp. 87–95, 2021.
- [10] T. Ma, X. Lin, B. Song, S. Y. Philip, and X. Zeng, "Kg-ntl: Knowledge graph enhanced multi-task learning for molecular interaction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 7, pp. 7068–7081, 2022.
- [11] J. Wu, K. He, R. Mao, C. Li, and E. Cambria, "Megacare: Knowledge-guided multi-view hypergraph predictive framework for healthcare," *Information Fusion*, vol. 100, p. 101939, 2023.
- [12] Y. Wu, Y. Chen, Z. Yin, W. Ding, and I. King, "A survey on graph embedding techniques for biomedical data: Methods and applications," *Information Fusion*, vol. 100, p. 101909, 2023.
- [13] H. Zhang, J. Wu, S. Liu, and S. Han, "A pre-trained multi-representation fusion network for molecular property prediction," *Information Fusion*, vol. 103, p. 102092, 2024.
- [14] Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, W. Kuang, J. Peng, L. Chen, and J. Zeng, "A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information," *Nature Communications*, vol. 8, no. 1, p. 573, 2017.
- [15] F. Wan, L. Hong, A. Xiao, T. Jiang, and J. Zeng, "Neodti: neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions," *Bioinformatics*, vol. 35, no. 1, pp. 104–111, 2019.
- [16] M. Saebi, S. Kreig, C. Zhang, M. Jiang, T. Kajdanowicz, and N. V. Chawla, "Heterogeneous relational reasoning in knowledge graphs with reinforcement learning," *Information fusion*, vol. 88, pp. 12–21, 2022.
- [17] S. K. Mohamed, V. Nováček, and A. Nounu, "Discovering protein drug targets using knowledge graph embeddings," *Bioinformatics*, vol. 36, no. 2, pp. 603–610, 2020.
- [18] R. Celebi, H. Uyar, E. Yasar, O. Gumus, O. Dikenelli, and M. Dumontier, "Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings," *BMC Bioinformatics*, vol. 20, no. 1, pp. 1–14, 2019.
- [19] X. Su, Z. You, D. Huang, L. Wang, L. Wong, B. Ji, and B. Zhao, "Biomedical knowledge graph embedding with capsule network for multi-label drug-drug interaction prediction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 6, pp. 5640–5651, 2022.
- [20] K. Liang, Y. Liu, S. Zhou, W. Tu, Y. Wen, X. Yang, X. Dong, and X. Liu, "Knowledge graph contrastive learning based on relation-symmetrical structure," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 1, pp. 226–238, 2023.
- [21] J. Huang, Y. Zhao, W. Hu, Z. Ning, Q. Chen, X. Qiu, C. Huo, and W. Ren, "Trustworthy knowledge graph completion based on multi-sourced noisy data," in *WWW*, 2022, pp. 956–965.
- [22] J. Zhu, C. Huang, and P. De Meo, "Dfmke: A dual fusion multi-modal knowledge graph embedding framework for entity alignment," *Information Fusion*, vol. 90, pp. 111–119, 2023.
- [23] J. Pujara, E. Augustine, and L. Getoor, "Sparsity and noise: Where knowledge graph embeddings fall short," in *EMNLP*, 2017, pp. 1751–1756.
- [24] V. N. Ioannidis, X. Song, S. Manchanda, M. Li, X. Pan, D. Zheng, X. Ning, X. Zeng, and G. Karypis, "Drkg - drug repurposing knowledge graph for covid-19," <https://github.com/gnn4dr/DRKG/>, 2020.
- [25] K. Ma, H. Fu, T. Liu, Z. Wang, and D. Tao, "Deep blur mapping: Exploiting high-level semantics by deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5155–5166, 2018.
- [26] T. Guo, C. Xu, B. Shi, C. Xu, and D. Tao, "Smooth deep image generator from noises," in *AAAI*, no. 01, 2019, pp. 3731–3738.
- [27] K. Kishan, R. Li, F. Cui, and A. R. Haake, "Predicting biomedical interactions with higher-order graph convolutional networks," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 2, pp. 676–687, 2021.
- [28] Y. Chen, T. Ma, X. Yang, J. Wang, B. Song, and X. Zeng, "Muffin: multi-scale feature fusion for drug-drug interaction prediction," *Bioinformatics*, vol. 37, no. 17, pp. 2651–2658, 2021.
- [29] K. Teru, E. Denis, and W. Hamilton, "Inductive relation prediction by subgraph reasoning," in *ICML*, 2020, pp. 9448–9457.
- [30] X. Xu, P. Zhang, Y. He, C. Chao, and C. Yan, "Subgraph neighboring relations infomax for inductive link prediction on knowledge graphs," in *IJCAI*, 2022, pp. 2341–2347.
- [31] K. Liang, L. Meng, S. Zhou, W. Tu, S. Wang, Y. Liu, M. Liu, L. Zhao, X. Dong, and X. Liu, "Mines: Message intercommunication for inductive relation reasoning over neighbor-enhanced subgraphs," in *AAAI*, vol. 38, no. 9, 2024, pp. 10 645–10 653.
- [32] Y. Zhang, Z. Zhou, Q. Yao, X. Chu, and B. Han, "Adaprop: Learning adaptive propagation for graph neural network based knowledge graph reasoning," in *SIGKDD*, 2023, pp. 3446–3457.

- [33] Z. Fan, K. Xu, Z. Dong, H. Peng, J. Zhang, and P. S. Yu, “Graph collaborative signals denoising and augmentation for recommendation,” in *SIGIR*, 2023, pp. 2037–2041.
- [34] Y. Quan, J. Ding, C. Gao, L. Yi, D. Jin, and Y. Li, “Robust preference-guided denoising for graph based social recommendation,” in *WWW*, 2023, pp. 1097–1108.
- [35] C. Tian, Y. Xie, Y. Li, N. Yang, and W. X. Zhao, “Learning to denoise unreliable interactions for graph collaborative filtering,” in *SIGIR*, 2022, pp. 122–132.
- [36] Y. Gao, Y. Du, Y. Hu, L. Chen, X. Zhu, Z. Fang, and B. Zheng, “Self-guided learning to denoise for robust recommendation,” in *SIGIR*, 2022, pp. 1412–1422.
- [37] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang, “Rotate: Knowledge graph embedding by relational rotation in complex space,” in *ICLR*, 2018, pp. 1–18.
- [38] X. Zhang and M. Zitnik, “Gnnguard: Defending graph neural networks against adversarial attacks,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 9263–9275, 2020.
- [39] Z. Li, X. Sun, Y. Luo, Y. Zhu, D. Chen, Y. Luo, X. Zhou, Q. Liu, S. Wu, L. Wang *et al.*, “Gslb: the graph structure learning benchmark,” *Advances in Neural Information Processing Systems*, vol. 36, pp. 1–13, 2024.
- [40] E. Jang, S. Gu, and B. Poole, “Categorical reparameterization with gumbel-softmax,” in *ICLR*, 2017, pp. 1–12.
- [41] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *ICLR*, 2017, pp. 1–13.
- [42] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, “Modeling relational data with graph convolutional networks,” in *ESWC*, 2018, pp. 593–607.
- [43] A. v. d. Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” *arXiv preprint arXiv:1807.03748*, 2018.
- [44] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda *et al.*, “Drugbank 5.0: a major update to the drugbank database for 2018,” *Nucleic Acids Research*, vol. 46, no. D1, pp. D1074–D1082, 2018.
- [45] S. Avram, T. B. Wilson, R. Curpan, L. Halip, A. Borota, A. Bora, C. G. Bologna, J. Holmes, J. Knockel, J. J. Yang *et al.*, “Drugcentral 2023 extends human clinical data and integrates veterinary drugs,” *Nucleic Acids Research*, vol. 51, no. D1, pp. D1276–D1287, 2023.
- [46] N. P. Tatonetti, P. P. Ye, R. Daneshjou, and R. B. Altman, “Data-driven prediction of drug effects and interactions,” *Science Translational Medicine*, vol. 4, no. 125, pp. 125–125, 2012.
- [47] M. Zitnik, M. Agrawal, and J. Leskovec, “Modeling polypharmacy side effects with graph convolutional networks,” *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, 2018.
- [48] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *ICAI*, 2010, pp. 249–256.
- [49] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *ICLR*, 2014, pp. 1–15.
- [50] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, and O. Yakhnenko, “Translating embeddings for modeling multi-relational data,” in *NeurIPS*, 2013, p. 2787–2795.
- [51] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, “Embedding entities and relations for learning and inference in knowledge bases,” *ICLR*, pp. 1–12, 2015.



Yujie Chen received her Master’s degree in Computer Science from Hunan University, Changsha, China, in 2022. She is presently pursuing her PhD at the College of Computer Science and Electronic Engineering, also at Hunan University. Her research primarily focuses on knowledge graph representation, natural language processing, and graph machine learning, particularly their applications in the scientific domain. She has published a paper in this area, notably in the journal *Bioinformatics*.



Wen Tao received her M.E. degree in Computer Science and Technology from the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China, in 2024. She received her B.E. degree in Computer Science and Technology from the School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China, in 2021. Her research interests include graph machine learning, drug discovery, and explainable AI.



Dashun Zheng obtained a Bachelor of Management degree in Information Management and Information Systems from Xinjiang University of Finance and Economics in 2022, and a Master’s degree in Big Data and the Internet of Things from the Macao Polytechnic Institute in 2024. He is currently pursuing a Ph.D. in Computer Application Technology at the Macao Polytechnic Institute. His research interests include Natural Language Processing, AI for Healthcare, and Multimodal Learning.



Xuan Lin received the Ph.D. degree in computer science from Hunan University, Changsha, China, in 2021. He is currently a lecturer at the College of Computer Science, Xiangtan University, Xiangtan, China. He was a visiting scholar with the University of Illinois at Chicago, from 2019 to 2020. His main research interests include machine learning, graph neural networks, and bioinformatics. He has published several research papers in these fields including *IEEE Transactions on Knowledge and Data Engineering*, *Briefings in Bioinformatics*, *IJCAI*, *AAAI*, *ECAI*, *BIBM*, etc.



Tengfei Ma received the Master’s degree in computer science from Hunan University, Changsha, China, in 2022. He is currently a PhD student at the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China. His main research interests include graph machine learning, knowledge graph representation, and their applications in science. He has published several research papers in these fields including *IEEE Transactions on Knowledge and Data Engineering*, *Bioinformatics*, *Knowledge-Base Systems*, *IJCAI*, etc.



Patrick Cheong-Iao Pang is currently an Assistant Professor and a Ph.D. supervisor in the Faculty of Applied Sciences at Macao Polytechnic University (MPU). He received his Ph.D. from the University of Melbourne and held appointments at the University of Melbourne and Victoria University before joining MPU. His research interests include natural language processing (NLP) applications, and digital health and educational technologies. His interdisciplinary work seeks to build AI applications for health and education, thanks to his many years of industry experience in designing and building IT systems.



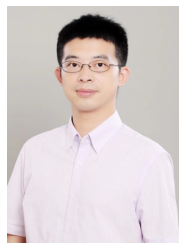
Yiping Liu received a B.Eng. degree in electrical engineering and automation and a Ph.D. degree in control theory and control engineering from the China University of Mining and Technology, Xuzhou, China in 2012 and 2017, respectively. He is currently an Associate Professor at the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China. During 2018–2020, he was a Research Assistant Professor at the Department of Computer Science and Intelligent Systems, at Osaka Prefecture University, Sakai, Japan. During 2016–2017, he was a Visiting Scholar with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, USA. His research interests include evolutionary computation, multi-objective optimization, and machine learning.



Yijun Wang received the B.E. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2014 and 2019, respectively. She is currently an Assistant Professor with the School of Computer Science and Electronic Engineering, at Hunan University, Changsha, China. She has authored or co-authored more than 10 papers in related conferences and journals. Her research interests include multimedia understanding, natural language processing, and data mining.



Longyue Wang (Senior Member, IEEE) has received a B.S. degree in network engineering in 2011, and an M.S. degree in software engineering in 2014. From 2015 to 2018, he pursued a Ph.D. at Dublin City University. He is a senior research fellow at Tencent AI Lab now. He has studied and practiced in a broad field of Artificial Intelligence, especially in Large Language Models, Multimodal, Language Agents, Natural Language Processing, Machine Translation, Deep Learning, and AI for Science.



Bosheng Song received a Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2015. He spent eighteen months working with the Research Group on Natural Computing, University of Seville, Seville, Spain, from November 2013 to May 2015. He worked as a post-doctoral researcher with the School of Automation, Huazhong University of Science and Technology, from 2016 to 2019. He is currently an associate professor with the College of Information Science and Engineering, Hunan University, Changsha, China. His current research interests include membrane computing and bioinformatics.



Xiangxiang Zeng (Senior Member, IEEE) received a Ph.D. degree in system engineering from the Huazhong University of Science and Technology, China, in 2011. He is a Yuelu distinguished professor at the College of Information Science and Engineering, Hunan University, Changsha, China. Before joining Hunan University, in 2019, he was with the Department of Computer Science, Xiamen University. He was a visiting scholar with Indiana University, The Chinese University of Hong Kong, Oklahoma State University, etc. His main research interests include computational intelligence, graph neural networks, and bioinformatics.



Philip S. Yu (Fellow, IEEE) received the B.S. degree in E.E. from National Taiwan University, the M.S. and Ph.D. degrees in E.E. from Stanford University, and the M.B.A. degree from New York University. He is a Distinguished Professor in Computer Science at the University of Illinois Chicago and holds the Wexler Chair in Information Technology. Before joining UIC, Dr. Yu was with IBM, where he was manager of the Software Tools and Techniques department at the Watson Research Center. His research interest is on big data, and artificial intelligence, including data mining, database and privacy. Dr. Yu is the recipient of ACM SIGKDD 2016 Innovation Award, the IEEE Computer Society's 2013 Technical Achievement Award, and the Research Contributions Award from IEEE Intl. Conference on Data Mining (ICDM) in 2003. He also received the VLDB 2022 Test of Time Award, ACM SIGSPATIAL 2021 10-year Impact Award, WSDM 2020 Honorable Mentions of the Test of Time Award, ICDM 2013 10-year Highest-Impact Paper Award, and the EDBT Test of Time Award (2014). He was the Editor-in-Chief of ACM Transactions on Knowledge Discovery from Data (2011-2017) and IEEE Transactions on Knowledge and Data Engineering (2001-2004). He is a Fellow of the ACM.