

# GRAPHORACLE: Efficient Fully-Inductive Knowledge Graph Reasoning via Relation-Dependency Graphs

Enjun Du<sup>1,2</sup>, Siyi Liu<sup>1</sup>, Yongqi Zhang<sup>1\*</sup>

<sup>1</sup>The Hong Kong University of Science and Technology (Guangzhou)

<sup>2</sup>Beijing Institute of Technology

{EnjunDu.cs, ssui.liu1022}@gmail.com, yongqizhang@hkust-gz.edu.cn

## Abstract

Knowledge graph reasoning in the fully-inductive setting—where both entities and relations at test time are unseen during training—remains an open challenge. In this work, we introduce GRAPHORACLE, a novel framework that achieves robust fully-inductive reasoning by transforming each knowledge graph into a Relation-Dependency Graph (RDG). The RDG encodes directed precedence links between relations, capturing essential compositional patterns while drastically reducing graph density. Conditioned on a query relation, a multi-head attention mechanism propagates information over the RDG to produce context-aware relation embeddings. These embeddings then guide a second GNN to perform inductive message passing over the original knowledge graph, enabling prediction on entirely new entities and relations. Comprehensive experiments on 60 benchmarks demonstrate that GRAPHORACLE outperforms prior methods by up to 25% in fully-inductive and 28% in cross-domain scenarios. Our analysis further confirms that the compact RDG structure and attention-based propagation are key to efficient and accurate generalization.

**Code** — <https://github.com/EnjunDu/GraphOracle>

**Extended version** — <https://arxiv.org/pdf/2505.11125>

## 1 Introduction

Knowledge graphs (KGs) encode structured knowledge as entity–relation–entity triples, serving as the backbone for scientific discovery, web-scale reasoning, and intelligent systems (Bai et al. 2025; Shen et al. 2024; Zhang, Zhang, and Yuan 2024; Yang et al. 2025; Zhang et al. 2019b, 2020; Yu et al. 2025; Wang et al. 2025b,c,a; Zhang et al. 2023a, 2025a). The central challenge in KG reasoning is *link prediction*: given an incomplete graph and a query  $(h, r, ?)$ , predict the missing tail entity  $t$  (Du, Liu, and Zhang 2025). Inductive KG reasoning requires models to generalize to facts that were not explicitly observed during training. In the most challenging scenarios, models must perform **fully-inductive** (Zhang et al. 2025b; Cui, Sun, and Hu 2025) reasoning—handling both unseen entities and relations during inference. **Cross-domain** generalization (Wang et al. 2025d;

Hong, Lee, and Whang 2025; Lin, Wang, and Tang 2025) further requires reasoning over entirely different knowledge graphs containing 100% novel entities and relations. Both settings pose a fundamental *compositional generalization* challenge: models must recombine learned relational patterns to infer new facts in completely unfamiliar contexts.

Recent approaches to fully-inductive reasoning, including INGRAM (Lee, Chung, and Whang 2023) and ULTRA (Galkin et al. 2024), attempt to address this challenge by constructing auxiliary relation graphs  $G_R = (R, E_R)$  that capture transferable relational patterns. However, as KGs scale, this paradigm exposes two fundamental limitations that severely hinder their effectiveness in real-world applications. (1) These methods rely on co-occurrence statistics to connect relations, creating dense graphs with  $|E_R| = \Theta(|R|^2)$  edges that inflate computational costs to  $O(|R|^3 \cdot L)$  for  $L$ -layer message passing. The proliferation of spurious connections obscures meaningful compositional signals, while symmetric treatment of relation pairs erases the inherent directionality of logical composition. (2) Current models compress each relation into a single fixed embedding, forcing individual representations to capture diverse semantic roles across vastly different query contexts. For instance, the relation “associated with” may connect proteins to diseases in biomedical contexts but link authors to topics in academic graphs, yet existing methods use the same representation regardless of context.

These observations raise a critical research question: *How to design a framework that captures meaningful relation dependencies while maintaining computational efficiency?* This requires moving beyond dense co-occurrence graphs to a sparse, directed structure that preserves compositional patterns, and replacing fixed relation embeddings with dynamic, query-dependent representations.

To realize this design, we propose **GRAPHORACLE**, a relation-centric framework that transforms entity–relation interactions into a compact *Relation-Dependency Graph (RDG)*. Unlike INGRAM and ULTRA, which generate excessive connections, our RDG retains only meaningful directed precedence links, yielding significantly fewer edges yet richer compositional patterns, with its effectiveness remaining stable as the number of relations increases. To capture relation dependencies for a specific query, we design a multi-head attention mechanism that recursively propa-

\*Corresponding author

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

gates information over the RDG, dynamically assembling relation recipes conditioned on the query. By pre-training on four KGs in the general domain, GRAPHORACLE only needs minimal finetuning to achieve exceptional adaptability across transductive, inductive, and cross-domain reasoning tasks, improving performance by over 16.8% on average compared to state-of-the-art methods. Our key contributions in this work can be summarized as follows:

- We introduce **GRAPHORACLE**, a relation-centric model that converts KGs into RDGs, explicitly encoding compositional patterns while reducing the number of edges on the relation graph compared to prior approaches.
- We develop a query-dependent multi-head attention mechanism that dynamically propagates information over the RDG, yielding domain-invariant relation embeddings that enable generalization to unseen graphs.
- Extensive experiments across 60 benchmarks show that GRAPHORACLE consistently outperforms SOTA methods, with particularly strong results in both fully-inductive and cross-domain settings, demonstrating its robustness and generalization capability in challenging scenarios.

## 2 Related Works

**Knowledge Graph Reasoning** A KG consists of sets of entities  $\mathcal{V}$ , relations  $\mathcal{R}$ , and fact triples  $\mathcal{F} \subseteq (\mathcal{V} \times \mathcal{R} \times \mathcal{V})$  as  $\mathcal{G} = (\mathcal{V}, \mathcal{R}, \mathcal{F})$ .  $(e_q, r_q, e_a)$  is a triple in KG where  $e_q, e_a \in \mathcal{V}$  and  $r_q \in \mathcal{R}$ . KG reasoning encompasses several increasingly challenging settings based on what information is available during training versus inference. In the **transductive** setting, both entities and relations remain fixed:  $(\mathcal{V}_{tra} = \mathcal{V}_{inf}) \wedge (\mathcal{R}_{tra} = \mathcal{R}_{inf})$ . This allows models to learn fixed embeddings for all components. The **entity-inductive** setting introduces unseen entities at inference while keeping relations fixed:  $(\mathcal{V}_{tra} \neq \mathcal{V}_{inf}) \wedge (\mathcal{R}_{tra} = \mathcal{R}_{inf})$ . Most challenging is the **fully-inductive** setting where both entities and relations are novel:  $(\mathcal{V}_{tra} \neq \mathcal{V}_{inf}) \wedge (\mathcal{R}_{tra} \neq \mathcal{R}_{inf})$ . Beyond these, **cross-domain** reasoning requires transferring to entirely different KGs with no shared entities or relations, demanding the most robust generalization capabilities.

### 2.1 Transductive Reasoning

Transductive methods assume all entities and relations at inference have been seen during training, enabling the use of fixed relation embeddings. Models like ConvE (Dettmers et al. 2017), RotatE (Sun et al. 2019) and DuASE (Li et al. 2024) learn low-dimensional relation and entity embeddings directly, while GNN variants such as R-GCN (Schlichtkrull et al. 2018) implement relation-specific message passing that effectively parameterizing relation influence via learned embedding-like transformations. These embedding-based approaches form strong baselines but fundamentally cannot generalize beyond their training vocabulary.

### 2.2 Entity Inductive Reasoning

Entity inductive KG reasoning relaxes the entity constraint while maintaining fixed relation embeddings. Early solutions leveraged auxiliary cues—text descriptions in content-masking models (Shi and Weninger 2018) or ontological

Method	Ent	Full	Cross	Relation Representation
	Ind.	Ind.	Dom.	
RotatE & DuASE	✗	✗	✗	Relation Embedding
A*Net & AdaProp	✓	✗	✗	Relation Embedding
DRUM	✓	✗	✗	Differentiable rule chaining
RLogic	✓	✗	✗	Symbolic rule matching
INGRAM	✓	✓	✗	Undirected RG
ULTRA	✓	✓	✗	Interaction-Conditioned RG
TRIX	✓	✓	✗	Adjacency Motifs RG
KG-ICL	✓	✓	✗	Prompt RG
GRAPHORACLE	✓	✓	✓	Relation-Dependency Graph

Table 1: Comparison of inductive capabilities and relation representations. “RG” is short for “relation graph”.

features in OntoZSL (Geng et al. 2021)—and symbolic rule learners such as AMIE (Galárraga et al. 2013) and NeuralLP (Yang, Yang, and Cohen 2017). More recent approaches like DRUM (Sadeghian et al. 2019) employ differentiable rule chaining, while RLogic (Cheng et al. 2022) uses symbolic rule matching. SOTA GNN-based methods including GraIL (Teru, Denis, and Hamilton 2020), PathCon (Wang, Ren, and Leskovec 2021), NBFNet (Zhu et al. 2021b), RED-GNN (Zhang and Yao 2022a), A\*Net (Zhu et al. 2023) and AdaProp (Zhang et al. 2023b) propagate messages along relational paths to accommodate new entities—yet they still rely on fixed relation embeddings, limiting their applicability to scenarios with novel relations.

### 2.3 Fully-Inductive Reasoning

Fully-inductive settings demand handling both unseen entities and relations, requiring explicit relation graph (RD) structures. RMPI (Geng et al. 2023) and INGRAM (Lee, Chung, and Whang 2023) pioneer this direction by constructing undirected RDs; however, RMPI is limited to sub-graph extraction, and INGRAM’s degree discretization hampers transfer across graphs with different relation distributions. ISDEA (Gao et al. 2025) and MTDEA (Zhou, Bevilacqua, and Ribeiro 2023) adopt double-equivariant GNNs, but their computational overhead restricts scalability. ULTRA (Galkin et al. 2024) advances this with interaction-conditioned RDs that adapt based on query context. TRIX (Zhang et al. 2025b) introduces expressive adjacency motifs for richer relation modeling, while KG-ICL (Cui, Sun, and Hu 2024) employs prompt-based RDs.

### 2.4 Cross-domain Reasoning

Cross-domain KG reasoning represents the frontier of generalization, transferring patterns to graphs with entirely new entities and relations. Early work relied on domain-agnostic logical rules; recent advances leverage pre-trained graph foundation models. MDGFM (Wang et al. 2025d) introduces multi-domain contrastive pre-training, while SAMGPT (Zhang, Chen, and Huang 2025) demonstrates strong transfer without textual signals. Stability-GNN (Hong, Lee, and Whang 2025) addresses structural shift through adversarial perturbations, and UnifiedGNN (Lin, Wang, and Tang 2025) jointly handles

multiple inductive settings via relation adapters. RiemannGFM (Liu, Pan, and Sun 2025) incorporates geometric regularization, while Text-Free MDGPT (Li, Wang, and Xu 2025) and GraphMFM (Cheng, Jiang, and Li 2024) scale cross-domain pre-training through modality-agnostic masked modeling.

As summarized in Table 1, KG reasoning methods progress from fixed relation embeddings (transductive) to rule-based reasoning (entity-inductive) to explicit relation graphs (fully-inductive). While existing fully-inductive methods construct undirected or interaction-conditioned graphs, they remain limited to single-domain scenarios. GRAPHORACLE uniquely introduces directed relation-dependency graphs that capture compositional patterns, enabling the first successful cross-domain generalization.

### 3 Preliminary

Given a query  $(e_q, r_q, ?)$ , the goal is to find an answer entity  $e_a$  such that  $(e_q, r_q, e_a)$  is true. Many methods leverage GNN to aggregate relational paths and can be formulated as the following recursive function, where each candidate entity  $e_y$  at step  $\ell$  accumulates information from its in-neighbors:

$$\mathbf{h}_{r_q}^\ell(e_q, e_y) = \bigoplus_{(e_x, r, e_y) \in \mathcal{N}(e_y)} \mathbf{h}_{r_q}^{\ell-1}(e_q, e_x) \otimes \phi(r, r_q), \quad (1)$$

where  $\mathbf{h}_{r_q}^0(e_q, e) = \mathbf{1}$  if  $e = e_q$ , and  $\mathbf{0}$  otherwise,  $\oplus/\otimes$  are learnable additive and multiplicative operators.  $\phi$  encodes relation-type compatibility. After  $L$  steps' iteration, the answer is ranked by the score  $s(e_q, r_q, e_a) = \mathbf{w}_s^\top \mathbf{h}_{r_q}^L(e_q, e_a)$ . The learning objective is formulated in a contrastive approach that maximizes the log-likelihood of correct triples in the training set, which amounts to minimizing:

$$\mathcal{L}_{\text{train}} = - \sum_{(e_q, r_q, e_a) \in \mathcal{F}_{\text{train}}} \left[ \log \sigma(s(e_q, r_q, e_a)) + \sum_{e'_n \in \mathcal{N}_{(e_q, r_q, e_a)}^-} \log(1 - \sigma(s(e_q, r_q, e'_n))) \right], \quad (2)$$

where  $\sigma$  is a sigmoid function and  $\mathcal{N}_{(e_q, r_q, e_a)}^-$  contains negative samples. GRAPHORACLE retains this framework and introduces an RDG pre-training objective that endows  $\phi$  with universal semantics, enabling zero-shot generalization to *both* unseen entities and unseen relation vocabularies.

## 4 The Proposed Method

In order to enable fully inductive KG reasoning and improve the generalization ability of models across KGs, the key is to generalize the dependencies among relations for different KGs. To achieve this goal, we firstly introduce Relation Dependency Graph (RDG), which explicitly models how relations depend on each other, in Section 4.1 Based on RDG, we propose a query-dependent multi-head attention mechanism to learn relation representations from a weighted combination of precedent relations in Section 4.2. Subsequently,

in Section 4.3, we introduce the approach where entity representations are represented with the recursive function (1) in the original KGs by using the relation representations just obtained. The overview of our approach is shown in Fig. 1.

### 4.1 RDG Construction

To enable fully-inductive and cross-KG generalization in knowledge graph reasoning, we must capture the fundamental dependency patterns through which one relation can be expressed as a composition of others. Our key idea is to transform entity–relation interactions into a relation-dependency interaction structure that explicitly encodes how relations influence one another.

Given a KG  $\mathcal{G} = (\mathcal{V}, \mathcal{R}, \mathcal{F})$ , we construct a RDG  $\mathcal{G}^{\mathcal{R}}$  through a structural transformation. First, we define a relation adjacency operator  $\Phi : \mathcal{F} \rightarrow \mathcal{R} \times \mathcal{R}$  that extracts transitive relation dependencies:

$$\Phi(\mathcal{F}) = \bigcup_{e, e', e'' \in \mathcal{V}} \{(r_i, r_j) \mid (e, r_i, e') \in \mathcal{F} \wedge (e', r_j, e'') \in \mathcal{F}\}.$$

Then, the RDG is defined as  $\mathcal{G}^{\mathcal{R}} = (\mathcal{R}, \mathcal{E}^{\mathcal{R}})$ , where the node set  $\mathcal{R}$  contains all the relations and the edge sets  $\mathcal{E}^{\mathcal{R}} = \Phi(\mathcal{F})$  includes relation dependencies induced by entity-mediated pathways. This transformation alters the conceptual framework, shifting from an entity-centric perspective to a relation-interaction manifold where compositional connections between relations become explicit. Each directed edge  $(r_i, r_j)$  in  $\mathcal{G}^{\mathcal{R}}$  represents a potential relation-dependency pathway, indicating that relation  $r_i$  preconditions relation  $r_j$  through their sequential interaction over a shared entity context. The edge structure encodes compositional relational semantics, capturing how one relation may influence the probability or applicability of another when they occur in sequence.

To incorporate the hierarchical and compositional nature of relation interactions, we define a partial ordering function  $\tau : \mathcal{R} \rightarrow \mathbb{R}$  that assigns each relation a position in a relation precedence structure. This ordering is derived from the KG's inherent structure through rigorous topological analysis of relation co-occurrence patterns and functional dependencies. Relations that serve as logical precursors in inference chains are assigned lower  $\tau$  values, thereby establishing a directed acyclic structure in the relation graph that reflects the natural flow of information propagation. Using this ordering, we define the set of preceding relations for any relation  $r_v$  as:

$$\mathcal{N}^{\text{past}}(r_v) = \{r_u \in \mathcal{R} \mid (r_u, r_v) \in \mathcal{E}^{\mathcal{R}} \text{ and } \tau(r_u) < \tau(r_v)\}. \quad (3)$$

This formulation enables us to capture the directional dependency patterns where relations with lower positions in the hierarchy systematically precede and inform relations with higher  $\tau$ . By explicitly modeling these precedence relationships, our framework can identify and leverage compositional reasoning patterns that remain invariant across domains, enhancing the generalization capabilities.

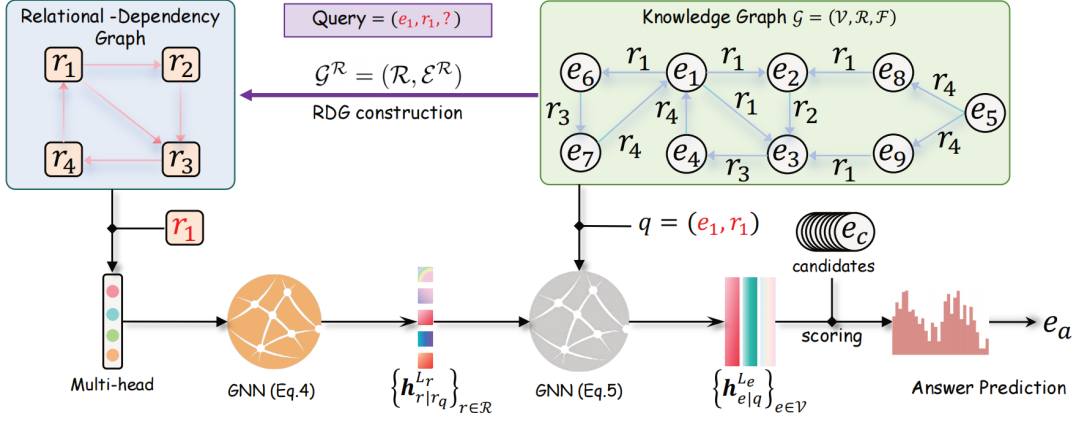


Figure 1: Overview of the GRAPHORACLE process that predicts the answer entity  $e_a$  from a given query  $(e_1, r_1, ?)$ : Given a KG, we first construct the RDG. Then, a GNN with multi-head attention is used to propagate messages among RDG to obtain relation representations  $\mathbf{h}_{r|r_q}^{L_r}$ , which are then used in another GNN with attention over entity representations. Finally, the candidate entities are scored based on the aggregated entity representations, and then ranked for answer entity prediction.

## 4.2 Relation Representation Learning on RDG

Building on the constructed RDG  $\mathcal{G}^{\mathcal{R}}$ , we develop a representation mechanism that captures the contextualized semantics of relations conditioned on a specific query. Given a query relation  $r_q$ , we introduce an RDG aggregation mechanism to compute  $d$ -dimensional relation-node representations  $\mathbf{R}_q \in \mathbb{R}^{|\mathcal{R}| \times d}$  conditional on  $r_q$ .

Following Eq.(1), we apply a labeling initialization to distinguish the query relation node  $r_q$  in  $\mathcal{G}^{\mathcal{R}}$ . Then employ multi-head attention relation-dependency message passing over the graph:

$$\mathbf{h}_{r_v|r_q}^0 = \text{INDICATOR}_r(r_v, r_q) = \delta_{r_v, r_q} \cdot \mathbf{1}^d, \quad r_v \in \mathcal{G}^{\mathcal{R}}$$

$$\mathbf{h}_{r_v|r_q}^\ell = \sigma \left( \frac{1}{H} \sum_{h=1}^H \left[ \mathbf{W}_1^{\ell, h} \sum_{r_u \in \mathcal{N}^{\text{past}}(r_v)} \hat{\alpha}_{r_u r_v}^{\ell, h} \mathbf{h}_{r_u|r_q}^{\ell-1} + \mathbf{W}_2^{\ell, h} \hat{\alpha}_{r_v r_v}^{\ell, h} \mathbf{h}_{r_v|r_q}^{\ell-1} \right] \right), \quad (4)$$

where  $\delta_{r_v, r_q} = 1$  if  $v = q$ , and 0 otherwise.  $H$  is the number of attention heads, and  $\mathbf{W}_1^{\ell, h}, \mathbf{W}_2^{\ell, h} \in \mathbb{R}^{d \times d}$  are head-specific parameter matrices. The relation-dependency attention weight  $\hat{\alpha}_{r_u r_v}^{\ell, h}$  captures the directional influence of relation  $r_u$  on relation  $r_v$ , computed as:

$$\hat{\alpha}_{r_u r_v}^{\ell, h} = \frac{\exp(\mathbf{a}^T (\mathbf{W}_\alpha^h \mathbf{h}_{r_u}^{\ell-1} \parallel \mathbf{W}_\alpha^h \mathbf{h}_{r_v}^{\ell-1}))}{\sum_{r_w \in \mathcal{N}^{\text{past}}(r_v)} \exp(\mathbf{a}^T (\mathbf{W}_\alpha^h \mathbf{h}_{r_w}^{\ell-1} \parallel \mathbf{W}_\alpha^h \mathbf{h}_{r_v}^{\ell-1}))},$$

where  $\mathbf{a} \in \mathbb{R}^{2d}$  is a learnable attention parameter vector,  $\parallel$  denotes vector concatenation, and  $\mathbf{W}_\alpha^h \in \mathbb{R}^{d \times d}$  are head-specific trainable projection matrix. The neighborhood function  $\mathcal{N}^{\text{past}}(r_v)$  enforces the relation-dependency ordering of relations as defined in Eq. (3).

After  $L_r$  layers of message passing, the final relation representation incorporates both local and higher-order dependencies  $\mathbf{R}_q = \{\mathbf{h}_{r|r_q}^{L_r} \mid r \in \mathcal{R}\}$ .

## 4.3 Entity Representation Learning on the Original KG

After obtaining the relation representations  $\mathbf{R}_q$  from RDG conditioned on  $r_q$ , we obtain query-dependent entity representations by conducting message passing over the original KG structures. This approach enables effective reasoning across both seen and unseen entities and relations.

For a given query  $(e_q, r_q, ?)$ , we compute entity representations recursively with Eq. (1) through the KG  $\mathcal{G}$ . The initial representations  $\mathbf{h}_{e|q}^0 = \mathbf{1}$  if  $e = e_q$ , and otherwise  $\mathbf{0}$ . At each layer  $\ell$ , the representation of an entity  $e$  is computed as:

$$\mathbf{h}_{e|q}^\ell = \delta \left( \mathbf{W}^\ell \cdot \sum_{(e_s, r, e) \in \mathcal{F}_{\text{train}}} \alpha_{e_s, r|q}^\ell (\mathbf{h}_{e_s|q}^{\ell-1} + \mathbf{h}_{r|r_q}^{L_r}) \right), \quad (5)$$

where  $\delta(\cdot)$  is a non-linear activation. The attention weight  $\alpha_{e_s, r|q}^\ell$  is computed as:

$$\alpha_{e_s, r|q}^\ell = \sigma \left( (\mathbf{w}_\alpha^\ell)^\top \text{ReLU}(\mathbf{W}_\alpha^\ell \cdot (\mathbf{h}_{e_s|q}^{\ell-1} \parallel \mathbf{h}_{r|r_q}^{L_r} \parallel \mathbf{h}_{r_q|r_q}^{L_r})) \right),$$

where  $\mathbf{w}_\alpha^\ell \in \mathbb{R}^d$  and  $\mathbf{W}_\alpha^\ell \in \mathbb{R}^{d \times 3d}$  are learnable parameters,  $\sigma$  is the sigmoid function and  $\cdot$  denotes the standard matrix-vector multiplication.

We iterate Eq. (5) for  $L_e$  steps and use the final layer representation  $\mathbf{h}_{e|q}^{L_e}$  for scoring each entity  $e \in \mathcal{V}$ . The critical idea here is replacing the learnable relation embeddings  $\mathbf{r}$  with the contextualized relation embedding  $\mathbf{h}_{r|r_q}^{L_r}$  from our RDG, enabling fully inductive reasoning (Time complexity of the GRAPHORACLE model is given in Appendix C, and the Theoretical analysis on its expressiveness and generalization is given in Appendix D).

## 4.4 Training Details

All the learnable parameters such as  $\{\mathbf{W}_O^h, \mathbf{W}_1^{\ell, h}, \mathbf{W}_2^{\ell, h}, \mathbf{W}_\alpha^h, \mathbf{a}, \mathbf{W}^\ell, \mathbf{W}_\alpha^\ell, \mathbf{w}_\alpha^\ell, \mathbf{w}^L\}$  are trained end-to-end by minimizing the loss function Eq. (2). GRAPHORACLE adopts a sequential multi-dataset

pre-train  $\rightarrow$  fine-tune paradigm to acquire a general relation-dependency graph representation across KGs  $\{\mathcal{G}_1, \dots, \mathcal{G}_K\}$ . For each graph  $\mathcal{G}_k$ , we optimize the regularized objective  $\mathcal{L}^{(k)} = \mathcal{L}_{\text{task}}^{(k)} + \lambda_k \|\Theta\|_2^2$ , where  $\mathcal{L}_{\text{task}}^{(k)}$  denotes the task-specific loss on  $\mathcal{G}_k$  (e.g., Eq. (2)),  $\Theta$  represents all learnable parameters, and  $\lambda_k$  controls the strength of L2 regularization. Early stopping technique is used for each graph once validation MRR fails to improve for several epochs. This iterative pre-train process, together with our relation-dependency graph encoder, equips GRAPHORACLE with strong cross-domain generalization. When adapting GRAPHORACLE to new KGs, we firstly build the RDG and then support two inference paradigms:

- **Zero-shot Inference.** The pre-trained model is directly applied to unseen KGs without tuning.
- **Fine-tuning.** For more challenging domains, we fine-tune the pre-trained parameters on the target KG  $\mathcal{G}_{\text{target}}$  for a limited number of epochs  $E_{\text{fine-tune}} \ll E_{\text{train}}$ .

## 5 Experiment

To evaluate the comprehensive capabilities of GRAPHORACLE as a Foundation Model for KG reasoning, we formulate the following research questions: **RQ1:** How does GRAPHORACLE model perform compared with state-of-the-art models on diverse KGs and cross-domain datasets? **RQ2:** How do different relation-dependency patterns contribute to GRAPHORACLE’s performance? **RQ3:** To what extent can external information enhance the performance of GRAPHORACLE? **RQ4:** How do the components and configurations contribute to the performance?

### 5.1 Experimental Setup

**Datasets** We conduct comprehensive experiments on 60 KGs, which we classify into three categories according to their properties (Details are given in Appendix D.):

- **Transductive and Inductive datasets.** To ensure fair comparison, we follow the same dataset settings as ULTRA (Galkin et al. 2024), TRIX (Zhang et al. 2025b), and KG-ICL (Cui, Sun, and Hu 2025), including 16 transductive, 18 entity-inductive, and 23 fully-inductive datasets, totaling 57 in all.
- **Cross domain datasets.** (i) **Biomedical Datasets:** We use biomedical KG PrimeKG (Chandak, Huang, and Zitnik 2023) to examine the cross-domain capabilities of GRAPHORACLE. We finetune with 80% samples in raw PrimeKG, and validate with 10% samples. When testing on the remaining 10%, we specially focus on the predictions for triplets: (Protein, *Interacts\_with*, BP/M-F/CC), (Drug, *Indication*, Disease), (Drug, *Target*, Protein), (Protein, *Associated\_with*, Disease), (Drug, *Contraindication*, Disease). (ii) **Recommendation domain:** We transform the Amazon-book (Wang et al. 2019) dataset into a pure KG reasoning dataset to adapt to the KGs Reasoning field by defining the interactions between users and items as a new relation in the KG. (iii) **Geographic datasets** (GeoKG) (GeoNames Team 2025) . (Detail process are given in Appendix F.)

**Pretrain and Finetune.** GRAPHORACLE is pre-trained on three general KGs (NELL-995, CoDEX-Medium, FB15k-237) to capture diverse relational structures and reasoning patterns. It takes 150,000 training steps with batch size of 32 using AdamW optimizer (Loshchilov and Hutter 2019) on a single A6000 (48GB) GPU. For cross-domain adaptation, we employ a lightweight fine-tuning approach that updates only the final layer parameters while freezing the pre-trained representations. The finetune process only takes 1  $\sim$  2 epochs to achieve the best results. The pre-training process takes approximately 36 hours, while fine-tuning requires only 15  $\sim$  60 minutes depending on the target dataset. Detailed hyperparameters, architecture specifications, and training configurations are provided in Appendix E.

**Baselines** We compare the proposed GRAPHORACLE with (i) **Transductive:** ConvE (Dettmers et al. 2017), QuatE (Zhang et al. 2019a), DuASE (Li et al. 2024) and BioBRIDGE (Wang et al. 2024); (ii) **Entity inductive:** MINERVA (Das et al. 2017), DRUM (Sadeghian et al. 2019), AnyBURL (Meilicke et al. 2020), RNNLogic (Qu et al. 2021), RLogic (Cheng et al. 2022) GraphRulRL (Mai et al. 2025), CompGCN (Vashishth et al. 2019), NBFNet (Zhu et al. 2021a), RED-GNN (Zhang and Yao 2022b), A\*Net (Zhu et al. 2023), Adaprop (Zhang et al. 2023c) and one-shot-subgraph (Zhou et al. 2024); (iii) **Fully inductive:** INGRAM (Lee, Chung, and Whang 2023), ULTRA (Galkin et al. 2024), TRIX (Zhang et al. 2025b) and KG-ICL (Cui, Sun, and Hu 2025). The results for the baseline methods were either directly obtained from the original publications or reproduced using the official source code provided by the authors. Due to page limitations, some other baselines can be found in INGRAM (Lee, Chung, and Whang 2023), BioBRIDGE (Wang et al. 2024) and KUCNet (Liu et al. 2024).

### 5.2 Overall Performance (RQ1)

The main experimental results demonstrate the performance of GRAPHORACLE after pre-training on three knowledge graphs and a brief fine-tuning phase of only two epochs across 60 diverse datasets (with comprehensive results reported in Appendix G). A key observation is that GRAPHORACLE consistently surpasses the supervised state-of-the-art across all evaluated baseline datasets and metrics, as summarized in Table 2. This consistent superiority highlights the overall effectiveness of the proposed approach, with particularly substantial gains in more challenging evaluation settings. Notably, the improvements are observed across all reasoning types, and are most pronounced in fully inductive scenarios, where the model is required to generalize to entirely unseen entities—representing the most stringent tests of reasoning capability.

### 5.3 Relation-Dependency Pattern Analysis (RQ2)

To investigate whether GRAPHORACLE truly internalizes the compositionality of the relations, that is, the way complex relations are constructed systematically from simpler ones, we performed a series of perturbation analyzes on the learned RDG. First, we calculated relation attention weights using Eq. (4.2), averaged over the WN18RR and YAGO3-10

Model	Transductive			Entity Inductive			Fully Inductive		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
Supervised SOTA	0.4185	0.4715	0.5771	0.4915	0.4730	0.6296	0.4593	0.2942	0.6246
GraphOracle	0.4486	0.5550	0.6111	0.5449	0.5684	0.6722	0.5203	0.3688	0.7279
Improvement	7.19%	17.70%	5.89%	10.86%	20.16%	6.77%	13.28%	25.36%	16.54%

Table 2: Average performance comparison between GRAPHORACLE and Supervised SOTA under four generalization settings.

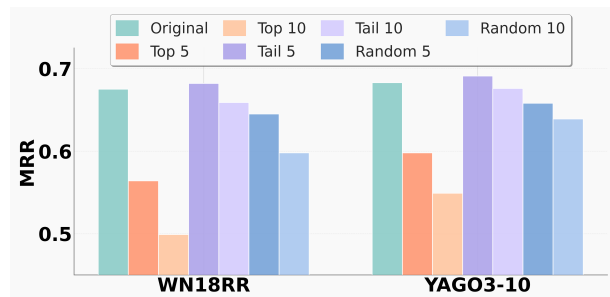


Figure 2: Perturbation Analysis of RDG Edges by Attention-Derived Importance Scores.

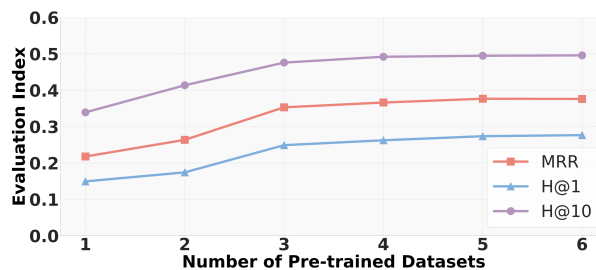


Figure 3: Impact of Number of Pre-trained Datasets on Zero-Shot Evaluation Metrics.

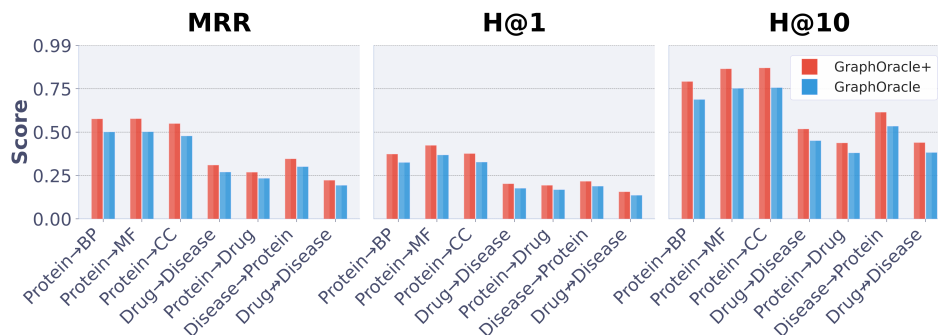


Figure 4: Comparison on PrimeKG: Evaluating GRAPHORACLE Enhanced by External Entity information (GRAPHORACLE+).

datasets, to assign an *importance score* to each relation pair, reflecting its contribution to downstream reasoning. Subsequently, during inference, we systematically disabled specific subsets of edges based on these attention weights: (i) the top-5 and top-10 most important (highest attention) compositional relation pairs; (ii) the bottom-5 and bottom-10 least important (lowest attention) pairs; and (iii) 5 and 10 randomly selected pairs.

As shown in Fig. 2, removing the highly-ranked compositional edges—those encoding key multi-hop templates essential for composing higher-level relations—causes a sharp decline in MRR on both WN18RR and YAGO3-10. This confirms that GRAPHORACLE heavily relies on these identified compositional pathways for its predictions. Conversely, suppressing a small number of low-importance edges sometimes leads to slight performance improvements, suggesting that these weaker compositional cues might act as semantic noise. Perturbations involving randomly removed edges result in only moderate performance degradation. This underscores the idea that it is not merely the quantity of re-

lations but the specific, learned compositional interactions between them that are crucial for GRAPHORACLE’s reasoning process. These findings collectively substantiate that GRAPHORACLE’s predictions are rooted in the compositional structure of its RDG, rather than relying on isolated relation statistics.

#### 5.4 Compatible with Additional Initial Information (RQ3)

To explore the potential of external information in enhancing KG reasoning, we introduced an improved entity initialization strategy. This involved incorporating modality-specific encoded features as initial entity vectors, moving beyond standard random initialization. The resulting model, denoted as **GRAPHORACLE+**, leverages foundation model embeddings to create more semantically rich entity representations (details are provided in Appendix H). As demonstrated in Fig. 4 (details are given in Appendix Tabel 17), experimental evaluations on the PrimeKG dataset show that GRAPHORACLE+ achieves consistent performance gains across all met-

Models	Nell-100			WK-100			FB-100			YAGO3-10		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
GRAPHORACLE	<b>0.702</b>	<b>0.623</b>	<b>0.905</b>	<b>0.417</b>	<b>0.192</b>	<b>0.711</b>	<b>0.576</b>	<b>0.407</b>	<b>0.812</b>	<b>0.696</b>	<b>0.672</b>	<b>0.807</b>
W/o RDG	0.376	0.278	0.493	0.132	0.102	0.243	0.237	0.149	0.432	0.593	0.545	0.712
W/o multi-head	0.598	0.534	0.817	0.302	0.39	0.619	0.492	0.321	0.724	0.629	0.576	0.722
Graph <sub>INGRAM</sub>	0.478	0.375	0.663	0.189	0.092	0.389	0.398	0.283	0.557	0.478	0.512	0.625
Graph <sub>ULTRA</sub>	0.548	0.490	0.720	0.201	0.105	0.466	0.465	0.297	0.679	0.587	0.583	0.725
Message <sub>INGRAM</sub>	0.517	0.426	0.726	0.271	0.111	0.518	0.428	0.289	0.635	0.539	0.473	0.674
Message <sub>ULTRA</sub>	0.569	0.502	0.741	0.282	0.124	0.597	0.468	0.305	0.672	0.503	0.539	0.669

Table 3: Ablation Analysis of GRAPHORACLE’s Core Architectural Components across Four Benchmark Datasets.

rics. Notably, MRR scores improved by 15% for protein-biological process prediction and 14% for protein-molecular function prediction. These results affirm that GRAPHORACLE’s framework significantly benefits from integrating external information. In an era increasingly influenced by large language models, the capacity for flexible incorporation of diverse information sources is crucial for advancing generalization and adaptability, especially within specialized and complex domains such as biomedicine.

### 5.5 Ablation Study (RQ4)

To rigorously evaluate the contribution of each architectural component within GRAPHORACLE, we conducted an extensive series of ablation experiments. We first investigate the impact of removing the RDG and the effect of reducing the number of attention heads  $H$  in Eq. (4) from eight to one. The results are detailed in Table 3. Clearly, eliminating the RDG or the multi-head attention mechanism causes a marked decline in all evaluation metrics, highlighting their indispensability to GRAPHORACLE’s performance.

In addition, we quantified how the breadth of pre-training data affects zero-shot performance. Specifically, we pre-trained on one to six heterogeneous datasets and evaluated the resulting checkpoints on unseen Table 3’s graphs. The averaged results, depicted in Fig. 3 (implementation details are reported in Appendix E), reveal that zero-shot performance saturates once three diverse datasets are included in the pre-training mixture. Incorporating additional datasets beyond this point yields no further significant gains. We conjecture that after this point the model has already encountered a sufficiently rich spectrum of relational patterns, and subsequent datasets may introduce largely redundant or potentially noisy signals.

Furthermore, to underscore the unique contributions of our proposed mechanisms, we compared GRAPHORACLE’s relation graph construction and message passing techniques against those employed by INGRAM and ULTRA. For this, we created variants where Graph<sub>INGRAM</sub> denotes using INGRAM’s method for relation graph construction, and Message<sub>INGRAM</sub> signifies adopting INGRAM’s message passing scheme (similarly for ULTRA). As detailed in Table 3, substituting either GRAPHORACLE’s graph construction or its message passing method with those from INGRAM or ULTRA resulted in a substantial reduction in performance. This comparative analysis further substantiates the effectiveness and integral role of each distinct component within the GRAPHORACLE framework.

## 6 Conclusion

In this work, we introduced GRAPHORACLE, a relation-centric foundation model for unifying reasoning across heterogeneous KGs. By converting KGs into RDG, our approach explicitly encodes compositional patterns among relations, yielding domain-invariant embeddings. Experiments on 60 diverse benchmarks showed consistent state-of-the-art performance, improving mean reciprocal rank by up to 16.8% over baselines with minimal adaptation. We also demonstrated that GRAPHORACLE’s performance can be enhanced by integrating external information through GRAPHORACLE+, which leverages foundation model embeddings for improved initialization. Ablation studies confirmed the essential contributions of the relation-dependency graph and multi-head attention components. These findings establish relation-dependency pre-training as a scalable approach toward universal KG reasoning.

## Acknowledgements

This work is supported by Guangdong Basic and Applied Basic Research Foundation 2025A1515010304, Guangzhou Science and Technology Planning Project 2025A03J4491.

## References

- Bai, J.; Fan, W.; Hu, Q.; Zong, Q.; Li, C.; Tsang, H. T.; Luo, H.; Yim, Y.; Huang, H.; Zhou, X.; Qin, F.; Zheng, T.; Peng, X.; Yao, X.; Yang, H.; Wu, L.; Ji, Y.; Zhang, G.; Chen, R.; and Song, Y. 2025. AutoSchemaKG: Autonomous Knowledge Graph Construction through Dynamic Schema Induction from Web-Scale Corpora. In *arXiv: 2505.23628*.
- Chandak, P.; Huang, K.; and Zitnik, M. 2023. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1): 67.
- Cheng, F.; Jiang, H.; and Li, X. 2024. GraphMFM: Modality-Agnostic Masked Graph Modelling for Cross-Graph Transfer. In *EMNLP*.
- Cheng, K.; Liu, J.; Wang, W.; and Sun, Y. 2022. RLogic: Recursive Logical Rule Learning from Knowledge Graphs. In *KDD*, 179–189.
- Cui, Y.; Sun, Z.; and Hu, W. 2024. A Prompt-Based Knowledge Graph Foundation Model for Universal In-Context Reasoning. In *NeurIPS*.
- Cui, Y.; Sun, Z.; and Hu, W. 2025. A Prompt-Based Knowledge Graph Foundation Model for Universal In-Context Reasoning.

- Das, R.; Dhuliawala, S.; Zaheer, M.; Vilnis, L.; Durugkar, I.; Krishnamurthy, A.; Smola, A.; and McCallum, A. 2017. Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. In *ICLR*.
- Dettmers, T.; Minervini, P.; Stenetorp, P.; and Riedel, S. 2017. Convolutional 2D knowledge graph embeddings. In *AAAI*.
- Du, E.; Liu, S.; and Zhang, Y. 2025. Mixture of Length and Pruning Experts for Knowledge Graphs Reasoning. *EMNLP*.
- Galárraga, L.; Teflioudi, C.; Hose, K.; and Suchanek, F. M. 2013. AMIE: Association Rule Mining under Incomplete Evidence in Ontological Knowledge Bases. In *WWW*, 413–422. ACM.
- Galkin, M.; Yuan, X.; Mostafa, H.; Tang, J.; and Zhu, Z. 2024. Towards Foundation Models for Knowledge Graph Reasoning. *ICLR*.
- Gao, J.; Zhou, Y.; Zhou, J.; and Ribeiro, B. 2025. Double Equivariance for Inductive Link Prediction for Both New Nodes and New Relation Types. *NeurIPS*.
- Geng, Y.; Chen, J.; Chen, Z.; Pan, J. Z.; Ye, Z.; Yuan, Z.; Jia, Y.; and Chen, H. 2021. OntoZSL: Ontology-Enhanced Zero-Shot Learning. In *WWW*, 3325–3336.
- 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 *ICDE*, 1221–1233. IEEE.
- GeoNames Team. 2025. GeoNames: The Geographical Database. <https://www.geonames.org/>. Maintained by unxos gmbh. Accessed: 2025-05-05.
- Hong, M.; Lee, J.; and Whang, J. J. 2025. Stability and Generalization Capability of Subgraph Reasoning Models for Inductive Knowledge Graph Completion. In *ICLR*.
- Lee, J.; Chung, C.; and Whang, J. J. 2023. INGRAM: Inductive Knowledge Graph Embedding via Relation Graphs. In *ICML*, volume 202, -. PMLR.
- Li, J.; Su, X.; Zhang, F.; and Gao, G. 2024. Learning Low-dimensional Multi-domain Knowledge Graph Embedding via Dual Archimedean Spirals. In *Findings of the Association for Computational Linguistics: ACL 2024*, 1982–1994. Association for Computational Linguistics.
- Li, X.; Wang, Q.; and Xu, X. 2025. Text-Free Multi-domain Graph Pre-training: Toward Graph Foundation Models. In *ACL*.
- Lin, Z.; Wang, R.; and Tang, J. 2025. UnifiedGNN: Bridging Transductive, Inductive and Fully-inductive KG Reasoning. In *ICML*.
- Liu, G.; Yao, Q.; Zhang, Y.; and Chen, L. 2024. Knowledge-Enhanced Recommendation with User-Centric Subgraph Network. *ICDE*.
- Liu, Y.; Pan, S.; and Sun, L. 2025. RiemannGFM: Geometric Graph Foundation Models for Cross-Domain Transfer. In *NeurIPS*.
- Loshchilov, I.; and Hutter, F. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations (ICLR)*.
- Mai, Z.; Wang, W.; Liu, X.; Feng, X.; Wang, J.; and Fu, W. 2025. A Reinforcement Learning Approach for Graph Rule Learning. *Big Data Mining and Analytics*, 8(1): 31–44.
- Meilicke, C.; Chekol, M. W.; Fink, M.; and Stuckenschmidt, H. 2020. Reinforced Anytime Bottom Up Rule Learning for Knowledge Graph Completion. *arXiv preprint arXiv:2004.04412*.
- Qu, M.; Chen, J.; Xhonneux, L.; Bengio, Y.; and Tang, J. 2021. RNNLogic: Learning Logic Rules for Reasoning on Knowledge Graphs. In *ICLR*.
- Sadeghian, A.; Armandpour, M.; Ding, P.; and Wang, D. 2019. DRUM: End-To-End Differentiable Rule Mining On Knowledge Graphs. In *NeurIPS*, 15347–15357.
- 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 the European Semantic Web Conference (ESWC)*, 593–607.
- Shen, J.; Qian, H.; Liu, S.; Zhang, W.; Jiang, B.; and Zhou, A. 2024. Capturing Homogeneous Influence among Students: Hypergraph Cognitive Diagnosis for Intelligent Education Systems. In *KDD*, 2628–2639.
- Shi, B.; and Wenginger, T. 2018. Open-World Knowledge Graph Completion. In *AAAI*, 1957–1964.
- Sun, Z.; Deng, Z.-H.; Nie, J.-Y.; and Tang, J. 2019. RotatE: Knowledge graph embedding by relational rotation in complex space. In *ICLR*.
- Teru, K. K.; Denis, E.; and Hamilton, W. L. 2020. Inductive Relation Prediction by Subgraph Reasoning. In *ICML*, 9448–9457.
- Vashishth, S.; Sanyal, S.; Nitin, V.; and Talukdar, P. 2019. Composition-based multi-relational graph convolutional networks.
- Wang, H.; Ren, H.; and Leskovec, J. 2021. Relational Message Passing for Knowledge Graph Completion. In *KDD*, 1697–1707.
- Wang, H.; Tian, Y.; Liu, M.; Zhang, Z.; and Zhu, X. 2025a. SDEval: Safety Dynamic Evaluation for Multimodal Large Language Models. *arXiv:2508.06142*.
- Wang, H.; Wang, S.; Zhong, Y.; Yang, Z.; Wang, J.; Cui, Z.; Yuan, J.; Han, Y.; Liu, M.; and Ma, Y. 2025b. Affordance-R1: Reinforcement Learning for Generalizable Affordance Reasoning in Multimodal Large Language Model. *arXiv:2508.06206*.
- Wang, H.; Zhang, Z.; Ji, K.; Liu, M.; Yin, W.; Chen, Y.; Liu, Z.; Zeng, X.; Gui, T.; and Zhang, H. 2025c. DAG: Unleash the Potential of Diffusion Model for Open-Vocabulary 3D Affordance Grounding. *arXiv:2508.01651*.
- Wang, S.; Wang, B.; Shen, Z.; Deng, B.; and Kang, Z. 2025d. Multi-Domain Graph Foundation Models: Robust Knowledge Transfer via Topology Alignment. *arXiv:2502.02017*.
- Wang, X.; He, X.; Cao, Y.; Liu, M.; and Chua, T.-S. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, 950–958. ACM.

- Wang, Z.; Wang, Z.; Srinivasan, B.; Ioannidis, V. N.; Rangwala, H.; and Anubhai, R. 2024. BioBRIDGE: Bridging Biomedical Foundation Models via Knowledge Graphs. In *ICLR*.
- Yang, F.; Yang, Z.; and Cohen, W. W. 2017. Differentiable Learning of Logical Rules for Knowledge Base Reasoning. In *NeurIPS*, 2319–2328.
- Yang, S.; Qin, Z.; Du, E.; Zhou, P.; and Huang, T. 2025. Dual Social View Enhanced Contrastive Learning for Social Recommendation. *IEEE Transactions on Computational Social Systems*, 12(5): 2156–2170.
- Yu, C.; Wang, H.; Shi, Y.; Luo, H.; Yang, S.; Yu, J.; and Wang, J. 2025. SeqAfford: Sequential 3D Affordance Reasoning via Multimodal Large Language Model. In *CVPR*.
- Zhang, G.; Yuan, G.; Cheng, D.; Liu, L.; Li, J.; and Zhang, S. 2025a. Mitigating propensity bias of large language models for recommender systems. *ACM Transactions on Information Systems*, 43(6): 1–26.
- Zhang, G.; Zhang, S.; and Yuan, G. 2024. Bayesian graph local extrema convolution with long-tail strategy for misinformation detection. *ACM Transactions on Knowledge Discovery from Data*, 18(4): 1–21.
- Zhang, S.; Tay, Y.; Yao, L.; and Liu, Q. 2019a. Quaternion knowledge graph embeddings. In *NeurIPS*.
- Zhang, Y.; Bevilacqua, B.; Galkin, M.; and Ribeiro, B. 2025b. TRIX: A More Expressive Model for Zero-shot Domain Transfer in Knowledge Graphs.
- Zhang, Y.; Chen, Q.; and Huang, J. 2025. SAMGPT: Text-Free Graph Foundation Model for Multi-domain Pre-training and Cross-domain Adaptation. In *SIGIR*.
- Zhang, Y.; and Yao, Q. 2022a. Knowledge Graph Reasoning with Relational Digraph. In *WWW*, 1–13. ACM. ISBN 978-1-4503-9096-5/22/04.
- Zhang, Y.; and Yao, Q. 2022b. Knowledge Graph Reasoning with Relational Directed Graph. In *Proceedings of TheWebConf*.
- Zhang, Y.; Yao, Q.; Dai, W.; and Chen, L. 2020. AutoSF: Searching Scoring Functions for Knowledge Graph Embedding. In *ICDE*, 433–444.
- Zhang, Y.; Yao, Q.; Shao, Y.; and Chen, L. 2019b. NSCaching: Simple and Efficient Negative Sampling for Knowledge Graph Embedding. In *ICDE*, 614–625.
- Zhang, Y.; Yao, Q.; Yue, L.; Wu, X.; Zhang, Z.; Lin, Z.; and Zheng, Y. 2023a. Emerging drug interaction prediction enabled by a flow-based graph neural network with biomedical network. *Nature Computational Science*, 3(12): 1023–1033.
- 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 the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*, 1–12. New York, NY, USA: ACM. ISBN 979-8-4007-0103-0/23/08.
- Zhang, Y.; Zhou, Z.; Yao, Q.; Chu, X.; and Han, B. 2023c. Adaprop: Learning adaptive propagation for knowledge graph reasoning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*.
- Zhou, J.; Bevilacqua, B.; and Ribeiro, B. 2023. An OOD Multi-Task Perspective for Link Prediction with New Relation Types and Nodes. *arXiv:2307.06046*.
- Zhou, Z.; Zhang, Y.; Yao, J.; Yao, Q.; and Han, B. 2024. LESS IS MORE: ONE-SHOT-SUBGRAPH LINK PREDICTION ON LARGE-SCALE KNOWLEDGE GRAPHS. In *ICLR*.
- Zhu, Z.; Yuan, X.; Galkin, M.; Xhonneux, S.; Zhang, M.; Gazeau, M.; and Tang, J. 2023. A\*Net: A Scalable Path-based Reasoning Approach for Knowledge Graphs. In *NeurIPS*.
- Zhu, Z.; Zhang, Z.; Xhonneux, L.; and Tang, J. 2021a. Neural Bellman-Ford Networks: A General Graph Neural Network Framework for Link Prediction. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- Zhu, Z.; Zhang, Z.; Xhonneux, L.-P.; and Tang, J. 2021b. Neural Bellman-Ford Networks: A General Graph Neural Network Framework for Link Prediction. In *NeurIPS*.