DUAL-RP: Dual-Source Structural and Textual Re-ranking for Relation Prediction

Siyan Wu[†1], Chenhua Zhu[†1], Jieyu Zhan[*1], Lihua Cai[*1,2]
[1]*South China Normal University*
Guangzhou, China
{siyan.wu, chenghua.zhu, lihua.cai}@m.scnu.edu.cn
zhanjieyu@scnu.edu.cn
[2]*Xiamen Rekey Medical Technology Co., LTD*
Xiamen, China

***Abstract*—Knowledge graph completion (KGC) is a fundamental task in database and AI systems, aiming to infer missing facts in knowledge graphs and thereby improve the completeness and usability of structured knowledge. However, existing methods often struggle to jointly capture structural reasoning and semantic precision within a unified framework. We present DUAL-RP, a relation prediction framework that bridges knowledge graph embedding (KGE) models and large language models (LLMs) through a semantic re-ranking paradigm. DUAL-RP first employs lightweight KGE models to generate a compact set of candidate relations for a given entity pair. A token translator module then aligns continuous structural embeddings with the token embedding space of LLMs, enabling effective interaction between structural priors and semantic reasoning. In addition, multi-hop relational paths are incorporated as natural-language context to provide complementary semantic cues for relation ranking and interpretability. Experiments on FB15k-237, CoDEx-S, and DBpedia50 show that DUAL-RP consistently delivers strong overall ranking performance across diverse baselines, achieving the most consistent improvements in Mean Reciprocal Rank and Hits@1.**

***Index Terms*—Knowledge Graph Completion, Relation Prediction, Knowledge Graph Embedding, Large Language Models, Semantic–Structural Fusion**

I. Introduction

Knowledge Graphs (KGs) provide structured representations of entities and relations and support a wide range of applications, such as search [1], recommendation [2], and question answering [3]. However, due to the high cost of manual curation and the open-world nature of knowledge acquisition, real-world KGs are often incomplete. Knowledge Graph Completion (KGC) addresses this issue by inferring missing facts, where relation prediction aims to identify the most plausible relation between a given entity pair [4].

Existing approaches to relation prediction can be broadly categorized into three paradigms:

Knowledge graph embedding (KGE) methods (Fig. 1(a)) encode entities and relations into continuous vector spaces and evaluate triple plausibility using geometric or algebraic scoring functions [4], [5]. While effective at capturing structural regularities, they mainly rely on graph topology and weakly incorporate semantic constraints, limiting their ability to model long-range dependencies and compositional patterns.

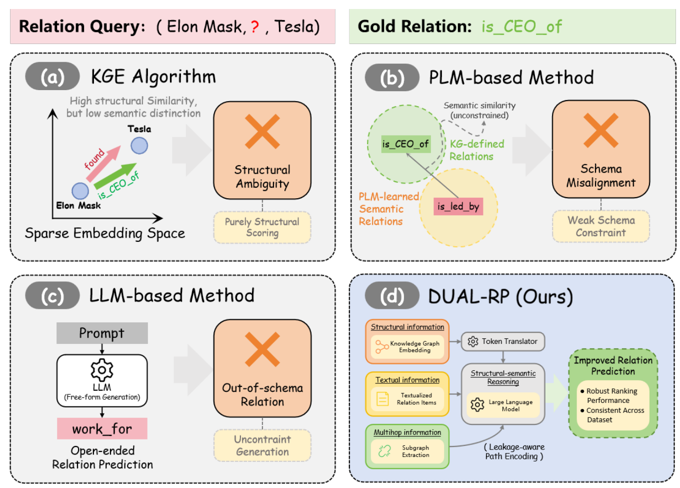

Fig. 1: Comparison of relation prediction paradigms: (a) KGE-based, (b) PLM-based, (c) LLM-based, and (d) the proposed **DUAL-RP** framework.

Pre-trained language model (PLM)-based methods (Fig. 1(b)) introduce textual semantics by encoding entity or relation descriptions and aligning language with knowledge representations [6]. Although they improve semantic expressiveness, structural information is typically incorporated through loose coupling, making it difficult to enforce precise relational constraints in complex graphs.

Large language model (LLM)-based methods (Fig. 1(c)) formulate relation prediction as a generative or prompting-based task. These methods leverage strong contextual reasoning and parametric knowledge but often suffer from limited controllability and may deviate from KG structure under open-ended generation [7].

Despite their strengths, these paradigms share several limitations. First, *structural–semantic disconnection* arises from separate representation spaces. Second, *limited controllability* affects the reliability of generative reasoning. Third, *multi-hop reasoning* remains insufficient for capturing complex relational dependencies.

[†]These two authors contributed equally to this work.
[*]Corresponding authors.

To address these challenges, we propose **DUAL-RP**, a unified framework that integrates structural representations from KGE with the semantic reasoning capability of LLMs (Fig. 1(d)). DUAL-RP formulates relation prediction as a structure-guided semantic re-ranking problem, combining candidate filtering with context-aware ranking to achieve accurate and controllable inference.

Our main contributions are summarized as follows:

- We propose **DUAL-RP**, a unified framework that bridges structural embeddings and language models for relation prediction.
- We design a token translator, a two-stage filtering and re-ranking pipeline, and a multi-hop semantic augmentation strategy.
- We conduct experiments on FB15k-237, CoDEx-S, and DBpedia50, demonstrating consistent improvements over representative baselines.

II. Related Work

A. KGE Methods

KGE methods learn low-dimensional representations for entities and relations, and evaluate triple plausibility via scoring functions. Representative models include translation-based methods such as TransE [5], tensor factorization methods such as DistMult [8] and ComplEx [9], and geometry-based methods such as RotatE [4]. These methods are efficient and structurally consistent, but mainly rely on graph topology and often struggle to capture implicit semantics and long-range compositional reasoning.

B. PLM-based Methods

PLM-based methods enhance KG representations by incorporating textual semantics. Representative approaches include KG-BERT [10], BERTRL [11], and LASS [12], which encode entity or relation descriptions and strengthen language–structure interaction through pretraining or joint modeling. Although these methods improve semantic expressiveness, they still struggle to fully exploit structural constraints and often suffer from misalignment between textual and relational representations.

C. LLM-based Methods

LLM-based methods enable relation prediction through prompting, in-context learning, or task-specific adaptation. Representative approaches include chain-of-thought prompting [13], KG-LLM [14], InstructGLM [15], and StructGPT [16]. More recent work has investigated tighter integration between knowledge graphs and LLMs across related tasks. In particular, BrikQA [3] uses structural and textual signals to support multi-hop question answering over knowledge graphs, whereas ST-KGLP [17] aligns structural and textual knowledge for LLM-based link prediction. Despite their strong contextual reasoning ability, these methods are often sensitive to prompt design and may not impose sufficient structural constraints, which can lead to unstable behavior on complex graphs [7].

To improve robustness and parameter efficiency, recent studies further introduce lightweight adaptation and structure-aware tuning, such as LoRA [18]. Overall, existing paradigms still reveal a gap between symbolic structure and semantic reasoning, motivating unified and structure-aware integration of embedding representations with language models.

III. Problem Definition

A KG is defined as a set of triples $G = (E, R, T)$, where E and R denote the sets of entities and relations, and $T \subseteq E \times R \times E$ denotes the set of observed facts. Given an entity pair (h, t), the goal of relation prediction is to identify the most plausible relation $r^* \in R$ between them.

In addition to direct connections, entities in a KG may also be linked through multi-hop relational paths. A path $p(h, t) = (r_1, \ldots, r_k)$ corresponds to a relational chain $h \xrightarrow{r_1} e_1 \xrightarrow{r_2} \cdots \xrightarrow{r_k} t$, which provides higher-order semantic evidence beyond direct relations.

In DUAL-RP, relation prediction is formulated as a candidate re-ranking task. Given an entity pair (h, t), an upstream structural retriever first produces a candidate relation set

$$R_{\text{cand}}(h, t) = \{r_1, r_2, \ldots, r_K\}, \quad R_{\text{cand}}(h, t) \subseteq R.$$

The model then reorders these candidates based on structural and semantic context, including optional multi-hop relational evidence, and outputs a ranked sequence

$$R_{\text{pred}}(h, t) = (\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_K),$$

where each $\hat{r}_i \in R_{\text{cand}}(h, t)$. The top-ranked relation is taken as the final prediction:

$$r^* = \hat{r}_1. \tag{1}$$

The main challenge is to effectively integrate structural and semantic context for candidate re-ranking, while maintaining alignment between KG-derived representations and language-based reasoning.

IV. DUAL-RP

DUAL-RP is a unified framework for relation prediction that integrates structural representations from KGEs with the semantic reasoning capability of LLMs. The overall workflow proceeds in a coarse-to-fine manner. Given a query entity pair, DUAL-RP first performs relation candidate filtering using KGE-based structural scoring. Subsequently, structure-aware semantic reasoning is conducted over the filtered candidates, including structural encoding, multi-hop semantic expansion, and semantic–structural ranking. Fig. 2 provides an overview of the DUAL-RP framework.

A. Relation Candidate Filtering

As shown in Fig. 2(a), performing inference over the entire relation set R is impractical for large-scale knowledge graphs. DUAL-RP therefore constructs a structural prior using KGE-based scoring [5] to narrow down the candidate space at an early stage.

Given a query entity pair (h, t), we denote their corresponding structural embeddings as $\mathbf{e}_h, \mathbf{e}_t \in \mathbb{R}^{d_e}$, and the embedding of a relation $r \in R$ as $\mathbf{e}_r \in \mathbb{R}^{d_r}$. A KGE scoring function

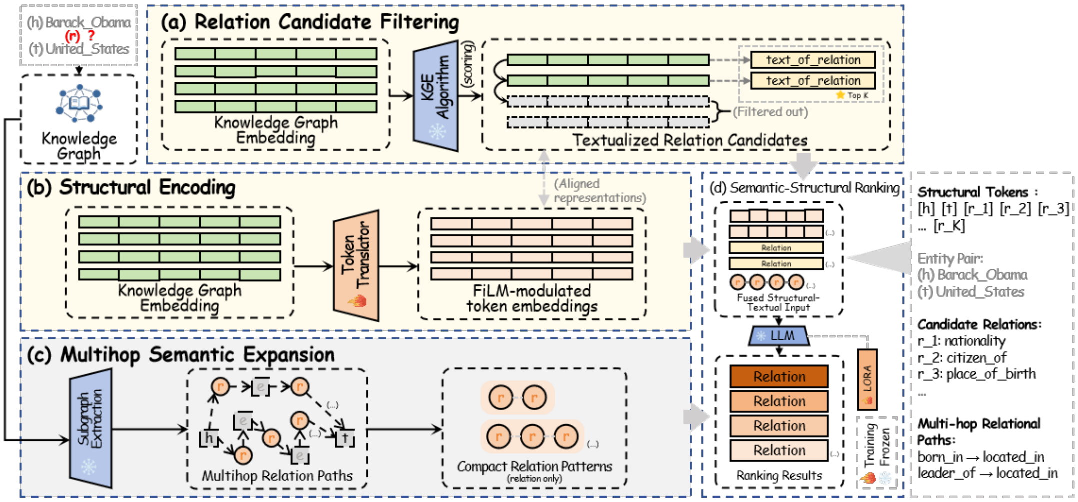

Fig. 2: Overview of the **DUAL-RP** framework.

$g(\cdot)$ is used to evaluate the plausibility of the triple (h, r, t), yielding a structural score $S(h, r, t) = g(\mathbf{e}_h, \mathbf{e}_r, \mathbf{e}_t)$.

Relations are ranked according to their scores $S(h, r, t)$, and the top-K relations are selected as candidates:

$$R_{\text{cand}} = \text{TopK}_{r \in R}\left(S(h, r, t), K\right). \tag{2}$$

Here, R_{cand} denotes the filtered set of top-K candidate relations for the entity pair (h, t). This candidate filtering step produces a compact relation set that serves as the input for subsequent structural encoding and semantic reasoning.

B. Structural Encoding

As illustrated in Fig. 2(b), the structural encoding module operates on the filtered candidate relations and maps entities and relations into continuous vector spaces to provide structural priors for downstream reasoning. Specifically, a knowledge graph embedding (KGE) model such as RotatE [4] encodes the head entity h, tail entity t, and each candidate relation $r_i \in R_{\text{cand}}$ into structural embeddings.

To bridge these structural embeddings with the language model space, we introduce a token translator implemented with a FiLM-style linear modulation module [19]. Given (h, t) and the candidate relation set R_{cand}, we form a structural sequence

$$(\Phi_E(h), \Phi_E(t), \Phi_R(r_1), \Phi_R(r_2), \ldots, \Phi_R(r_K)),$$

which is transformed into modulation parameters and applied to the corresponding role-specific token embeddings:

$$\tilde{\mathbf{x}}_g = \boldsymbol{\gamma}_g \odot \mathbf{x}_g + \boldsymbol{\beta}_g, \tag{3}$$

where $\mathbf{x}_g$ is the base token embedding, and $\boldsymbol{\gamma}_g, \boldsymbol{\beta}_g$ are FiLM-style scaling and shifting terms derived from the structural representations.

The resulting structural tokens are prepended to the textual prompt at the embedding level, injecting structural information without modifying the prompt template or vocabulary.

C. Multi-hop Semantic Expansion

As illustrated in Fig. 2(c), entity pairs may be connected through multi-hop relational paths. DUAL-RP incorporates such paths as lightweight semantic evidence without explicit path modeling.

Given an entity pair (h, t), we extract valid multi-hop relation sequences $p(h, t) = (r_1, \ldots, r_\ell)$ connecting h to t. Intermediate entities are omitted, and sequences are treated as relation-level patterns.

We adopt a localized subgraph extraction strategy inspired by GraIL [20] to efficiently identify such sequences, without constructing subgraph representations or performing graph neural reasoning.

For efficiency, we restrict $\ell \leq L_{\max}$ and retain at most M sequences per pair. All sequences are treated as symbolic patterns and are neither embedded nor explicitly scored.

Each sequence is mapped to natural language and injected into the prompt as auxiliary context, providing complementary semantic cues for relation ranking.

D. Semantic–Structural Ranking

As shown in Fig. 2(d), DUAL-RP formulates relation prediction as a list-wise candidate re-ranking task once the candidate relations and contextual information are prepared. The input to the language model consists of a fixed task instruction, embedding-level structural tokens, a candidate relation list, and auxiliary multi-hop relational context, forming a unified input representation X.

Given the unified input X, the language model M_θ directly generates a reordered relation list, rather than assigning an explicit scalar score to each candidate independently. That is, DUAL-RP treats ranking as a sequence generation problem, where the output is an ordered permutation of the candidate set R_{cand}:

$$\mathcal{R}_{\text{pred}} = \left(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_{|R_{\text{cand}}|}\right), \tag{4}$$

where $\hat{r}_i \in R_{\text{cand}}$, and candidates appearing earlier in the generated sequence are interpreted as being ranked higher by the model. The top-ranked relation $\hat{r}_1$ is taken as the final prediction.

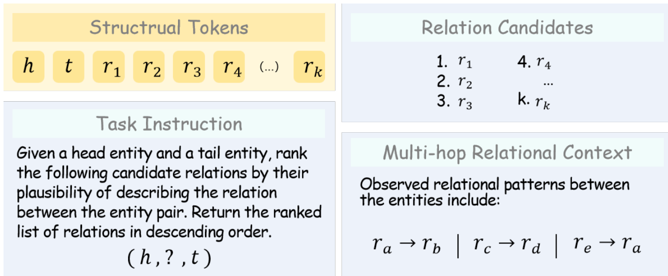

Fig. 3: Illustration of the unified LLM input used for semantic–structural relation ranking.

This formulation avoids the need for manually designed candidate-wise scoring functions and allows the LLM to perform holistic ranking over the full candidate list under the joint semantic–structural context. In this way, structural priors and multi-hop evidence are integrated directly into the generation of the final ranking order.

To ensure stable inference and consistent interaction between structural and semantic signals, DUAL-RP adopts a standardized instruction-style prompt template adapted from the Alpaca format [21]. The template provides a unified and fixed structure for task instructions, candidate relations, and contextual cues, reducing sensitivity to prompt variations and enabling fair and reproducible comparison across inputs.

E. Training and Parameter Optimization

DUAL-RP is trained in a sequence-to-sequence manner. For each training query (h, t), we construct a candidate set of size K by combining the ground-truth relation with $K-1$ negative relations retrieved by the KGE-based retriever. The target ranking sequence is defined by placing the ground-truth relation at the first position, while preserving the relative order of the remaining negative candidates according to their KGE scores:

$$Y = (r^{+}, r_1^{-}, r_2^{-}, \dots, r_{K-1}^{-}).$$

Given the unified input X and target sequence Y, the training objective is the autoregressive negative log-likelihood:

$$\mathcal{L} = -\sum_{j=1}^{K} \log P_\theta(y_j \mid X, y_{<j}), \tag{5}$$

where y_j is the j-th relation in the target sequence.

For parameter-efficient fine-tuning, we adopt Low-Rank Adaptation (LoRA) [18]:

$$W = W_0 + AB^\top, \tag{6}$$

where W_0 is the frozen pretrained weight, and A, B are low-rank trainable matrices.

V. Experiments

A. Experimental Settings

We evaluate DUAL-RP on three public knowledge graph benchmarks: **FB15k-237** [22], **CoDEx-S** [23], and **DBpedia50** [1], which differ in scale, relational diversity, and semantic complexity. Dataset statistics are summarized in TABLE II.

TABLE II: Statistics of the datasets used in our experiments.

Dataset	#Entities	#Relations	#Train	#Valid	#Test
FB15k-237	14,541	237	272,115	17,535	19,496
CoDEx-S	2,034	42	32,888	1,827	1,828
DBpedia50	24,638	351	35,077	4,384	10,969

a) Task setting and evaluation: Relation prediction aims to identify the correct relation for a given entity pair (h, t) from the full relation set R. Performance is measured using MRR and Hits@k ($k = 1, 3, 10$).

For DUAL-RP, inference is implemented in a top-K candidate re-ranking setting. Given (h, t), a KGE-based retriever first selects the top-K candidate relations from R, and the LLM-based re-ranker then predicts the final relation within this candidate set. Although inference is restricted to the retrieved candidates, evaluation is still conducted against the original full relation space. That is, a prediction is counted as correct only when the gold relation is successfully retrieved into the candidate set and ranked correctly by the re-ranker. Therefore, the reported results reflect both retrieval coverage and re-ranking accuracy, rather than candidate-set ranking alone.

Unless otherwise specified, we set $K = 25$ in all experiments. We choose $K = 25$ as a practical trade-off between candidate coverage and re-ranking efficiency. As shown in TABLE III, increasing K beyond 25 yields only marginal coverage gains on FB15k-237 and DBpedia50, while leading to longer candidate lists and higher inference overhead for the LLM re-ranker. On CoDEx-S, larger K values further improve coverage, but $K = 25$ already retains more than 90% of gold relations. Therefore, we use $K = 25$ as the default setting to balance retrieval recall and computational cost.

b) Baselines: We compare DUAL-RP with representative baselines from three paradigms. Embedding-based methods include TransE [5], ComplEx [9], DistMult [8], RotatE [4], HAKE [24], and PairRE [25]. PLM-based methods include KG-BERT [10], BERTRL [11], and LASS [12]. LLM-based baselines include prompt-based methods (IO Prompting [26], CoT [13], and StructGPT [16]), which do not require task-specific training, as well as task-adapted methods (KG-LLM [14] and InstructGLM [15]), which are trained following their original protocols. For task-adapted baselines, the training objective is reformulated for relation prediction while preserving the original learning paradigm.

c) Implementation details: Unless otherwise specified, Qwen2.5-3B is used as the backbone language model for all LLM-based experiments. For DUAL-RP and other task-adapted LLM baselines, parameter-efficient fine-tuning with LoRA is applied. For semantic expansion, we extract relation paths with a maximum length of $L_{\max} = 3$ and retain up to $M = 5$ paths per entity pair. These paths are converted into natural-language descriptions and used as additional context. All experiments are implemented with PyTorch and Transformers and conducted on two NVIDIA RTX 4090 GPUs.

TABLE I: Main experiment results (MRR and Hits@k).

Method	FB15k-237				CoDEx-S				DBpedia50			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
KGE Methods												
TransE	0.3803	0.2041	0.3487	0.5068	0.3460	0.2200	0.4016	0.5940	0.0838	0.0714	0.0876	0.1063
ComplEx	0.2711	0.1865	0.2959	0.4444	0.2470	0.1506	0.3017	0.4640	0.0509	0.0491	0.0653	0.0827
DistMult	0.2593	0.1815	0.2289	0.4148	0.3374	0.2180	0.3912	0.5693	0.0507	0.0453	0.0530	0.0624
RotatE	0.3241	0.2314	0.3573	0.5903	0.4500	0.3410	0.5063	0.6530	0.1062	0.0511	0.1430	0.1958
HAKE	0.3495	0.2623	0.3821	0.6085	0.4875	0.3783	0.5560	0.6840	0.1345	0.0640	0.1682	0.2215
PairRE	0.3538	0.2685	0.3880	0.6150	0.4750	0.3620	0.5410	0.6710	0.1290	0.0619	0.1620	0.2161
PLM Methods												
KG-BERT	0.3952	0.3257	0.4150	0.5201	0.5125	0.3350	0.6354	**0.8850**	0.1857	0.1159	0.2350	**0.3807**
BERTRL	0.4321	0.3456	0.4558	**0.6802**	0.6259	0.5500	0.7204	0.8500	0.1753	0.0950	0.1751	0.3658
LASS	0.4158	0.3305	0.4302	0.6401	0.5810	0.4802	0.6852	0.8228	0.1800	0.1052	0.2054	0.3752
LLM Methods												
IO Prompting	0.4952	0.4545	0.5430	0.5730	0.6533	0.5762	0.6549	0.7564	0.2150	0.1850	0.2252	0.2459
CoT	0.5051	0.4653	0.5556	0.5850	0.6751	0.5950	0.6757	0.7850	0.2180	0.1900	0.2300	0.2552
StructGPT	0.5183	0.4781	0.5702	0.6024	0.6951	0.6250	0.7150	0.8101	0.2250	0.1989	0.2491	0.2683
KG-LLM	0.5120	0.4722	0.5651	0.5953	0.6880	0.6150	0.6953	0.8053	0.2223	0.1950	0.2354	0.2620
InstructGLM	0.5258	0.4850	0.5750	0.6050	0.7055	0.6354	0.7251	0.8150	0.2280	0.2021	0.2452	0.2701
Ours (DUAL-RP)	**0.5871**	**0.5550**	**0.6170**	0.6330	**0.7627**	**0.7216**	**0.8020**	0.8239	**0.2392**	**0.2141**	**0.2584**	0.2750

TABLE III: Coverage (%) of gold relations within top-k candidates.

Dataset	20	25	30	35	40
FB15k-237	85.9	87.0	88.0	88.9	90.0
CoDEx-S	87.8	90.2	92.8	95.6	98.7
DBpedia50	32.3	33.0	33.5	34.1	34.5

B. Main Results

TABLE I summarizes the performance of DUAL-RP and all baselines on three benchmark datasets. Overall, DUAL-RP achieves the best average performance across datasets, consistently ranking first in terms of MRR and Hits@1, while remaining competitive on other metrics. This demonstrates the effectiveness of integrating structural embeddings with semantic reasoning for relation prediction.

On FB15k-237, embedding-based methods perform reasonably well due to the dominance of structural regularities. However, DUAL-RP still yields consistent improvements by incorporating semantic signals, leading to higher ranking accuracy. PLM-based approaches benefit from textual representations but show less stable gains across datasets.

On CoDEx-S and DBpedia50, where semantic information plays a more important role, LLM-based methods significantly outperform traditional KGE and PLM baselines. Among them, instruction-tuned and structure-aware models such as KG-LLM, InstructGLM, and StructGPT achieve strong results, indicating the advantage of leveraging pretrained language models for semantic alignment.

Compared with these methods, DUAL-RP further improves overall ranking performance by explicitly fusing structural priors with semantic reasoning. While some LLM-based baselines achieve competitive results on individual metrics, DUAL-RP consistently delivers the highest mean ranking performance across datasets, highlighting its robustness and effectiveness under diverse relational and semantic conditions.

To support reproducibility, we provide the code, detailed experimental results, and supplementary case analyses at: https://github.com/TheBlueBanisters/DUAL-RP.

C. Ablation Study

To assess the contribution of each component, we conduct ablation studies under the same settings as the main experiments, with results summarized in TABLE IV. We evaluate four variants by removing structural information, the Token Translator, multi-hop relational information, and supervised fine-tuning (SFT), respectively.

Removing any component consistently degrades performance across all datasets, indicating that each module contributes to the overall effectiveness of DUAL-RP. In particular, excluding structural information or supervised fine-tuning leads to substantial performance drops, highlighting the critical role of structural priors and task-specific adaptation. Removing the Token Translator and multi-hop relational information results in moderate but consistent degradation, suggesting their complementary roles in aligning structural and semantic representations and enriching contextual reasoning.

VI. Conclusion

This paper presents **DUAL-RP**, a unified framework that bridges structured knowledge modeling and semantic reasoning for relation prediction. DUAL-RP leverages lightweight KGE models to efficiently filter candidate relations and introduces a token translator to align continuous embeddings with the input space of LLMs, while incorporating multi-hop path information to provide additional contextual support for relation reasoning. Experiments on multiple public benchmarks show that DUAL-RP achieves strong overall performance relative to embedding-based, PLM-based, and LLM-based baselines, with particularly consistent gains in MRR

TABLE IV: Ablation study results (MRR and Hits@k).

Model Variant	FB15k-237				CoDEx-S				DBpedia50			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
w/o Structural Info	0.5123	0.4631	0.5412	0.5678	0.6765	0.6279	0.7142	0.7483	0.2083	0.1764	0.2260	0.2457
w/o Token Translator	0.5284	0.4860	0.5541	0.5789	0.6920	0.6443	0.7306	0.7615	0.2135	0.1821	0.2324	0.2538
w/o Multi-hop Info	0.5570	0.5221	0.5843	0.6071	0.7312	0.6900	0.7735	0.7984	0.2281	0.2029	0.2473	0.2661
w/o SFT	0.4692	0.4215	0.4958	0.5237	0.6104	0.5610	0.6513	0.6891	0.1914	0.1622	0.2103	0.2295
Full Model	**0.5871**	**0.5550**	**0.6170**	**0.6330**	**0.7627**	**0.7216**	**0.8020**	**0.8239**	**0.2392**	**0.2141**	**0.2584**	**0.2750**

and Hits@1. These results demonstrate the effectiveness of integrating structural priors with semantic reasoning. Despite its strong performance, DUAL-RP still faces limitations in inference efficiency and adaptability on extremely large-scale or multilingual KGs. Future work will focus on improving efficiency and deployability, and extending the framework to broader downstream tasks.

ACKNOWLEDGEMENTS

This work was supported by the Student Extracurricular Research Program of South China Normal University. The authors also disclose the use of an AI-based language model (ChatGPT [27], OpenAI) for language editing and clarity improvement.

REFERENCES

[1] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A nucleus for a web of open data," in *Proceedings of the 6th International Semantic Web Conference (ISWC)*, 2007, pp. 722–735.

[2] H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, and M. Guo, "RippleNet: Propagating user preferences on the knowledge graph for recommender systems," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, 2018, pp. 172–181.

[3] S. Luo, X. Lu, Q. Zhao, and W. Rao, "Bridging the gap between knowledge graphs and LLMs for multi-hop question answering," in *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM)*, 2025, pp. 5006–5010.

[4] Z. Sun, Z. Deng, J. Nie, and J. Tang, "RotatE: Knowledge graph embedding by relational rotation in complex space," in *International Conference on Learning Representations (ICLR)*, 2019.

[5] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2013, pp. 2787–2795.

[6] X. Wang, T. Gao, Z. Zhu, Z. Liu, J. Li, and J. Tang, "KEPLER: A unified model for knowledge embedding and pre-trained language representation," *Transactions of the Association for Computational Linguistics (TACL)*, vol. 9, pp. 176–194, 2021.

[7] J. Maynez, S. Narayan, B. Bohnet, and R. McDonald, "On faithfulness and factuality in abstractive summarization," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020, pp. 1906–1919.

[8] B. Yang, W. tau Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," in *International Conference on Learning Representations (ICLR)*, 2015.

[9] T. Trouillon, J. Welbl, S. Riedel, E. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 2016, pp. 2071–2080.

[10] L. Yao, C. Mao, and Y. Luo, "KG-BERT: BERT for knowledge graph completion," *arXiv preprint arXiv:1909.03193*, 2019.

[11] H. Zha, Z. Chen, and X. Yan, "Inductive relation prediction by BERT," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2022, pp. 5923–5931.

[12] J. Shen, C. Wang, L. Gong, and D. Song, "LASS: Joint language semantic and structure embedding for knowledge graph completion," in *Proceedings of the 29th International Conference on Computational Linguistics (COLING)*, 2022, pp. 1976–1988.

[13] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou, "Chain-of-Thought prompting elicits reasoning in large language models," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, 2022.

[14] L. Yao, J. Peng, C. Mao, and Y. Luo, "Exploring large language models for knowledge graph completion," *arXiv preprint arXiv:2308.13916*, 2023.

[15] R. Ye, C. Zhang, R. Wang, S. Xu, and Y. Zhang, "Language is all a graph needs," in *Findings of the Association for Computational Linguistics: EACL 2024*, 2024, pp. 1955–1973.

[16] J. Jiang, K. Zhou, Z. Dong, K. Ye, X. Zhao, and J.-R. Wen, "StructGPT: A general framework for large language model to reason over structured data," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023, pp. 9237–9251.

[17] S. Luo, X. Lu, Q. Zhao, and W. Rao, "Combining structural and textual knowledge for knowledge graph link prediction via large language models," in *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM)*. ACM, 02 2026, pp. 479–488.

[18] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in *International Conference on Learning Representations (ICLR)*, 2022.

[19] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville, "FiLM: Visual reasoning with a general conditioning layer," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[20] K. K. Teru, E. Denis, and W. L. Hamilton, "Inductive relation prediction by subgraph reasoning," in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 2020, pp. 9448–9457.

[21] R. Taori, I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, and T. Hashimoto, "Stanford alpaca: An instruction-following LLaMA model," 2023, GitHub repository.

[22] K. Toutanova, D. Chen, P. Pantel, H. Poon, P. Choudhury, and M. Gamon, "Representing text for joint embedding of text and knowledge bases," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015, pp. 1499–1509.

[23] T. Safavi and D. Koutra, "CoDEx: A comprehensive knowledge graph completion benchmark," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 8328–8350.

[24] Z. Zhang, J. Cai, Y. Zhang, and J. Wang, "Learning hierarchy-aware knowledge graph embeddings for link prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 34, no. 3, 2020, pp. 3065–3072.

[25] L. Chao, J. He, T. Wang, and W. Chu, "PairRE: Knowledge graph embeddings via paired relation vectors," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2021, pp. 4360–4369.

[26] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 1877–1901.

[27] OpenAI, "ChatGPT," 2024, https://chatgpt.com/, accessed Jan. 2026.