

# THOR: Inductive Link Prediction over Hyper-Relational Knowledge Graphs

Weijian Yu

University of Macau  
Macau, China  
yc47946@um.edu.mo

Yuhuan Lu

Khalifa University  
Abu Dhabi, UAE  
yuhuan.lu@ku.ac.ae

Dingqi Yang\*

University of Macau  
Macau, China

## Abstract

Knowledge graphs (KGs) have become a key ingredient supporting a variety of applications. Beyond the traditional triplet representation of facts where a relation connects two entities, modern KGs observe an increasing number of hyper-relational facts, where an arbitrary number of qualifiers associated with a triplet provide auxiliary information to further describe the rich semantics of the triplet, which can effectively boost the reasoning performance in link prediction tasks. However, existing link prediction techniques over such hyper-relational KGs (HKGs) mostly focus on a transductive setting, where KG embedding models are learned from the specific vocabulary of a given KG and subsequently can only make predictions within the same vocabulary, limiting their generalizability to previously unseen vocabularies. Against this background, we propose THOR, an inducTive link prediction technique for Hyper-relational knOwledge gRaphs. Specifically, we first introduce both relation and entity foundation graphs, modeling their fundamental inter- and intra-fact interactions in HKGs, which are agnostic to any specific relations and entities. Afterward, THOR is designed to learn from the two foundation graphs with two parallel graph encoders followed by a transformer decoder, which supports efficient masked training and fully-inductive inference. We conduct a thorough evaluation of THOR in hyper-relational link prediction tasks on 12 datasets with different settings. Results show that THOR outperforms a sizable collection of baselines, yielding 66.1%, 55.9%, and 20.4% improvement over the best-performing rule-based, semi-inductive, and fully-inductive techniques, respectively. A series of ablation studies also reveals our key design factors capturing the structural invariance transferable across HKGs for inductive tasks.

## CCS Concepts

- Computing methodologies → Knowledge representation and reasoning.

## Keywords

Hyper-relation; Knowledge graph; Inductive link prediction

## 1 Introduction

Knowledge graphs (KGs) have been widely used to support various real-world applications ranging from question answering to recommendation systems [16, 36]. In a typical KG, factual knowledge is traditionally represented as a set of triplets and each triplet  $(h, r, t)$  consists of a head entity  $h$ , a tail entity  $t$ , and a relation  $r$  connecting them, such as *(Albert Einstein, educated at, ETH Zurich)*. To efficiently reason over such KGs, link prediction tasks are widely

adopted to predict a missing entity in a query fact  $(h, r, [MASK])$ , where the  $[MASK]$  token represents the missing entity. However, it has been widely revealed that the traditional triplet-based KGs oversimplify the real-world facts, which are often more complex beyond the triplet structure [24]. Specifically, modern KGs contain an increasing number of hyper-relational facts, where a primary triplet  $(h, r, t)$  is associated with several qualifiers, represented as key-value (relation-entity) pairs  $(k, v)$ , providing auxiliary information to further describe the rich semantics of the primary triplet. For example, *[Albert Einstein, educated at, ETH Zurich), (academic degree, Bachelor of Science), (academic major, mathematics education)]* further indicates the academic degree and academic major beyond the aforementioned triple fact. Such qualifiers are very useful in boosting link prediction performance over KGs, by providing additional clues for predicting the missing entity [24].

Despite the recent advancement in link prediction over hyper-relational KGs (HKG) [6, 11, 12, 21, 24, 37], existing techniques mostly consider a transductive setting, where models learn entity and relation embeddings from the specific vocabulary of a given KG to capture the structural and semantic properties, and subsequently can only make prediction within the same vocabulary, failing to generalize to unseen entities and relations. However, the dynamics of modern KGs imply continuously emerging entities and relations over time, while these transductive link prediction techniques thus show limited generalizability by failing to generalize to previously unseen vocabularies. In this context, a few existing works design inductive link prediction techniques under either a *semi-inductive setting* with a consistent relation vocabulary but arbitrary (even disjoint) entity vocabularies [41], or a more generalized *fully-inductive setting* with both arbitrary (even disjoint) relation and entity vocabularies [7]. Different from the transductive link prediction techniques that rely on specific embeddings of entities and relations in a KG, the inductive techniques need to learn the structural invariance transferable across KGs and can thus generalize to arbitrary entity and relation vocabularies.

In the current literature, existing inductive link prediction techniques face the following key challenges when handling HKGs. First, these techniques mainly consider triple facts by designing a relation graph modeling the fundamental interactions (edges) between relations, such as head-to-tail or tail-to-head interactions [7, 10] or membership of certain graph motifs [14]. As the edge of this relation graph is agnostic about any specific relations, models learnt from this graph thus support inductive inference. However, this relation graph overlooks the sophistication of the HKGs with qualifiers, which significantly enrich yet also complicate the interactions between entities and relations beyond the triple facts,

\*Corresponding author

as the qualifiers associated with the primary triplet are also represented as relation-entity pairs. Second, existing approaches mostly design graph encoders to learn structural invariance from the relation graph by contrasting a training triple fact against randomly corrupted negative samples (randomly corrupting one of the two entities in the triple fact) [7, 8, 10, 17, 47, 52]. However, lessons learnt from transductive techniques [44] revealed that such a random corruption negative sampling scheme is significantly inefficient for HKGs due to the large corruption space, where the corruption space increases exponentially with the number of elements in a hyper-relational fact.

To address the above challenges, we propose THOR, a fully-inductive link prediction technique for Hyper-relational knOwledge gRaphs. First, we design two foundation graphs capturing both fundamental interactions between relations and entities in an HKG, respectively. Since both foundation graphs model the position-wise interactions of elements in the HKG, they are agnostic to any specific relations and entities, thus supporting fully-inductive inference. Second, for inductive learning from the relation and entity foundation graphs, we first design two parallel graph encoders based on Neural Bellman-Ford Networks (NBFNet) [52], which is a path-based graph neural architecture for generic inductive link prediction, followed by a transformer decoder, fully supporting masked training without the need for the random corruption negative sampling process. Our contributions are as follows:

- We introduce both relation and entity foundation graphs for HKGs, which model their fundamental inter- and intra-fact interactions and are agnostic to any specific relations and entities, thus being capable of supporting fully-inductive inference.
- We propose an inductive link prediction technique THOR, learning from the two foundation graphs with two parallel NBFNet-based graph encoders, followed by a transformer decoder, fully supporting efficient masked training without the need for the expensive random corruption negative sampling process.
- We conduct a thorough evaluation of THOR in different settings on 12 datasets from different domains. Results show that THOR outperforms a sizable collection of 16 baselines in different link prediction tasks, yielding 66.1%, 55.9%, and 20.4% improvement over the best-performing rule-based, semi-inductive, and fully-inductive techniques, respectively. In particular, in cross-domain fully-inductive tasks, it outperforms the best baselines by 32.5%.
- A series of ablation studies further verify our design choices of THOR, revealing key design factors capturing the structural invariance transferable across HKG: 1) qualifiers are indeed useful for inductive settings, but defining their fundamental interactions is tricky; we found that using relation-key interactions and exhaustive entity interactions is most effective; 2) avoiding negative samples is also very useful, yielding 57.9% improvement.

## 2 Preliminary

We first define our key terminology, followed by problem settings.

**Hyper-relational Knowledge Graph (HKG).** Given an entity set  $E$  and a relation set  $R$ , a hyper-relational KG  $\mathcal{G} = (E, R, F)$  consists of a set of fact  $F = \left\{ \left[ (h, r, t), \{(k_i, v_i)\}_{i=1}^n \right] \mid h, t, v_i \in E, r, k_i \in R \right\}$ , where  $(h, r, t)$  refers to the primary triplet of this fact, and  $\{(k_i, v_i)\}_{i=1}^n$

refer to  $n$  key-value pairs providing additional information about the primary triplet.

**Link prediction over HKGs.** Given a query hyper-relational fact with a missing entity represented by a [MASK] token, the link prediction task aims at predicting the masked entity. In our setting, the masked entity can be at any position in a fact, i.e.,  $h$ ,  $t$ , or  $v$ .

**Transductive link prediction task.** In the transductive setting, models learn from a training KG  $\mathcal{G}_{train} = (E_{train}, R_{train}, F_{train})$ , and make predictions on query facts from an inference KG  $\mathcal{G}_{inf} = (E_{inf}, R_{inf}, F_{inf})$ , where the inference KG are defined on the same entity and relation vocabularies as the training KG, i.e.,  $E_{inf} \subseteq E_{train}$  and  $R_{inf} \subseteq R_{train}$ .

**Inductive link prediction task.** In the inductive setting, the training KG and the inference KG do not share the same entity and relation vocabularies. According to their shared elements, there are mainly two settings of inductive link prediction tasks<sup>1</sup>.

- **Semi-inductive link prediction task.** The semi-inductive setting assumes the same relation vocabulary for the training and inference KGs  $R_{train} = R_{inf}$ , but different entity vocabularies between the training and inference KGs. The entity vocabulary in the inference graph contains unseen entities in the training stage, where the inference entity set could be an extension of the training entity set  $E_{train} \subset E_{inf}$ , or they can be completely disjoint  $E_{train} \cap E_{inf} = \emptyset$ .
- **Fully-inductive link prediction task.** The fully-inductive setting assumes both different relation and entity vocabularies between the training and inference KGs, where both relation and entity vocabularies in the inference graph contain unseen relations and entities in the training stage, respectively. In the hardest setting, both relation and entity vocabularies are disjoint between the training and inference KGs, i.e.,  $R_{train} \cap R_{inf} = \emptyset$  and  $E_{train} \cap E_{inf} = \emptyset$ .

In this work, we focus on the (hardest) fully-inductive case with both disjoint relation and entity vocabularies between the training and inference KGs. Note that since the harder inductive setting is strictly a superset of the easier inductive (or transductive) settings, fully-inductive link prediction techniques are thus by design applicable in all other settings.

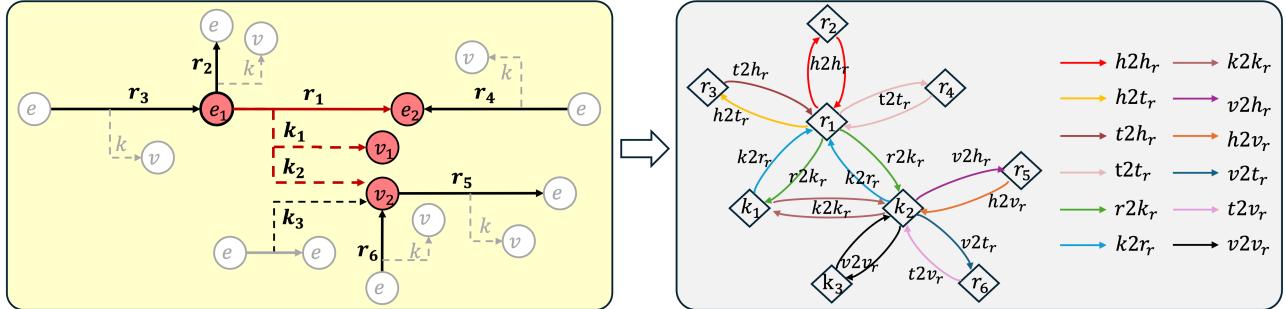
## 3 Method

In this section, we introduce THOR, a fully-inductive link prediction technique for Hyper-relational knOwledge gRaphs. In the following, we first introduce the relation and entity foundation graphs, and then present THOR learning from both foundation graphs for fully-inductive link prediction over HKGs.

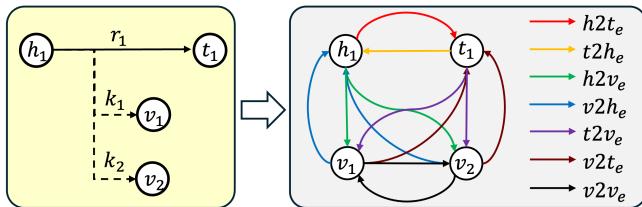
### 3.1 Relation and Entity Foundation Graphs

Our relation and entity foundation graphs are designed to capture the fundamental interactions between elements in an HKG, to support the fully-inductive link prediction. To this end, we consider *position-wise interactions* as fundamental interactions, which are agnostic to specific entities and relations.

<sup>1</sup>We note that a few works adopt different definitions of inductive settings; please see Appendix A for details.



**Figure 1: Relation foundation graph with different fundamental interactions.** The plot focuses on the relations of the fact highlighted in red, illustrating all the fundamental interactions involving its relations  $r_1$ ,  $k_1$ , and  $k_2$  only. Interactions among other relations also exist (e.g.,  $h2t_r$  and  $t2h_r$  interactions between  $r_2$  and  $r_3$ ), but are not shown for the clarity of visualization.



**Figure 2: Entity foundation graph with different types of fundamental intra-fact interactions.**

**Relation Foundation Graph.** Inspired by [7, 38], we design a relation foundation graph  $\mathcal{G}_{fund}^r$  modeling the fundamental interaction between relations, where two relations are connected according to their semantic positions in the facts. First, we consider  $\{h2h_r, h2t_r, t2h_r, t2t_r\}$  to depict the relation interactions anchored with primary triplets across different facts. For example, Figure 1 shows when two facts share the same head entity, their primary relations  $r_1$  and  $r_2$  are connected by two reciprocal edges  $h2h_r$ , indicating their head-to-head interactions while the subscript  $r$  implies the edge belongs to the relation foundation graph. Second, the qualifiers (represented as k-v pairs) in a hyper-relational fact significantly enrich yet also complicate the relation interactions. Considering the fact that the k-v pairs serve as auxiliary descriptions for the primary triplet, we define  $\{r2k_r, k2r_r\}$  to depict the interaction between the primary relation and the keys within a fact, as shown in Figure 1. Meanwhile, we also consider the interaction between keys  $k2k_r$  within a fact. Third, we further consider fundamental interactions connecting a key to the relations of other facts via its value entity  $v$ , i.e.,  $\{h2v_r, v2h_r, t2v_r, v2t_r, v2v_r\}$ , as shown in Figure 1. For example, the primary relation  $r_5$  and the key relation  $k_2$  are connected by two reciprocal edges  $h2v_r$  and  $v2h_r$ , indicating their head-to-value and value-to-head interactions, respectively.

**Entity Foundation Graph.** We design an entity foundation graph  $\mathcal{G}_{fund}^e$  modeling the intra-fact interactions within a hyper-relational fact. Specifically, we first define entity fundamental interactions  $\{h2t_e, t2h_e\}$  for the primary triplet, indicating tail-to-tail and tail-to-head interactions, respectively, while the subscript  $e$  implies the edge belongs to the entity foundation graph, as shown in Figure 2. Second, we also consider the interactions involving the qualifiers  $\{h2v_e, v2h_e, t2v_e, v2t_e, v2v_e\}$ . For example,  $h_1$  and  $v_1$  are

connected by two reciprocal edges  $h2v_e$  and  $v2h_e$ , indicating their head-to-value and value-to-head interactions, respectively.

In summary, the relation and entity foundation graphs model the fundamental interactions between relations and entities in an HKG, respectively. These position-wise interactions are agnostic to any specific relations and entities, capturing the structural invariance of HKGs and thus being capable of supporting fully-inductive link prediction. In our experiments later, we conducted a thorough ablation study to systematically evaluate the usefulness of different types of fundamental interactions, and subsequently select the optimal set of relation fundamental interactions  $\mathcal{R}_{fund}^r = \{h2h_r, h2t_r, t2h_r, t2t_r, r2k_r, k2r_r\}$  and entity fundamental interactions  $\mathcal{R}_{fund}^e = \{h2t_e, t2h_e, h2v_e, v2h_e, t2v_e, v2t_e, v2v_e\}$ .

### 3.2 THOR

Based on both relation and entity foundation graphs, we design THOR with two parallel graph encoders based on neural Bellman-Ford Networks (NBFNet) [52], followed by a transformer decoder, which supports efficient masked training and fully-inductive inference. Figure 3 shows the architectural overview of THOR, which consists of three components.

**Relation and Entity Foundation Graph Encoders.** We adopt NBFNet [52] as graph encoders for both relation and entity foundation graphs. Specifically, different from traditional GNNs that encode node features under the graph automorphism property, where different but automorphic nodes receive the same unconditional feature in message passing (i.e., neighborhood symmetry), NBFNet uses a labeling trick [45] for relational graphs that provides conditional message passing based on given query nodes. It uses an indicator function to initialize query node features, and then performs message passing to obtain all node representations for the subsequent tasks. In other words, it generates node representations conditioned on a given query, which is more expressive than standard unconditional GNN encoders [52]. We advocate for this labeling trick and adopt NBFNet as encoders in THOR. Formally, given a query hyper-relational fact  $q$  with a masked entity, the relation foundation graph encoder first labels all its relations (primary relation and keys) using an indication function with all-ones values, and then performs message passing:

$$\overrightarrow{h_{r|q}^0} = \text{INDICATOR}(r, q) = \mathbb{1}_{r \in q} * 1^d, r \in \mathcal{G}_{fund}^r \quad (1)$$

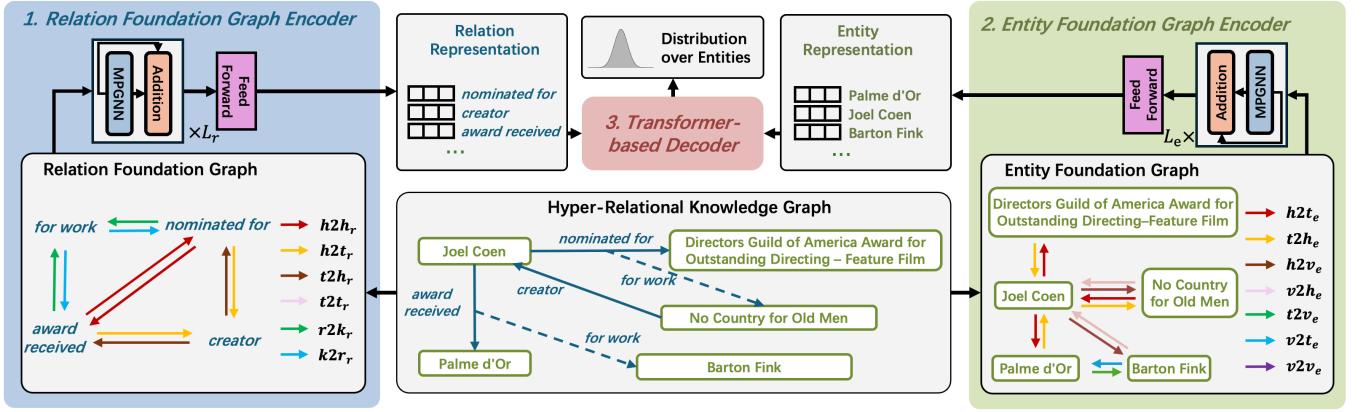


Figure 3: Architectural overview of THOR with 1) a relation foundation graph encoder, 2) an entity foundation graph encoder, and 3) a transformer-based decoder.

$$\vec{h}_{r|q}^{l+1} = \text{UPDATE} \left( \vec{h}_{r|q}^l, \text{AGGR} \left( \text{MESSAGE} \left( \vec{h}_{w|q}^l, \theta^r \right) | w \in N_{\mathcal{G}_{fund}^r}(r), \theta^r \in \mathcal{R}_{fund}^r \right) \right) \quad (2)$$

where  $d$  is the embedding size and  $N_{\mathcal{G}_{fund}^r}(r)$  refers to the neighbors of relation  $r$  in the relation foundation graph. The message passing functions follow the design of NBFNet. Note that we use the labeling trick with all-ones values because it is more generalizable in inductive settings than learnable embeddings, which tend to overfit the training graph [15]. Similarly, the entity foundation graph encoder first labels all entities (head, tail, and values of qualifiers) of the query fact  $q$  using the indication function with all-ones values, and then performs message passing as follows:

$$\vec{h}_{e|q}^0 = \text{INDICATOR}(e, q) = \mathbb{1}_{e \in q} * \mathbb{1}^d, e \in \mathcal{G}_{fund}^e \quad (3)$$

$$\vec{h}_{e|q}^{l+1} = \text{UPDATE} \left( \vec{h}_{e|q}^l, \text{AGGR} \left( \text{MESSAGE} \left( \vec{h}_{w|q}^l, \theta^e \right) | w \in N_{\mathcal{G}_{fund}^e}(e), \theta^e \in \mathcal{R}_{fund}^e \right) \right) \quad (4)$$

Following this design, our relation and entity foundation graph encoders learn the embeddings of fundamental relation interactions  $\theta^r \in \mathcal{R}_{fund}^r$  and entity interactions  $\theta^e \in \mathcal{R}_{fund}^e$ , rather than embeddings of specific relations or entities, thus supporting fully-inductive tasks.

**Transformer-based Decoder.** After obtaining the conditional entity and relation representations for a given query fact, we adopt a transformer-based decoder for hyper-relational link prediction tasks. Specifically, we adopt an edge-biased self-attention network as a decoder [27], utilizing learnable edge biases to capture the positional interactions of elements (entities and relations) within a hyper-relational fact. Considering the representation of each element (an entity or relation representation learnt above) in a hyper-relational fact as  $\vec{u}_i$ , the pair-wise similarity between the elements is computed as follows:

$$\beta_{ij} = \frac{\left( W^Q \vec{u}_i \right)^T \left( W^K \vec{u}_j + c_{ij}^K \right)}{\sqrt{d}} \quad (5)$$

where  $W^Q$  and  $W^K$  are linear transformation matrices for attention query and key, and  $c_{ij}^K$  indicates learnable edge bias for the key. Following [37], we consider five types edge biases, including  $(h, r)$ ,  $(t, r)$ ,  $(r, k)$ ,  $(k, v)$ , and others not included. After a Softmax layer

for normalizing similarity score  $\beta_{ij}$ , the updated element representation for the next layer can be calculated as follows:

$$\vec{u}_i = \sum_{j=1}^N \frac{\exp(\beta_{ij})}{\sum_{k=1}^N \exp(\beta_{ik})} \left( W^V \vec{u}_j + \vec{c}_{ij}^V \right) \quad (6)$$

where  $N$  is the number of elements in a hyper-relational fact,  $W^V$  is the linear transformation matrix for attention value, and  $\vec{c}_{ij}^V$  is the learnable edge bias for value, similar to  $c_{ij}^K$ . After generating the representation of the [MASK] token as  $\vec{x}_m$ , the probabilities of the candidates (entities) for the masked entity can be computed as:

$$\vec{p} = \text{Softmax} \left( U_0 \vec{x}_m + b_m \right) \quad (7)$$

where  $U_0$  is the entity representation matrix generated from the entity foundation graph encoder, and  $b_m$  is a learnable bias.

**Training and Inference.** THOR can be trained in an end-to-end manner by feeding the encoded entity and relation representations to the decoder and minimizing the cross-entropy loss. Notably, our transformer-based decoder can directly score all candidate entities in the entity foundation graphs without the need for expensive negative sampling process such as [7, 52] in the training process.

In the inference stage, given an inference graph and a query hyper-relational fact, THOR first builds the corresponding relation and entity foundation graphs, then encodes both graphs conditioned on the query facts, and finally feeds encoded entity and relation representations to the decoder for predicting the missing entity within the inference graph. The inference stage supports fully-inductive settings by design, because THOR resolves the link prediction task for a given query on a given KG through the relation and entity foundation graphs built from that KG by learning the fundamental relation and entity interactions, which are agnostic to any specific relations and entities.

## 4 Experiments

### 4.1 Experiment Setup

**Dataset.** We evaluate THOR on 12 datasets with different settings. We first use an existing semi-inductive HKG data series WD20K, which consists of 6 datasets of different versions. Due to the lack of fully-inductive HKG datasets in the current literature, we also crawl

**Table 1: Dataset Statistics**

Dataset		Train		Inference		Valid		Test	
		#Fact	#Ent/Rel	#Fact	#Ent/Rel	#Fact	#Ent/Rel	#Fact	#Ent/Rel
Semi-inductive datasets	WD20K100(V1)	7785	5785/91	2667	4286/74	295	642/42	365	777/42
	WD20K100(V2)	4146	3226/56	4274	5593/53	539	975/41	678	1215/42
	WD20K66(V1)	9020	6537/178	6949	8367/151	910	1515/110	1114	1797/109
	WD20K66(V2)	4553	4286/147	8922	9936/119	1481	2331/78	1841	2714/88
	WD20K33(V1)	23355	16284/321	8001	10277/245	885	1208/57	1095	1392/59
	WD20K33(V2)	12438	11335/278	12822	14165/228	1617	1910/49	2034	2200/50
Fully-inductive datasets	WDSPLIT100(V1)	12117	8927/95	7676	9207/303	1279	2085/109	1280	2103/114
	WDSPLIT100(V2)	7958	6964/132	11652	9551/170	1942	2645/76	1942	2628/81
	JFFI100(V1)	5325	3503/147	7467	4522/73	1666	802/26	2079	870/24
	JFFI100(V2)	6330	2840/126	9320	5014/96	1778	1146/41	2217	1239/41
	JFFI(V1)	21207	5360/194	15375	6495/119	1708	872/33	2137	966/29
	JFFI(V2)	9012	3328/156	18152	7515/157	2017	1464/50	2523	1613/51

and process another 6 datasets from two sources with different versions. Table 1 summarizes the dataset statistics.

- **WD20K** series [1] are semi-inductive HKG datasets collected from Wikidata, with consistent relation vocabulary and disjoint entity vocabularies between training and inference graphs. There are three variants of this dataset with different proportions of hyper-relational facts; WD20K100, WD20K66, and WD20K33 contain 100%, 66%, and 33% hyper-relational facts, respectively. Note that the original WD20K33 has overlapped entities between training and inference graph; we thus filter our own WD20K33 without overlapped entities following the strategy in [1] (see Appendix C for details) to avoid data leakage. Following [1], *each dataset variant has two versions, V1 and V2, where V1 has a larger training graph while V2 has a larger inference graph.*
- **WDSPLIT100** is our own collected fully-inductive dataset, which retains disjoint entity vocabularies but partially overlapped relation vocabularies between training and inference graphs. Following the inductive KG dataset processing guideline in [30] (originally proposed for triple facts), we first crawl over 1 million hyper-relational facts from WikiData [34], and then SPLIT this raw dataset with clustering algorithms on its entity graph to get disjoint entity clusters, and then select the hyper-relational facts whose entities belongs to two different clusters as training and inference graphs (see Appendix C for details). We also adopt a similar two-version design, resulting in WDSPLIT100(V1) and WDSPLIT100(V2). Note that this dataset has 100% hyper-relational facts, as we crawl only hyper-relational facts by design.
- **JFFI** is our own processed dataset filtered from JF17K [46] (originally collected from Freebase) for the hardest Fully Inductive setting, with both disjoint entity and relation vocabularies between training and inference graphs. Specifically, we use the same process as for the WD20K33 dataset (see Appendix C) to first create training and inference graphs with disjoint entity vocabularies, and then further remove the overlapped relations between training and inference graphs, resulting in a fully-inductive dataset with both disjoint entity and relation vocabularies. We also adopt a similar version design as for other datasets, resulting in JFFI(V1) and JFFI(V2). Moreover, we also consider their variants with 100% hyper-relational facts, JFFI100(V1) and JFFI100(V2).

**Baselines.** We compare THOR against a sizable collection of baselines in four categories: 1) Transductive hyper-relational link prediction techniques **StarE** [6], **GRAN** [37], and **HAHE** [21]; 2) Rule-based inductive techniques **Neural LP** [40] and **DRUM** [25]; 3) Semi-inductive techniques **NBFNet** [52], **QBLP** [1], and **NS-HART** [41]; 4) Fully-inductive techniques **INGRAM** [17], **RMPI** [10], **ULTRA** [7], **MOTIF** [14], **KG-ICL** [4], **TRIX** [47], **MAYPL** [18] and **MetaNIR** [38]. Please see Appendix E for more details. Note that among all inductive techniques, only QBLP, NS-HART, MAYPL, and MetaNIR can handle hyper-relational facts; we thus drop qualifiers for all other baselines. Moreover, the original QBLP model assumes pretrained entity and relation embeddings available [23] (only provided for Wikidata) in the training and inference datasets; this setting contradicts the core idea of inductive settings, where no information except for the inference KG itself is available. Therefore, we also consider the variant of QBLP by removing its pretrained embeddings, denoted as **QBLP (w/o pretrain)**. In all experiments, we report MRR as the metric.

## 4.2 Performance Comparison with Baselines

**Comparison with Inductive Techniques.** Table 2 shows the comparison with inductive techniques, where THOR achieves the best performance in most cases, yielding 66.1%, 55.9%, and 20.4% improvement over the best-performing rule-based, semi-inductive, and fully-inductive techniques, respectively. Specifically, we see that rule-based techniques and semi-inductive techniques show low performance on fully-inductive datasets (in particular in the hardest inductive setting on JFFI), because these techniques highly rely on the shared relations as anchors for link prediction tasks, which are not always available in fully-inductive settings (completely unavailable on JFFI). Note that rule-based techniques are semi-inductive by nature because they learn the logical rules on relations. Compared to the fully-inductive baselines, which are all designed for triple facts, THOR shows superior performance by effectively learning from the relation and entity foundation graphs designed exclusively for HKGs. Finally, by comparing the performance across different datasets, we observe that THOR’s performance generally increases with a higher portion of hyper-relational facts and with larger training graphs.

**Comparison with Inductive Techniques in cross-domain inference.** We consider cross-domain fully-inductive settings where the training and inference graphs are from different KG domains. As our datasets come from two different KG domains (Wikidata

**Table 2: Performance comparison with inductive techniques.** “H/T” refers to the results in head and tail entity prediction; “ALL” refers to the results in predicting all entities (including value entity); non-applicable techniques are marked with N/A.

Method		WD20K100(V1)		WD20K66(V1)		WD20K33(V1)		WDSPLIT100(V1)		JFFI100(V1)		JFFI(V1)	
		H/T	All	H/T	All	H/T	All	H/T	All	H/T	All	H/T	All
Rule-based	Neural LP	0.2471	N/A	0.1584	N/A	0.1638	N/A	0.0873	N/A	0.0146	N/A	0.0170	N/A
	DRUM	0.2834	N/A	0.1637	N/A	0.1549	N/A	0.0759	N/A	0.0152	N/A	0.0093	N/A
Semi-inductive	NBFNet	0.2280	N/A	0.1124	N/A	0.1050	N/A	0.0550	N/A	0.0021	N/A	0.0029	N/A
	QBLP (w/o pretrain)	0.0052	N/A	0.0012	N/A	0.0026	N/A	0.0034	N/A	0.0358	N/A	0.0023	N/A
	QBLP	0.1028	N/A	0.0368	N/A	0.1045	N/A	0.0045	N/A	N/A	N/A	N/A	N/A
	NS-HART	0.5325	0.5694	0.1868	0.2622	0.1865	0.2116	0.2390	0.2272	0.0609	0.0472	0.0211	0.0223
Fully-inductive	INGRAM	0.2420	N/A	0.0930	N/A	0.0190	N/A	0.0680	N/A	0.0430	N/A	0.1640	N/A
	RMPI	0.4583	N/A	0.1791	N/A	0.1823	N/A	0.2671	N/A	0.0595	N/A	0.0047	N/A
	ULTRA	0.4380	N/A	0.2359	N/A	0.1980	N/A	0.3358	N/A	0.3144	N/A	0.3073	N/A
	MOTIF	0.4264	N/A	0.2507	N/A	0.1108	N/A	0.3227	N/A	0.3362	N/A	0.2765	N/A
	KG-ICL	0.5284	N/A	0.2390	N/A	0.2644	N/A	0.2786	N/A	0.0675	N/A	0.0746	N/A
	TRIX	0.2064	N/A	0.1984	N/A	0.2205	N/A	0.3579	N/A	<b>0.3869</b>	N/A	0.3122	N/A
	MAYPL	0.4421	0.4761	0.1816	0.2468	0.2934	0.3077	0.2453	0.2854	0.1262	0.1773	0.1331	0.0967
	MetaNIR	0.3052	0.2749	0.2045	0.2401	0.2587	0.2560	0.1376	0.1359	0.2374	0.1773	0.2490	0.1914
	THOR	<b>0.6365</b>	<b>0.6557</b>	<b>0.2734</b>	<b>0.3486</b>	<b>0.3066</b>	<b>0.3669</b>	<b>0.4913</b>	<b>0.5125</b>	0.3687	<b>0.2612</b>	<b>0.3683</b>	<b>0.2638</b>
Method		WD20K100(V2)		WD20K66(V2)		WD20K33(V2)		WDSPLIT100(V2)		JFFI100(V2)		JFFI(V2)	
Rule-based	Neural LP	0.1964	N/A	0.2053	N/A	0.1490	N/A	0.0839	N/A	0.0394	N/A	0.0281	N/A
	DRUM	0.1982	N/A	0.2274	N/A	0.1338	N/A	0.0887	N/A	0.0237	N/A	0.0249	N/A
Semi-inductive	NBFNet	0.1647	N/A	0.0845	N/A	0.0707	N/A	0.0612	N/A	0.0029	N/A	0.0081	N/A
	QBLP (w/o pretrain)	0.0024	N/A	0.0017	N/A	0.0042	N/A	0.0023	N/A	0.0051	N/A	0.0016	N/A
	QBLP	0.0521	N/A	0.0178	N/A	0.0509	N/A	0.0091	N/A	N/A	N/A	N/A	N/A
	NS-HART	0.3628	0.4185	0.1454	0.2270	0.1328	0.1613	0.1885	0.2184	0.0343	0.0386	0.0129	0.0165
Fully-inductive	INGRAM	0.0760	N/A	0.0130	N/A	0.0040	N/A	0.0360	N/A	0.0730	N/A	0.0400	N/A
	RMPI	0.3615	N/A	0.1831	N/A	0.1571	N/A	0.2654	N/A	0.0702	N/A	0.0881	N/A
	ULTRA	0.3769	N/A	0.1945	N/A	0.2042	N/A	0.2922	N/A	0.2167	N/A	0.2105	N/A
	MOTIF	0.1843	N/A	0.1939	N/A	0.1788	N/A	0.2658	N/A	0.2481	N/A	0.1792	N/A
	KG-ICL	0.4041	N/A	0.2342	N/A	0.2012	N/A	0.3578	N/A	0.1022	N/A	0.1098	N/A
	TRIX	0.4235	N/A	0.2154	N/A	0.1758	N/A	0.3230	N/A	0.2090	N/A	0.2445	N/A
	MAYPL	0.2482	0.3445	0.1625	0.2388	0.2176	0.2712	0.2420	0.3049	0.1863	0.1613	0.1859	0.1633
	MetaNIR	0.2103	0.2310	0.1360	0.1571	0.1942	0.2068	0.1856	0.2387	0.2887	0.2155	0.2376	0.1768
	THOR	<b>0.5097</b>	<b>0.5904</b>	<b>0.3014</b>	<b>0.3767</b>	<b>0.2284</b>	<b>0.3083</b>	<b>0.4413</b>	<b>0.5411</b>	<b>0.3069</b>	<b>0.2323</b>	<b>0.2777</b>	<b>0.2309</b>

**Table 3: Performance on cross-domain fully-inductive settings (training → inference)**

Method	WD20K33(V1) → JFFI(V1)		WD20K33(V1) → JFFI(V2)		WD20K33(V2) → JFFI(V1)		WD20K33(V2) → JFFI(V2)		JFFI(V1) → WD20K33(V1)		JFFI(V1) → WD20K33(V2)		JFFI(V2) → WD20K33(V1)		JFFI(V2) → WD20K33(V2)	
	H/T	All														
ULTRA	0.3075	N/A	0.2288	N/A	0.2891	N/A	0.2121	N/A	0.1196	N/A	0.1350	N/A	0.0904	N/A	0.0775	N/A
MOTIF	0.0767	N/A	0.0777	N/A	0.1811	N/A	0.1903	N/A	0.1533	N/A	0.1234	N/A	0.1051	N/A	0.1273	N/A
KG-ICL	0.0803	N/A	0.1126	N/A	0.0864	N/A	0.1148	N/A	0.2548	N/A	0.2098	N/A	0.2164	N/A	0.2026	N/A
TRIX	0.2957	N/A	0.1952	N/A	0.2784	N/A	0.2178	N/A	0.1976	N/A	0.1556	N/A	0.1498	N/A	0.1858	N/A
MAYPL	0.1653	0.1176	0.1407	0.1158	0.1088	0.0803	0.0896	0.0772	0.0884	0.1221	0.0809	0.1352	0.0590	0.0975	0.0571	0.1113
MetaNIR	0.0671	0.0621	0.0815	0.0840	0.0817	0.0532	0.0786	0.0814	0.0284	0.0419	0.0415	0.0518	0.0269	0.0289	0.0265	0.0380
THOR	<b>0.3165</b>	<b>0.2245</b>	<b>0.2470</b>	<b>0.1776</b>	<b>0.3265</b>	<b>0.2292</b>	<b>0.2444</b>	<b>0.1676</b>	<b>0.2849</b>	<b>0.3399</b>	<b>0.2186</b>	<b>0.3004</b>	<b>0.2796</b>	<b>0.3305</b>	<b>0.2040</b>	<b>0.2842</b>

and Freebase)<sup>2</sup>, we select the largest datasets from the two domains (WD20K33 and JFFI) to perform cross-domain fully-inductive link prediction. Table 3 shows the results on all exhaustive dataset combinations, compared to the top fully-inductive baselines. We see that THOR consistently outperforms the baselines, yielding 32.5% improvement over the best-performing ones, further showing the superiority of THOR in the cross-domain fully-inductive setting.

**Comparison with Transductive Techniques.** We also compare THOR against SOTA transductive hyper-relational link prediction techniques, in both inductive and transductive settings for

the baselines. The results are in Appendix F due to space limitations. First, we find that all transductive baselines show very low performance in inductive settings, as expected, because they learn nothing about unseen entities in the inference graphs from the training graph. Second, in the transductive setting where the baselines are trained on the inference graph and evaluated on the test graph, THOR trained in the inductive setting even surprisingly shows a slightly higher performance than the transductive baselines on average, due to its superior generalizability.

### 4.3 Ablation Study

We conduct a series of ablation studies to systematically validate our key design choices in the aspects of 1) relation foundation

<sup>2</sup>Although facts from different KG sources may overlap as they all boil down to general human knowledge, the specific fact representations are different across KGs without any given alignment. We thus still refer to them as the cross-domain setting.

**Table 4: Ablation Study**

Method		WD20K100(V1)		WD20K66(V1)		WD20K33(V1)		WDSPLIT100(V1)		JFFI100(V1)		JFFI(V1)	
		H/T	ALL	H/T	ALL	H/T	ALL	H/T	ALL	H/T	ALL	H/T	ALL
Relation foundation graph	THOR (noR2K)	0.6222	0.6462	0.2659	0.3247	0.2915	0.3545	0.4820	0.4979	0.3292	0.2310	0.3609	0.2566
	THOR (noPrim)	0.6162	0.6378	0.2653	0.3298	0.2839	0.3459	0.4819	0.5005	0.3232	0.2289	0.3405	0.2503
	THOR (addK2K)	0.6267	0.6508	0.2725	0.3433	0.2912	0.3573	0.4901	0.5067	0.2818	0.2027	0.3580	0.2573
	THOR (addShareV)	0.5890	0.6366	0.2551	0.3180	0.2940	0.3586	0.4793	0.4962	0.2485	0.1836	0.3398	0.2427
	THOR (addAllIFI)	0.6205	0.6485	0.2619	0.3333	0.2831	0.3466	0.4859	0.5006	0.3176	0.2301	0.3478	0.2514
	THOR (add3Path)	0.6059	0.6402	0.2677	0.3391	0.2857	0.3511	0.4805	0.5040	0.2829	0.2047	0.3552	0.2571
	THOR (add3Star)	0.6072	0.6246	0.2575	0.3197	0.3002	0.3618	0.4679	0.4965	0.3408	0.2398	0.3611	0.2630
	THOR (addAllMotif)	0.6301	0.6517	0.2600	0.3479	0.2889	0.3560	0.4834	0.5038	0.3559	0.2501	0.3358	0.2383
Entity foundation graph	THOR (noV2V)	0.6290	0.6421	0.2652	0.3414	0.3061	0.3646	0.4831	0.5060	0.3639	0.2574	0.3600	0.2556
	THOR (noP2V)	0.3218	0.2705	0.1210	0.1456	0.1643	0.1686	0.2283	0.2322	0.2674	0.1828	0.2595	0.1839
	THOR (noV)	0.3462	0.2289	0.1156	0.0938	0.1554	0.1346	0.2364	0.1964	0.1792	0.1263	0.2397	0.1633
Structure	THOR (ULTRA-alike)	0.1721	0.1463	0.1517	0.1365	0.1189	0.1152	0.2289	0.1840	0.1422	0.1139	0.1014	0.0850
	THOR	<b>0.6365</b>	<b>0.6557</b>	<b>0.2734</b>	<b>0.3486</b>	<b>0.3066</b>	<b>0.3669</b>	<b>0.4913</b>	<b>0.5125</b>	<b>0.3687</b>	<b>0.2612</b>	<b>0.3683</b>	<b>0.2638</b>
Method		WD20K100(V2)		WD20K66(V2)		WD20K33(V2)		WDSPLIT100(V2)		JFFI100(V2)		JFFI(V2)	
		H/T	ALL	H/T	ALL	H/T	ALL	H/T	ALL	H/T	ALL	H/T	ALL
Relation foundation graph	THOR (noR2K)	0.4394	0.4548	0.2466	0.3324	0.2176	0.2770	0.3915	0.4605	0.2949	0.2305	0.1258	0.1319
	THOR (noPrim)	0.4581	0.4978	0.2514	0.3253	0.2231	0.2982	0.3582	0.4424	0.2504	0.1900	0.1867	0.1626
	THOR (addK2K)	0.4648	0.4918	0.2190	0.2891	0.2272	0.3019	0.3619	0.4250	0.1943	0.1695	0.2396	0.1939
	THOR (addShareV)	0.3699	0.3898	0.2318	0.3392	0.2283	0.2991	0.3414	0.3934	0.2687	0.2064	0.2314	0.1771
	THOR (addAllIFI)	0.4697	0.5044	0.2362	0.3134	0.2117	0.2908	0.3931	0.4502	0.2716	0.2233	0.1972	0.1710
	THOR (add3Path)	0.4864	0.5197	0.2738	0.3508	0.2139	0.2897	0.3798	0.4165	0.2567	0.1999	0.2530	0.1929
	THOR (add3Star)	0.4893	0.5541	0.2759	0.3670	0.2253	0.2985	0.3565	0.3954	0.2932	0.2190	0.2626	0.2072
	THOR (addAllMotif)	0.4878	0.5635	0.2803	0.3562	0.2273	0.3019	0.3861	0.4522	0.2607	0.2164	0.2125	0.1836
Entity foundation graph	THOR (noV2V)	0.5082	0.5548	0.2965	0.3717	0.2260	0.2987	0.4369	0.5108	0.2984	0.227	0.2637	0.2072
	THOR (noP2V)	0.2561	0.2645	0.1236	0.1698	0.1255	0.1596	0.1702	0.2184	0.1953	0.1420	0.1783	0.1317
	THOR (noV)	0.2457	0.1547	0.1123	0.0852	0.1265	0.1060	0.1879	0.1849	0.1448	0.1001	0.1787	0.1237
Structure	THOR (ULTRA-alike)	0.2377	0.1818	0.1429	0.1282	0.1561	0.1472	0.1576	0.1403	0.1957	0.1445	0.1554	0.1212
	THOR	<b>0.5097</b>	<b>0.5904</b>	<b>0.3014</b>	<b>0.3767</b>	<b>0.2284</b>	<b>0.3083</b>	<b>0.4413</b>	<b>0.5411</b>	<b>0.3069</b>	<b>0.2323</b>	<b>0.2777</b>	<b>0.2309</b>

graphs, 2) entity foundation graphs, and 3) the structure of parallel foundation graph encoders.

**Ablation Study on Relation Foundation Graph.** To validate our design choice for the relation foundation graph, we design the following variants by removing or expanding our selected fundamental interactions  $\mathcal{R}_{fund}^r = \{h2h_r, h2t_r, t2h_r, t2t_r, r2k_r, k2r_r\}$ . **THOR (noR2K)** removes within-fact relation-key interactions connecting the primary Relation and Keys within the same fact  $\{r2k_r, k2r_r\}$ . **THOR (noPrim)** removes cross-fact primary relation interactions connecting the Primary relations across different facts  $\{h2h_r, h2t_r, t2h_r, t2t_r\}$ . **THOR (addK2K)** adds within-fact key-key interactions connecting Keys in different qualifiers within the same fact  $k2k_r$ . **THOR (addShareV)** adds cross-fact relation-key interactions connecting relations and keys via Shared Values across different facts  $\{h2v_r, v2h_r, t2v_r, v2t_r, v2v_r\}$ . **THOR (addAllIFI)** uses All relation Fundamental Interactions exhaustively, as shown in Figure 1, which is a combination of THOR (addK2K) and THOR (addShareV).

In addition, out of curiosity, we also consider extending the graph motifs, which have recently been shown to be expressive for building triplet-based KG foundation models [14], to our HKG. To this end, following the original MOTIF framework [14], we inject additional fundamental interactions if two relations belong to the same motif. We adopt the following suggested motifs (please see [14] for the definition of these motifs). **THOR (add3Path)** adds interactions connecting relations by sharing 3-Path motifs. **THOR (add3Star)** adds interactions connecting relations by sharing 3-Star motifs. **THOR (addAllMotif)** adds interactions connecting relations by

sharing either the 3-path and 3-star motif, which is a combination of THOR (add3Path) and THOR (add3Star).

Results are shown in Table 4. On one hand, we observe that THOR outperforms THOR (noR2K) and THOR (noPrim) by 11.0% and 11.1%, respectively. On the other hand, THOR (addK2K), THOR (addShareV), and THOR (addAllIFI) underperform THOR by 11.5%, 13.6%, and 9.9%, respectively. The lessons learnt here are 1) *the cross-fact primary relation interactions and within-fact relation-key interactions encode the key structural invariance transferable across HKGs*, which can effectively support the inductive link prediction tasks; 2) *the within-fact key-key interactions and cross-fact relation-key (also key-key) interactions are redundant and noisy for capturing such structural invariance*, thus resulting in a performance drop in link prediction tasks.

In addition, we also see that adding motifs surprisingly yields a performance drop of 9.0%, 6.3%, and 6.8% for THOR (add3Path), THOR (add3Star), and THOR (addAllMotif), respectively. This differs from the positive results on triplet-based KGs as reported in [14]. We believe this is probably due to *the motif for triplet-based KGs mismatching the structural invariance encoded in HKGs*, which deserves further investigation in future work.

**Ablation Study on Entity Foundation Graph.** To validate our design choice for the entity foundation graph, we design the following variants by ablating our selected set of fundamental interactions  $\mathcal{R}_{fund}^e = \{h2t_e, t2h_e, h2v_e, v2h_e, t2v_e, v2v_e\}$ . Specifically, **THOR (noV2V)** removes the interaction  $v2v_e$  connecting two Values within the same fact. **THOR (noP2V)** removes the interactions connecting head/tail in Primary triplet to any Values

in qualifiers  $\{h2v_e, v2h_e, t2v_e, v2t_e\}$ . **THOR (noV)** removes all interactions involving Values in qualifiers  $\{h2v_e, v2h_e, t2v_e, v2t_e, v2v_e\}$ .

Results are shown in Table 4. THOR outperforms THOR (noV2V), THOR (noP2V), and THOR (noV) by 2.5%, 47.3%, and 55.6%, respectively, which verifies our design choice for exhaustive entity interactions in entity fundamental interactions. First, the significant improvement of THOR over THOR (noP2V) and THOR (noV) reveals the predominant role of entity interactions between primary triplets and qualifiers. Moreover, we observe a slight yet noticeable performance decrease when removing  $v2v_e$ , both from THOR to THOR (noV2V) by 2.5% and from THOR (noP2V) to THOR (noV) by 15.5%, which implies that the entity interactions between qualifiers are also important for inductive link prediction.

#### Ablation Study on the Structure of Parallel Graph Encoders.

To verify the architectural design of THOR, we reorganize the three components of THOR following the design of a state-of-the-art inductive technique ULTRA [7], denoted as **THOR (ULTRA-alike)**. Specifically, it encodes the relation foundation graph first, and then uses the encoded relation representation as the fundamental interactions in the entity foundation graph encoder; subsequently, its training process resorts to negative samples by randomly corrupting entities in positive hyper-relational facts. Note that THOR (ULTRA-alike) uses the same relation foundation graph and the same encoder as THOR. Results are shown in Table 4. THOR significantly outperforms THOR (ULTRA-alike) by 57.9%, showing the superiority of the THOR architecture in learning HKGs for inductive link prediction.

## 5 Related Work

### 5.1 Transductive link prediction over HKGs

Traditional link prediction techniques for HKGs mainly focus on the transductive setting, predicting within the training vocabulary. Early methods use an n-ary format to represent hyper-relational facts, splitting the primary relation  $r$  into  $r_h$  and  $r_t$ , and representing facts as sets of relation-entity pairs. NaLP [12] and RAM [20] model the relatedness among all role-value pairs, leveraging structural information for improved performance. Recent works highlight the drawbacks of the n-ary representation, proposing a hyper-relational representation with an unaltered primary triplet and additional qualifiers [24]. Recently, given the superiority of transformers [37], StarE [6], Hy-Transformer [42], QUAD [28], and HAHE [21] adopt transformer-based decoders and design various encoders to capture correlations among entities and relations. Although these works show superior performance, they fail to accommodate the dynamics of the modern KGs with newly emerging entities and relations.

### 5.2 Inductive link prediction over KGs

Existing works on inductive link prediction can be classified into two categories based on their settings regarding overlapping vocabularies [7].

**Semi-inductive techniques** assume a consistent relation vocabulary but arbitrary (even disjoint) entity vocabularies between training and inference KGs. MorsE [3] represents unseen entities by averaging the representation of connected relations. RED-GNN[48] encodes shared relational digraphs using dynamic programming and employs query-dependent attention to select relevant edges,

followed by message passing where unseen entities are initialized with all-zero vectors. NBFNet [52] designs a unified path-based framework using a generalized Bellman-Ford algorithm. Following NBFNet, several path-based techniques integrate selective message passing inspired by the  $A^*$  algorithm [13, 51], information bottleneck [43], gated message function [19], and path-length-based cutoff [29]. Other semi-inductive techniques encode unseen entities using relational message passing [26, 31], local relative position within subgraphs [22, 32], and global anchors [39]. While most semi-inductive techniques focus on triple facts, some also consider HKGs. QBLP [1] extends semi-inductive techniques to HKGs, integrating relational message passing from StarE [6]. NS-HART [41] introduces semantic hypergraph sampling and iterative message passing between hyperedges and entities. Although the semi-inductive techniques address the link prediction with unseen entities, their generalizability is still limited by the assumption of a fixed relation vocabulary.

**Fully-inductive techniques** assume arbitrary (even disjoint) relation and entity vocabularies between training and inference KGs, making them the most generalized case for link prediction. INGRAM [17] builds a relation foundation graph by connecting relations sharing the same head or tail, updating relation representation via relation-graph-based aggregation. RMPI [10] proposes four foundational relations and uses relational GAT [33] for entity representation. TARGI [5] introduces two extra foundation relations, *pair* and *loop*, to better describe complex topologies. ULTRA [7] adopts RMPI’s foundation relation graph and uses neural Bellman-Ford convolutional layers [52] for latent embeddings. TRIX [47] extends ULTRA by designing iterative graph convolutional layers with different orders. MOTIF [14] also extends the relation foundation graph in ULTRA with k-star motif and k-path motif to capture high-order graph features. [9] unifies INGRAM and ULTRA with a double permutation equivariant graph representation framework. [49] further extends the double equivariance concept to multi-task double equivariance by representing relations with conflicting predictive patterns. KG-ICL [4] tokenizes the k-hop subgraph of entities with specific local structures, and SCORE [35] introduces a semantic unifier for features with diversified origins. In this paper, our THOR is also designed under the double equivariance theorem [50], but focuses on accommodating hyper-relational facts, which are increasingly prominent in modern KGs. We present the theoretical analysis of THOR under the double equivariance theorem in Appendix B.

## 6 Conclusion

This paper proposes THOR, an inductive link prediction technique for Hyper-relational knOwledge gRaphs. We first introduce both relation and entity foundation graphs, modeling their fundamental inter- and intra-fact interactions in HKGs, and then design THOR learning from the two foundation graphs with two parallel NBFNet-based graph encoders followed by a transformer decoder, efficiently supporting fully-inductive inference. Our extensive experiments show that THOR outperforms a sizable collection of baselines, yielding 66.1%, 55.9%, and 20.4% improvement over the best-performing rule-based, semi-inductive, and fully-inductive techniques, respectively, and also reveal the key design factors capturing the structural

invariance transferable across HKGs for inductive tasks, which we believe shed light on building HKG foundation models.

In the future, we plan to extend the fundamental interactions to accommodate more complex hyper-relational motifs in HKGs.

## References

- [1] Mehdi Ali, Max Berrendorf, Mikhail Galkin, Veronika Thost, Tengfei Ma, Volker Tresp, and Jens Lehmann. 2021. Improving inductive link prediction using hyper-relational facts. In *The Semantic Web–ISWC 2021: 20th International Semantic Web Conference, ISWC 2021, Virtual Event, October 24–28, 2021, Proceedings* 20. Springer, 74–92.
- [2] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.
- [3] Mingyang Chen, Wen Zhang, Yushan Zhu, Hongting Zhou, Zonggang Yuan, Changliang Xu, and Huajun Chen. 2022. Meta-knowledge transfer for inductive knowledge graph embedding. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, 927–937.
- [4] Yuanning Cui, Zequn Sun, and Wei Hu. 2024. A prompt-based knowledge graph foundation model for universal in-context reasoning. *Advances in Neural Information Processing Systems* 37 (2024), 7095–7124.
- [5] Ling Ding, Lei Huang, Zhizhi Yu, Di Jin, and Dongxiao He. 2025. Towards Global-Topology Relation Graph for Inductive Knowledge Graph Completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39, 11581–11589.
- [6] Mikhail Galkin, Priyansh Trivedi, Gaurav Maheshwari, Ricardo Usbeck, and Jens Lehmann. 2020. Message passing for hyper-relational knowledge graphs. *arXiv preprint arXiv:2009.10847* (2020).
- [7] Mikhail Galkin, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu. 2023. Towards foundation models for knowledge graph reasoning. *arXiv preprint arXiv:2310.04562* (2023).
- [8] Michael Galkin, Jincheng Zhou, Bruno Ribeiro, Jian Tang, and Zhaocheng Zhu. 2024. A foundation model for zero-shot logical query reasoning. *Advances in Neural Information Processing Systems* 37 (2024), 54137–54160.
- [9] Jianfei Gao, Yangze Zhou, Jincheng Zhou, and Bruno Ribeiro. 2023. Double equivariance for inductive link prediction for both new nodes and new relation types. *arXiv preprint arXiv:2302.01313* (2023).
- [10] Yuxia Geng, Jiaoyan Chen, Jeff Z Pan, Mingyang Chen, Song Jiang, Wen Zhang, and Huajun Chen. 2023. Relational message passing for fully inductive knowledge graph completion. In *2023 IEEE 39th international conference on data engineering (ICDE)*. IEEE, 1221–1233.
- [11] Saiping Guan, Xiaolong Jin, Jiafeng Guo, Yuanzhuo Wang, and Xueqi Cheng. 2020. Neuinfer: Knowledge inference on n-ary facts. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, 6141–6151.
- [12] Saiping Guan, Xiaolong Jin, Yuanzhuo Wang, and Xueqi Cheng. 2019. Link prediction on n-ary relational data. In *The world wide web conference*, 583–593.
- [13] Peter E Hart, Nils J Nilsson, and Bertram Raphael. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics* 4, 2 (1968), 100–107.
- [14] Xingyu Huang, Pablo Barceló, Michael M Bronstein, Ismail Ilkan Ceylan, Mikhail Galkin, Juan L Reutter, and Miguel Romero Orth. 2025. How Expressive are Knowledge Graph Foundation Models? *arXiv preprint arXiv:2502.13339* (2025).
- [15] Xingyu Huang, Miguel Romero, Ismail Ceylan, and Pablo Barceló. 2023. A theory of link prediction via relational weisfeiler-leman on knowledge graphs. *Advances in Neural Information Processing Systems* 36 (2023), 19714–19748.
- [16] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S Yu. 2021. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems* 33, 2 (2021), 494–514.
- [17] Jaejun Lee, Chanyoung Chung, and Joyce Jiyoung Whang. 2023. InGram: Inductive knowledge graph embedding via relation graphs. In *International Conference on Machine Learning*. PMLR, 18796–18809.
- [18] Jaejun Lee and Joyce Jiyoung Whang. [n.d.] Structure Is All You Need: Structural Representation Learning on Hyper-Relational Knowledge Graphs. In *Forty-second International Conference on Machine Learning*.
- [19] Tianyu Liu, Qitan Lv, Jie Wang, Shuling Yang, and Hanzhu Chen. 2023. Learning rule-induced subgraph representations for inductive relation prediction. *Advances in Neural Information Processing Systems* 36 (2023), 3517–3535.
- [20] Yu Liu, Quanming Yao, and Yong Li. 2021. Role-aware modeling for n-ary relational knowledge bases. In *Proceedings of the Web Conference 2021*, 2660–2671.
- [21] Haoran Luo, Yuhao Yang, Yikai Guo, Mingzhi Sun, Tianyu Yao, Zichen Tang, Kaiyang Wan, Meina Song, Wei Lin, et al. 2023. Hahe: hierarchical attention for hyper-relational knowledge graphs in global and local level. *arXiv preprint arXiv:2305.06588* (2023).
- [22] Sijie Mai, Shuangjia Zheng, Yuedong Yang, and Haifeng Hu. 2021. Communicative message passing for inductive relation reasoning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35, 4294–4302.
- [23] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).
- [24] Paolo Rosso, Dingqi Yang, and Philippe Cudré-Mauroux. 2020. Beyond triplets: hyper-relational knowledge graph embedding for link prediction. In *Proceedings of the web conference 2020*, 1885–1896.
- [25] Ali Sadeghian, Mohammadreza Armandpour, Patrick Ding, and Daisy Zhe Wang. 2019. Drum: End-to-end differentiable rule mining on knowledge graphs. *Advances in neural information processing systems* 32 (2019).
- [26] Zhoutian Shao, Yuanning Cui, and Wei Hu. 2024. Expanding the Scope: Inductive Knowledge Graph Reasoning with Multi-starting Progressive Propagation. In *International Semantic Web Conference*. Springer, 41–58.
- [27] Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. 2018. Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155* (2018).
- [28] Harry Shomer, Wei Jin, Juanhui Li, Yao Ma, and Hui Liu. 2023. Learning representations for hyper-relational knowledge graphs. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, 253–257.
- [29] Harry Shomer, Yao Ma, Juanhui Li, Bo Wu, Charu C Aggarwal, and Jiliang Tang. 2023. Distance-based propagation for efficient knowledge graph reasoning. *arXiv preprint arXiv:2311.01024* (2023).
- [30] Harry Shomer, Jay Revolinsky, and Jiliang Tang. 2024. Towards Better Benchmark Datasets for Inductive Knowledge Graph Completion. *arXiv preprint arXiv:2406.11898* (2024).
- [31] Kai Sun, Jiapu Wang, Huajie Jiang, Yongli Hu, and Baocai Yin. 2024. Query-Enhanced Adaptive Semantic Path Reasoning for Inductive Knowledge Graph Completion. *arXiv preprint arXiv:2406.02205* (2024).
- [32] Komal Teru, Etienne Denis, and Will Hamilton. 2020. Inductive relation prediction by subgraph reasoning. In *International conference on machine learning*. PMLR, 9448–9457.
- [33] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).
- [34] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.
- [35] Kai Wang and Siqiang Luo. 2024. Towards Graph Foundation Models: The Perspective of Zero-shot Reasoning on Knowledge Graphs. *arXiv preprint arXiv:2410.12609* (2024).
- [36] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE transactions on knowledge and data engineering* 29, 12 (2017), 2724–2743.
- [37] Quan Wang, Haifeng Wang, Yajuan Lyu, and Yong Zhu. 2021. Link prediction on n-ary relational facts: A graph-based approach. *arXiv preprint arXiv:2105.08476* (2021).
- [38] Jiayao Wei, Saiping Guan, Xiaolong Jin, Jiafeng Guo, and Xueqi Cheng. 2025. Inductive link prediction in n-ary knowledge graphs. In *Proceedings of the 31st International Conference on Computational Linguistics*, 8885–8896.
- [39] Zhiwen Xie, Yi Zhang, Guangyou Zhou, Jin Liu, Xinhui Tu, and Jimmy Xiangji Huang. 2024. One Subgraph for All: Efficient Reasoning on Opening Subgraphs for Inductive Knowledge Graph Completion. *IEEE Transactions on Knowledge and Data Engineering* (2024).
- [40] Fan Yang, Zhilin Yang, and William W Cohen. 2017. Differentiable learning of logical rules for knowledge base reasoning. *Advances in neural information processing systems* 30 (2017).
- [41] Gongzhu Yin, Hongli Zhang, Yuchen Yang, and Yi Luo. 2025. Inductive Link Prediction on N-ary Relational Facts via Semantic Hypergraph Reasoning. *arXiv preprint arXiv:2503.20676* (2025).
- [42] Donghan Yu and Yiming Yang. 2021. Improving hyper-relational knowledge graph completion. *arXiv preprint arXiv:2104.08167* (2021).
- [43] Han Yu, Kai Chen, Ziniu Liu, Hongkui Tu, and Aiping Li. 2024. Inductive relation prediction with information bottleneck. *Neurocomputing* 609 (2024), 128503.
- [44] Weijian Yu, Jie Yang, and Dingqi Yang. 2024. Robust link prediction over noisy hyper-relational knowledge graphs via active learning. In *Proceedings of the ACM Web Conference 2024*, 2282–2293.
- [45] Muhan Zhang, Pan Li, Yinglong Xia, Kai Wang, and Long Jin. 2021. Labeling trick: A theory of using graph neural networks for multi-node representation learning. *Advances in Neural Information Processing Systems* 34 (2021), 9061–9073.
- [46] Richong Zhang, Junpeng Li, Jiajie Mei, and Yongyi Mao. 2018. Scalable instance reconstruction in knowledge bases via relatedness affiliated embedding. In *Proceedings of the 2018 world wide web conference*, 1185–1194.
- [47] Yucheng Zhang, Beatrice Bevilacqua, Mikhail Galkin, and Bruno Ribeiro. 2025. Trix: A more expressive model for zero-shot domain transfer in knowledge graphs. *arXiv preprint arXiv:2502.19512* (2025).
- [48] Yongqi Zhang and Quanming Yao. 2022. Knowledge graph reasoning with relational digraph. In *Proceedings of the ACM web conference 2022*, 912–924.
- [49] Jincheng Zhou, Beatrice Bevilacqua, and Bruno Ribeiro. 2023. A multi-task perspective for link prediction with new relation types and nodes. *arXiv preprint arXiv:2307.06046* (2023).

- [50] Jincheng Zhou, Yucheng Zhang, Jianfei Gao, Yangze Zhou, and Bruno Ribeiro. 2023. Double equivariance for inductive link prediction for both new nodes and new relation types. *arXiv preprint arXiv:2302.01313* (2023).
- [51] Zhaocheng Zhu, Xinyu Yuan, Michael Galkin, Louis-Pascal Khonneux, Ming Zhang, Maxime Gazeau, and Jian Tang. 2023. A\* net: A scalable path-based reasoning approach for knowledge graphs. *Advances in Neural Information Processing Systems* 36 (2023), 59323–59336.
- [52] Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal Khonneux, and Jian Tang. 2021. Neural bellman-ford networks: A general graph neural network framework for link prediction. *Advances in neural information processing systems* 34 (2021), 29476–29490.

## A Different Definitions of Inductive Settings

In the current literature, a few works such as QBLP [1] and NS-HART [41] have a different definition of semi- and fully-inductive link prediction tasks, where their semi-inductive setting refers to the case of a consistent relation vocabulary and *partially overlapped but not completely disjoint* entity vocabularies, while their fully-inductive setting further assumes a consistent relation vocabulary but *completely disjoint* entity vocabularies. In this paper, we follow the mainstream definition of semi- and fully-inductive link prediction settings [7, 10, 17, 30, 47] as we presented in Section 2, and list both QBLP [1] and NS-HART [41] as semi-inductive techniques.

## B Theoretical Analysis

Our design rationale stems from the double equivariance theorem [50] (same for ULTRA [7], MOTIF [14], and TRIX [47]), where the knowledge graph foundation model should be equivariant to entity or relation permutation, thus generating the same representation for isomorphic subgraphs. Formally, permutation on entities and relations is defined as:

**Definition A.1 (Entity and relation permutation).** Given a hyper-relational KG  $\mathcal{G} = (E, R, F)$  with  $F = \{[(h, r, t), \{(k_i, v_i)\}_{i=1}^n] \mid h, t, v_i \in E, r, k_i \in R\}$ , entity or relation permutation is  $\phi \in \mathbb{S}_e$  and  $\tau \in \mathbb{S}_r$  in symmetric groups  $\mathbb{S}_e$  and  $\mathbb{S}_r$ . The operation  $\phi \circ \tau \circ \mathcal{G}$  is defined as  $\phi \circ \tau \circ \mathcal{G} = (\phi \circ E, \tau \circ R, \phi \circ t \circ F)$ , where  $(\phi \circ E)_{\phi \circ i} = E_i, (\tau \circ R)_{\tau \circ j} = R_j$  and  $\phi \circ \tau \circ F = \{[(\phi \circ h, \tau \circ r, \phi \circ t), \{(\tau \circ k_i, \phi \circ v_i)\}_{i=1}^n] \mid [(h, r, t), \{(k_i, v_i)\}_{i=1}^n] \in F\}$ .

As reviewed in [50], a fully inductive KG model should generate double equivariant entity and relation representations, defined as:

**Definition A.2 (Double equivariant knowledge graph representations).** Given a hyper-relational KG  $\mathcal{G} = (E, R, F)$  with node representations  $\Gamma(\mathcal{G})$ ,  $\Gamma(\mathcal{G})$  is double equivariant knowledge graph representations if for any  $\phi \in \mathbb{S}_e$  and  $\tau \in \mathbb{S}_r$  in relation and entity symmetric groups,  $\Gamma(\phi \circ \tau \circ \mathcal{G}) = \phi \circ \tau \circ \Gamma(\mathcal{G})$ .

Following these theorems, our THOR supports fully-inductive reasoning with double-permutation-equivariant representations by 1) designing the entity and relation foundation graphs, where the two graphs' fundamental interactions do not rely on specific entities and relations, respectively, but only according to their semantic positions, thus being *permutation-equivariant*; 2) using NBFNet (an *equivariant* GNN) as graph encoders; and 3) using a transformer decoder (again, *permutation-equivariant by nature*).

## C Dataset Processing

The detailed data processing steps for the **WD20K33** and **JFFI** datasets are as follows. Specifically, different from the WDSPLIT100

dataset that we crawl by ourselves, WD20K33 and JFFI are filtered from the existing datasets WD50K [6] and JF17K [46], respectively. Subsequently, we adopt the filtering strategy proposed in [1]. We first sample a number of seed facts from the whole dataset (including train, valid, and test) and their k-hop neighbors, thus forming the training entity set  $E_{train}$  and the inductive entity set  $E_{ind} = E / E_{train}$ . Then we filter facts whose entities belong to  $E_{train}$  and  $E_{ind}$  into the training dataset  $\mathcal{G}_{train}$  and inductive dataset  $\mathcal{G}_{ind}$ , respectively. At last we split the inductive dataset into inference dataset  $\mathcal{G}_{inf}$ , valid dataset  $\mathcal{G}_{valid}$  and test dataset  $\mathcal{G}_{test}$ .

The detailed data crawling and processing steps for **WDSPLIT100** are as follows. Following the inductive KG dataset processing guideline in [30], we first extract over 1 million hyper-relational facts from WIKIDATA [34], then split this raw dataset with the Louvain clustering algorithm [2] over the graph built by the primary triplets (neglecting their qualifiers at this moment) to get the disjoint entity sets. Afterward, we select the facts whose entities belong to an arbitrary cluster of those entity clusters to form dataset pieces. Finally, we select relatively larger dataset pieces as  $\mathcal{G}_{train}$  and  $\mathcal{G}_{ind}$ . We also adopt a similar two-version design as for WD20K, resulting in WDSPLIT100(V1) and WDSPLIT100(V2). Note that this dataset has 100% hyper-relational facts because we crawl only hyper-relational facts by design.

## D Time Complexity

THOR's time complexity is  $O(a^2ed + vd^2)$  composed of NBFNet-based relation and entity encoders with time complexity  $O(ed+vd^2)$  and  $O(a^2ed + vd^2)$  as well as transformer decoder with time complexity  $O(a^2ed)$ , where  $a, e, v$ , and  $d$  are the average arity of hyper-relational facts, the number of facts, the number of entities, and the embedding dimension, respectively. We ignore the relation foundation graph, which is much smaller than the entity foundation graph and also processed in parallel. Compared to ULTRA's complexity  $O(e* d + v * d^2)$ , THOR is reasonably more complex as it also learns from qualifiers. As an example, on JFFIv1, THOR requires 2 3x more computation than ULTRA(104s/epoch v.s. 48s/epoch in training, 18s v.s. 6s in inference), but yields 19.9% performance improvement.

## E Baselines

We present the details on our baselines below.

**Transductive hyper-relational link prediction techniques.** **StarE** [6] represents a hyper-relational fact as a directed heterogeneous graph on triplets and designs an attributed message passing GNN where qualifiers are viewed as edge attributes and merged into the primary relation. **GRAN** [37] is a graph-based approach to link prediction on n-ary relational facts, which effectively captures rich associations within facts by representing them as heterogeneous graphs and utilizing edge-biased fully-connected attention to model local dependencies. **HAHE** [21] maintains the main structure of GRAN while introducing bidirectional hyper-edge message passing to capture the global structural information.

**Rule-based link prediction techniques.** **Neural LP** [40] combines the parameter and structure learning of first-order logical rules in an end-to-end differentiable model, where inference tasks can be compiled into sequences of differentiable operations. A neural controller system is then developed to compose these operations

**Table 5: Comparison with transductive baselines in the inductive setting**

Method	WD20K100(V1)		WD20K66(V1)		WD20K33(V1)		WDSPLIT(V1)		JFFI100(V1)		JFFI(V1)	
	H/T	All										
StarE	0.0523	N/A	0.0106	N/A	0.0042	N/A	0.0053	N/A	0.0458	N/A	0.0327	N/A
GRAN	0.0417	0.0482	0.0176	0.0234	0.0081	0.0096	0.0047	0.0058	0.0677	0.0493	0.0487	0.0373
HAHE	0.0518	0.0548	0.0217	0.0264	0.0096	0.0124	0.0088	0.0099	0.0715	0.0644	0.0677	0.0549
THOR	<b>0.6365</b>	<b>0.6557</b>	<b>0.2734</b>	<b>0.3486</b>	<b>0.3066</b>	<b>0.3669</b>	<b>0.4913</b>	<b>0.5125</b>	<b>0.3687</b>	<b>0.2612</b>	<b>0.3683</b>	<b>0.2638</b>

Method	WD20K100(V2)		WD20K66(V2)		WD20K33(V2)		WDSPLIT(V2)		JFFI100(V2)		JFFI(V2)	
	H/T	All										
StarE	0.0437	N/A	0.0083	N/A	0.0007	N/A	0.0103	N/A	0.0562	N/A	0.0289	N/A
GRAN	0.0319	0.0354	0.0153	0.0207	0.0013	0.0017	0.0062	0.0073	0.0624	0.0536	0.0365	0.0301
HAHE	0.0436	0.0497	0.0233	0.0319	0.0058	0.0105	0.0053	0.0084	0.0852	0.0697	0.0438	0.0362
THOR	<b>0.5097</b>	<b>0.5904</b>	<b>0.3014</b>	<b>0.3767</b>	<b>0.2284</b>	<b>0.3083</b>	<b>0.4413</b>	<b>0.5411</b>	<b>0.3069</b>	<b>0.2323</b>	<b>0.2777</b>	<b>0.2309</b>

**Table 6: Comparison with transductive baselines in the transductive setting**

Method	WD20K100(V1)		WD20K66(V1)		WD20K33(V1)		WDSPLIT(V1)		JFFI100(V1)		JFFI(V1)	
	H/T	ALL										
StarE	0.3606	N/A	0.3021	N/A	0.2514	N/A	0.5226	N/A	0.3336	N/A	0.3345	N/A
GRAN	0.3964	0.4257	<b>0.3155</b>	0.3374	0.2062	0.2347	0.5159	0.5371	0.3279	0.2761	0.3027	<b>0.2662</b>
HAHE	0.4238	0.4562	0.3081	0.3205	0.2217	0.2508	<b>0.5436</b>	<b>0.5709</b>	0.3401	<b>0.3017</b>	0.3256	0.2598
THOR	<b>0.6365</b>	<b>0.6557</b>	0.2734	<b>0.3486</b>	<b>0.3066</b>	<b>0.3669</b>	0.4913	0.5125	<b>0.3687</b>	0.2612	<b>0.3683</b>	0.2638

Method	WD20K100(V2)		WD20K66(V2)		WD20K33(V2)		WDSPLIT(V2)		JFFI100(V2)		JFFI(V2)	
	H/T	ALL										
StarE	0.2290	N/A	<b>0.3367</b>	N/A	0.2166	N/A	0.5099	N/A	0.2933	N/A	0.2672	N/A
GRAN	0.2036	0.2537	0.3182	0.3410	0.1874	0.2113	0.4936	0.5183	0.2957	0.2384	0.2389	0.2006
HAHE	0.2217	0.2496	0.3264	0.3459	0.2054	0.2183	<b>0.5211</b>	0.5398	<b>0.3128</b>	<b>0.2740</b>	0.2673	0.2245
THOR	<b>0.5097</b>	<b>0.5904</b>	0.3014	<b>0.3767</b>	<b>0.2284</b>	<b>0.3083</b>	0.4413	<b>0.5411</b>	0.3069	0.2323	<b>0.2777</b>	<b>0.2309</b>

for link prediction. **DRUM** [25] is motivated by making a connection between learning confidence scores for each rule and low-rank tensor approximation. It uses bidirectional RNNs to share useful information across the tasks of learning rules for different relations. Note that in rule-based methods, the filter operation eliminates all unseen entities prior to evaluation, which artificially inflates accuracy in an inductive setting. To ensure a fair comparison, we modify this operation to retain the ground-truth entities present in the test set for ranking only.

**Semi-inductive link prediction techniques.** **NBFNet** [52] generalizes the Bellman-Ford algorithm and proposes a path-based message passing mechanism with labeling trick [45], which can be applied to both transductive and inductive settings. **QBLP** [1] adopts a StarE encoder with a transformer decoder and can reason over an inference hyper-relational graph. **NS-HART** [41]. NS-HART defines an n-ary semantic hypergraph with a specifically designed k-hop neighborhood for HKG, and introduces a subgraph aggregator network to model complicated interactions in the subgraphs.

**Fully-inductive link prediction techniques.** **INGRAM** [17] generates a latent representation with relation graphs whose adjacent matrix  $E_h$  is defined as  $E_h[i, j]$  for times of the  $i$ th entity connected with the  $j$ th relation as head and similar for  $E_t$ . GNN is conducted over relation graphs at both entity-level and relation-level to generate the representation of entities and relations for further link prediction. **RMPI** [10] defines a relation foundation graph with foundation interaction  $h2h_r$ ,  $h2t_r$ ,  $t2h_r$ , and  $t2t_r$  similar to those in our relation foundation graph. Relational graph convolution is conducted over the relation foundation graph and the

original KG for the latent representation of relations and entities. **ULTRA** [7] inherits the structure design of the relation foundation graph from RMPI, replaces the convolution layers with NBFNet for path-based convolution, and enhances the model with labeling trick [45]. **MOTIF** [14] extends relation foundation graph of ULTRA with k-star and k-path motif, thus capturing high-order features of KGs. **KG-ICL** [4] extracts prompt graphs for query relations and designs a unified graph tokenizer representing the entities and relations into pre-defined tokens, followed by two message passing neural networks for prompt encoding and KG reasoning. **TRIX** [47] is a fully-inductive technique that improves expressivity and relation prediction efficiency by constructing a relation adjacency matrix incorporating entity information and employing an iterative embedding update mechanism to refine representations of both entities and relations. **MAYPL** [18] initializes all entities and relations with two learnable entity and relation tokens with neighborhood aggregation to generate entity and relation representations, followed by hyper-edge level message passing for reasoning. **MetaNIR** [38] builds complicated inter- and intra-fact relation foundation graph for relational GNN to generate entity and relation representations, followed with GRAN [37] decoder.

## F Performance Comparison with Transductive Baselines

We compare our THOR with transductive baselines StarE [6], GRAN [37], and HAHE [21] in both inductive and transductive settings.

In the inductive setting, unseen entities and relations are not learnt in the training stage and thus have randomly initialized embeddings in the inference stage. The results are shown in Table 5,

where the transductive baselines yield very low performance as expected. Those transductive techniques highly rely on the well-trained embeddings on the fixed entity and relation vocabulary while learn nothing about unseen entities and relations in the training procedure, thus losing the generalizability to unseen entities or relations in the inductive setting.

In the transductive setting, the baselines are trained on the inference graph  $\mathcal{G}_{inf}$ , then validated and evaluated on the valid and test graphs<sup>3</sup>, respectively, without using the training graph at all. Table 6 shows the results compared to our THOR (the results of THOR are still in the fully-inductive setting). *We observe that THOR even surprisingly shows a slightly higher performance (with an improvement of 17.1% on average) than the transductive baselines, yielding the best results in the majority of the cases, due to its superior generalizability.*

## G Limitations

Although the proposed THOR demonstrates superior performance in fully-inductive link prediction over HKGs, it still exhibits certain limitations. First, THOR may experience performance degradation on transductive datasets, in particular on domain-specific KGs. This issue could be mitigated by exploring hybrid approaches that integrate both inductive and transductive link prediction paradigms. Second, the stability of THOR warrants further enhancement, as inductive models are generally prone to overfitting, a trend observed in our experiments across both semi- and fully-inductive baselines.

---

<sup>3</sup>Note that these three graphs have consistent relation and entity vocabularies.