The Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25)

S^2DN: Learning to Denoise Unconvincing Knowledge for Inductive Knowledge Graph Completion

Tengfei Ma[1], Yujie Chen[1], Liang Wang[2,3], Xuan Lin[4], Bosheng Song[1*], Xiangxiang Zeng[1]
[1]College of Computer Science and Electronic Engineering, Hunan University, China
[2]NLPR, MAIS, Institute of Automation, Chinese Academy of Sciences
[3]School of Artificial Intelligence, University of Chinese Academy of Sciences
[4]College of Computer Science, Xiangtan University, China
{tfma,yjchen}@hnu.edu.cn, liang.wang@cripac.ia.ac.cn, jack_lin@xtu.edu.cn, {boshengsong, xzeng}@hnu.edu.cn

Abstract

Inductive Knowledge Graph Completion (KGC) aims to infer missing facts between newly emerged entities within knowledge graphs (KGs), posing a significant challenge. While recent studies have shown promising results in inferring such entities through knowledge subgraph reasoning, they suffer from (i) *the semantic inconsistencies of similar relations*, and (ii) *noisy interactions inherent in KGs* due to the presence of unconvincing knowledge for emerging entities. To address these challenges, we propose a Semantic Structure-aware Denoising Network (S^2DN) for inductive KGC. Our goal is to learn adaptable general semantics and reliable structures to distill consistent semantic knowledge while preserving reliable interactions within KGs. Specifically, we introduce a semantic smoothing module over the enclosing subgraphs to retain the universal semantic knowledge of relations. We incorporate a structure refining module to filter out unreliable interactions and offer additional knowledge, retaining robust structure surrounding target links. Extensive experiments conducted on three benchmark KGs demonstrate that S^2DN surpasses the performance of state-of-the-art models. These results demonstrate the effectiveness of S^2DN in preserving semantic consistency and enhancing the robustness of filtering out unreliable interactions in contaminated KGs.

Code — https://github.com/xiaomingaaa/SDN
Extended version — https://arxiv.org/abs/2412.15822

Introduction

Knowledge graphs (KGs) represent the relations between real-world entities as facts providing a general way to store semantic knowledge (Wang et al. 2017; Zhang et al. 2023a; Li et al. 2023b). KGs have been successfully used in various applications, including recommendation systems (Yang et al. 2023; Li et al. 2024a), question answering (Cao et al. 2022; Liu et al. 2023), and drug discovery (Lin et al. 2020; Ma et al. 2022). However, KGs often suffer from incompleteness (Geng et al. 2023; Li et al. 2023a) and newly emerging entities (Xu et al. 2022). Thus, inductive knowledge graph completion (KGC), proposed to predict missing facts on unseen entities in KGs, has been a hot area of research (Zhang et al. 2023b; Bai et al. 2023) and industry (Du et al. 2021; Ji et al. 2022).

To improve the generalization ability of the KGC task on unknown entities, some researchers propose rule-based methods (Meilicke et al. 2018; Yang, Yang, and Cohen 2017; Cheng, Ahmed, and Sun 2023). These methods enable effective reasoning under emerging entities by mining common reasoning rules, while they are limited by predictive performance (Xu et al. 2022). Recently, there has been a surge in inductive KGC methods based on Graph Neural Networks (GNNs), inspired by the ability of GNNs to aggregate local information. For instance, GraIL (Teru, Denis, and Hamilton 2020) models enclose subgraphs of the target link to capture local topological structure, thereby possessing the inductive ability of emerging entities. Motivated by GraIL and the message passing mechanism of GNN, some works (Mai et al. 2021; Chen et al. 2021) have further utilized the enclosing subgraph structure and designed effective propagation strategies to model informative neighbors for inductive KGC. To explicitly enhance the prediction ability of unseen entities through semantic knowledge in the KGs, SNRI (Xu et al. 2022) proposes a relational paths module to improve the performance of inductive KGC, while RMPI (Geng et al. 2023) designs a novel relational message-passing network for fully inductive KGC with both unseen entities and relations. Despite the promising results yielded by these models, they are limited to information redundancy in modeling irrelevant entities and relations. AdaProp (Zhang et al. 2023b) is developed to learn an adaptive propagation path and filter out irrelevant entities, achieving powerful performance. However, these models ignore the unconvincing knowledge (e.g., semantic inconsistencies of similar relations in context and inherent noise within KGs). For example, the same semantic "*the location is*" with different relations *located_in* and *lie_in* may lead to inconsistent knowledge expressions in the context "(Alibaba, *lie_in*, Hangzhou), and (Hangzhou, *located_in*, China)", increasing the complexity of subgraph reasoning, which deduces the prediction performance on inductive situations. Additionally, logical reasoning may be misled to conclude a confused fact (Obama, *live_in*, New York) from the reasoning chain (Obama, *work_in*, The White House, *located_in*, New York) due to the presence of the

---

[*] Corresponding Author

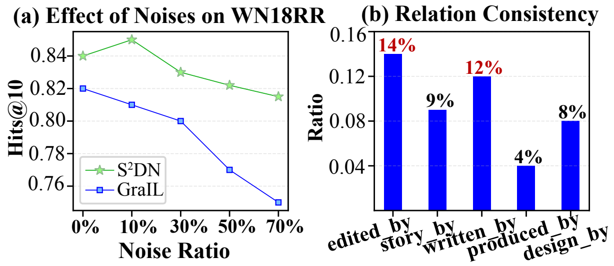


Figure 1: (a) S^2DN outperforms GraIL in terms of Hits@10 on noisy KGs with different noise ratios (i.e., *high robustness*). (b) The relation *edited_by* shows a high percentage of being converted to other relations (enumerated on the x-axis) with similar semantics (i.e., *high semantic consistency*).

false positive fact: The White House is *located in* New York.

Motivated by the above observations, the inductive KGC presents two main challenges: (i) ***Inconsistency***: inconsistent representations of relations with the same semantics in context, and (ii) ***Noisy***: the presence of inevitable noise in KGs that is difficult to ignore. In response to these challenges, we introduce S^2DN, a semantic structure-aware denoising network designed to maintain consistent semantics and filter out noisy interactions, thereby enhancing robustness in inductive KGC. Drawing inspiration from the successful application of smoothing technologies in image denoising (Ma et al. 2018; Guo et al. 2019), we have developed a semantic smoothing module to generalize similar relations with blurred semantics. Additionally, to eliminate task-irrelevant noise and provide supplementary knowledge, we introduce a structure refining module to retain reliable interactions in a learnable manner. By integrating general semantics and reliable structure, S^2DN denoise unconvincing knowledge from a semantic-structure synergy perspective. As depicted in Figure 1, S^2DN demonstrates superior inductive prediction performance (Hits@10) compared to GraIL across different KGs under various noise levels. Additionally, S^2DN ensures improved semantic consistency, as evidenced by a high percentage of the relation *edited_by* being smoothed to others with similar semantics.

The contributions of S^2DN are summarized: 1) We address inductive KGC from a novel perspective by adaptively reducing the negative impact of semantic inconsistency and noisy interactions; 2) We innovatively propose a semantic smoothing module to generalize KG relations dynamically by blurring similar relations into consistent knowledge semantics; 3) To emphasize reliable interactions, we introduce the structure refining module to filter out noise adaptively and offer additional knowledge. 4) Extensive experiments on benchmark datasets and contaminated KGs demonstrate that S^2DN outperforms the baselines in inductive KGC.

Related Works

Inductive Knowledge Graph Completion. Previous methods fall into two main categories: *Rule-based* and *GNN-based* approaches. The rule-based methods are independent of entities by mining logical rules for explicit reasoning. For instance, NeuralLP (Yang, Yang, and Cohen 2017) and DRUM (Sadeghian et al. 2019) learn logical rules and their confidence simultaneously in an end-to-end differentiable manner. Similarly, IterE (Zhang et al. 2019) and RNNLogic (Qu et al. 2020) treat logic rules as a latent variable, concurrently training a rule generator alongside a reasoning predictor utilizing these rules. While rule-based methods have shown comparable prediction performance, they often overlook the surrounding structure of the target link, resulting in limited expressive ability in inductive scenarios (Xu et al. 2022). Recently, there has been a shift towards leveraging GNNs for KGC tasks (Teru, Denis, and Hamilton 2020; Zhang et al. 2023b). For example, LAN (Wang et al. 2019) learns the embeddings of unseen entities by aggregating information from neighboring nodes, albeit restricted to scenarios where unseen entities are surrounded by known entities. GraIL (Teru, Denis, and Hamilton 2020) and CoMPILE (Mai et al. 2021) address this limitation by modeling the enclosing subgraph structure around target facts. However, these approaches neglect the neighboring relations surrounding the target triple. Thus, SNRI (Xu et al. 2022) and RED-GNN (Zhang and Yao 2022) fully exploit complete neighboring relations and paths within the enclosing subgraph. While LogCo (Pan et al. 2022) tackles the challenge of deficient supervision of logic by leveraging relational semantics. RMPI (Geng et al. 2023) proposes a relational message-passing network to utilize relation patterns for subgraph reasoning. These methods ignore the negative impact of task-irrelevant entities. AdaProp (Zhang et al. 2023b) designs learning-based sampling mechanisms to identify the semantic entities. Although these methods have achieved promising results, they suffer from (i) inconsistent semantics of similar relations, and (ii) inherent noisy associations within KGs. Our work proposes to smooth semantic relations and learn reliable structures to tackle these limitations.

Denoising Methods in Knowledge Graphs. The presence of noise within KGs is a common issue stemming from the uncertainty inherent in learning-based construction methods (Xue and Zou 2023; Pujara, Augustine, and Getoor 2017). Denoising on KGs has been applied to the recommendation (Fan et al. 2023) and social networks (Quan et al. 2023). For instance, ADT (Wang et al. 2021) and KRDN (Zhu et al. 2023) designed a novel training strategy to prune noisy interactions and implicit feedback during training. RGCF (Tian et al. 2022) and SGDL (Gao et al. 2022) proposed a self-supervised robust graph collaborative filtering model to denoise unreliable interactions and preserve the diversity for recommendations. However, these methods are limited in their ability to denoise noisy interactions in domain-specific networks (e.g., recommendation and social networks). Some approaches (Zhang et al. 2024; Hong, Bu, and Wu 2021) attempt to adopt rule-based triple confidence and structural entity attributes to capture noise-aware KG embedding. Despite achieving promising results, these methods overlook inconsistent semantic relations and work for transductive KGC reasoning. Inspired by the smoothing insight in image denoising (Ma et al. 2018; Guo et al. 2019), we propose a method to smooth the complex relations within KGs for inductive KGC. By doing so, we aim to eliminate inconsistent semantics and extract reliable structures in local subgraphs, particularly effective in inductive scenarios.

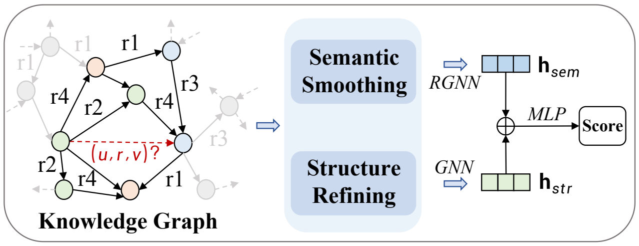


Figure 2: The S^{2}DN framework comprises two modules for inductively predicting links in a given KG : (1) Smoothing relational semantics by blurring similar relations adaptively; (2) Refining the structure of subgraphs by learning reliable interactions dynamically.

Preliminary

Knowledge Graphs. KGs contain structured knowledge about real-world facts, including common concepts, entity attributes, or external commonsense. We define a KG as a heterogeneous graph $\mathcal{G} = \{(h, r, t)|h, t \in \mathcal{E}, r \in \mathcal{R}\}$ where each triple (h, r, t) describes a relation r between the head entity h and tail entity t as a fact.

Enclosing Subgraph. Following GraIL (Teru, Denis, and Hamilton 2020), when given a KG $\mathcal{G}$ and a triple (u, r, v), we extract an enclosing subgraph $g = (V, E)$ surrounding the target triple. 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 shortest path distance between given nodes on $\mathcal{G}$. Subsequently, we obtain the nodal intersection $V = \{s|s \in \mathcal{N}_k(u) \cap \mathcal{N}_k(v)\}$ as vertices of the subgraph. Finally, we draw the triples linked by the set of nodes V from $\mathcal{G}$ as $g = (V, E)$.

Problem Definition. We concentrate on predicting missing links between emerging entities within a knowledge graph $\mathcal{G}$ (i.e., inductive KGC). This prediction is achieved by adaptively smoothing relational semantics and refining reliable structures. We define the problem of inductive KGC as a classification task, aiming to estimate the interaction probability of various relations inductively. Specifically, given an unknown fact (u, r, v) where either u or v is an emerging entity, we propose a model to predict the interaction probability denoted as $p_{(u,r,v)} = \mathcal{F}((u, r, v)|\Theta, \mathcal{G}, g)$, where Θ is the trainable parameters of S^{2}DN.

Proposed Method

The Framework of S^{2}DN

Overview. S^{2}DN reasons on the enclosing subgraph surrounding the target link inductively from both semantic and structure perspectives, as illustrated in Figure 2. To identify the unknown link (u, r, v), S^{2}DN incorporates two key components: *Semantic Smoothing* and *Structure Refining*. Semantic Smoothing adaptively merges relations with similar semantics into a unified representation, ensuring consistency in representation space. In parallel, Structure Refining dynamically focuses on filtering out task-irrelevant facts surrounding the target link and incorporates additional knowledge, thus improving the reliability of interactions. The refining process works in tandem with the previously smoothed relations to predict unknown links involving new entities more effectively. After obtaining the smoothed and refined subgraphs, we model them using a Relational Graph Neural Network (RGNN, (Schlichtkrull et al. 2018)) and a Graph Neural Network (GNN, (Kipf and Welling 2017)), respectively. Finally, the embeddings of smoothed and refined subgraphs are concatenated and fed into a classifier to predict the interaction probability of the target link (u, r, v).

Semantic Smoothing. KGs often suffer from semantic inconsistencies in their relationships. For example, in the contexts (Alibaba, *lie_in*, Hangzhou) and (Hangzhou, *located_in*, China), the relations “*located_in*” and “*lie_in*” represent the same semantic meaning, “*the location is*”. These inconsistencies lead to discrepancies in the representation space (Pujara, Augustine, and Getoor 2017; Xue and Zou 2023). To mitigate this limitation, inspired by the pixel smoothing insight of image denoising (Ma et al. 2018; Guo et al. 2019), we propose a semantic smoothing module depicted in Figure 3A. This module blurs similar relations while preserving the smoothed relational semantics, aiming to alleviate the negative impact of potential inconsistency. To achieve this, we first identify the subgraph $g = (V, E)$ that surrounds the missing link (u, r, v). Then a trainable strategy is employed to smooth embeddings $\mathbf{E} \in \mathbb{R}^{|\mathcal{R}| \times dim}$ of similar relations into consistent representation space, where $|\mathcal{R}|$ denotes the count of original relations and dim is the embedding size of relation embedding. Subsequently, we define the smooth operation as follows:

$$w = \text{softmax}(\mathbf{E} \otimes \mathbf{W}^{\mathrm{T}} + b),$$
$$\tilde{\mathcal{R}} = \frac{\exp((\log w + G)/\tau)}{\sum_{r \in \mathcal{R}} \exp((\log w_r + G_r)/\tau)}, \tag{1}$$

where w denotes smoothing weights and $\tilde{\mathcal{R}}$ is the smoothed relations from original relations $\mathcal{R}$. $\mathbf{W} \in \mathbb{R}^{|\mathcal{R}| \times dim}$ is trainable parameters, G is a noise sampled from a Gumbel distribution, and τ represents the temperature parameter. We adopt the Gumbel Softmax trick (Jang, Gu, and Poole 2017), facilitating differentiable learning over discrete outputs w. This operation learns to categorize relations with consistent semantics in the context of g into the same relation index. We refine the embeddings of relations $\tilde{\mathcal{R}}$ as follows:

$$\tilde{\mathbf{E}} = \tilde{\mathcal{R}} \otimes \mathbf{E}, \tag{2}$$

where $\tilde{\mathbf{E}}$ denotes the smoothed embeddings from original representation $\mathbf{E}$ and $\otimes$ represents the operation of matrix multiplication. This process enables similar relations to be mapped into consistent representation space guided by downstream tasks. To prevent the loss of information caused by excessive smoothing of relations and contain further consistencies, we incorporate a trade-off objective designed to preserve generic information during the optimization process. After obtaining the smoothed relations, we refined the enclosing subgraph g by the new relations $\tilde{\mathcal{R}}$. Then a L-layer

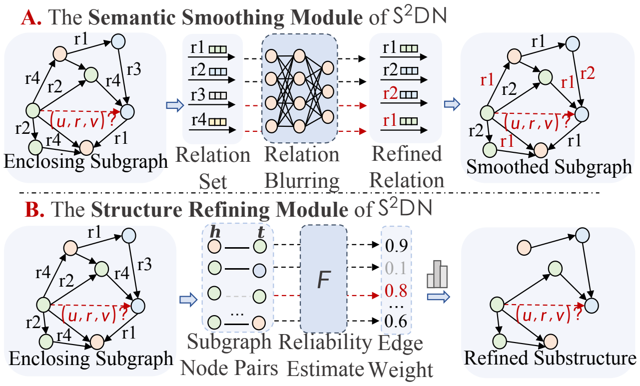


Figure 3: The architecture of **Semantic Smoothing** and **Structure Refining** modules of S^{2}DN.

RGNN (Schlichtkrull et al. 2018; Xu et al. 2022) is introduced to capture the global semantic representation of the refined g. Specifically, the updating function of the nodes over the blurred relation embedding $\tilde{\mathbf{E}}$ in l-th layer is defined as:

$$\begin{aligned} \mathbf{x}_i^l &= \sum_{r\in\tilde{R}} \sum_{j\in\mathcal{N}_r(i)} \alpha_{i,r} \mathbf{W}_r^l \phi(\tilde{\mathbf{e}}_r^{l-1}, \mathbf{x}_j^{l-1}), \\ \alpha_{i,r} &= \text{sigmoid}\left(\mathbf{W}_1 \left[\mathbf{x}_i^{l-1} \oplus \mathbf{x}_j^{l-1} \oplus \tilde{\mathbf{e}}_r^{l-1}\right]\right), \end{aligned} \tag{3}$$

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, $\mathbf{x}_i$ stands for the embedding of node i, $\tilde{\mathbf{e}}_r$ denotes the smoothed embedding under relation r, and ϕ is the aggregation operation to fuse the hidden features of nodes and relations. In addition, we initialize the entity (i.e., node) embedding $\mathbf{x}_i^0$ with the designed node features (Teru, Denis, and Hamilton 2020) and the original relation embedding $\mathbf{E}$ is initialized by Xavier initializer (Glorot and Bengio 2010). The details of designed node features refer to Appendix B.1.3. Finally, we obtain the global representation $\mathbf{h}_{sem}$ of the smoothed subgraph g as follows:

$$\mathbf{h}_{sem} = \frac{1}{|V|} \sum_{i\in V}^{V} \sigma(f(\mathbf{x}_i^L)), \tag{4}$$

where V is the node set of smoothed subgraph g and $f(\cdot)$ denotes the feature transformation function. σ indicates the activation function ReLU (Nair and Hinton 2010).

Structure Refining. To improve the precise estimation of noisy interactions within KGs, we propose a structure refining module specifically designed for the local enclosing subgraph, as depicted in Figure 3B. This module dynamically adapts the reliable subgraph structure based on both node features and feedback from downstream tasks. The underlying concept is that nodes with similar features or structures are more prone to interact with each other compared to those with dissimilar attributes (Zhang and Zitnik 2020; Li et al. 2024b). Our objective is to assign weights to all edges connecting the nodes using a reliability estimation function denoted as $F(\cdot,\cdot)$, which relies on the learned node features. Following this, the refined local subgraph is generated by removing low-weight noisy edges while retaining the more significant and reliable connections. To elaborate, when presented with an extracted enclosing subgraph $g = (V, E)$ surrounding the missing link (u, r, v), we prioritize the degree of interaction over relations, thereby enriching the structural information of semantic smoothing modules. We conceptualize all potential edges between nodes as a collection of mutually independent Bernoulli random variables, parameterized by the learned attention weights π.

$$\tilde{g} = \bigcup_{i,j\in V} \{(i,j) \sim \text{Ber}\left(\pi_{i,j}\right)\}. \tag{5}$$

In this context, V denotes the set of nodes within the enclosing subgraph, and $(i,j) \in E$ denotes the edge connecting nodes i and j. We optimize the reliability probability π concurrently with the downstream inductive KGC task. The value of $\pi_{i,j}$ characterizes the task-specific reliability of the edge (i,j) where a smaller $\pi_{i,j}$ suggests that the edge (i,j) is more likely to be noisy and thus should be assigned a lower weight or removed altogether. The reliable probability $\pi_{i,j} = F(i,j)$ for each edge between node pair (i,j) can be calculated as follows:

$$\begin{aligned} \pi_{i,j} &= \text{sigmoid}\left(Z(i)Z(j)^{\mathrm{T}}\right), \\ Z(i) &= \mathbf{MLP}\left(\mathbf{X}\left(i\right)\right), \end{aligned} \tag{6}$$

where $\mathbf{X}\left(i\right)$ represents the 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 (More detailed choices of $F(\cdot,\cdot)$ are discussed in Appendix C.4.4). Since the extracted enclosing subgraph g is not differentiable when the probability π is modeled as a Bernoulli distribution, we employ the reparameterization trick. This allows us to relax the binary entries $\text{Ber}(\pi_{i,j})$ for updating the π:

$$\text{Ber}(\pi_{i,j}) \approx \text{sigmoid}\left(\frac{1}{t}\left(\log\frac{\pi_{i,j}}{1-\pi_{i,j}} + \log\frac{\epsilon}{1-\epsilon}\right)\right), \tag{7}$$

where $\epsilon \sim \mathit{Uniform}(0,1)$, $t \in \mathbb{R}^+$ indicates the temperature for the concrete distribution. With $t \geq 0$, the function undergoes smoothing with a well-defined gradient $\frac{\partial \text{Ber}(\pi_{i,j})}{\partial \pi_{i,j}}$, which facilitates the optimization of learnable subgraph structure throughout the training process. After post-concrete relaxation, the subgraph structure becomes a weighted fully connected graph, which is computationally intensive. To address this, edges with a probability of less than 0.5 are removed from the subgraph, yielding the refined graph $\tilde{g}$. Following this refinement, L-layer GCNs (Kipf and Welling 2017) are applied to the refined subgraph using the designed node features (see Appendix B.1.3) to derive its global representation $\mathbf{h}_{str}$ as follows:

$$\begin{aligned} h^l &= \mathbf{GCN}\left(h^{l-1}, \tilde{g}\right), \\ \mathbf{h}_{str} &= \frac{1}{|V|} \sum_{i\in V}^{V} \sigma(f(h^L(i))), \end{aligned} \tag{8}$$

where $\sigma(\cdot)$ represents the sigmoid activation function. $f(\cdot)$ is a multi-layer perceptron that denotes the feature transformation operation.

Theoretical Discussion of Smoothing

The smoothing operation can be used as a way to minimize the information bottleneck between the original semantic relations and the downstream supervised signals (i.e., labels). Following the standard practice in the method (Tishby, Pereira, and Bialek 2000), given original relation embedding $\mathbf{E}$, smoothed relation embedding $\tilde{\mathbf{E}}$, and target Y, they follow the Markov Chain $< Y \rightarrow \mathbf{E} \rightarrow \tilde{\mathbf{E}} >$.

Definition 1 (Information Bottleneck). *For the input relation embedding* $\mathbf{E}$ *and the label of downstream task* Y*, the* ***Information Bottleneck*** *principle aims to learn the minimal sufficient representation* $\tilde{\mathbf{E}}$*:*

$$\tilde{\mathbf{E}} = \arg\min_{\tilde{\mathbf{E}}} -I(\tilde{\mathbf{E}}; Y) + I(\tilde{\mathbf{E}}; \mathbf{E}), \tag{9}$$

where $I(A; B) = H(A) - H(A|B)$ denotes the Shannon mutual information (Cover 1999). Intuitively, the first term $-I(\tilde{\mathbf{E}}; Y)$ is the reasoning objective, which is relevant to downstream tasks. The second term $I(\tilde{\mathbf{E}}; \mathbf{E})$ encourages the task-independent information of the original relational semantic dropped. Suppose $\mathbf{E}_n \in \mathbb{R}$ is a task-irrelevant semantic noise in original subgraph g, the learning process of $\tilde{\mathbf{E}}$ follows the Markov Chain $< (Y, \mathbf{E}_n) \rightarrow \mathbf{E} \rightarrow \tilde{\mathbf{E}} >$. The smoothed relation embedding $\tilde{\mathbf{E}}$ only preserves the task-related information in the observed embedding $\mathbf{E}$.

Lemma 1 (Smoothing Objective). *Given the original relation embedding* $\mathbf{E}$ *within the enclosing subgraph* g *and the label* $Y \in \mathbb{Y}$*, let* $\mathbf{E}_n \in \mathbb{R}$ *be a task independent semantic noise for* Y*. Denote* $\tilde{\mathbf{E}}$ *as the smoothed relations learned from* $\mathbf{E}$*, then the following inequality holds:*

$$I(\tilde{\mathbf{E}}; \mathbf{E}_n) \leq I(\tilde{\mathbf{E}}; \mathbf{E}) - I(\tilde{\mathbf{E}}; Y). \tag{10}$$

The detailed proof refers to Appendix A.1. Lemma 1 shows that optimizing the objective in Eq. (10) is equivalent to encouraging $\tilde{\mathbf{E}}$ to be more related to task-relevant information by minimizing the terms $I(\tilde{\mathbf{E}}; \mathbf{E})$ and $-I(\tilde{\mathbf{E}}; Y)$. Therefore, we introduce a Kullback–Leibler (KL) loss (Sun et al. 2022) to minimize the difference between original and smoothed relation embeddings and adopt the cross-entropy loss to maximize the mutual information between the smoothed relations and downstream tasks.

Training and Optimization

This section delves into the prediction and optimization details of S^{2}DN within the framework of the inductive KGC task. Here, we view the inductive KGC as a classification task. Given a predicted link (u, r, v), the link probability $p_{(u,r,v)}$ for the given link is computed using representations from both semantic and structural perspectives as follows:

$$p_{(u,r,v)} = \sigma(\mathbf{MLP}([\mathbf{h}_{sem} \oplus \mathbf{h}_{str}])), \tag{11}$$

where $\oplus$ denotes the concatenate operation, $\mathbf{MLP}(\cdot)$ indicates a classifier here and $\sigma(\cdot)$ is the sigmoid activate function. Subsequently, we adopt the cross-entropy loss and introduce an objective to balance the difference between smoothed and original relations:

$$\ell = -\sum_{r \in \mathcal{R}} \log(p_{(u,r,v)}) y_{(u,r,v)} + \mathcal{D}(\tilde{\mathbf{E}} || \mathbf{E}) + \lambda \|\Theta\|_2, \tag{12}$$

where $y_{(u,r,v)}$ is the real label of the given link, Θ represents the trainable parameters of S^{2}DN, $\mathcal{D}$ denotes the KL loss, and λ is a hyperparameter denoting the coefficient of the regular term. $\tilde{\mathbf{E}}$ and $\mathbf{E}$ denote the representations before and after relation smoothing in the subgraph of the current sample, respectively.

Experiments

We carefully consider the following *key* research questions: **RQ1)** Does S^{2}DN outperform other state-of-the-art inductive KGC baselines? **RQ2)** Are the proposed *Semantic Smoothing* and *Structure Refining* modules effective? **RQ3)** Can S^{2}DN enhance the semantic consistency of the relations and refine reliable substructure surrounding the target facts?

Experiment Setup

Dataset & Evaluation. We utilize three widely-used datasets: WN18RR (Dettmers et al. 2018), FB15k-237 (Toutanova et al. 2015), and NELL-995 (Xiong, Hoang, and Wang 2017), to evaluate the performance of S^{2}DN and baseline models. Following (Teru, Denis, and Hamilton 2020; Zhang et al. 2023b), we use the same four subsets with increasing size of the three datasets. Each subset comprises distinct training and test sets. We measure the filtered ranking metrics **Hits@1**, **Hits@10**, and mean reciprocal rank (**MRR**), where larger values indicate better performance.

Baselines. We compare S^{2}DN against the rule- and GNN-based methods. The rule-based methods are **NeuralLP** (Yang, Yang, and Cohen 2017), **DRUM** (Sadeghian et al. 2019), and **A*Net** (Zhu et al. 2024). The GNN-based models are **CoMPILE** (Mai et al. 2021), TAGT (Chen et al. 2021), **SNRI** (Xu et al. 2022), and **RMPI** (Geng et al. 2023). Furthermore, we design two variants of S^{2}DN to verify the effectiveness of each module by removing: (i) the Semantic Smoothing module (called **S^{2}DN w/o SS**), (ii) the Structure Refining module (called **S^{2}DN w/o SR**).

Comparison with Baselines (RQ1)

To address **RQ1**, we present the performance comparison of S^{2}DN and baseline models in predicting missing links for emerging entities as shown in Tables 1 and 2. Our results demonstrate that S^{2}DN achieves comparable performance to the baseline models across all datasets.

Specifically, we make the following observations: (1) S^{2}DN shows improved average performance over rule-based inductive methods with Hits@10 metrics of 9.2%, 4.7%, and 6.8% on WN18RR, FB15k-237, and NELL-995 respectively. Furthermore, GNN-based methods like CoMPILE and TAGT outperform various rule-based approaches in ranking tasks on most datasets, indicating the effectiveness of GNN-based methods in leveraging neighboring information and structures for inductive KGC. (2) SNRI, which integrates local semantic relations, outperforms CoMPILE and TAGT on multiple datasets, highlighting the importance of utilizing local semantic relations within KGs for inductive KGC. (3) RMPI, through efficient message passing between relations to leverage relation patterns for KGC based on graph transformation and pruning, outperforms SNRI,

Methods	Avg.	V1			V2			V3			V4		
	Hits@10	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
DRUM	64.15	57.92	74.37	64.27	54.71	68.93	59.46	33.98	46.18	37.63	55.66	67.13	60.11
NeuralLP	64.15	55.32	74.35	62.04	51.99	68.91	57.26	29.96	46.23	36.13	55.61	67.13	59.19
A*Net	72.06	70.51	80.58	74.68	70.99	79.11	74.45	46.47	54.87	49.68	63.92	73.67	66.24
CoMPILE	74.43	74.20	82.97	78.59	78.11	79.84	79.01	50.33	59.75	54.44	72.71	75.17	72.71
TAGT	73.28	69.15	82.45	75.45	75.42	78.68	77.43	50.08	58.60	54.29	71.97	73.41	73.22
SNRI	79.19	70.47	84.84	76.31	72.68	82.09	77.04	50.99	67.52	57.46	74.28	**82.33**	77.76
RMPI	73.34	**75.53**	82.45	79.43	75.85	78.68	77.64	52.64	58.84	55.97	71.48	73.41	72.98
S^2DN	**81.23**	74.73	**87.64**	**79.89**	**78.23**	**85.60**	**81.16**	**52.89**	**69.52**	**58.10**	**75.33**	82.15	**78.04**
S^2DN w/o SS	78.49	74.46	84.11	78.02	77.21	83.01	78.87	49.92	68.34	57.11	74.21	78.53	76.01
S^2DN w/o SR	76.31	73.31	84.04	77.61	77.14	81.63	80.28	51.24	63.22	55.75	73.93	76.34	75.58

Table 1: The performance (i.e., **Hits@1**, **Hits@10**, **MRR**, in percentage) of S^2DN on WN18RR dataset. The boldface denotes the highest score and the underline indicates the second-best one.

Methods	Avg.	V1			V2			V3			V4		
	Hits@10	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
DRUM	55.11	30.47	52.92	39.33	27.84	58.73	39.95	26.08	52.90	47.02	27.09	55.88	38.41
NeuralLP	55.16	31.43	52.92	38.25	28.88	58.94	38.39	26.09	52.90	37.40	25.89	55.88	35.96
A*Net	76.57	37.09	57.66	43.63	50.47	77.84	60.33	**61.45**	**85.32**	60.39	56.38	85.46	68.29
CoMPILE	78.95	41.70	63.17	49.92	54.54	81.07	64.14	55.63	84.45	**67.18**	54.77	87.13	67.89
TAGT	77.61	38.05	63.41	48.20	50.84	81.07	61.48	51.21	80.87	61.83	49.89	85.08	62.16
SNRI	79.87	30.98	64.29	42.81	50.52	81.37	**66.58**	53.29	84.87	64.70	54.21	88.97	62.01
RMPI	78.00	41.93	63.41	50.57	54.92	80.54	64.46	53.87	81.33	63.67	52.91	86.73	64.31
S^2DN	**81.25**	**43.68**	**67.34**	**52.10**	**55.45**	**82.38**	64.80	56.31	83.97	65.07	**60.96**	**91.31**	**68.44**
S^2DN w/o SS	79.68	40.48	67.07	48.97	52.41	80.96	63.13	53.34	80.91	63.66	57.21	89.77	66.12
S^2DN w/o SR	78.97	39.76	64.15	48.58	47.59	78.35	58.74	54.16	85.14	64.94	56.32	88.23	67.01

Table 2: The performance (i.e., **Hits@1**, **Hits@10**, **MRR**, in percentage) of S^2DN on FB15k-237 dataset. We mark the best score with bold font and the second best with underline.

emphasizing the significance of fully exploiting relational patterns and pruning links to enhance subgraph reasoning effectiveness. (4) S^2DN surpasses other GNN-based subgraph reasoning methods on most datasets, indicating that denoising unconvincing knowledge by promoting the consistency of relations and reliability of structures significantly enhances inductive KGC performance. In summary, S^2DN enhances the inductive reasoning capabilities of GNN-based models by effectively keeping relational semantics consistent and eliminating unreliable links, unlike previous GNN-based methods that overlook the impact of noise within KGs.

Ablation Study (RQ2)

We undertake an ablation study across all datasets for the inductive KGC. The results are depicted in Table 1 and Table 2, confirming the effectiveness of each module.

S^2DN w/o SS. After removing the semantic smoothing module, there is a notable performance decline across most datasets for inductive subgraph reasoning. This finding underscores the efficacy of maintaining relation consistency within the encompassing subgraph. Consequently, an informative enclosing subgraph featuring semantically consistent relations holds the potential to enhance S^2DN.

S^2DN w/o SR. The exclusion of the structure refining module results in performance degradation across various datasets. This observation highlights the inadequacy of unreliable subgraph structures in conveying information effectively for the downstream KGC task, failing to mitigate the impact of noisy interactions. Conversely, a dependable structure or pristine subgraph enhances inductive reasoning capabilities by disregarding potential noise and preserving reliable interactions.

Robustness of S^2DN (RQ3)

Semantic Consistency of Relations. We conduct an experiment to analyze the semantic consistency. Specifically, as shown in Section , relations with similar semantics tend to be categorized into the same category, which benefits the semantic consistency of relations. During the inference process on the test dataset of WN18RR_V1, FB15k-237_V1, and NELL-995_V1, we count the proportion m_{ij} of relation i that is categorized into another relation j as the degree of semantic consistency. Thus, we can visualize the proportions $M = \{m_{ij} | 0 \leq i, j \leq |\mathcal{R}|\}$ of all relation pairs as shown in Figure 4. For FB15k-237_V1 and NELL-995_V1, we singled out relationships related to the topic of *movie* and *sport*, respectively. We observe from Figure 4 that relations with similar semantics have a higher transformation rate than those with dissimilar semantics (e.g, the similar relations *instance* and *meronym* in Figure 4a, *edited by* and *written by* in Figure 4b). This phenomenon also proves that the Semantic Smoothing module can unify relations with similar semantics, thus maintaining the semantic consistency of the knowledge graph relations.

Type	Model	0%	15% (↓)	35% (↓)	50% (↓)
Semantic	RMPI	82.46	81.07(1.7)	78.98(4.2)	75.31(8.7)
	S^2DN w/o SS	86.11	85.31(0.9)	82.34(4.4)	80.09(7.0)
	S^2DN w/o SR	84.04	83.00(1.2)	82.13(2.3)	80.98(3.6)
	S^2DN	87.64	86.03(1.8)	84.89(3.1)	83.12(5.2)
Structure	RMPI	82.46	80.79(2.0)	78.31(5.1)	76.24(7.5)
	S^2DN w/o SS	86.11	85.07(1.2)	83.87(2.6)	82.05(4.7)
	S^2DN w/o SR	84.04	82.88(1.4)	80.32(4.4)	78.02(7.2)
	S^2DN	87.64	86.88(0.9)	85.79$_{(2.1)}$	84.98(3.0)

Table 3: The results (**Hits@10**) of S^2DN on **WN18RR_V1** under different noise ratios. The underline indicates the drop rate (%) of performance over noisy KGs.

Type	Model	0%	15% (↓)	35% (↓)	50% (↓)
Semantic	RMPI	63.41	61.76(2.6)	59.01(6.9)	57.29(9.7)
	S^2DN w/o SS	67.07	65.22(2.7)	63.03(6.0)	61.89(7.7)
	S^2DN w/o SR	64.15	63.01(1.7)	62.33(2.8)	60.14(6.3)
	S^2DN	67.34	66.05(1.9)	65.78(2.3)	64.97(3.5)
Structure	RMPI	63.41	62.79(0.9)	60.54(4.5)	58.37(7.9)
	S^2DN w/o SS	67.07	66.09(1.4)	64.98(3.1)	63.03(6.0)
	S^2DN w/o SR	64.15	62.21(3.0)	60.07(6.3)	58.89(8.1)
	S^2DN	67.34	66.19(1.7)	65.53(2.6)	64.89(3.6)

Table 4: The performance (**Hits@10**) of S^2DN on **FB15k-237_V1** under various noise ratios. The underline denotes the degree of performance decline (%) over noisy KGs.

Reliability of S^2DN on Noisy KGs. We generate different proportions of *semantic* and *structural* negative interactions (i.e., 5%, 15%, 35%, and 50%) to contaminate the training KG. Semantic noises are created by randomly replacing the relations with others from known triples (e.g, the relation *edited_by* is replaced by *written_by*), while the structure noises are generated by sampling unknown triples from all *entity-relation-entity* combinations. We then compare the performance of RMPI, S^2DN, and its variants on noisy KGs. As shown in Tables 3 and 4, the performance of inductive KGC and their corresponding degradation ratio under different noise levels.

As we introduce more noise, the performance of all methods declines under both semantic and structural settings. This decline is attributed to the dilution of expressive power caused by the randomly introduced false facts. Notably, S^2DN and its variants demonstrate less deduction compared to RMPI across most noisy KGs, suggesting that filtering out irrelevant interactions aids inductive reasoning on subgraphs. The variant S^2DN w/o SS exhibits superior performance on structurally noisy KGs compared to semantically noisy ones, highlighting the efficacy of the semantic smoothing module in modeling consistent relations. Conversely, the variant S^2DN w/o SR outperforms others on semantic noisy KGs rather than structural ones, indicating the importance of the structure refining module in uncovering informative interactions. Furthermore, S^2DN demonstrates comparable performance to other methods on both types of noisy KGs, indicating that S^2DN, equipped with semantic smoothing and structure refining modules, possesses enhanced subgraph reasoning capabilities. This observation shows S^2DN can effectively mitigate unconvincing knowledge while providing reliable local structure.

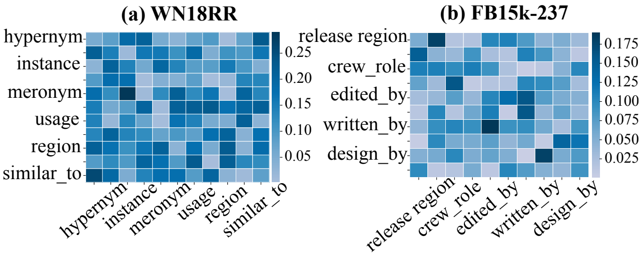
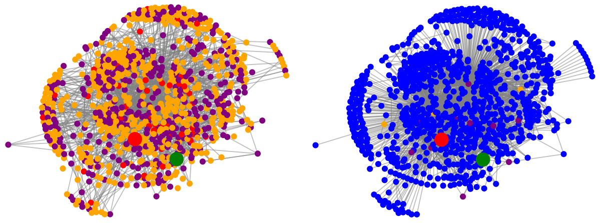


Figure 4: The transition ratio between the original and blurred relations on three datasets (V1 version). The element m_{ij} in the matrix represents the proportion of the relation i is smoothed to relation j.



(a) GraIL, FB15k-237 (b) S^2DN, FB15k-237

Figure 5: The big red and green nodes represent the source and target entities. The small nodes in red, orange, and blue are shared entities involved in $1-3$ hops between source and target nodes. The purple nodes indicate unshared entities.

Visualization of Subgraphs. To explicitly demonstrate the ability of S^2DN to provide reliable links to downstream tasks, we designed a case study on FB15k-237. We visualize the exemplar reasoning subgraph of S^2DN (i.e., the refined subgraph) and GraIL (i.e., the original subgraph) models for different queries in Figure 5. As illustrated in Figure 5, we observe that compared to GraIL, S^2DN can provide more knowledge for enhanced subgraph reasoning while retaining the original reliable information. For example, Figure 5(a) and Figure 5(b) show the subgraphs from GraIL and S^2DN have a similar layout, while S^2DN offers more links and filter out irrelevant interaction between source and target entities. This indicates that S^2DN is effective in subgraph reasoning inductively by a structure-refined mechanism.

Conclusion

We introduced S^2DN to address the challenges posed by the semantic inconsistencies and inevitable noisy interactions in KGs for inductive KGC, emphasizing semantic consistency and structural reliability. Experimental results show that S^2DN surpasses SOTA baselines by keeping relational semantic consistency and offering robust associations. In future work, we will transfer S^2DN to more noise-sensitive domain applications such as biology.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (62272151, 62122025, 62425204, 62450002, 62432011 and U22A2037), the National Science and Technology Major Project (2023ZD0120902), Hunan Provincial Natural Science Foundation of China (2022JJ20016), The science and technology innovation Program of Hunan Province (2022RC1099), and the Beijing Natural Science Foundation (L248013).

References

Bai, J.; Luo, C.; Li, Z.; Yin, Q.; Yin, B.; and Song, Y. 2023. Knowledge graph reasoning over entities and numerical values. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 57–68.

Cao, Q.; Li, B.; Liang, X.; Wang, K.; and Lin, L. 2022. Knowledge-Routed Visual Question Reasoning: Challenges for Deep Representation Embedding. *IEEE Transactions on Neural Networks and Learning Systems*, 33(7): 2758–2767.

Chen, J.; He, H.; Wu, F.; and Wang, J. 2021. Topology-aware correlations between relations for inductive link prediction in knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 6271–6278.

Cheng, K.; Ahmed, N. K.; and Sun, Y. 2023. Neural compositional rule learning for knowledge graph reasoning. In *International Conference on Learning Representations*.

Cover, T. M. 1999. *Elements of information theory*. John Wiley & Sons.

Dettmers, T.; Minervini, P.; Stenetorp, P.; and Riedel, S. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Du, Z.; Zhou, C.; Yao, J.; Tu, T.; Cheng, L.; Yang, H.; Zhou, J.; and Tang, J. 2021. Cogkr: Cognitive graph for multi-hop knowledge reasoning. *IEEE Transactions on Knowledge and Data Engineering*.

Fan, Z.; Xu, K.; Dong, Z.; Peng, H.; Zhang, J.; and Yu, P. S. 2023. Graph Collaborative Signals Denoising and Augmentation for Recommendation. In *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, 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 *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1412–1422.

Geng, Y.; Chen, J.; Pan, J. Z.; Chen, M.; Jiang, S.; Zhang, W.; and Chen, H. 2023. Relational message passing for fully inductive knowledge graph completion. In *Proceedings of IEEE International Conference on Data Engineering (ICDE)*, 1221–1233. IEEE.

Glorot, X.; and Bengio, Y. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of International Conference on Artificial Intelligence and Statistics*, 249–256. JMLR Workshop and Conference Proceedings.

Guo, T.; Xu, C.; Shi, B.; Xu, C.; and Tao, D. 2019. Smooth deep image generator from noises. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 01, 3731–3738.

Hong, Y.; Bu, C.; and Wu, X. 2021. High-Quality Noise Detection for Knowledge Graph Embedding with Rule-Based Triple Confidence. In *Pacific Rim International Conference on Artificial Intelligence, PRICAI 2021*, 572–585.

Jang, E.; Gu, S.; and Poole, B. 2017. Categorical reparameterization with gumbel-softmax. In *Proceedings of International Conference on Learning Representations*.

Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; and Yu, P. S. 2022. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2): 494–514.

Kipf, T. N.; and Welling, M. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of International Conference on Learning Representations*, 1–13.

Li, Q.; Guo, S.; Wu, J.; Li, J.; Sheng, J.; Peng, H.; and Wang, L. 2023a. Event Extraction by Associating Event Types and Argument Roles. *IEEE Trans. Big Data*, 9(6): 1549–1560.

Li, Q.; Li, J.; Wu, J.; Peng, X.; Ji, C.; Peng, H.; Wang, L.; and Yu, P. S. 2024a. Triplet-aware graph neural networks for factorized multi-modal knowledge graph entity alignment. *Neural Networks*, 179: 106479.

Li, Y.; Yu, K.; Zhang, Y.; Liang, J.; and Wu, X. 2023b. Adaptive Prototype Interaction Network for Few-Shot Knowledge Graph Completion. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14.

Li, Z.; Sun, X.; Luo, Y.; Zhu, Y.; Chen, D.; Luo, Y.; Zhou, X.; Liu, Q.; Wu, S.; Wang, L.; et al. 2024b. GSLB: the graph structure learning benchmark. *Advances in Neural Information Processing Systems*, 36.

Lin, X.; Quan, Z.; Wang, Z.-J.; Ma, T.; and Zeng, X. 2020. KGNN: Knowledge Graph Neural Network for Drug-Drug Interaction Prediction. In *Proceedings of International Joint Conference on Artificial Intelligence*, 2739–2745.

Liu, L.; Chen, Y.; Das, M.; Yang, H.; and Tong, H. 2023. Knowledge Graph Question Answering with Ambiguous Query. In *Proceedings of the ACM Web Conference*, 2477–2486.

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.

Mai, S.; Zheng, S.; Yang, Y.; and Hu, H. 2021. Communicative message passing for inductive relation reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 4294–4302.

Meilicke, C.; Fink, M.; Wang, Y.; Ruffinelli, D.; Gemulla, R.; and Stuckenschmidt, H. 2018. Fine-grained evaluation of rule-and embedding-based systems for knowledge graph completion. In *Proceedings of International Semantic Web Conference*, 3–20. Springer.

Nair, V.; and Hinton, G. E. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of International Conference on Machine Learning*, 807–814.

Pan, Y.; Liu, J.; Zhang, L.; Zhao, T.; Lin, Q.; Hu, X.; and Wang, Q. 2022. Inductive relation prediction with logical reasoning using contrastive representations. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, 4261–4274.

Pujara, J.; Augustine, E.; and Getoor, L. 2017. Sparsity and noise: Where knowledge graph embeddings fall short. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, 1751–1756.

Qu, M.; Chen, J.; Xhonneux, L.-P.; Bengio, Y.; and Tang, J. 2020. Rnnlogic: Learning logic rules for reasoning on knowledge graphs. In *Proceedings of International Conference on Learning Representations*.

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

Sadeghian, A.; Armandpour, M.; Ding, P.; and Wang, D. Z. 2019. Drum: End-to-end differentiable rule mining on knowledge graphs. *Advances in Neural Information Processing Systems*, 32.

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 *Proceedings of Extended Semantic Web Conference*, 593–607.

Sun, Q.; Li, J.; Peng, H.; Wu, J.; Fu, X.; Ji, C.; and Philip, S. Y. 2022. Graph structure learning with variational information bottleneck. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 4165–4174.

Teru, K.; Denis, E.; and Hamilton, W. 2020. Inductive relation prediction by subgraph reasoning. In *Proceedings of International Conference on Machine Learning*, 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 *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, 122–132.

Tishby, N.; Pereira, F. C.; and Bialek, W. 2000. The information bottleneck method. *arXiv preprint physics/0004057*.

Toutanova, K.; Chen, D.; Pantel, P.; Poon, H.; Choudhury, P.; and Gamon, M. 2015. Representing text for joint embedding of text and knowledge bases. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1499–1509.

Wang, P.; Han, J.; Li, C.; and Pan, R. 2019. Logic attention based neighborhood aggregation for inductive knowledge graph embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 7152–7159.

Wang, Q.; Mao, Z.; Wang, B.; and Guo, L. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12): 2724–2743.

Wang, W.; Feng, F.; He, X.; Nie, L.; and Chua, T.-S. 2021. Denoising implicit feedback for recommendation. In *Proceedings of ACM International Conference on Web Search and Data Mining*, 373–381.

Xiong, W.; Hoang, T.; and Wang, W. Y. 2017. Deeppath: A reinforcement learning method for knowledge graph reasoning. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 564–573.

Xu, X.; Zhang, P.; He, Y.; Chao, C.; and Yan, C. 2022. Subgraph neighboring relations infomax for inductive link prediction on knowledge graphs. In *Proceedings of International Joint Conference on Artificial Intelligence*, 2341–2347.

Xue, B.; and Zou, L. 2023. Knowledge Graph Quality Management: A Comprehensive Survey. *IEEE Transactions on Knowledge and Data Engineering*, 35(5): 4969–4988.

Yang, F.; Yang, Z.; and Cohen, W. W. 2017. Differentiable learning of logical rules for knowledge base reasoning. *Advances in Neural Information Processing Systems*, 30.

Yang, Y.; Huang, C.; Xia, L.; and Huang, C. 2023. Knowledge graph self-supervised rationalization for recommendation. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3046–3056.

Zhang, Q.; Dong, J.; Tan, Q.; and Huang, X. 2024. Integrating Entity Attributes for Error-Aware Knowledge Graph Embedding. *IEEE Transactions on Knowledge and Data Engineering*, 36(4): 1667–1682.

Zhang, W.; Paudel, B.; Wang, L.; Chen, J.; Zhu, H.; Zhang, W.; Bernstein, A.; and Chen, H. 2019. Iteratively learning embeddings and rules for knowledge graph reasoning. In *The World Wide Web Conference*, 2366–2377.

Zhang, W.; Zhu, Y.; Chen, M.; Geng, Y.; Huang, Y.; Xu, Y.; Song, W.; and Chen, H. 2023a. Structure Pretraining and Prompt Tuning for Knowledge Graph Transfer. In *Proceedings of the ACM Web Conference*, 2581–2590.

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

Zhang, Y.; and Yao, Q. 2022. Knowledge graph reasoning with relational digraph. In *Proceedings of the ACM Web Conference*, 912–924.

Zhang, Y.; Zhou, Z.; Yao, Q.; Chu, X.; and Han, B. 2023b. Adaprop: Learning adaptive propagation for graph neural network based knowledge graph reasoning. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3446–3457.

Zhu, X.; Du, Y.; Mao, Y.; Chen, L.; Hu, Y.; and Gao, Y. 2023. Knowledge-refined Denoising Network for Robust Recommendation. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*, 362–371.

Zhu, Z.; Yuan, X.; Galkin, M.; Xhonneux, L.-P.; Zhang, M.; Gazeau, M.; and Tang, J. 2024. A* net: A scalable path-based reasoning approach for knowledge graphs. *Advances in Neural Information Processing Systems*, 36.