Lecture#3: Foundation Models for Knowledge Graph Reasoning

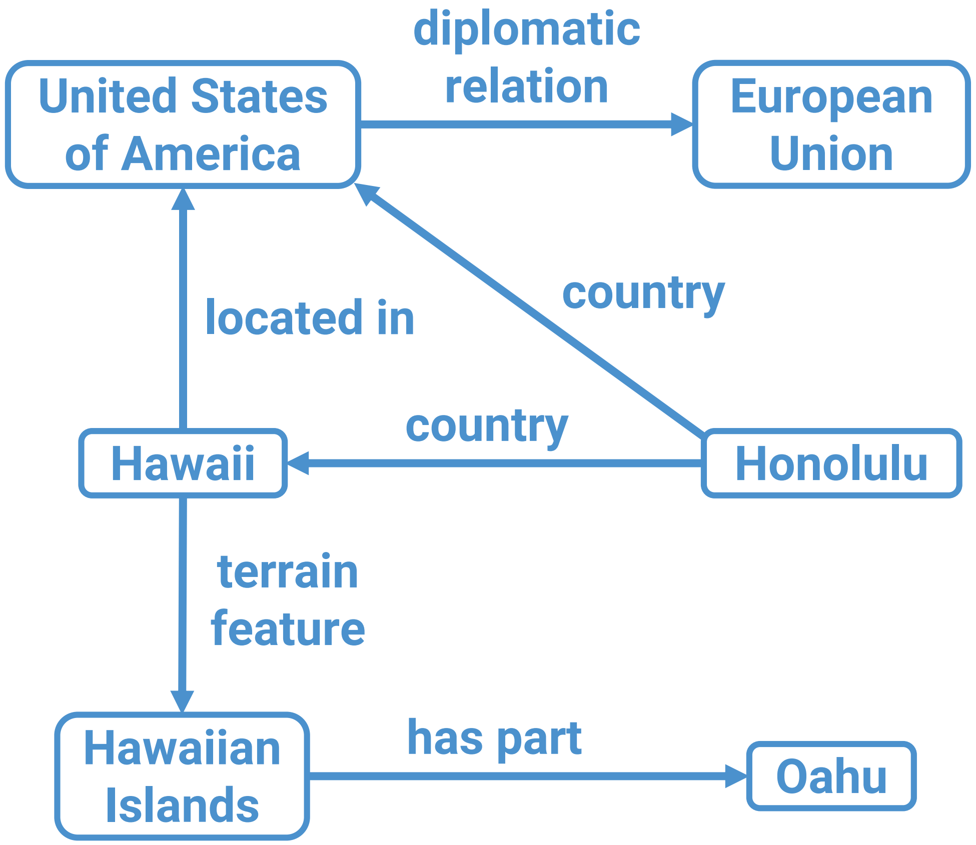
Joyce Jiyoung Whang
School of Computing, KAIST

Key Facets in Modern Knowledge Graph Representation Learning (KeyKGRL), ISWC 2025 Tutorial

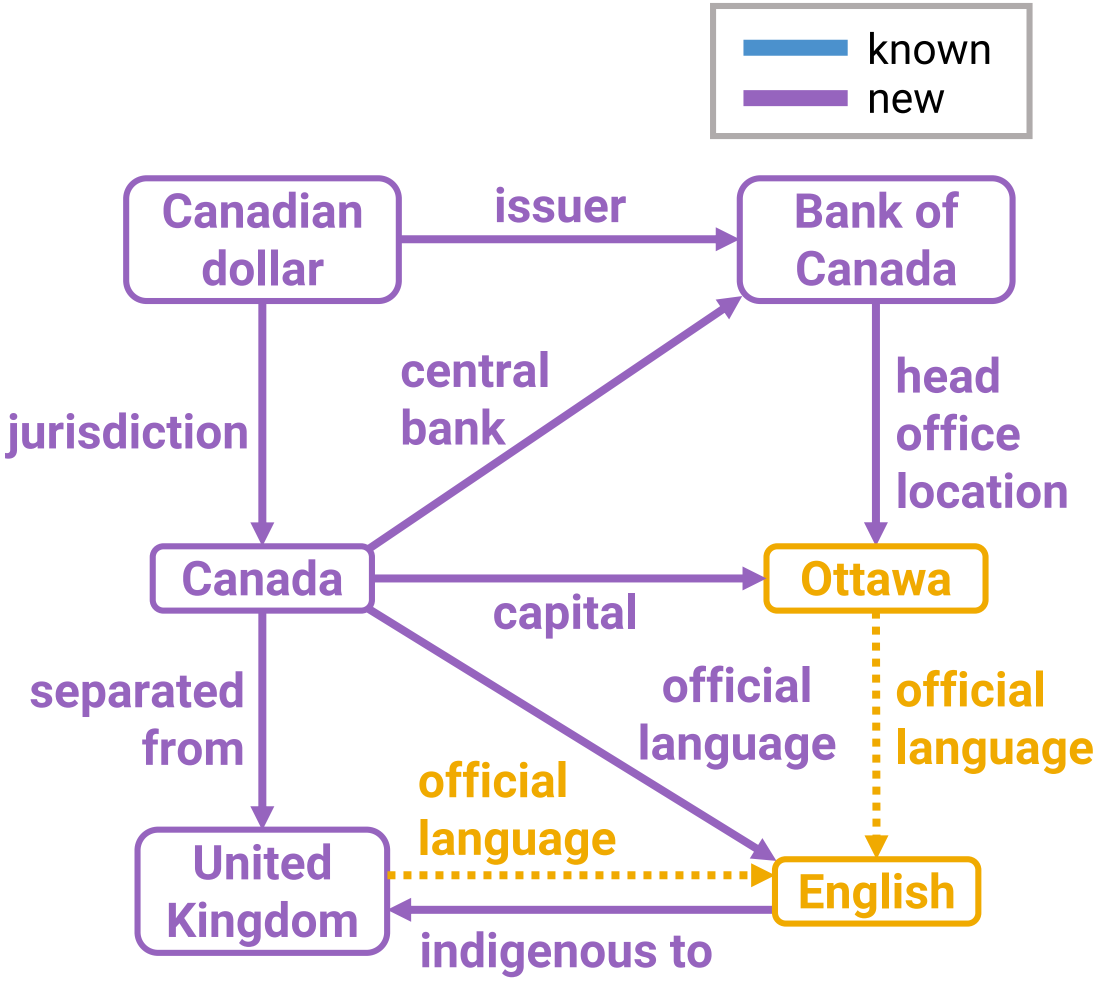
https://bdi-lab.kaist.ac.kr

KAIST BDILab
BIG DATA INTELLIGENCE

01 Inductive Inference

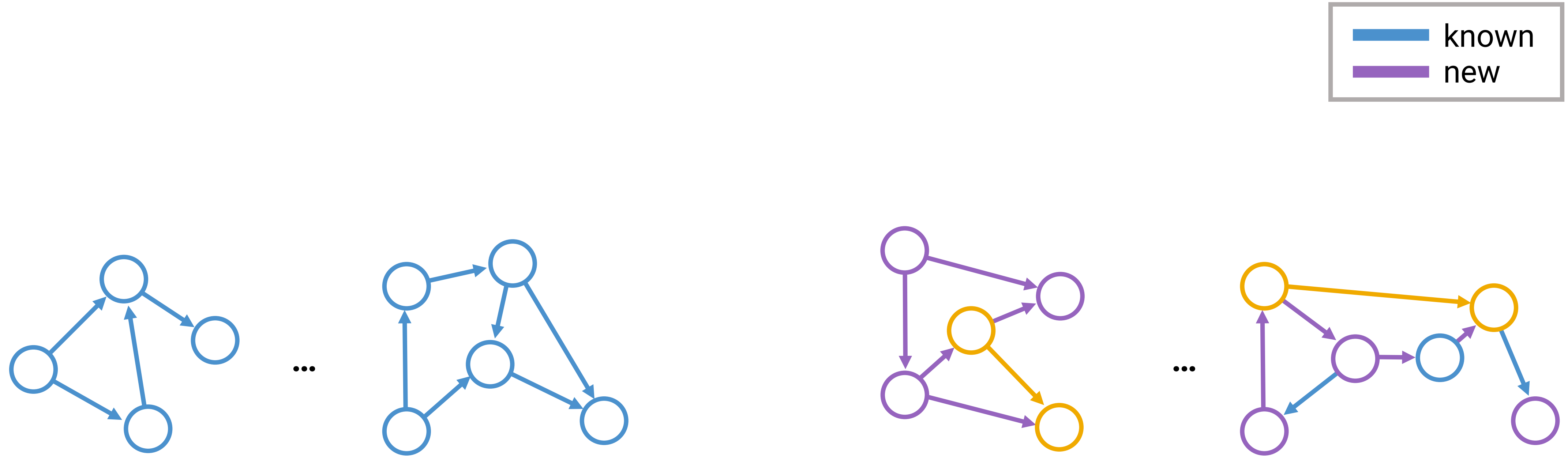


Training Graph



Inference Graph

01 Foundation Models for KG Reasoning



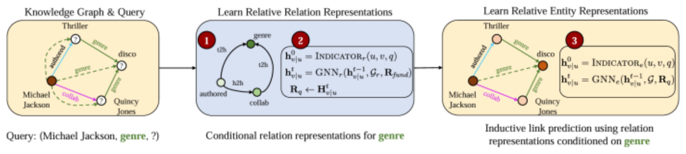
Training Graph #1 Training Graph #N Inference Graph #1 Inference Graph #M

Training Graphs **Inference Graphs**

01 Foundation Models for KG Reasoning

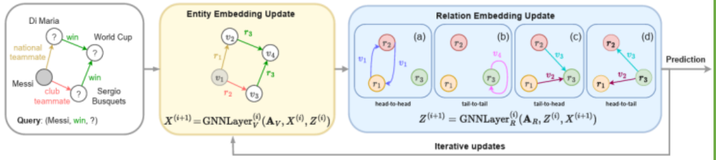
ULTRA (ICLR 2024)

- Proposes an approach for learning universal and transferable graph representations
- ULTRA builds relational representations as a function conditioned on their interactions
  - Can inductively generalize to any unseen KG with any relations



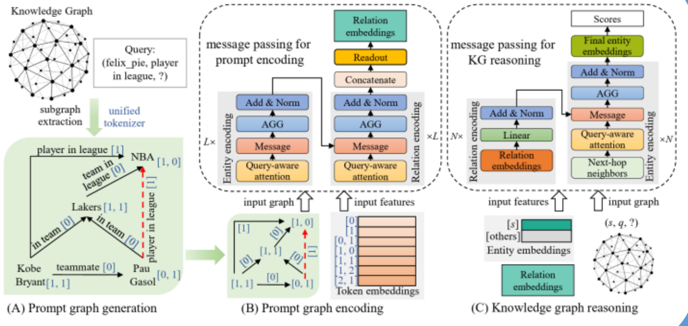
TRIX (LoG 2024)

- Introduces a more expressive and fully inductive model
  - Yields strictly more expressive triplet embeddings compared to the state-of-the-art methods
- TRIX can directly handle both entity and relation prediction tasks in inductive settings



KG-ICL (NeurIPS 2024)

- Proposes a prompt-based KG foundation model via an in-context learning for universal reasoning
- Introduces a prompt graph centered with a query-related example fact
- Proposes two message passing neural networks to perform prompt encoding and KG reasoning

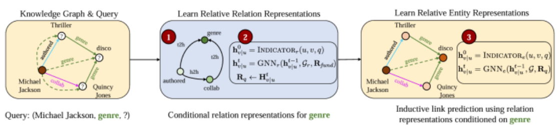


I

Towards Foundation Models for Knowledge Graph Reasoning

Mikhail Galkin*, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu

ICLR 2024

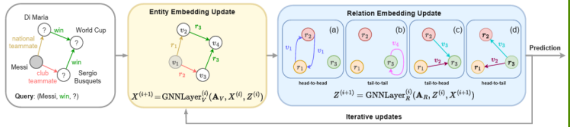


II

TRIX: A More Expressive Model for Zero-shot Domain Transfer in Knowledge Graphs

Yucheng Zhang, Beatrice Bevilacqua, Mikhail Galkin, and Bruno Ribeiro

LoG 2024

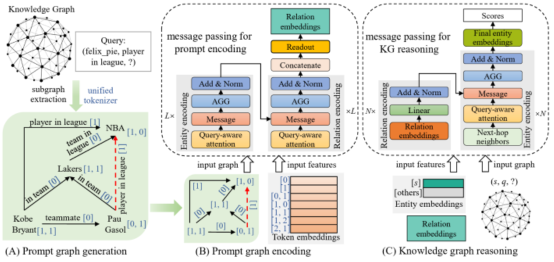


III

A Prompt-Based Knowledge Graph Foundation Model for Universal In-Context Reasoning

Yuanning Cui, Zequn Sun, and Wei Hu*

NeurIPS 2024

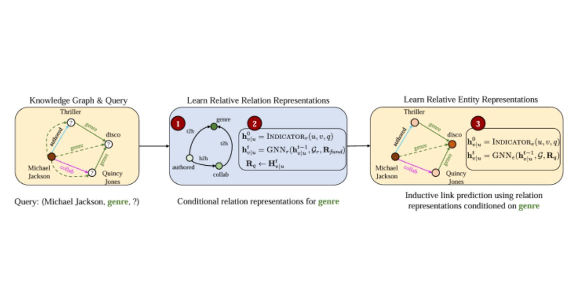


I

Towards Foundation Models for Knowledge Graph Reasoning

Mikhail Galkin*, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu

ICLR 2024

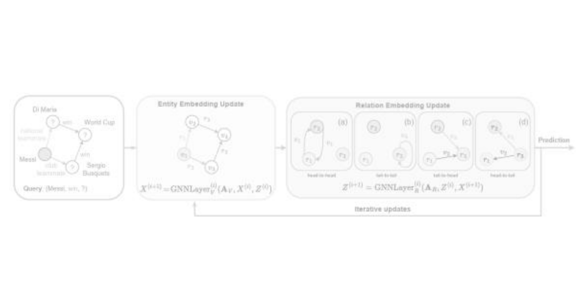


II

TRIX: A More Expressive Model for Zero-shot Domain Transfer in Knowledge Graphs

Yucheng Zhang, Beatrice Bevilacqua, Mikhail Galkin, and Bruno Ribeiro

LoG 2024

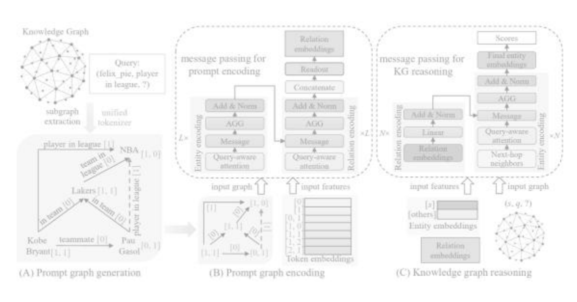


III

A Prompt-Based Knowledge Graph Foundation Model for Universal In-Context Reasoning

Yuanning Cui, Zequn Sun, and Wei Hu*

NeurIPS 2024



02 Motivation

- ML applications increasingly rely on the **pre-training and fine-tuning paradigm**
  - Foundation model refers to a model pre-trained on large datasets in a self-supervised fashion

- Foundation model **leverages certain invariances** pertaining to a domain of interest
  - Large language models like BERT, GPT-4, Llama operate on a fixed vocabulary of tokens

- Representation learning on KGs has **not yet witnessed the benefits of transfer learning** despite a wide range of downstream applications
  - Precision medicine, materials science, virtual assistants, or product graphs in e-commerce

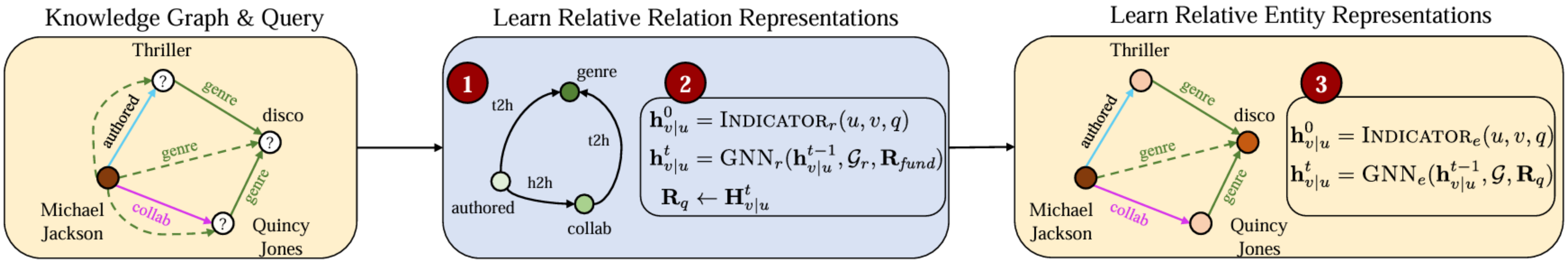
02 Contributions

- Find the **invariances** transferable across graphs with arbitrary entities and relations
  - Leveraging and learning such invariances would enable the pre-train and fine-tune paradigm of foundation models for KG reasoning

- Propose **ULTRA**, method for **u**nified, **l**earnable, and **tra**nsferable KG representations
  - Leverages the invariance of the relational structure and employs relative relation representations on top of this structure for parameterizing any unseen relation
  - Enables zero-shot generalization to any other KG of any size and any relations

- Experimentally, ULTRA generalizes to 50+ different KGs with sizes of 1,000-120,000 nodes and 5K-1M edges

02 Overview of ULTRA



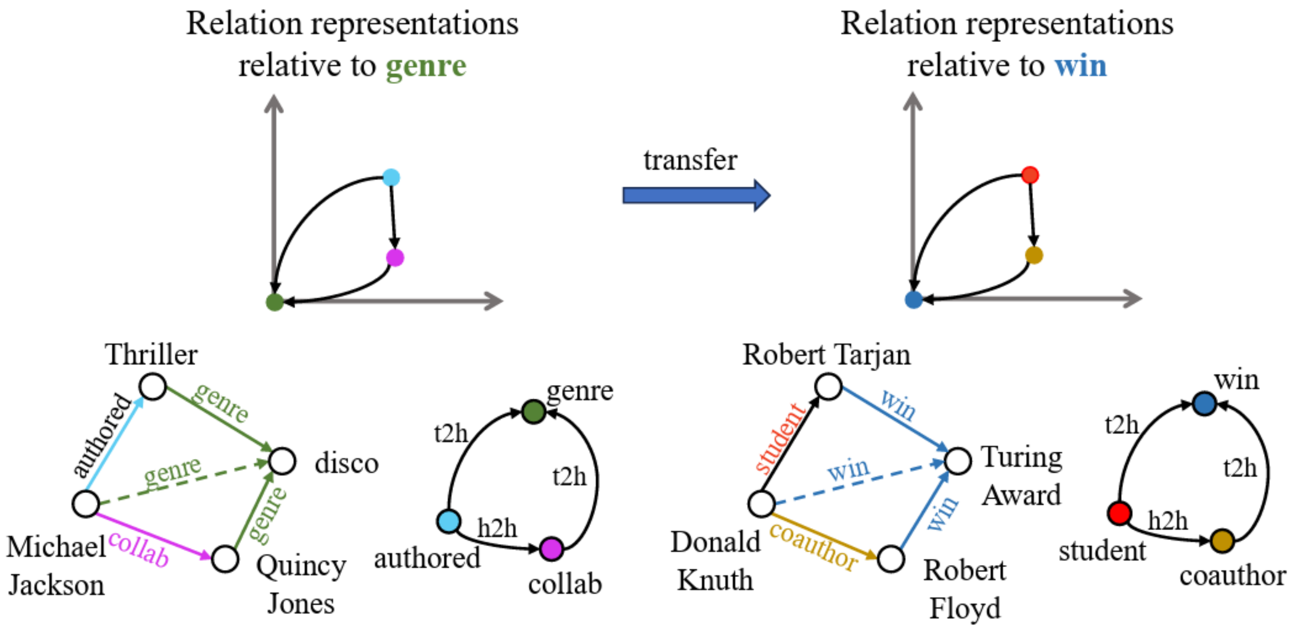
Query: (Michael Jackson, **genre**, ?)

Conditional relation representations for **genre**

Inductive link prediction using relation representations conditioned on **genre**

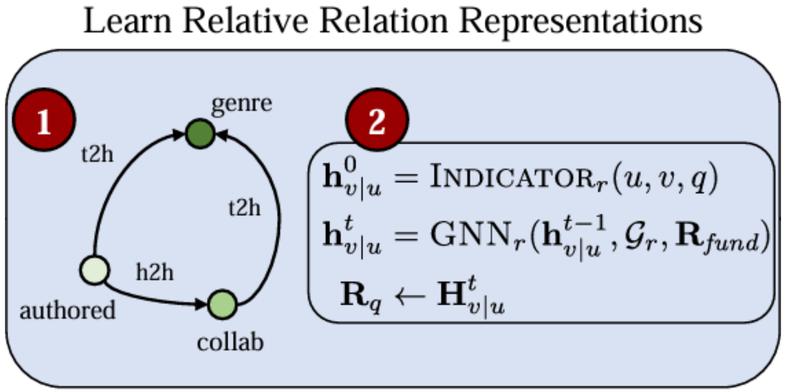
02 Relation Graph Construction

- Generalize KG reasoning to new entities and relations by **leveraging a graph of relations**, where each node corresponds to a distinct relation in the original graph
- Distinguish four **relation-to-relation interactions**
  - Tail-to-head, head-to-head, head-to-tail, and tail-to-tail



02 Conditional Relation Representations

- Given a query $(h, q, ?)$, ULTRA computes the representations of nodes in the relation graph **conditioned on the query relation q**
  - Each node corresponds to a relation in the original KG
  - Use **NBFNet** with a non-parametric DistMult message function and sum aggregation
- Each layer only learns the **embeddings of the four interactions**, a linear layer for representation update, and an optional layer normalization

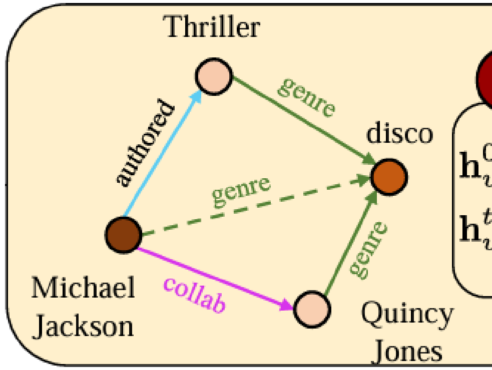


Conditional relation representations for **genre**

02 Entity-level Link Prediction

- Given a query and **conditional relation representations**, any off-the-shelf inductive link predictor that only needs relational features can be utilized
  - Another instance of **NBFNet** is used to account for separate relation representations per query
- After message passing, the final MLP maps the node states to logits denoting the score of a node to be a tail of the initial query

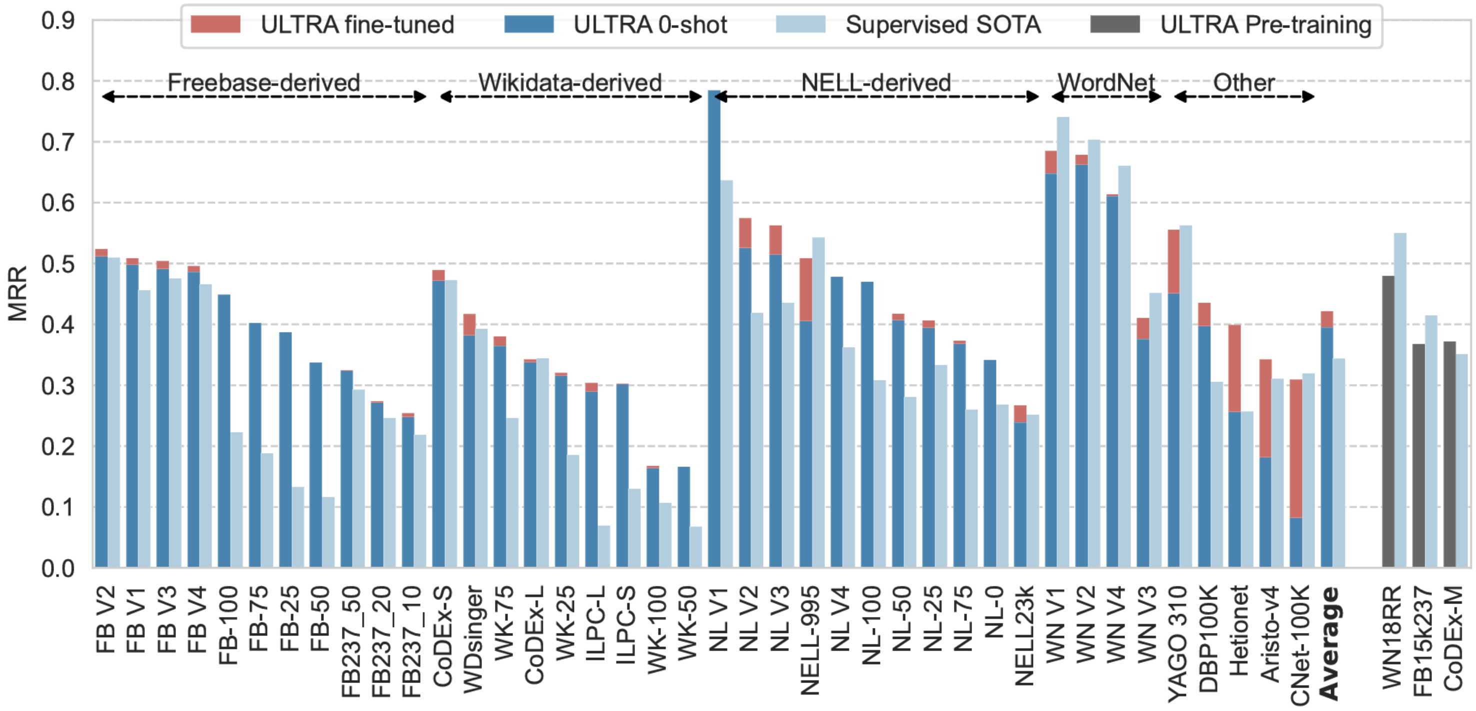
Learn Relative Entity Representations



$$\mathbf{h}^0_{v|u} = \textsc{Indicator}_e(u, v, q)$$
$$\mathbf{h}^t_{v|u} = \mathrm{GNN}_e(\mathbf{h}^{t-1}_{v|u}, \mathcal{G}, \mathbf{R}_q)$$

Inductive link prediction using relation representations conditioned on **genre**

Experiments



Model	Inductive $(e) + (e, r)$ (27 graphs)		Transductive e (13 graphs)		Total Avg (40 graphs)		Pretraining (3 graphs)		Inductive $(e) + (e, r)$ (8 graphs)
	MRR	**H@10**	**MRR**	**H@10**	**MRR**	**H@10**	**MRR**	**H@10**	**Hits@10 (50 negs)**
Supervised SOTA	0.342	0.482	0.348	0.494	0.344	0.486	**0.439**	**0.585**	0.731
ULTRA 0-shot	0.435	0.603	0.312	0.458	0.395	0.556	-	-	0.859
ULTRA fine-tuned	**0.443**	**0.615**	**0.379**	**0.543**	**0.422**	**0.592**	0.407	0.568	**0.896**

02 Conclusion

- Present **ULTRA**, an approach to learn universal and transferable graph representations that can serve as one of the methods towards building foundation models for KG reasoning
  - ULTRA enables training and inference on any multi-relational graph without any input features leveraging the invariance of the relational structure and conditional relation representations

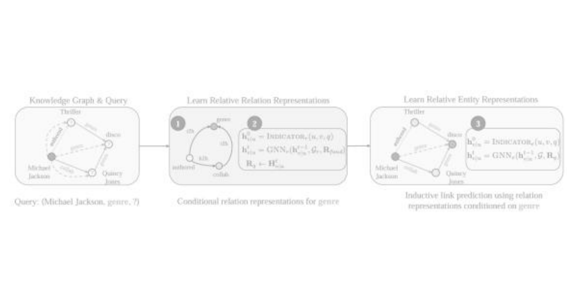
- Experimentally, a single pre-trained ULTRA model **outperforms state-of-the-art tailored supervised baselines** on 50+ graphs of 1k-120k nodes even in the zero-shot regime by average 15%

I

Towards Foundation Models for Knowledge Graph Reasoning

Mikhail Galkin*, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu

ICLR 2024

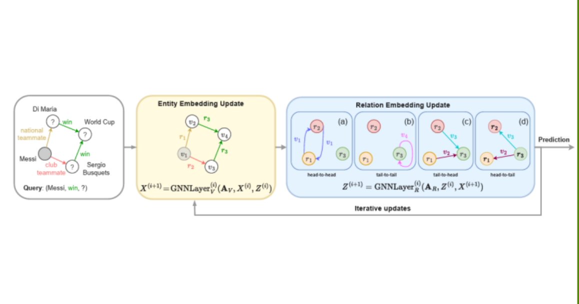


II

TRIX: A More Expressive Model for Zero-shot Domain Transfer in Knowledge Graphs

Yucheng Zhang, Beatrice Bevilacqua, Mikhail Galkin, and Bruno Ribeiro

LoG 2024

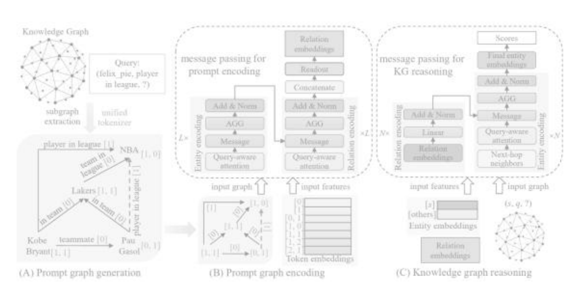


III

A Prompt-Based Knowledge Graph Foundation Model for Universal In-Context Reasoning

Yuanning Cui, Zequn Sun, and Wei Hu*

NeurIPS 2024



03 Motivation

- Challenge #1: **Limited expressivity of existing methods**
  - Existing state-of-the-art foundation models such as ULTRA have expressivity limitations
- Challenge #2: **Insufficient support for relation prediction tasks**
  - Existing foundation models are primarily designed for entity prediction tasks
- Challenge #3: **Underexploration of the abilities of LLMs to perform the same tasks**
  - Large-context LLMs have demonstrated notable performance in KGC tasks, but their effectiveness in the inductive setting, where test KGs come from new domain, remains largely underexplored

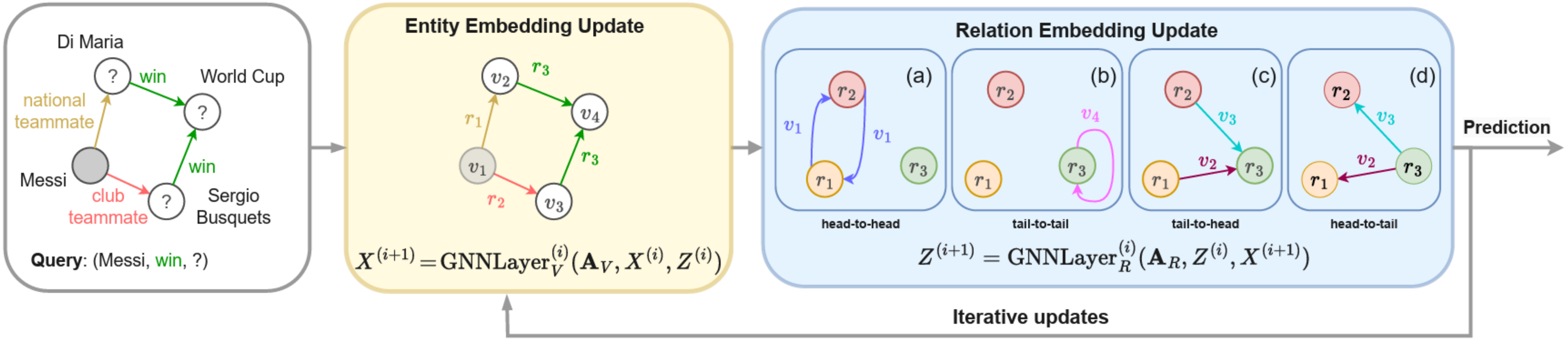
03 Contributions

- Show that the limited expressive power of state-of-the-art foundation model ULTRA arises from its approach in **capturing relation interactions rather than which entities share those relations**

- Propose **TRIX**, **T**ransferable **R**elation-Entity **I**nteractions in crossing patterns (**X**-patterns), which returns more expressive triplet representations than ULTRA

- Demonstrate that while LLMs can do KGC accurately given enough context about the background knowledge, they **rely on textual information**, and therefore fail to utilize the actual graph information, given in the context

03 Overview of TRIX



03 Relation Adjacency Matrix

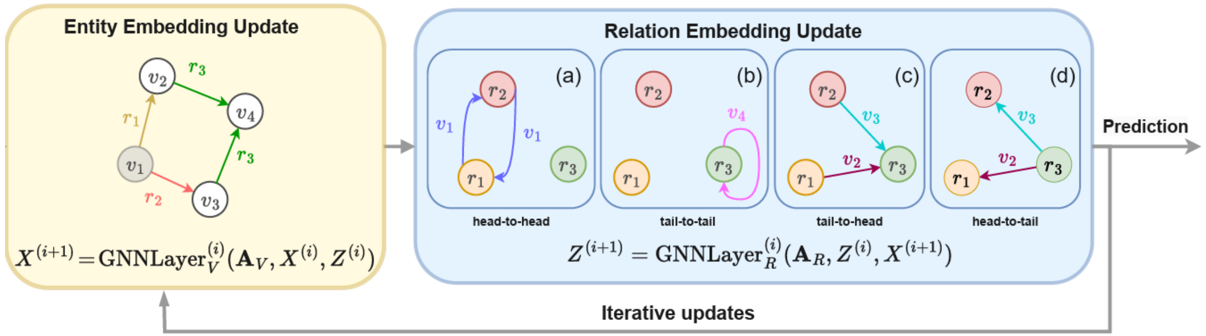
- Increase the expressiveness of relation graphs by **including entity information**
  - Edges between two relations contain information of which entities share these relations

- For each pair of relations (r_1, r_2), count how many times an entity v is part of triplets involving these two relations as:
  - The head entity in both, the tail entity in both, the head in r_1 and the tail in r_2, and the tail in r_1 and the head in r_2
  - Label of edge in the relation graph corresponds to an entity in the original KG

03 Iterative Entity and Relation Embedding Updates

- At each layer, TRIX **iteratively** performs **message passing on the original KG** to update the entity representations and **message passing on the relation graph** to update the relation representations
  - For entity prediction, the message passing is first performed on the original KG
  - For relation prediction, the message passing is first performed on the relation graph

- The message passing is performed using **two separate NBFNets**



03 Experiments

Model	Inductive e, r (23 graphs)		Inductive e (18 graphs)		Transductive (13 graphs)		Total Avg (54 graphs)		Pretraining (3 graphs)	
	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
ULTRA zero-shot	0.345	0.513	0.431	0.566	0.312	0.458	0.366	0.518	N/A	N/A
TRIX zero-shot	**0.368**	**0.540**	**0.455**	**0.592**	**0.339**	**0.500**	**0.390**	**0.548**	N/A	N/A
ULTRA fine-tuned	0.397	**0.556**	0.442	0.582	0.379	0.543	0.408	0.562	0.407	**0.568**
TRIX fine-tuned	**0.401**	**0.556**	**0.459**	**0.594**	**0.390**	**0.558**	**0.418**	**0.569**	**0.415**	0.563

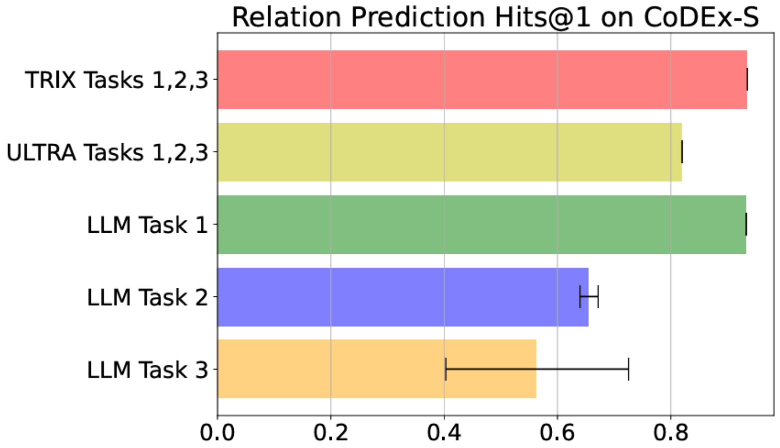
Entity Prediction Results

Model	Inductive e, r (23 graphs)		Inductive e (18 graphs)		Transductive (13 graphs)		Total Avg (54 graphs)		Pretraining (3 graphs)	
	MRR	H@1	MRR	H@1	MRR	H@1	MRR	H@1	MRR	H@1
ULTRA zero-shot	0.785	0.691	0.714	0.590	0.629	0.507	0.724	0.613	N/A	N/A
TRIX zero-shot	**0.842**	**0.770**	**0.756**	**0.611**	**0.752**	**0.647**	**0.792**	**0.687**	N/A	N/A
ULTRA fine-tuned	0.823	0.741	0.716	0.591	0.707	0.608	0.759	0.659	0.876	**0.817**
TRIX fine-tuned	**0.850**	**0.785**	**0.759**	**0.615**	**0.785**	**0.693**	**0.804**	**0.706**	**0.879**	0.797

Relation Prediction Results

03 Link Prediction with LLMs

- All triplets of the KG are provided in the prompt



- Task 1: In-domain LLM predictions
  - Entities and relations are expressed by their names in natural language
- Task 2: Out-of-domain LLM predictions
  - The neighbor entities of the head entity and the relations that connect the head entity with its neighbors are replaced with metasyntactic words to simulate information from a new domain
- Task 3: Double-equivariant LLM predictions
  - All the entities and relations are expressed with IDs, and the mappings from entities and relations to their IDs are given to the LLM and the IDs are shuffled in each test run

03 Conclusion

- Propose **TRIX**, an architecture designed to improve expressiveness and support efficient relation prediction tasks

- Demonstrate that increased expressiveness translates into better performance
  - 57 KG datasets were used in the experiments

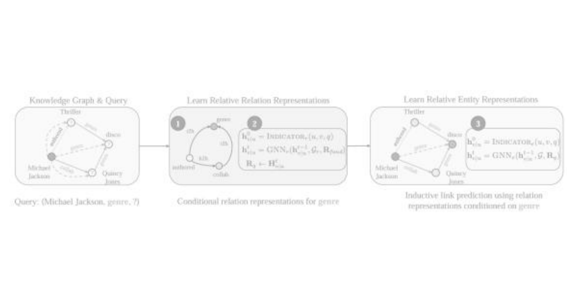
- Shed light on the **limitations of LLMs In exploiting graph information** in new domains for entity and relation prediction tasks

I

Towards Foundation Models for Knowledge Graph Reasoning

Mikhail Galkin*, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu

ICLR 2024

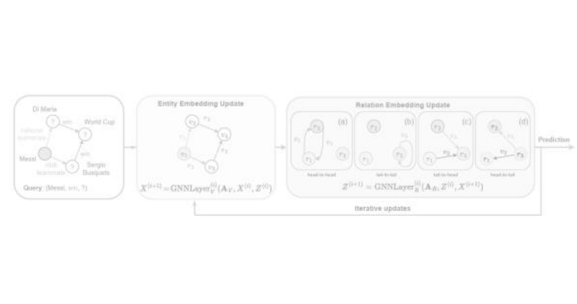


II

TRIX: A More Expressive Model for Zero-shot Domain Transfer in Knowledge Graphs

Yucheng Zhang, Beatrice Bevilacqua, Mikhail Galkin, and Bruno Ribeiro

LoG 2024

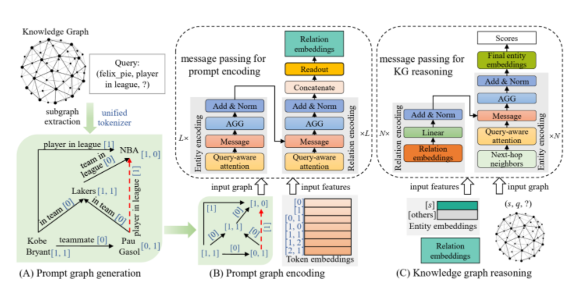


III

A Prompt-Based Knowledge Graph Foundation Model for Universal In-Context Reasoning

Yuanning Cui, Zequn Sun, and Wei Hu*

NeurIPS 2024



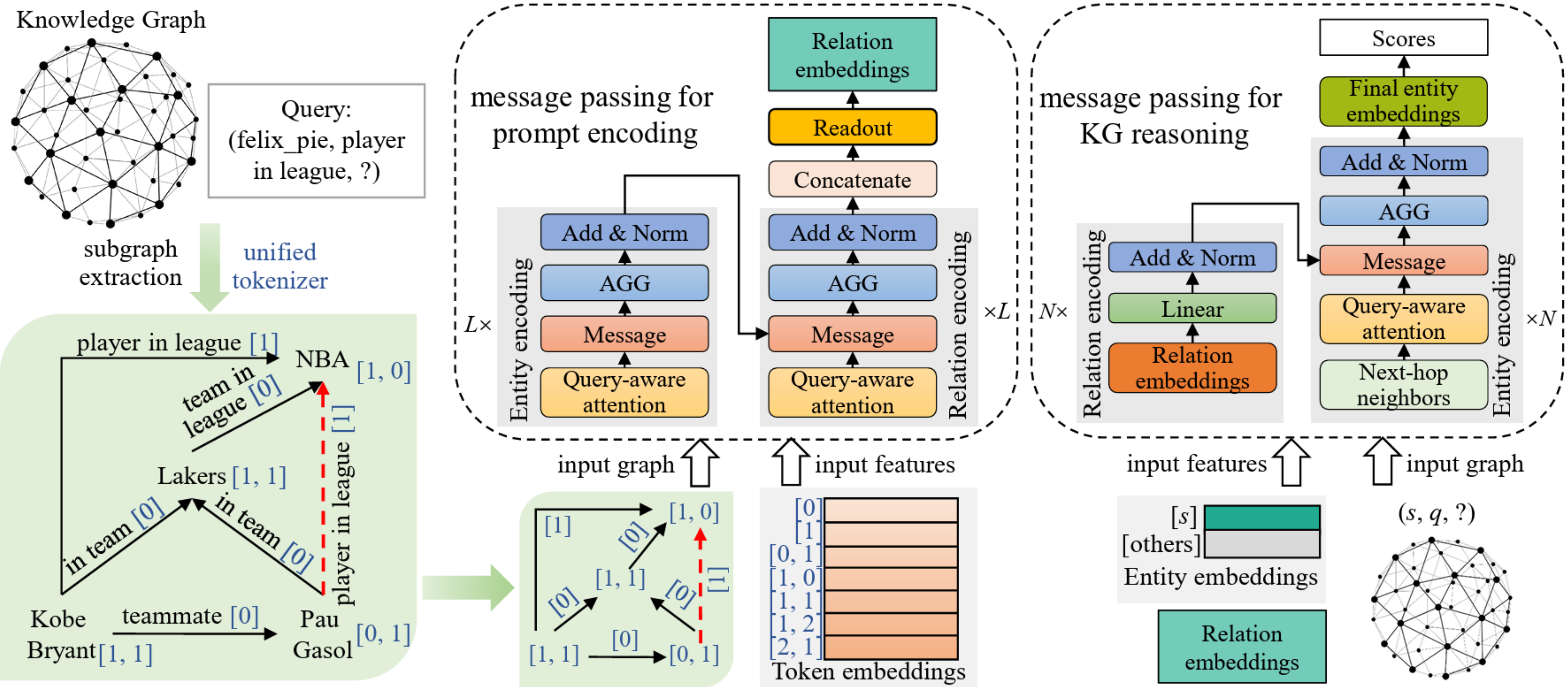
04 Motivation

- Existing Foundation models for KG reasoning model relation interactions using a query-conditioned relation graph
- The relation graph **only describes the connectivity of relations** in the KG, with less attention to the local context of the entity and relation in a query
  - As a result, these methods usually fail to generate discriminative relation representations
- Need to capture the local contexts and **highlight the important relations relevant to queries**, rather than relying on a global relation graph

04 Contributions

- Propose **KG-ICL**, a KG reasoning foundation model with in-context learning
  - Prompts the per-trained model to engage in relational reasoning over diverse KGs
- Propose a **prompt graph as context** to support in-context learning
  - A prompt graph consists of an example fact about the query relation and its relevant subgraphs and paths
  - A unified tokenizer maps entities and relations in prompt graphs to predefined tokens
- Propose two message passing networks for **prompt graph encoding** and **KG reasoning**

04 Overview of KG-ICL

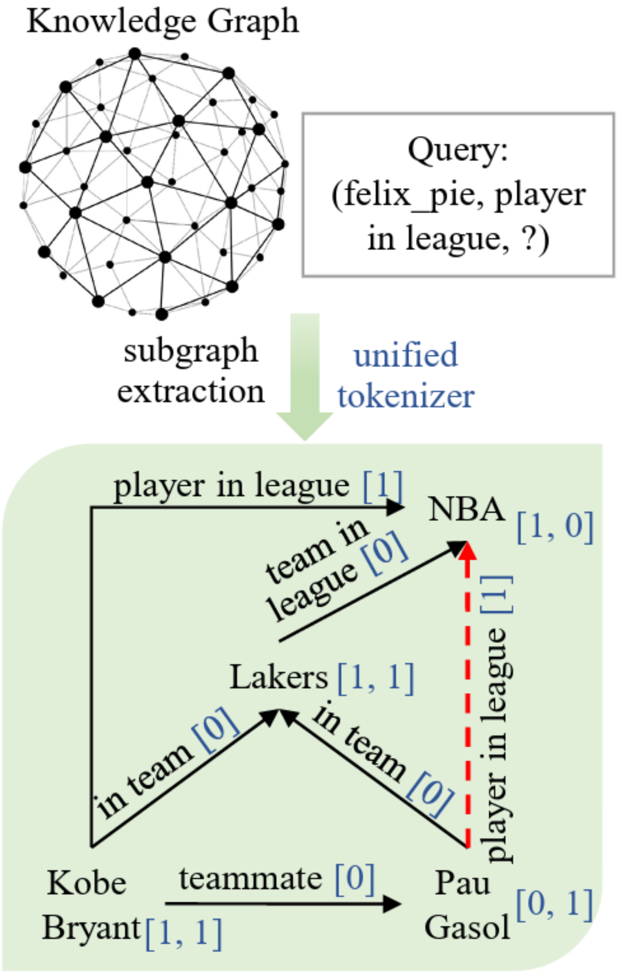


(A) Prompt graph generation (B) Prompt graph encoding (C) Knowledge graph reasoning

04 Prompt Graph Generation

- For a query relation q, **randomly sample M example facts** from the original graph

- A prompt graph of an example fact (h, q, t) is **generated using paths between h and t**
  - Include one-hop neighbors of the head and tail entities
  - Include all entities in a k-hop path between the head and tail entities
  - Extracts an induced subgraph from the original KG



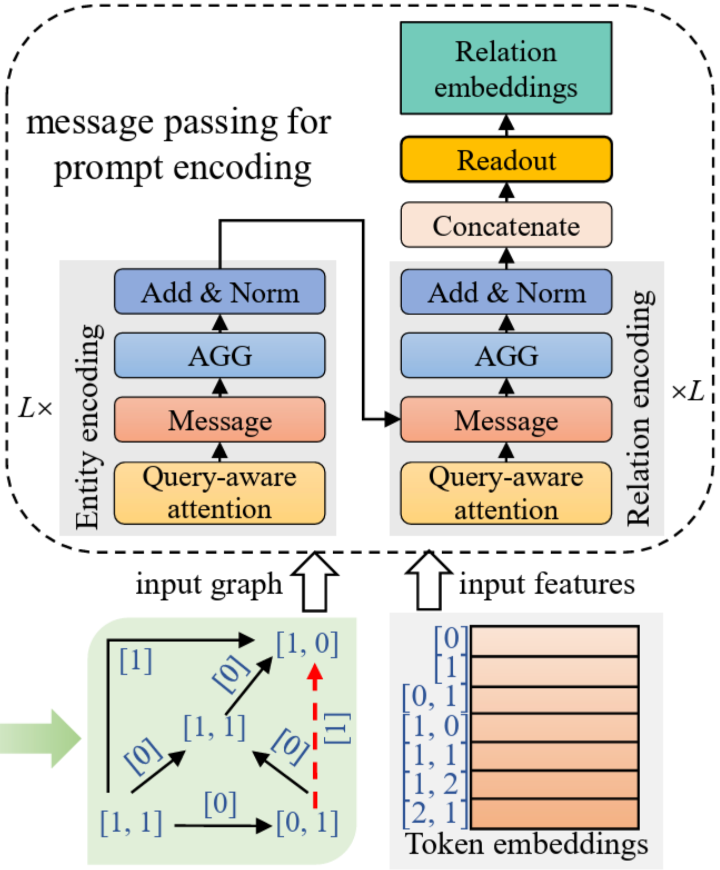
04 Prompt Encoding

- The entities and relations in prompt graphs are **mapped to the predefined tokens**
  - Each entity is mapped based on the length of its shortest paths to h and t
  - Each relation is mapped by whether it is the same as the query relation

- Each token is assigned a learnable vector representation

- **Iteratively updates** the entity representations and the relation representations using message passing

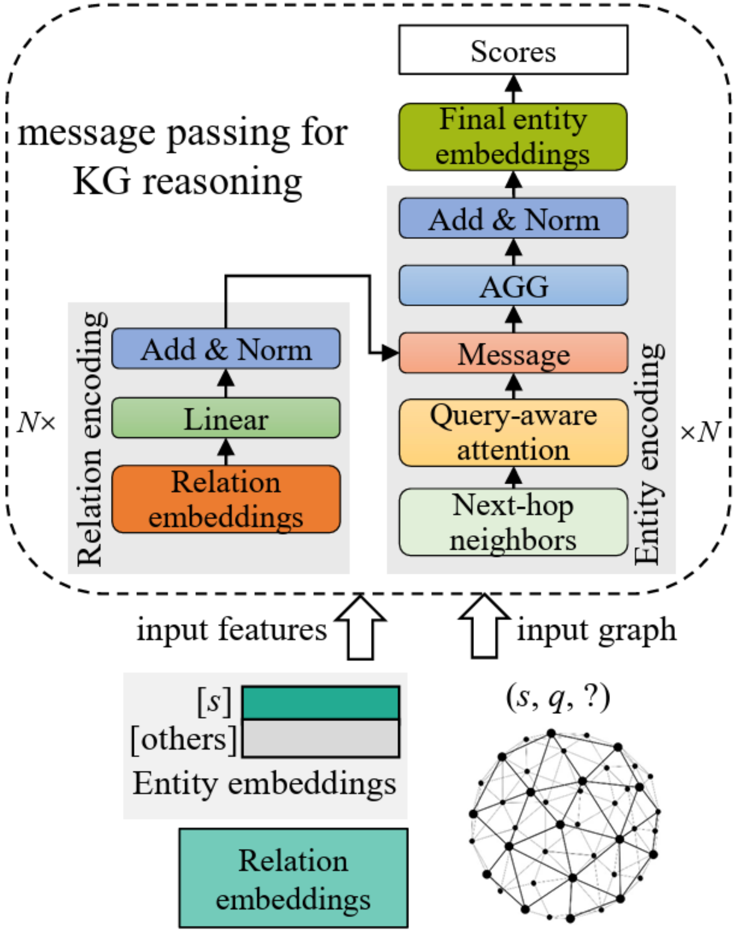
- Only uses the relation representations in the subsequent module



(B) Prompt graph encoding

04 In-Context KG Encoding and Reasoning

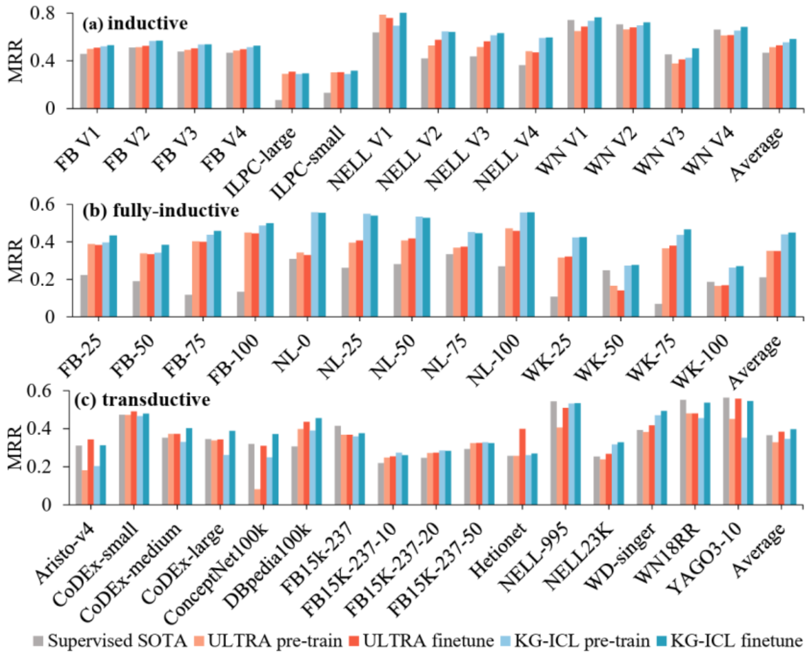
- Draw inspiration from the **conditional message passing neural network** to achieve a KG-independent encoding
  - Separately encodes entities based on the query
- Given a query fact (s, q, x), initialize **the representation of s as the representation of query relation q**
  - The relation representations are set to the representations computed by prompt encoding
- Use a message passing neural network to aggregate multi-hop information for updating entity representations



(C) Knowledge graph reasoning

04 Experiments

Models	Inductive (14 KGs)		Fully-inductive (13 KGs)		Transductive (16 KGs)		Average (43 KGs)	
	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
Supervised SOTA	0.466	0.607	0.210	0.347	0.365	0.511	0.351	0.493
ULTRA pre-train	0.513	0.664	0.352	0.536	0.329	0.479	0.396	0.557
ULTRA finetune	0.528	0.684	0.350	0.542	0.384	0.548	0.421	0.590
KG-ICL pre-train	0.554	0.707	0.439	0.635	0.346	0.493	0.442	0.606
KG-ICL finetune	**0.582**	**0.727**	**0.449**	**0.647**	**0.397**	**0.554**	**0.473**	**0.638**



04 Conclusion

- Introduce **KG-ICL**, a KG foundation model with in-context learning to improve the effectiveness and transferability of KG reasoning
  - Introduce **a prompt graph and a unified tokenizer** as the bridge to knowledge transfer between different KGs

- Propose a prompt graph generation module, a prompt encoding module, and a KG reasoning module to **achieve in-context learning**

- Evaluate the in-context reasoning ability on 43 different KGs in both transductive and inductive settings

05 References

- Some slides are made based on the following references.

  - M. Galkin et al., "Towards Foundation Models for Knowledge Graph Reasoning", ICLR, 2024.
  - Y. Zhang et al., "TRIX: A More Expressive Model for Zero-shot Domain Transfer in Knowledge Graphs", LoG, 2024.
  - Y. Cui et al., "A Prompt-Based Knowledge Graph Foundation Model for Universal In-Context Reasoning", NeurIPS, 2024.