Improving Knowledge Graph Completion with Structure-Aware Supervised Contrastive Learning

Jiashi Lin[1], Lifang Wang[1*], Xinyu Lu[2], Zhongtian Hu[1], Wei Zhang[1], Wenxuan Lu[1]
[1]School of Computer Science, Northwestern Polytechnical University, Xi'an, China,
[2]School of Information, Xi'an University of Finance and Economics, Xi'an, China
jslin@mail.nwpu.edu.cn

Abstract

Knowledge Graphs (KGs) often suffer from incomplete knowledge, which restricts their utility. Recently, Contrastive Learning (CL) has been introduced to Knowledge Graph Completion (KGC), significantly improving the discriminative capabilities of KGC models and setting new benchmarks in performance. However, existing contrastive methods primarily focus on individual triples, overlooking the broader structural connectivities of KGs. This narrow focus hampers a more comprehensive understanding of the graph's structural knowledge. To address this gap, we propose **StructKGC**, a novel contrastive learning framework designed to flexibly accommodate the diverse topologies inherent in KGs. We introduce four contrastive tasks tailored to KG data: Vertex-level CL, Neighbor-level CL, Path-level CL, and Relation composition level CL. These tasks are trained synergistically during the fine-tuning of pre-trained language models (PLMs), allowing for a more nuanced capture of subgraph semantics. To validate the effectiveness of our method, we perform a comprehensive set of experiments on several real-world datasets. The experimental results demonstrate that our approach achieves SOTA performance under standard supervised and low-resource settings. Furthermore, we observe that the various structure-aware tasks introduced can mutually reinforce each other, resulting in consistent performance enhancements.

1 Introduction

Knowledge graphs, such as Freebase and Wikidata, are stores of relational facts that have become crucial sources of knowledge in knowledge-intensive applications. A KG typically comprises a collection of triples, where each triple (h, r, t) signifies the relationship between a head entity, a tail entity, and the corresponding relation. Factual knowledge is virtually infinite and subject to frequent changes, leading to concerns about the incompleteness of KGs.

To tackle this issue, researchers have focused on Knowledge Graph Completion (KGC) models that automatically fill in missing triples. These models fall into two main categories: embedding-based methods and text-based methods. Embedding-based methods learn low-dimensional vectors for entities and relations by minimizing a loss function (Bordes et al., 2013; Trouillon et al., 2016; Sun et al., 2019; Balazevic et al., 2019). Text-based methods (Wang et al., 2021, 2022, 2023) leverage available text to gather textual information for entities and relations. Recent approaches have applied Contrastive Learning (CL) to text-based KGC models, significantly improving their discriminative power (Wang et al., 2022, 2023). Typically, they use a dual-tower architecture that utilizes Pre-trained Language Models (PLMs) to produce textual embeddings and then use the InfoNCE contrastive objective (Oord et al., 2018) to perform instance discrimination. Despite their effectiveness, current contrastive approaches are not explicitly designed to identify the graph structure in KGs. They typically pair the entity-relation pair (h, r) with one positive tail entity t from the same triple (as shown in 2a). However, using just one positive sample, this approach fails to capture the broader connectivities and diverse topologies in KGs, which are crucial for understanding complex relation mappings (e.g., one-to-many, many-to-one) (Ji et al., 2015) and long-range dependencies (Lin et al., 2015a). Entities in a KG are often surrounded by multiple neighboring entity-relation pairs that enrich their profiles. Additionally, paths between entities can reveal meaningful patterns and dependencies, offering insights into intricate relations.

Furthermore, we observe that entities with similar positions in a KG share underlying attributes,

[*] Corresponding author.

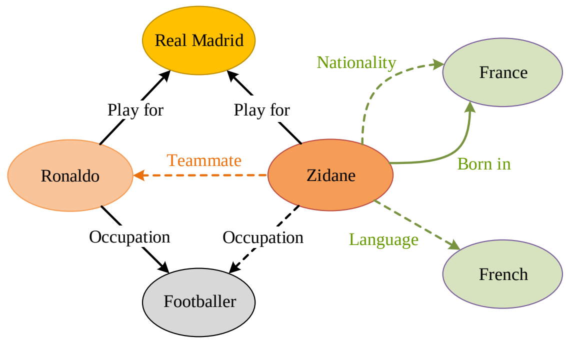


Figure 1: An example of knowledge graph completion. Structural contexts help infer missing facts. In the subgraph, entities *Ronaldo* and *Zidane*, sharing the context $(Play\_for, Real\_Madrid)$, exhibit analogous roles as $Footballer$. The entity-relation pairs $(Language, French)$, $(Nationality, French)$, and $(Born\_in, France)$ share relevant semantic implications for $Zidane$. Moreover, the path $Play\_for \wedge Play\_for^{-1}$ shares similar properties with relation $Teammate$ due to similar semantic interaction.

consistent with the distributional semantics assumption that similar words appear in similar contexts (Suresh et al., 2023). Assume that entities represented by e_i occupy the same structural position as $(e_i \xrightarrow{category} animal)$. Regardless of whether these entities pertain to dogs, cats, or lions, they exhibit analogous roles, such as respiration and sustenance requirements. Establishing such semantic relevance is essential for inferring incomplete facts. Fig.1 illustrates an example of knowledge graph completion, the co-occurrence of $Ronaldo$ and $Zidane$ in the context of $(Play\_for, Real\_Madrid)$ suggests that they possess analogous conceptual attributes. Therefore, inferring Zidane's accurate profession becomes more plausible when considering Ronaldo's occupation. This principle extends to entity-relation pairs and multi-hop paths. For example, the path $Play\_for \wedge Play\_for^{-1}$ shares similar properties with the relation $Teammate$, as both involve similar semantic interactions, mirroring the underlying relational patterns. However, current text-based methods primarily focus on internal links within triples, disregarding the external semantic relevance beyond the triples. This results in incomplete representations that miss the holistic structural knowledge. Therefore, integrating contextual structural information is crucial for text-based KGC, yet it remains largely underexplored.

Building on our analysis, we systematically investigate the effects of various structural forms. Instead of focusing on individual triples, we consider the structural context around an anchor as weakly positive instances. Our key idea is to maximize the mutual information between the central anchor and its context. To achieve this, we explore various subgraph topologies surrounding each triple, including vertex, neighbor, and path structures, as shown in Fig.2. Then, we propose a novel structure-aware contrastive learning framework, named **StructKGC**, which offers the flexibility to effortlessly handle a wide range of topology types and quantities within KGs, without introducing additional parameters. Following this, we introduce four supervised contrastive learning tasks tailored for KG data: *Vertex-level CL*, *Neighbor-level CL*, *Path-level CL*, and *Relation Composition Level CL*, each designed to capture distinct knowledge properties. During the fine-tuning stage of the PLMs, we jointly train these tasks, which facilitates a collaborative reinforcement effect among the different tasks and enables the model to effectively capture the underlying semantics within subgraphs.

In summary, the contributions of this work are as follows:

1. To the best of our knowledge, we are the first to systematically investigate the impact of different structures within KG for dual-tower-based KGC.

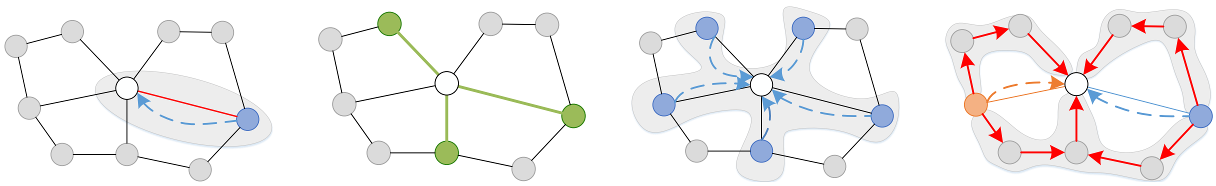
2. We propose a novel contrastive learning framework that incorporates four structure-aware tasks, enabling comprehensive structure awareness at the vertex, neighbor, path, and relation composition levels.

3. Experiments and analysis demonstrate the effectiveness and efficiency of this work against state-of-the-art approaches in standard supervised and low-resource settings.

2 Related Work

2.1 Embedding-based methods

Knowledge graph completion aims to infer missing facts or relationships in a knowledge graph. Embedding-based methods represent entities and relations in a continuous vector space. TransE (Bordes et al., 2013) operates on the translation assumption, i.e., $h + r \approx t$, but struggles with complex relations. TransR (Lin et al., 2015b) embeds entities and relations in different semantic spaces. RotatE (Sun et al., 2019) uses complex-valued embeddings and achieves semantic transformation via rotation



(a) Triple-based Positive (b) Vertex-based Positives (The green edges represent the same type of relation) (c) Neighbor-based Positives (d) Path-based Positives

Figure 2: An example of a subgraph around a triple. The triple-based method creates a single positive for each anchor from a sample triple (i.e., the object entity in the same triple as the anchor). The vertex-based positives represent multiple entities that connect to the same entity-relation pair. The neighbor-based positives refer to multiple entity-relation pairs surrounding the same entity. The path-based positives represent a collection of routes connecting the initial entity to the final entity.

on the complex plane. These methods, however, treat each triple separately, neglecting global graph information.

To address this, researchers have proposed methods considering the graph structure in KGs. R-GCN (Schlichtkrull et al., 2018) introduces a Graph Convolutional Network (GCN) variant for relational data. CompGCN (Vashishth et al., 2020) uses composition operations to jointly embed entities and relations. The Path Ranking Algorithm (PRA) (Lao and Cohen, 2010) captures contextual relationships via indirect paths. PtransE (Lin et al., 2015a) treats multi-hop paths as new relations. These studies have showcased the efficacy of incorporating graph context into knowledge graph completion. Nevertheless, these methodologies have limitations as they fail to consider the potential semantic correlations among the knowledge graph contexts. Additionally, they struggle considerably in formulating predictive models within an inductive paradigm.

2.2 Text-based methods

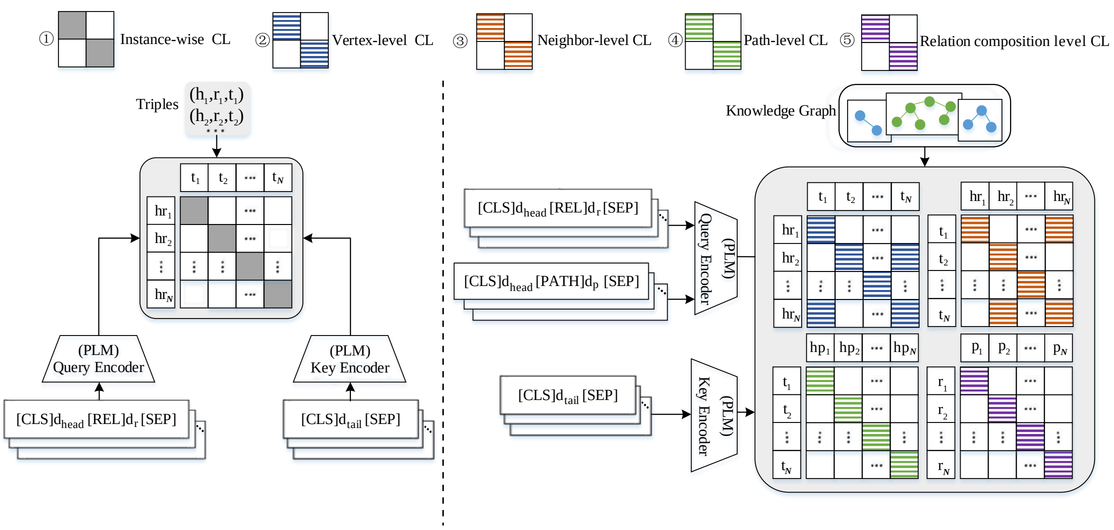
Text-based methods use descriptions to capture the semantics of knowledge graph components, enabling the inference of unseen entities. KG-BERT (Yao et al., 2019) first proposed using PLMs to model KGs by simply concatenating textual descriptions for binary classification. However, this approach suffers from low efficiency due to combinatorial explosion. StAR (Wang et al., 2021) addresses this by using two encoders to decouple the triple. MKGformer (Chen et al., 2022b) transforms link prediction into masked language modeling, improving inference efficiency. SimKGC (Wang et al., 2022) and C-LMKE (Wang et al., 2023) propose a contrastive learning framework for more discriminative KGC models, while Jiang et al. (Jiang et al., 2023) explore various negative sampling strategies. Recent research (Zhang et al., 2023) has introduced Large Language Models (LLMs) as sequence-to-sequence generators for KGC. AutoKG (Zhu et al., 2024) uses prompt engineering and evaluates GPT-3.5 and GPT-4 (Achiam et al., 2023) on KGC tasks. Although promising, text-based methods face challenges in capturing the abundant structural information inherent in knowledge graphs.

3 Methodology

In this section, we first introduce Knowledge Graph Completion (KGC), and then we investigate the existing contrastive learning loss functions employed in the KGC, analyzing its potential drawbacks. Following this, we propose a novel structure-aware contrastive learning framework. Fig.3b presents an overview of our method, which consists of four contrastive learning objectives tailored to the characteristics of knowledge graphs.

3.1 Task Formulation: Knowledge Graph Completion

Previous works often treat KG as a composition of triples, which represent individual facts $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$. Each triple in $\mathcal{T}$ has the form (h, r, t), where the head entity and tail entity $h, t \in \mathcal{E}$ and the relation $r \in \mathcal{R}$. In order to leverage the rich structural knowledge contained in KG, we formally represent the set of triples as a directed graph $\mathbb{G}$ that includes vertices (entities) and directed edges (relations). Each directed link in the graph, denoted as $l = (v_i, e_j, v_k) \in \mathbb{G}$, represents a fact. Given a query $q = (v, e, ?)$, where v is the source vertex and e is the relational edge, knowledge graph completion aims to enable effi-



(a) C-LMKE & SimKGC. (b) StructKGC.

Figure 3: (a) C-LMKE and SimKGC use instance-wise contrastive learning to train dual-tower PLMs, where each query (h, r) is paired with a positive instance t. (b) Our StructKGC overview: Vertex-level CL contrasts each query (h, r) with multiple vertice-based positives $P_v(h, r)$; Neighbor-level CL contrasts each key t with multiple neighbors $P_n(t)$; Path-level CL captures long-range dependencies by contrasting t with head-path pairs; Relation composition level CL contrasts multi-hop composite relations with direct relations. The combination of the four tasks allows PLMs to sufficiently perceive structural knowledge within KG.

cient retrieval and gather a set of candidate entities $V_o = \{v_o\}$ s.t. $(v, e, v_o) \notin \mathbb{G}$ due to the incompleteness of KG.

3.2 Revisiting Contrastive Loss of Knowledge Graph Completion

Contrastive learning has been proven successful in the task of knowledge graph completion. Given a batch of triple samples $(h, r, t)_{i=1}^{N}$, a dual-tower architecture with PLMs is employed to separately encode decoupled triples, as shown in Fig.2a. Specifically, the relation-aware embedding e_{hr} for (h, r) is computed by query encoder BERT_{hr}, while the tail entity embedding e_t is calculated by key encoder BERT_t. Then, cosine similarity is used as the scoring function to measure the distance between the two components. Following the InfoNCE loss (Chen et al., 2020), the general loss function can be represented as:

$$\mathcal{L} = -\log \frac{e^{\phi(hr,t)/\tau}}{\sum_{i=1}^{|\mathcal{N}|} e^{\phi(hr,t_i)/\tau}} \tag{1}$$

Here, $\phi(hr, t) = cos(e_{hr}, e_t) \in [-1, 1]$ represents the cosine similarity. $\mathcal{N}$ represents a set of negative examples in the same batch. The learnable temperature parameter τ is introduced to control the relative significance of these negatives. Note that the instance-wise contrastive loss described in Eq.1 only contrasts the entity-relation pair (h, r) exclusively with a single positive tail entity. As a result, it learns little about the structural information preserved in the knowledge graph.

3.3 Structure-Aware Contrastive Learning Framework

In this work, we consider the structural context beyond individual triples to enable PLMs to perceive more structural knowledge. A straightforward approach is to consider these structures as additional contrastive supervision. In this case, the weak positives are derived from the subgraph of a given anchor rather than from a single triple sample. Our primary objective is to maximize the mutual information between the anchor and its structural context, thus necessitating knowledge representation to integrate the underlying shared semantics. However, the original InfoNCE loss cannot handle scenarios with multiple positive instances. Inspired by (Khosla et al., 2020), we generalize Eq.1 to support structure-aware multi-positive contrastive learning:

$$\mathcal{L}_{sup} = -\frac{1}{|P(q)|} \sum_{q_p \in P(q)} \log \frac{e^{\phi(q,q_p)/\tau}}{\sum_{i=1}^{|\mathcal{N}|} e^{\phi(q,q_i)/\tau}} \tag{2}$$

Generally, $\phi(q, q_p) = cos(e_q, e_{q_p}) \in [-1, 1]$, q denotes the query (i.e., an entity-relation pair hr or an entity t); q_p is the positive representation of q derived from $P(q)$; $P(q)$ consists of multiple positive samples surrounding q, which are context-specific and categorized into three types in our approach: vertex-based positives, neighbor-based positives, and path-based positives, as depicted in Fig.2.

Since diverse structures depict different knowledge perspectives, we integrate four structure-aware contrastive tasks into the fine-tuning paradigm of PLMs, as illustrated in Fig.3b. Next, we will introduce these tasks and different types of structural positives in detail.

3.3.1 Vertex-level CL

The intricate mapping properties of relations in knowledge graphs (KGs), often result in multiple vertices connecting to the same head-relation query, offering rich semantics. To harness this, we formally construct vertex-based positives as $P_v(v, e) = \{v_k \mid \forall(v, e, v_k) \in \mathbb{G}\}$, which represents the set of vertices that can be reached from a given vertex v through a specific outgoing edge e. Based on this, a Vertex-level CL task is proposed to align the entity-relation pair with its vertex-based positives. Following Eq.2, the Vertex-level CL loss can be defined as:

$$\mathcal{L}_{VC} = -\frac{1}{|P_v(h,r)|} \sum_{q_p \in P_v(h,r)} \log \frac{e^{\phi(hr,q_p)/\tau}}{\sum_{i=1}^{|\mathcal{N}|} e^{\phi(hr,q_i)/\tau}} \tag{3}$$

3.3.2 Neighbor-level CL

Neighbor-level CL focuses on modeling the neighbors of an entity, which comprises incoming edges and adjacent vertices connected through those edges. Given a target entity v, we formally define its neighbor-based positives as a set of tuples: $P_n(v) = (v_i, e_j) \mid \forall(v_i, e_j, v) \in \mathbb{G}$. By examining this local structure, we can gain valuable insights into the entity's nature and neighbors' relevance. Similarly, the loss of Neighbor-level CL is defined as:

$$\mathcal{L}_{NC} = -\frac{1}{|P_n(t)|} \sum_{q_p \in P_n(t)} \log \frac{e^{\phi(t,q_p)/\tau}}{\sum_{i=1}^{|\mathcal{N}|} e^{\phi(t,q_i)/\tau}} \tag{4}$$

3.3.3 Path-level CL

A L-hop path $p_i(v_0, v_j)$ from head $h(v_0)$ to tail $t(v_L)$ is defined as: $h(v_0) \xrightarrow{e_0} v_1 \xrightarrow{e_1} v_2 \cdots (v_{L-1}) \xrightarrow{e_{L-1}} t(v_L)$, where v_i and v_{i+1} are connected by edge e_i. The path-based positives $P_p(v_0, v_j)$ is then defined as the set of paths $\{p_1, p_2, \cdots, p_n\}$. Path-level CL distinguishes whether an entity pair matches the multi-hop path. Similar to the PCRA algorithm (Lin et al., 2015a), the path reliability $R(p|h, r)$ is based on the flow of resources from the initial entity to the final entity. Then, we propose a weighted contrastive loss as follows:

$$\mathcal{L}_{PC} = -\frac{1}{|P_p(h,t)|} \sum_{p_p \in P_p(h,t)} R(p|h,r) \log \frac{e^{\phi(t,hp_p)/\tau}}{\sum_{i=1}^{N} e^{\phi(t,hp_i)/\tau}} \tag{5}$$

3.3.4 Relation Composition Level CL

Multi-hop paths enable logical inference of direct relations, forming complex queries and uncovering meaningful connections between entities. We introduce a relation composition level CL task to capture this. Based on the path reliability factor, the loss is defined as follows:

$$\mathcal{L}_{RC} = -\frac{1}{|P_p(h,t)|} \sum_{p_p \in P_p(h,t)} R(p|h,r) \log \frac{e^{\phi(r,p_p)/\tau}}{\sum_{i=1}^{N} e^{\phi(r,p_i)/\tau}} \tag{6}$$

3.4 Structural Positives Encoding

Following C-LMKE (Wang et al., 2023) and SimKGC (Wang et al., 2022), we utilize a pair of BERT-style architectures, with a query encoder to encode (h, r) and (h, p), and a key encoder for t, respectively. To mitigate the time-consuming nature of path extraction, we perform multi-hop path extraction during the preprocessing phase, decoupling it from the training process. Inspired by (Lin et al., 2015a), we limit path length and apply pruning techniques, retaining only paths with a reliability score greater than 0.01. To reduce the computational cost of encoding phase, we use an in-batch strategy that reuses potential positive and negative samples within the same batch, improving data efficiency and enabling practical training. Substantially, we convert the corresponding textual descriptions into input sequences and then input these sequences into the BERT encoder. Similar to SimKGC, we use mean pooling followed by L2 normalization to obtain the knowledge graph embeddings.

3.5 Model Training

Different structure-aware CL tasks capture distinct aspects of structural knowledge. To facilitate knowledge sharing across tasks, we train our KGC models by jointly performing these tasks. Specifically, we introduce a weighted combination of contrastive loss functions, each tailored to a specific task. The overall loss function is formulated as Eq.7.

$$\mathcal{L}_{overall} = w_1\mathcal{L}_{VC} + w_2\mathcal{L}_{NC} + w_3\mathcal{L}_{PC} + w_4\mathcal{L}_{RC} \tag{7}$$

Where w_i is tunable hyper-parameters for adapting to specific knowledge graph characteristics.

4 Experiments

4.1 Datasets

To assess the effectiveness of our approach, we evaluate it on two popular benchmark datasets: WN18RR (Dettmers et al., 2018) and FB15k-237 (Toutanova and Chen, 2015). The dataset statistics are shown in Table 1. WN18RR is a subset of WordNet, obtained by removing reversible relation data and filtering out facts related to inverse relations to avoid information leakage. FB15k-237 is a subset of Freebase, created by removing a significant amount of reversible relation data and filtering out trivial triples. We incorporated the textual information from KG-BERT (Yao et al., 2019) for the WN18RR and FB15k-237 datasets.

Dataset	# Ent	# Rel	# train	# valid	# test
FB15k-237	14541	237	272,115	17,535	20,466
WN18RR	40943	11	86,835	3,034	3,134

Table 1: Statistics of the benchmark datasets.

4.2 Baselines

In our study, we conducted a comparative analysis of our methods against both embedding-based and text-based approaches. The embedding-based methods we considered encompass TransE, ComplEx, RotatE, ConvE, CompGCN, Tucker, CompoundE, KRACL and RotatE-VLP. On the other hand, the text-based methods we evaluated include KG-BERT, MTL-KGC, StAR, KG-S2S, C-LMKE, SimKGC, LP-BERT and GHN.

4.2.1 Evaluation Metrics

Following previous work, we evaluate StructKGC using the knowledge graph completion task. Our evaluation involved all test triples (h, r, t), and our trained model aimed to rank all entities related to the predicted tail entity pairs $(h, r, ?)$ or head entity pairs $(t, r^{-1}, ?)$ for predicting t or h, respectively. To assess the performance, we employ four evaluation metrics: Mean Reciprocal Rank (MRR) and Hits@k (H@k, where $k \in \{1, 3, 10\}$). MRR represents the average reciprocal rank of all test triples, while H@k measures the proportion of correctly ranked entities within the top-k predicted entities.

4.3 Implementation Detail

Our knowledge graph completion model is implemented based on Pytorch (Paszke et al., 2019). The dual-tower encoders are initialized from the pretrained BERT-based-uncased model. For fair competition, we adhere to the setup of SimKGC (Wang et al., 2022) and use the same hyperparameters as presented in the oringinal paper. For newly introduced coefficients w_i, we use grid search to tune with a search range of $\{0.2, 0.4, 0.6, 0.8, 1\}$. All the experiments are executed on 2 A100 GPU. For further details, please refer to Appendix A.

4.4 Main Result

We compare our model with established links prediction task approaches on standard benchmarks, including FB15k-237 and WN18RR. Table 2 reports the link prediction performance of the baselines and our method with standard deviation from three runs using different random seeds. Based on the MRR, which most accurately depicts a model's total performance, our method achieved significant improvements over the embedding-based methods, with a margin of 2.1% on FB15k-237 and 16.9% on WN18RR, respectively. Additionally, our StructKGC performs better than text-based SOTA methods on both datasets, with a margin of 4.4% on FB15k-237 and 1.8% on WN18RR, respectively. This indicates that simultaneously leveraging structural knowledge and implicit textual knowledge within pre-trained language models can effectively improve the performance of KGC tasks. Overall, StructKGC markedly improves upon existing SOTA baselines.

4.5 Low-Resource Evaluation

To evaluate the performance of our method compared to baseline models in a low-resource setting, we randomly select factual triples to create a training subset for FB15k-237 and assess the models using the complete test set.

Fig.4 shows the MRR metrics for various competitive baselines and our proposed model in low-resource link prediction scenarios. As expected,

Method	FB15k-237				WN18RR			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TransE(Bordes et al., 2013)†	27.9	19.8	37.6	44.1	24.3	4.3	44.1	53.2
ComplEx(Trouillon et al., 2016)†	27.8	19.4	29.7	45.0	44.9	40.9	46.9	53.0
RotatE(Sun et al., 2019)†	33.8	24.1	37.5	53.3	47.6	42.8	49.2	57.1
ConvE(Dettmers et al., 2018)†	31.2	22.5	34.1	49.7	45.6	41.9	47.0	53.1
CompGCN(Vashishth et al., 2020)	35.5	26.4	39.0	53.5	48.1	44.8	49.2	54.8
TuckER(Balazevic et al., 2019)†	35.8	26.6	39.4	54.4	47.0	44.3	48.2	52.6
CompoundE(Ge et al., 2023)	35.0	26.2	39.0	54.7	49.2	45.2	51.0	57.0
KRACL(Tan et al., 2023)	36.0	26.6	39.5	<u>54.8</u>	52.7	48.2	54.7	61.3
RotatE-VLP(Li et al., 2023b)	<u>36.2</u>	<u>27.1</u>	<u>39.7</u>	54.2	49.8	45.5	51.4	58.2
KG-BERT(Yao et al., 2019)†	-	-	-	42.0	21.6	4.1	30.2	52.4
MTL-KGC(Kim et al., 2020)	26.7	17.2	29.8	45.8	33.1	20.3	38.3	59.7
StAR(Wang et al., 2021)	29.6	20.5	32.2	48.2	40.1	24.3	49.1	70.9
KG-S2S(Chen et al., 2022a)	33.6	25.7	37.3	49.8	57.4	53.1	59.5	66.1
C-LMKE(Wang et al., 2023)	30.6	21.8	33.1	48.4	61.9	52.3	67.1	78.9
$\text{SimKGC}_{\text{IB}}$(Wang et al., 2022)	33.3	24.6	36.2	51.0	67.1	58.7	<u>73.1</u>	81.7
LP-BERT (Li et al., 2023a)	31.0	22.3	33.6	49.0	48.2	34.3	56.3	75.2
GHN(Qiao et al., 2023)	33.9	25.1	36.4	51.8	<u>67.8</u>	<u>59.6</u>	71.9	<u>82.1</u>
StructKGC	**38.3(±0.10)**	**28.9(±0.08)**	**41.9(±0.06)**	**56.6(±0.14)**	**69.6(±0.10)**	**62.3(±0.12)**	**74.1(±0.07)**	**82.7(±0.09)**

Table 2: Main results on FB15k-237 and WN18RR datasets, †: results are from (Wang et al., 2021), and the other results are taken from the corresponding papers. Bold numbers represent the best and underlined numbers represent the second best.

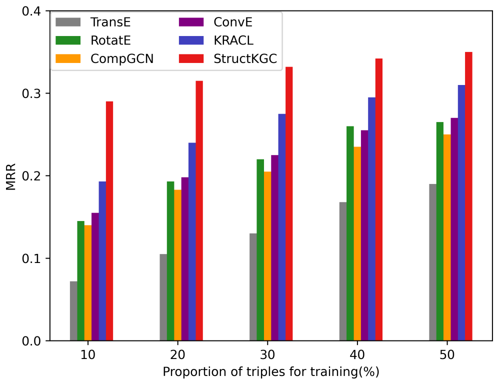


Figure 4: Link prediction performance on the FB15k-237 dataset under low-resourse setting.

performance declines across all models as the training data decreases. However, despite this trend, our model consistently outperforms the baselines, demonstrating its superior data efficiency in leveraging KG data. Moreover, the results highlight the robustness and stability of StructKGC, evidenced by the relatively low standard deviations. This strong performance can be attributed to the comprehensive supervision provided by our structure-aware framework, which enables the model to fully perceive and leverage structural data.

4.6 Study of Relation Catergory

Knowledge graphs contain complex relation mappings, categorized into four groups: one-to-one (1-to-1), one-to-many (1-to-M), many-to-one (M-to-1), and many-to-many (M-to-M). To further analyze the performance of StructKGC across different relation categories, we use the categorization approach proposed by (Bordes et al., 2013). Table 3 presents the statistical results in FB15k-237.

Relation Category	Numbers of Triples	Propotion (%)
One-to-One	192	0.94
One-to-Many	1293	6.32
Many-to-One	4185	20.45
Many-to-Many	14796	72.29

Table 3: Statistics of relation categories on FB15k-237 dataset.

We report the performance of our model compared to the baselines across four different relationship categories. Table 4 shows our findings: Firstly, When it comes to triples with one on the tail side, text-based approaches exhibit a notable advantage. This advantage can be ascribed to the capability of text-based methods to encompass supplementary textual knowledge, thereby mitigating the constraints of a singular structure. Secondly, predicting multiple entities presents a notably greater challenge, leading to a decline in the performance of all methods on M-side prediction. This highlights the critical role of effectively learning complex relation mapping. Despite these challenges, our method generally outperforms baselines across all metrics. Compared to the baseline method SimKGC, our approach shows substantial improvements in M-side predictions.

Category	TransE		DistMult		ConvE		CompGCN		SimKGC		StructKGC	
	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10
Forward Prediction												
1-to-1	47.6	58.8	25.7	31.2	36.6	51.0	45.3	58.9	<u>70.1</u>	<u>87.9</u>	**70.4**	**88.5**
1-to-M	6.0	11.8	3.2	6.7	6.9	15	7.6	15.1	<u>9.4</u>	<u>18.6</u>	**10.2**	**19.8**
M-to-1	53.6	84.6	57.5	75.0	76.2	87.8	77.9	88.5	<u>78.8</u>	<u>88.4</u>	**79.6**	**88.9**
M-to-M	28.7	55.3	18.4	37.6	37.5	60.3	<u>39.5</u>	**61.6**	34.8	56.31	**39.7**	<u>60.6</u>
Backward Prediction												
1-to-1	48.4	59.3	25.5	30.7	37.4	50.5	45.7	60.4	<u>73.3</u>	<u>92.7</u>	**74.3**	**93.8**
1-to-M	32.9	58.9	32.2	55.8	44.4	64.4	<u>47.1</u>	<u>65.6</u>	46.2	63.8	**47.4**	**65.7**
M-to-1	8.0	15.2	3.8	7.1	9.1	17.0	11.2	19.0	<u>15.3</u>	<u>27.7</u>	**19.5**	**34.3**
M-to-M	21.9	43.6	13.1	25.5	26.1	45.9	<u>27.5</u>	<u>47.4</u>	24.1	43.9	**28.0**	**48.8**

Table 4: Link prediction performance by relation category on FB15k-237 dataset

Method	FB15k-237				WN18RR			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
$\text{SimKGC}_{\text{IB}}$	33.3	24.6	36.2	51.0	67.1	58.5	73.1	81.7
$\text{SimKGC}_{\textbf{IB+VC}}$	34.3	25.1	37.9	53.1	68.5	61.3	72.7	82.0
$\text{SimKGC}_{\textbf{IB+VC+NC}}$	35.2	25.5	38.6	53.6	69.3	**62.6**	73.5	82.5
$\text{SimKGC}_{\textbf{IB+VC+NC+PC+RC}}$	**38.3**	**28.9**	**41.9**	**56.6**	**69.6**	62.3	**74.1**	**82.7**

Table 5: Ablation on structure-aware contrastive learning supervisions. VC/NC/PC/RC denotes *Vertex-level CL*, *Neighbor-level CL*, *Path-level CL* and *Relation composition level CL* respectively.

4.7 Ablation on Various Contrastive Tasks

We conduct an ablation study to investigate the effectiveness of our proposed contrastive tasks using SimKGC with in-batch negatives as a baseline. Table 5 presents the results. Incorporating VC enhances performance across all metrics on both datasets, which is attributed to better modeling of entity-relation queries, especially those with multiple tail entities. NC further improves results, suggesting it captures implied associations between entities and their neighbors more effectively. The path-related tasks PC+RC show a substantial 8.8% MRR increase on the FB15k-237 dataset, though the gain on WN18RR is less significant. This disparity may be due to dataset characteristics. As shown in Fig.7c, FB15k-237 has more paths, offering abundant training signals. WN18RR, derived from WordNet, has fewer relations but many transitive relationships. Encoding too many redundant paths in a text encoder with limited token length could harm the expressive capacity of knowledge representations. Although expanding token length could help, it raises encoding overhead and falls outside the scope of this paper.

5 Ablation on positive sample quantity

In this paper, we use an in-batch strategy to effectively reuse samples, resulting in an increase in both negative and positive samples as the batch size grows. This makes it difficult to analyze the contribution of positive samples alone, as the model's performance is also affected by changes in the number of negatives. To address this, we fix the batch size and limit the maximum sampling of positive samples per batch to analyze their impact on FB15k-237. Fig.5 shows that the model's performance can be significantly enhanced by including more positive samples, especially when the number of positive samples is small. This suggests that integrating further structural context can augment the model's performance. With increasing positive instances, the model can grasp a wider structural context, facilitating a deeper understanding of structural semantics. Nonetheless, we have observed that the advantage of adding more positive samples becomes less significant as the number of samples increases.

6 Conclusion

In this work, we propose a simple yet effective framework (**StructKGC**) that learns knowledge

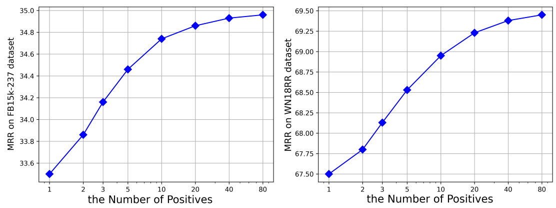


Figure 5: Effect of positive quantity on FB15k-237 and WN18RR datasets.

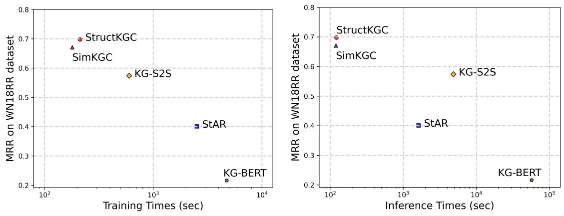


Figure 6: Training and inference time of StructKGC and text-based counterparts on WN18RR.

representations by efficiently utilizing the structure information. In particular, we propose a novel structure-aware supervised contrastive learning and design four CL tasks specifically designed for KGs. By jointly training these tasks, our StructKGC can sufficiently perceive diverse structural knowledge. Experiments show that our method achieves overall state-of-the-art performance better than other baselines in the link prediction task on benchmark datasets. The primary focus of this study is to leverage structural information for mining weak positive samples. In the future, we are interested in incorporating the negative sampling strategies, especially in hard negative mining, to further improve the discriminative ability in contrastive learning.

7 Limitations

Due to the additional entity-path textual pair encoding, we acknowledge that our StructKGC incurs higher training costs than SimKGC, as shown in Fig.6. Specifically, our proposed method requires approximately 1.2 times the iteration time of SimKGC. However, considering the significant gains achieved, this cost is deemed acceptable. Furthermore, by eliminating the encoding of paths, our method can achieve the same training time while yielding better results due to its enhanced data efficiency. Moreover, during the inference phase, our method demonstrates superior efficiency compared to most text-based methods. Although we are not the first to achieve fast inference—models like StAR (Wang et al., 2021) and SimKGC (Wang et al., 2022) already offer similar benefits—we want to underscore the indispensable role that fast inference plays in driving the advancement of new model developments.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Ivana Balazevic, Carl Allen, and Timothy Hospedales. 2019. TuckER: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5185–5194, Hong Kong, China. Association for Computational Linguistics.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Chen Chen, Yufei Wang, Bing Li, and Kwok-Yan Lam. 2022a. Knowledge is flat: A Seq2Seq generative framework for various knowledge graph completion. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4005–4017, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR.

Xiang Chen, Ningyu Zhang, Lei Li, Shumin Deng, Chuanqi Tan, Changliang Xu, Fei Huang, Luo Si, and Huajun Chen. 2022b. Hybrid transformer with multi-level fusion for multimodal knowledge graph completion. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 904–915.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Xiou Ge, Yun Cheng Wang, Bin Wang, and C.-C. Jay Kuo. 2023. Compounding geometric operations for knowledge graph completion. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages

6947–6965, Toronto, Canada. Association for Computational Linguistics.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 687–696, Beijing, China. Association for Computational Linguistics.

Fan Jiang, Tom Drummond, and Trevor Cohn. 2023. Don't mess with mister-in-between: Improved negative search for knowledge graph completion. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1818–1832, Dubrovnik, Croatia. Association for Computational Linguistics.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 18661–18673. Curran Associates, Inc.

Bosung Kim, Taesuk Hong, Youngjoong Ko, and Jungyun Seo. 2020. Multi-task learning for knowledge graph completion with pre-trained language models. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1737–1743, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Ni Lao and William W Cohen. 2010. Relational retrieval using a combination of path-constrained random walks. *Machine learning*, 81:53–67.

Da Li, Boqing Zhu, Sen Yang, Kele Xu, Ming Yi, Yukai He, and Huaimin Wang. 2023a. Multi-task pre-training language model for semantic network completion. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 22(11).

Rui Li, Xu Chen, Chaozhuo Li, Yanming Shen, Jianan Zhao, Yujing Wang, Weihao Han, Hao Sun, Weiwei Deng, Qi Zhang, and Xing Xie. 2023b. To copy rather than memorize: A vertical learning paradigm for knowledge graph completion. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6335–6347, Toronto, Canada. Association for Computational Linguistics.

Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. 2015a. Modeling relation paths for representation learning of knowledge bases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 705–714, Lisbon, Portugal. Association for Computational Linguistics.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015b. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Zile Qiao, Wei Ye, Dingyao Yu, Tong Mo, Weiping Li, and Shikun Zhang. 2023. Improving knowledge graph completion with generative hard negative mining. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5866–5878, Toronto, Canada. Association for Computational Linguistics.

Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The Semantic Web*, pages 593–607, Cham. Springer International Publishing.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Siddharth Suresh, Kushin Mukherjee, Xizheng Yu, Wei-Chun Huang, Lisa Padua, and Timothy Rogers. 2023. Conceptual structure coheres in human cognition but not in large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 722–738, Singapore. Association for Computational Linguistics.

Zhaoxuan Tan, Zilong Chen, Shangbin Feng, Qingyue Zhang, Qinghua Zheng, Jundong Li, and Minnan Luo. 2023. Kracl: Contrastive learning with graph context modeling for sparse knowledge graph completion. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 2548–2559, New York, NY, USA. Association for Computing Machinery.

Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 57–66, Beijing, China. Association for Computational Linguistics.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2071–2080, New York, New York, USA. PMLR.

Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha Talukdar. 2020. Composition-based multi-relational graph convolutional networks. In *International Conference on Learning Representations*.

Bo Wang, Tao Shen, Guodong Long, Tianyi Zhou, Ying Wang, and Yi Chang. 2021. Structure-augmented text representation learning for efficient knowledge graph completion. In *Proceedings of the Web Conference 2021*, WWW '21, page 1737–1748, New York, NY, USA. Association for Computing Machinery.

Liang Wang, Wei Zhao, Zhuoyu Wei, and Jingming Liu. 2022. SimKGC: Simple contrastive knowledge graph completion with pre-trained language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4281–4294, Dublin, Ireland. Association for Computational Linguistics.

Xintao Wang, Qianyu He, Jiaqing Liang, and Yanghua Xiao. 2023. Language models as knowledge embeddings. *Preprint*, arXiv:2206.12617.

Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Kg-bert: Bert for knowledge graph completion. *arXiv preprint arXiv:1909.03193*.

Yichi Zhang, Zhuo Chen, Wen Zhang, and Huajun Chen. 2023. Making large language models perform better in knowledge graph completion. *arXiv preprint arXiv:2310.06671*.

Yuqi Zhu, Xiaohan Wang, Jing Chen, Shuofei Qiao, Yixin Ou, Yunzhi Yao, Shumin Deng, Huajun Chen, and Ningyu Zhang. 2024. Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities. *Preprint*, arXiv:2305.13168.

A Hyperparameters

The maximum number of tokens in the description sequence is limited to 50. We conduct a grid search to identify the optimal learning rate within the range of $\{5e^{-4}, 5e^{-5}, 1e^{e}{-}5\}$. The initial temperature τ is set to 0.05. For the coefficients w_i, we use grid search to tune with a search range of $\{0.2, 0.4, 0.6, 0.8, 1\}$. Training is carried out using the AdamW optimizer with linear learning rate decay, and the models are trained with a batch size of 1024. The training epochs are set to 100 for the WN18RR dataset and 10 for the FB15k-237 dataset. Table 6 provides a summary of the training hyperparameters.

Hyperparameters	**FB15k-237**	**WN18RR**
Learning rate	1e-5	5e-5
LR Scheduler	Linear Warmup	Linear Warmup
Warmup steps	400	400
initial temperature	0.05	0.05
Epochs	10	100
Batch size	1024	1024
Gradient clipping	10.0	10.0
max # of tokens	50	50
Combination weights	[1, 0.2, 1, 1]	[1, 1, 0.6, 0.4]

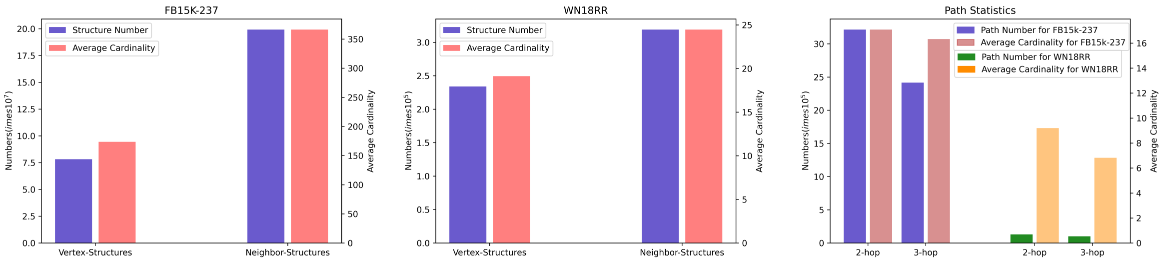
Table 6: Hyperparameters for our proposed StructKGC model.

B Structure Statistical Analysis

In order to determine the universality of the various structures discussed in this paper within KGs, we analyze the numbers and proportions of different types of structures across various static benchmark datasets. Our findings, as shown in Fig.7, indicate that there are high numbers and proportions of these structures across different datasets, which could be potentially weakly positive indicators of the feasibility of our method. We also find that FB15k-237 has a higher number of structures, particularly in terms of paths, when compared to WN18RR. This suggests that FB15k-237 provides more structural signals for PLMs during training, which could explain why StructKGC shows more improvement over the text-based model on FB15k-237 datasets.

C Case Study

We conducted a comprehensive case study examining a diverse range of structural strengths. The detailed results, covering various scenarios, are presented in Tables 7, 8, and 9. The model consistently assigns high similarity scores to both the golden answers and their surrounding contexts, revealing the semantic connections between them. Furthermore, the relation composition case studies highlight the effectiveness of our approach in uncovering underlying logical rules. These analyses demonstrate how the model effectively leverages contextual cues to generate accurate explanations and predictions.



(a) (b) (c)

Figure 7: Statistics of various structures of FB15k-237 and WN18RR training set.

	Entity	Relation	Golden Answer : Score	Relevant Contextual Entities in Training Set : Score
FB15k-237	Pontefract	contains/location^{-1}	England: 0.59	West Yorkshire: 0.64, United Kingdom: 0.54
	Thomas Lennon	profession	Screenwriter: 0.65	Television producer: 0.64, Film Producer: 0.58
	Garland	contains/location^{-1}	Texas: 0.72	Dallas County: 0.71, Collin County: 0.65
	Superhero movie	genre/film^{-1}	Superman II: 0.44	Spider-Man: 0.56, Green Hornet: 0.44
WN18RR	iris_family	member_meronym	spartium_NN_1: 0.64	genus_belamcanda: 0.72, ixia: 0.72, sisyrinchium: 0.64
	subfamily_papilionoideae	member_meronym	spartium_NN_1: 0.64	templetonia_NN_1: 0.70, lablab_NN_1: 0.70
	polish_NN_1	derivationally_related_form	furbish_VB_1: 0.70	gloss_VB_1: 0.75, smoothen_VB_2: 0.68
	africa_NN_1	has_part	senegal_NN_1: 0.58	republic_of_kenya: 0.679, republic_of_guinea: 0.68

Table 7: Case study of entity prediction. The notation $^{-1}$ indicates the inverse operation of a relation.

	Entity	Golden Neighbors : Score	Relevant Contextual Neighbors in Training Set : Score
FB15k-237	Paul Robeson	(nationality, America): 0.50	(ethnicity, African American): 0.52, (graduate, Columbia University): 0.37
	The Ghost Writer	(genre/film, Mystery): 0.44	(netflix/genre, Political thriller): 0.46, (genre/film, Crime Fiction): 0.45
	Jason Flemyng	(profession, Actor): 0.46	(award, Screen Actors Guild Award): 0.41, (perform, The Red Violin): 0.26
WN18RR	mindfulness_NN_1	(hypernym^{-1} , attentiveness): 0.53	(derivationally_related_form, thoughtful_JJ_4): 0.60
	wheeled_vehicle_NN_1	(hypernym, scooter_NN_3): 0.59	(has_part^{-1}, axle_NN_1): 0.46, (hypernym, self-propelled_vehicle): 0.56

Table 8: Case study of neighbor prediction.

	Direct Relation	Relevant Multi-hop Paths in Training Set	Confidence
FB15k-237	award_nominations./award/award_nomination	film/actor $\wedge$ award/award_category/nominated_for^{-1}	0.099
	music/genre/artists	music/genre/parent_genre $\wedge$ music/genre/artists	0.093
	people/person/place_of_birth	people/place_lived/location $\wedge$ location/hud_county_place	0.069
WN18RR	derivationally_related_form	derivationally_related_form $\wedge$ similar_to $\wedge$ similar_to^{-1}	0.071
	has_part	has_part $\wedge$ has_part	0.06
	also_see	also_see $\wedge$ also_see	0.059

Table 9: Case study of relation composition.