Learning to Denoise Unreliable Interactions for Link Prediction on Biomedical Knowledge Graph

Tengfei Ma[1], Yujie Chen[1], Wen Tao[1], Dashun Zheng[2], Xuan Lin[3], Patrick Cheong-Iao Pang[2], Yiping Liu[1], Yijun Wang[1], Bosheng Song[1], Xiangxiang Zeng[1]

[1]School of Computer Science and Engineering, Hunan University
[2]Faculty of Applied Sciences, Macao Polytechnic University
[2]Key Laboratory of Intelligent Computing and Information Processing, Xiangtan University
tfma@hnu.edu.cn, yjchen@hnu.edu.cn, taowen@hnu.edu.cn, p2212871@mpu.edu.mo, jack_lin@xtu.edu.cn, mail@patrickpang.net, ypliu@hnu.edu.cn, wyjun@hnu.edu.cn, boshengsong@hnu.edu.cn, xzeng@hnu.edu.cn

Abstract

Link prediction in biomedical knowledge graphs (KGs) aims at predicting unknown interactions between entities, including drug-target interaction (DTI) and drug-drug interaction (DDI), which is critical for drug discovery and therapeutics. Previous methods prefer to utilize the rich semantic relations and topological structure of the KG to predict missing links, yielding promising outcomes. However, all these works only focus on improving the predictive performance without considering the inevitable noise and unreliable interactions existing in the KGs, which limits the development of KG-based computational methods. To address these limitations, we propose a **Denoised L**ink **P**rediction framework, called **DenoisedLP**. DenoisedLP obtains reliable interactions based on the local subgraph by denoising noisy links in a learnable way, providing a universal module for mining underlying task-relevant relations. To collaborate with the smoothed semantic information, DenoisedLP introduces the semantic subgraph by blurring conflict relations around the predicted link. By maximizing the mutual information between the reliable structure and smoothed semantic relations, DenoisedLP emphasizes the informative interactions for predicting relation-specific links. Experimental results on real-world datasets demonstrate DenoisedLP outperforms state-of-the-art methods on DTI and DDI prediction tasks, and verify the effectiveness and robustness of denoising unreliable interactions on the contaminated KGs.

1 Introduction

Identifying missing links in biomedical knowledge graphs (KGs) is pivotal to drug discovery and therapeutics, including drug-target interaction (DTI) prediction (Ye et al. 2021) and drug-drug interaction (DDI) prediction (Lin et al. 2020; Yu et al. 2021). The success of link prediction based on KGs in social networks (He, Yang, and Shi 2020) and recommendations (Wang et al. 2019; Yang et al. 2022) encourages researchers to develop various computational methods to accelerate drug development. However, accurately recognizing the unknown interactions between various 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 edges (Luo et al. 2017; Wan et al. 2019). These methods adopted network-based models, which cannot model semantic relations between various entities (e.g., drug, pathway, disease). Subsequently, a line of works applied the knowledge graphs embedding methods to learn the semantic relations with low-dimensional embeddings for predicting DTI (Mohamed, Nováček, and Nounu 2020) and DDI (Celebi et al. 2019). However, these methods are limited to efficiently learning the topological structure of complex biomedical KGs. Recently, various models (Lin et al. 2020; Yu et al. 2021; Ma et al. 2022) based on hetergeneous graph neural networks have achieved promising results. These methods focus on learning the semantic knowledge and local structure of the neighboring relational paths, which can obtain the semantic relations and tractable pathways around the predicted link.

Despite their effectiveness, existing KG-based models suffer from noisy interactions and conflict relations in biomedical KGs. Most biomedical knowledge graphs are constructed from unstructured text and multi-source databases by using the technology of nature language processing (Pujara, Augustine, and Getoor 2017; Ioannidis

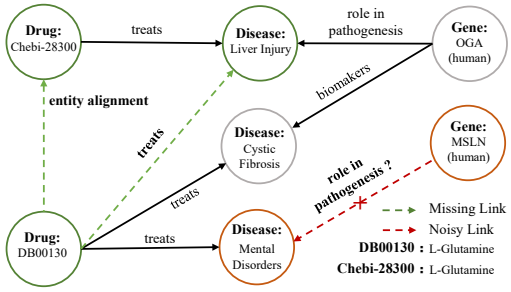

Figure 1: The explanatory case of entity unalignment and factual errors existed in DRKG. ***Chebi:28300*** and ***DB00130*** indicate the same drug ***L-Glutamine***, but they are treated as different entities, which results in entity unalignment and facts missing. Meanwhile, the source text represents 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.

et al. 2020), which results in conflicts and factual errors in the KGs. For example, as shown in Figure 1, the ***Chebi:28300*** and ***DB00130*** are different entities in DRKG (Ioannidis et al. 2020) but they indicate the same drug ***L-Glutamine***[1], which results in facts missing (e.g., the fact of the drug *DB00130* treats the disease *Liver Injury*). In addition, the source text indicates the gene ***MSLN*** as a predictor for cancer patients (Okła et al. 2018), and no evidence supports it has a role in the disease ***Mental Disorders***. This may introduce a fact error into DRKG. In this case, existing KG-based methods are ineffective due to these noisy interactions. Based on the above observations, we propose a novel denoising model, called DenoisedLP. Intuitively, we design a structure reliability learning module for the local subgraph guided by downstream tasks, to contain reliable interactions. Inspired by the success of smooth technologies in image denosing (Ma et al. 2018; Guo et al. 2019) by blurring noisy pixels, we develop a smooth semantic preservation module, which blurs the sparse relations and filtering out task-irrelevant edges. This reduces the negative impact of conflict relations existing in the KG. To focus on informative interactions between reliable structure and smoothed semantic relations, we maximize the mutual information between the representations of them globally.

In summary, the contributions of DenoisedLP can be summarized as follows:

- To the best of our knowledge, this is the first work that proposes a denoising method to learn the reliable interactions and smooth semantics by blurring the sparse relations of biomedical KGs, which alleviates the negative influence of noisy interactions and conflicts.
- We emphasize reliable interactions by maximizing the mutual information between the learned subgraph structure and smoothed semantic relations to efficiently drop information irrelated to downstream tasks.
- Extensive experiments of the DTI and DDI prediction on benchmark datasets and contaminated KG demonstrate DenoisedLP outperforms the state-of-the-art baselines.

2 Related Work

Link Prediction on Biomedical KG

Link prediction is increasingly adopted in biomedical knowledge graphs to identify unknown biological relations and interactions between various entities (Kishan et al. 2021). The line of works mainly focus on the completion of DTI and DDI relations on KGs. TriModel (Mohamed, Nováček, and Nounu 2020) and KG-DDI (Celebi et al. 2019) 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 (Lin et al. 2020) proposed a graph neural network to learn the structural relations, which enhances the prediction of the DDI relations. Subsequently, KGE-NFM (Ye et al. 2021) 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 (Ma et al. 2022) and DDI (Chen et al. 2021) prediction. To focus on the local structure of the predicted entity pairs, SumGNN (Yu et al. 2021) designed a new method to efficiently emphasize the subgraph structure of the biomedical KG, which aids the drug interaction prediction. GraIL (Teru, Denis, and Hamilton 2020) and SNRI (Xu et al. 2022) proposed to model the enclosing subgraph structure and neighboring relational paths around the target triple to effectively predict unknown links. 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 relations.

Denoising Methods on Graphs

Denoising on graphs has been successfully applied to the recommendation (Fan et al. 2023) and social networks (Quan et al. 2023). RGCF (Tian et al. 2022) 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 (Gao et al. 2022) provided a universal solution using self-guided learning to denoise implicit noisy feedback that can generalize 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 cannot consider the complex relationships in biomedical KGs. To tackle these limitations, inspired by the smoothing insight in image denoising (Ma et al. 2018; Guo et al. 2019), we blur the complex relations to ignore task-irrelevant edges and learn the reliable interactions of the local subgraph.

3 Methodology

Preliminaries

Biomedical Knowledge Graph. We define a biomedical knowledge graph DRKG (Ioannidis et al. 2020) 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 with different node types e_h, e_t (e.g., drugs and genes) as a fact.

Local Subgraph. Based on GraIL (Teru, Denis, and Hamilton 2020), when given a KG $\mathcal{G}_{kg}$ and a link (u, v) with relation r, 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 $\mathcal{N}_k(v) = \{s|d(v, s) \leq k\}$ for both u and v, where $d(\cdot, \cdot)$ represents the distance metric 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 $\mathcal{G}_{sub} = (V, E)$.

[1]https://go.drugbank.com/drugs/DB00130

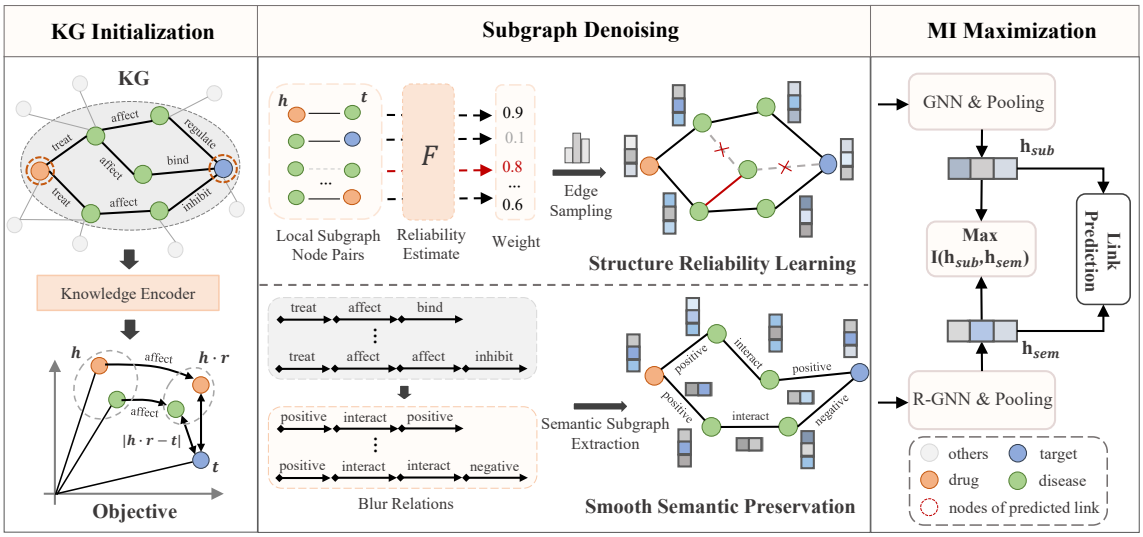

Figure 2: The DenoisedLP framework comprises three modules for predicting links in a given KG: (1) Initializing the entity and relation embeddings of the KG using RotatE; (2) Denoising the noisy interactions around the predicted link by learning a reliable subgraph structure and preserving the smooth semantic relations; (3) Maximizing the mutual information between refined and semantic subgraphs to focus on the informative interactions.

Semantic Subgraph. Given a link (u, v) with the scheme (e_u, r, e_v), we define a set of key metapaths $\mathcal{P} = \{(r_{e_u}, r_1, ..., r_k, r_{e_v}) | r_{e_u} \in \mathcal{N}^{rel}(e_u), r_k \in \mathcal{R}, r_{e_v} \in \mathcal{N}^{rel}(e_v)\}$ between nodes u and v, where $\mathcal{N}^{rel}(\cdot)$ are the neighboring relations around target entity type and k denotes the path length. According to the defined set of metapaths, we can extract various relational paths around the given link from original KG $\mathcal{G}_{kg}$ that construct a semantic subgraph $\mathcal{G}_{sem} = \{(h, r_{(h,t)}, t) | h, t \in \mathcal{E}, r_{(h,t)} \in \mathcal{R}_{(e_h, e_t)}\}$, where $\mathcal{R}_{(e_h, e_t)}$ denotes the relations between entity type e_h and e_t.

Problem Definition. In this paper, we focus on predicting the missing links of the biomedical KG $\mathcal{G}_{kg}$ by a learnable model to learn reliable interactions and filter out task-independent noise. Building upon the approach outlined in (Yin et al. 2022), we convert the link prediction into a classification task and take the relations classification as downstream tasks. Our goal is to estimate the link probability of various relations (e.g., DTI and DDI). For a given unknown link (u, v) with relation r, we propose a model to predict the probability of r 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 refined and semantic subgraphs.

Initialization of Knowledge Graph

In this paper, we utilize DRKG (Ioannidis et al. 2020) 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 (Sun et al. 2018). Given a triple (h, r, t), we expect that $\mathbf{x}_t = \mathbf{x}_h \odot \mathbf{e}_r$, where $\mathbf{x}_h, \mathbf{x}_t$ and $\mathbf{e}_r$ represent the embeddings of entities h, r and the relation r, respectively. The score function is defined as follows:

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

where the $\odot$ represents the element-wise product. 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 biomedical KG.

Subgraph Denoising

Structure Reliability Learning. To enable robust estimation of noisy interactions in KGs, we propose a structural reliability learning module for the local subgraph, which can dynamically adjust the reliable subgraph structure by using the pre-trained node features and the feedback of the downstream link 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 (Zhang and Zitnik 2020). To accomplish this, 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 Figure 2. Specifically, given an extracted local subgraph $\mathcal{G}_{sub} = (V, E)$ around the predicted link (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.

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

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 link 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 noised that 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:

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

where $\mathbf{X}\left(i\right)$ represents the pretrained feature of node i, $Z(i)$ is the learned embedding of node feature $\mathbf{X}\left(i\right)$, and $\mathbf{MLP}\left(\cdot\right)$ 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 binary entries $\text{Ber}(\pi_{i,j})$ to update the π:

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

where $\epsilon \sim$ *Uniform*$(0,1)$, $t \in \mathbb{R}^{+}$ indicates the temperature for the concrete distribution. 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 GCNs (Kipf and Welling 2017) on the refined subgraph with pretrained node features to obtain its global representation $\mathbf{h}_{sub}$ as follows:

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

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

Smooth Semantic Preservation. KGs often contain some conflict entities and relations that introduce a lot of noise into the downstream link prediction tasks (Pujara, Augustine, and Getoor 2017). Inspired by the smoothing insight of image denoising (Guo et al. 2019; Ma et al. 2018), we design a smooth semantic preservation module, which blurs the sparse relations and preserves the smoothed relational semantic to reduce the negative impact of the conflicts. Specifically, we smooth the KG by leveraging prior knowledge to generalize the interactions between biological entities into positive, interaction, and negative (i.e., the relations positive, interaction, and negative) according to the semantic similarity of the relations. Subsequently, we develop a semantic subgraph extraction module to explore the neighboring relations by extracting the paths with predefined metapaths. Specifically, given the link (u,v) with relation r, we use the defined metapaths to extract relational paths and construct a semantic subgraph $\mathcal{G}_{sem}$. After obtaining the semantic subgraph, we design a L-layer relational graph neural network (R-GNN) inspired by (Schlichtkrull et al. 2018; Xu et al. 2022) to obtain the global semantic representation of $\mathcal{G}_{sem}$. Specifically, we define the updating function of the nodes in l-th layer as:

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

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, and ϕ is the aggregation operation $\phi(\mathbf{x},\mathbf{e}) = \mathbf{x} - \mathbf{e}$ to fuse the hidden features of nodes and relations. 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:

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

where V_{sem} is the node set of semantic subgraph $\mathcal{G}_{sem}$. For more information about the smoothed KG and defined metapaths, please refer to the Technical Appendix.

MI Maximization in Subgraphs

To cooperate in denoising errors contained in the KG from different views of the local structure and smoothed semantic relations, we design an auxiliary self-supervised task based on mutual information (MI) maximization. We seek that the smoothed relations blur task-irrelevant edges and cooperate with the structure reliability learning for denoising noisy interactions. Specifically, we utilize InfoNCE (Oord, Li, and Vinyals 2018) to estimate mutual information between the representations of local structure and semantic subgraphs globally. In a formal context, when addressing the concept of subgraph mutual information, we treat the representations originating from both the reliable structure and the semantic subgraph, both of which are 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:

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

where $d(\cdot)$ 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.

Optimization

Similar to (Yin et al. 2022), we convert the link prediction to a classification task. For a given link (u,v) with relation r, we model the interaction probability $p_{(u,r,v)}$ of the link by

adopting the learned representations from the structure and semantic views as follows:

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

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

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

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:

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

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

4 Experiments

In this section, DenoisedLP[2] performs the link prediction task for two key relations (i.e., *drug-target-interaction* and *drug-drug-interaction*) based on biomedical KG.

Experimental Setups

Datasets. For link prediction of the relation DTI, we empirically perform experiments on two real-world datasets: (1) **DrugBank** (Wishart et al. 2018) collects the unique bioinformatics and cheminformatics resources that contains 12,063 drug-target pairs with 2,515 drugs, and 2,972 targets. (2) **DrugCentral** (Avram et al. 2023) 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 DenoisedLP on two wide-used datasets: (1) **DrugBank** (Wishart et al. 2018) contains 191,984 drug pairs with 86 types associated pharmacological relations for 1,703 drugs (e.g., *increase of cardiotoxic activity*). (2) **TWOSIDES** (Tatonetti et al. 2012) dataset contains 335 drugs with 26,443 drug pairs for 200 various side effect types. Following (Zitnik, Agrawal, and Leskovec 2018), we ensure each DDI type has at least 900 drug pairs by keeping 200 commonly-occurring types. In addition, we adopt the DRKG (Ioannidis et al. 2020) 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.

Evaluation strategy. For the DrugBank and DrugCentral datasets in predicting relation DTI, we ensure the positive and negative samples for each drug are balanced by random generating. Following SumGNN (Yu et al. 2021), we keep the train/valid/test sets of the DrugBank dataset containing samples of all types for predicting DDI links. For the TWOSIDES dataset, we follow the method in (Zitnik, Agrawal, and Leskovec 2018) 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 on Table 1 and Table 2.

Baselines. To verify the performance of DenoisedLP, we compare it against various baselines as follows:

- **GCN-KG** and **RotatE** (Sun et al. 2018) adopted the graph neural network (Kipf and Welling 2017) 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** (Teru, Denis, and Hamilton 2020) utilized a local subgraph for inductive relation prediction on KGs. To model neighboring relations effectively, **SNRI** (Xu et al. 2022) adopted the semantic subgraphs by extracting semantic relational paths to learn informative embedding.
- **TriModel** (Mohamed, Nováček, and Nounu 2020) and **KGE-NFM** (Ye et al. 2021) developed new methods to learn the relational representation of entities and relations, then predict the unknown links for DTI.
- **KGNN** (Lin et al. 2020) aggregated neighborhood information for each node from their local receptive via GNN on the biomedical knowledge graph for link prediction of relation DDI. **SumGNN** (Yu et al. 2021) focused on extracting information from the local subgraph of external KG in a learnable way and converts the link prediction into a multi-type and multi-class classification for DDI.

All baselines are implemented based on the office code and we tuned their hyperparameters to achieve optimal results.

Comparison with Baselines

We report the performance of our model and baselines for predicting links of the relations DTI and DDI in Table 1 and Table 2, respectively. As shown in Table 1, we observe DenoisedLP achieves the best prediction results in DTI links on both DrugBank and DrugCentral datasets. Specifically, DenoisedLP improves the AUC-ROC and AUC-PR by at least 2.19% and 2.43% respectively on the DrugBank dataset, and achieves 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 DenoisedLP 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 DenoisedLP 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 information and structures achieves better performance on DTI and DDI prediction, which indicates that the neighboring structure benefits the downstream link prediction tasks. (2) Compared with TriModel and KGNN, the SNRI using local semantic relations performs better than them on DrugBank for predicting DTI and DDI, which implies that the local semantic relations are more effective than the global structure and relations in predicting unknown links. (3) Among the subgraph-based methods (i.e., GraIL,

[2]Code and data are available at https://github.com/xxx/xxx

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
KG-MTL	82.55 ± 0.31	81.79 ± 0.52	84.39 ± 0.55	83.13 ± 0.74
KGE-NFM	82.71 ± 0.22	82.09 ± 0.52	86.34 ± 0.16	84.65 ± 0.14
DenoisedLP	**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
our w/o MI	82.05 ± 0.09	82.18 ± 0.18	86.34 ± 0.09	86.98 ± 0.17

Table 1: The performance on DrugBank and DrugCentral for *DTI* prediction. The **boldface** denotes the highest score and underline indicates the second highest score.

SNRI, and SumGNN), DenoisedLP 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. Unlike these models, the DenoisedLP reduces the negative influence of noise by denoising unreliable interactions in a learnable way and removing task-irrelevant relations, which achieves superior results. The overall performance compared with the baselines demonstrates the effectiveness and robustness of our model.

Ablation Study

To investigate the impact of each module in DenoisedLP, 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 DenoisedLP perform worse than the original model in Table 1 and Table 2, which verifies the effectiveness of each component.

ours w/o SRL. We observe that the performance value has a lot of reduction on all datasets for DTI and DDI prediction respectively, after removing the structure reliability learning module. The reason may be that the unreliable subgraph structure is less expressive for downstream tasks, which cannot effectively eliminate the negative influence of noisy interactions. In contrast, a complete reliable structure can improve the performance of the original model by removing possible noise and retaining trusty interactions.

ours w/o SSP. From the results reported in Table 1 and Table 2, we notice that the performance has a great degradation on all datasets when omitting the semantic subgraph of the smooth semantic preservation module. The observation demonstrates the semantic subgraph extracted by predefined metapaths is effective by ignoring task-irrelevant 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.

Methods	DrugBank (Multi-class)		TWOSIDES (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
SumGNN	85.58 ± 0.10	82.79 ± 0.19	87.42 ± 0.16	82.65 ± 0.07
DenoisedLP	**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
our w/o MI	85.07 ± 0.31	82.87 ± 0.24	85.96 ± 0.05	82.81 ± 0.32

Table 2: The results comparison on DrugBank and TWOSIDES for DDI prediction. The best is marked with **boldface** and the second best is with underline.

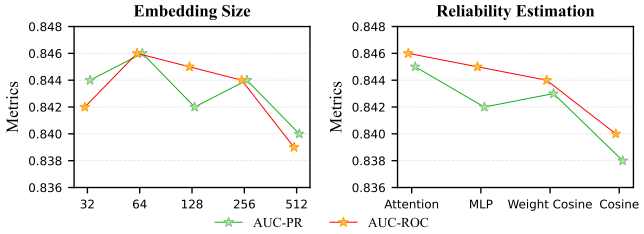

Figure 3: Hyper-parameter sensitivity analysis of DTI prediction based on the DrugBank dataset.

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

Hyper-parameter Sensitivity

We conduct hyper-parameter sensitivity analysis on the Drugbank dataset for DTI prediction to study the influence of several hyper-parameters.

Impact of embedding size. We explore the effect of hidden embedding size by varying it from 32 to 516. The left of Figure 3 depicts the changing trend of the AUC-ROC and AUC-PR values on the DrugBank dataset evaluated on DenoisedLP. Based on the results, we observe that the AUC-ROC values of DenoisedLP variation across different embedding sizes collapsed into a hunchback shape. The reason could be that enough embedding size can represent more information, while the larger one will progressively introduce a lot of noise with the degradation of DenoisedLP.

Impact of reliability estimation. To investigate the impact of various reliability estimation function $F(\cdot,\cdot)$ 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

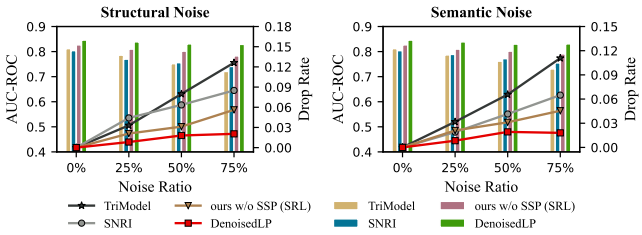

Figure 4: Performance comparison over various noisy types with different ratios. The bar represents the AUC-ROC and the line indicates the degradation ratio on the value.

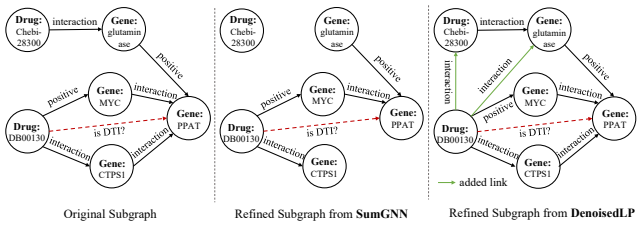

Figure 5: The original noisy subgraph and the refined subgraphs constructed from SumGNN and DenoisedLP.

the right of Figure 3, 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 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 coefficient guided by the downstream prediction tasks is more efficient. As a result, we set the *Attention* estimator to learn the reliability weights effectively.

Robustness of Interaction Noises

To verify the effectiveness of structure reliability learning and smooth semantic preservation modules in denoising interaction noise, 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 Figure 4. 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 DenoisedLP 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 SNIR and global KG-based TriModel. As shown in Figure 4, the AUC-ROC values on DrugBank and the corresponding performance degradation ratio 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, DenoisedLP and its variants exhibit smaller degradation than other methods for both types of noises. The variants **ours w/o SSP** and **ours w/o SRL** have 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 DenoisedLP and SNRI grew larger with increasing noise. This is because our model pays more attention to the informative interactions by maximizing the mutual information between the refined local and semantic subgraphs. This phenomenon shows that DenoisedLP can effectively mitigate noises using the reliable structure and smoothed semantics.

Case Study

We conduct a case study for predicting DTI relation between the drug *DB00130* and the gene *PPAT* to demonstrate the effectiveness of DenoisedLP, shown in Figure 5. 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, interaction, glutaminase*). By learning the reliable structure of the original subgraph, DenoisedLP 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, SumGNN uses pre-trained node features to calculate the attention weights of existing edges and prune them, potentially ignoring reliable information and failing to build missing edges to bring information interaction. This case shows that DenoiseLP can effectively learn a reliable structure, enhancing the performance of link prediction.

5 Conclusion

In this paper, we proposed a novel denoising model called DenoisedLP for predicting missing links with DTI and DDI relations on biomedical KGs. To mitigate the negative influence of noise in the KG, we proposed a structural reliability learning module to denoise unreliable links. Additionally, we adopted expert knowledge to blur the noisy relations and preserve the semantics. Moreover, we modeled the learned reliable and smoothed semantic subgraphs by MI maximization to emphasize the informative interactions for the downstream tasks. Our experiments on four datasets for DTI and DDI link predictions demonstrate that DenoisedLP significantly outperforms several existing state-of-the-art methods. These results verify the robustness of our model against interaction noises. In the future, we plan to construct a new high-quality biomedical KG based on the end-to-end denoising model as the foundation for accelerating drug discovery.

References

Avram, S.; Wilson, T. B.; Curpan, R.; Halip, L.; Borota, A.; Bora, A.; Bologa, C. G.; Holmes, J.; Knockel, J.; Yang, J. J.; et al. 2023. DrugCentral 2023 extends human clinical data and integrates veterinary drugs. *Nucleic Acids Research*, 51(D1): D1276–D1287.

Celebi, R.; Uyar, H.; Yasar, E.; Gumus, O.; Dikenelli, O.; and Dumontier, M. 2019. Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. *BMC Bioinformatics*, 20(1): 1–14.

Chen, Y.; Ma, T.; Yang, X.; Wang, J.; Song, B.; and Zeng, X. 2021. MUFFIN: multi-scale feature fusion for drug-drug interaction prediction. *Bioinformatics*, 37(17): 2651–2658.

Fan, Z.; Xu, K.; Dong, Z.; Peng, H.; Zhang, J.; and Yu, P. S. 2023. Graph Collaborative Signals Denoising and Augmentation for Recommendation. In *SIGIR*, 2037–2041.

Gao, Y.; Du, Y.; Hu, Y.; Chen, L.; Zhu, X.; Fang, Z.; and Zheng, B. 2022. Self-guided learning to denoise for robust recommendation. In *SIGIR*, 1412–1422.

Guo, T.; Xu, C.; Shi, B.; Xu, C.; and Tao, D. 2019. Smooth deep image generator from noises. In *AAAI*, 01, 3731–3738.

He, Q.; Yang, J.; and Shi, B. 2020. Constructing knowledge graph for social networks in a deep and holistic way. In *WWW*, 307–308.

Ioannidis, V. N.; Song, X.; Manchanda, S.; Li, M.; Pan, X.; Zheng, D.; Ning, X.; Zeng, X.; and Karypis, G. 2020. DRKG - Drug Repurposing Knowledge Graph for Covid-19. https://github.com/gnn4dr/DRKG/.

Kipf, T. N.; and Welling, M. 2017. Semi-supervised classification with graph convolutional networks. In *ICLR*, 1–13.

Kishan, K.; Li, R.; Cui, F.; and Haake, A. R. 2021. Predicting biomedical interactions with higher-order graph convolutional networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(2): 676–687.

Lin, X.; Quan, Z.; Wang, Z.-J.; Ma, T.; and Zeng, X. 2020. KGNN: Knowledge Graph Neural Network for Drug-Drug Interaction Prediction. In *IJCAI*, 2739–2745.

Luo, Y.; Zhao, X.; Zhou, J.; Yang, J.; Zhang, Y.; Kuang, W.; Peng, J.; Chen, L.; and Zeng, J. 2017. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature Communications*, 8(1): 573.

Ma, K.; Fu, H.; Liu, T.; Wang, Z.; and Tao, D. 2018. Deep blur mapping: Exploiting high-level semantics by deep neural networks. *IEEE Transactions on Image Processing*, 27(10): 5155–5166.

Ma, T.; Lin, X.; Song, B.; Philip, S. Y.; and Zeng, X. 2022. Kg-mtl: Knowledge graph enhanced multi-task learning for molecular interaction. *IEEE Transactions on Knowledge and Data Engineering*, 35(7): 7068–7081.

Mohamed, S. K.; Nováček, V.; and Nounu, A. 2020. Discovering protein drug targets using knowledge graph embeddings. *Bioinformatics*, 36(2): 603–610.

Okła, K.; Surówka, J.; Frąszczak, K.; Czerwonka, A.; Kaławaj, K.; Wawruszak, A.; Kotarski, J.; and Wertel, I. 2018. Assessment of the clinicopathological relevance of mesothelin level in plasma, peritoneal fluid, and tumor tissue of epithelial ovarian cancer patients. *Tumor Biology*, 40(10): 1010428318804937.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Pujara, J.; Augustine, E.; and Getoor, L. 2017. Sparsity and noise: Where knowledge graph embeddings fall short. In *EMNLP*, 1751–1756.

Quan, Y.; Ding, J.; Gao, C.; Yi, L.; Jin, D.; and Li, Y. 2023. Robust Preference-Guided Denoising for Graph based Social Recommendation. In *WWW*, 1097–1108.

Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; Van Den Berg, R.; Titov, I.; and Welling, M. 2018. Modeling relational data with graph convolutional networks. In *ESWC*, 593–607.

Sun, Z.; Deng, Z.-H.; Nie, J.-Y.; and Tang, J. 2018. Rotate: Knowledge graph embedding by relational rotation in complex space. In *ICLR*, 1–18.

Tatonetti, N. P.; Ye, P. P.; Daneshjou, R.; and Altman, R. B. 2012. Data-driven prediction of drug effects and interactions. *Science Translational Medicine*, 4(125): 125–125.

Teru, K.; Denis, E.; and Hamilton, W. 2020. Inductive relation prediction by subgraph reasoning. In *ICML*, 9448–9457.

Tian, C.; Xie, Y.; Li, Y.; Yang, N.; and Zhao, W. X. 2022. Learning to denoise unreliable interactions for graph collaborative filtering. In *SIGIR*, 122–132.

Wan, F.; Hong, L.; Xiao, A.; Jiang, T.; and Zeng, J. 2019. NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions. *Bioinformatics*, 35(1): 104–111.

Wang, X.; Wang, D.; Xu, C.; He, X.; Cao, Y.; and Chua, T.-S. 2019. Explainable reasoning over knowledge graphs for recommendation. In *AAAI*, 01, 5329–5336.

Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1): D1074–D1082.

Xu, X.; Zhang, P.; He, Y.; Chao, C.; and Yan, C. 2022. Subgraph neighboring relations infomax for inductive link prediction on knowledge graphs. In *IJCAI*, 2341–2347.

Yang, Y.; Huang, C.; Xia, L.; and Li, C. 2022. Knowledge graph contrastive learning for recommendation. In *SIGIR*, 1434–1443.

Ye, Q.; Hsieh, C.-Y.; Yang, Z.; Kang, Y.; Chen, J.; Cao, D.; He, S.; and Hou, T. 2021. A unified drug-target interaction prediction framework based on knowledge graph and recommendation system. *Nature Communications*, 12(1): 6775.

Yin, H.; Zhang, M.; Wang, Y.; Wang, J.; and Li, P. 2022. Algorithm and system co-design for efficient subgraph-based graph representation learning. In *VLDB*, 2788–2796.

Yu, Y.; Huang, K.; Zhang, C.; Glass, L. M.; Sun, J.; and Xiao, C. 2021. SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics*, 37(18): 2988–2995.

Zhang, X.; and Zitnik, M. 2020. Gnnguard: Defending graph neural networks against adversarial attacks. *Advances in Neural Information Processing Systems*, 33: 9263–9275.

Zitnik, M.; Agrawal, M.; and Leskovec, J. 2018. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13): i457–i466.