Prior Relational Schema Assists Effective Contrastive Learning for Inductive Knowledge Graph Completion

Ruilin Luo[1], Jiayi Li[2], Jianghangfan Zhang[3], Jing Xiao[4], Yujiu Yang[1†]
[1]Tsinghua University
[2]Baidu Inc.
[3]National University of Singapore
[4]Ping An Technology (Shenzhen) Co., Ltd.
lrl23@mails.tsinghua.edu.cn

Abstract

Knowledge Graph Completion (KGC) is a task aimed at uncovering the inherent relationships among known knowledge triplets in a Knowledge Graph (KG) and subsequently predicting missing links. Presently, there is a rising interest in inductive knowledge graph completion, where missing links may pertain to previously unobserved entities. Previous inductive KGC methods mainly rely on descriptive information of entities to improve the representation of unseen entities, neglecting to provide effective prior knowledge for relation modeling. To tackle this challenge, we capture prior schema-level interactions related to relations by leveraging entity type information, thereby furnishing effective prior constraints when reasoning with newly introduced entities. Moreover, We employ normal in-batch negatives and introduce schema-guided negatives to bolster the efficiency of normal contrastive representation learning. Experimental results demonstrate that our approach consistently achieves state-of-the-art performance on various established metrics across multiple benchmark datasets for link prediction. Notably, our method achieves a 20.5% relative increase in Hits@1 on the HumanWiki-Ind dataset. Our code is available at https://github.com/lrlbbzl/PReSA.

Keywords: Inductive Knowledge Graph Completion, Schema-level Learning, Contrastive Learning

1. Introduction

Knowledge Graphs (KGs) store a vast repository of factual knowledge in the form of triplets. A triplet (h, r, t) consists of a head entity h, a tail entity t, and a relation r linking them. There are many existing large-scale knowledge graphs, e.g. FreeBase (Bollacker et al., 2008), Wikidata (Tanon et al., 2016), and WordNet (Miller, 1994). KGs find applications in diverse domains, including recommendation systems (Yang et al., 2022), question answering (Liu et al., 2022), and web search (Xiong et al., 2017). Nonetheless, KGs consistently grapple with incompleteness (Carlson et al., 2010), necessitating Knowledge Graph Completion (KGC) to enhance their utility. With the emergence of various Large Language Models (LLMs), their capabilities in KGC have also attracted significant attention.

The main task of KGC is link prediction. Based on the kinds of links involved in testing, link prediction tasks can be divided into two categories: i) transductive setting calls for links between visible entities that have been observed during the training process; ii) inductive setting requires judgement of links in the event that fresh entities are added to the KG (Ali et al., 2021). Moreover, semi-inductive scenarios will occur more frequently in the real-world (Anil et al., 2023), meaning considering connections between known and newly introduced entities.

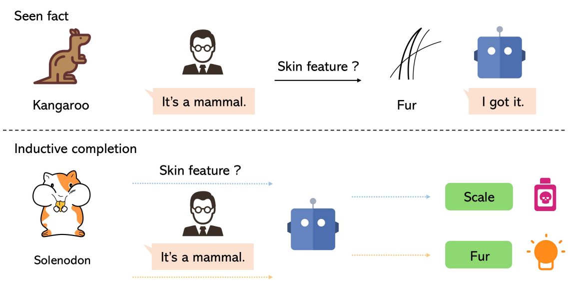


Figure 1: An example showcasing the effective utilization of schema learning for inductive link prediction.

In contrast to transductive setting, inductive KGC relies more on external facts to enhance the model's generalization. For example, BLP (Daza et al., 2021) emphasizes incorporating textual descriptions of entities to enhance their representations and subsequently as input to the Knowledge Graph Embedding (KGE) model. Many common knowledge queries can yield accurate responses by leveraging factual information and the extensive prior knowledge ingrained within pretrained language models (PLMs). Nevertheless, previous work underestimate the modeling of relations, leading to two noteworthy issues. Firstly, real-

† Corresponding author.

LREC-COLING 2024, pages 13014–13025

world relations frequently lack descriptions, constraining the model's ability to generalize relation representations across diverse contexts. Secondly, relations in KGs exhibit a substantial scale inferiority compared to entities, potentially resulting in imbalances during fine-tuning process of PLMs. As a result, the model's ability to perform high-quality reasoning in an inductive setting is limited.

Therefore, we harness schema information to address these issues, with schema essentially being the prior constraints formed by entity types and relations. On the one hand, schema is a bridge to bring high-level entity information into relations. On the other hand, each relation can be derived into numerous schemas in different contexts. We introduce PReSA, an acronym denoting **P**rior **Re**lational **S**chema **A**ssists Effective Contrastive Learning, which employs schema-level learning to enhance the model's generalization in an inductive setting. As depicted in Figure 1, once the schema *mammal_skin_feature* has been assimilated, the model can rectify the inference about the skin feature of *solenodon* from the inaccurate *scale* to *fur*. From this perspective, the training process captures the synergy between entity types and relations, facilitating the modeling of interactions with similar types of unobserved entities in inductive settings.

Moreover, we introduce a contrastive learning framework to ensure high-quality learning within embedding spaces for entities and schemas. Specifically, it involves the use of traditional **in-batch negatives** and proposes **schema-guided negatives** to serve as hard samples.

In summary, the contributions of this paper can be delineated as follows:

- We propose PReSA, which leverages schema information to underscore the significance of relation modeling in the process of link prediction inference, thereby mitigating the issue of weak generalization in the model. Additionally, we employ an efficient contrastive learning framework to acquire embeddings for both entities and schemas.

- We conduct a comprehensive evaluation of link prediction on HumanWiki, WN18RR and FB15k-237. Experimental results demonstrate that PReSA achieves SOTA performance across various metrics. Remarkably, in the HumanWiki-Ind dataset, PReSA exhibits a 10.2% relative improvement in MRR and a 20.5% relative enhancement in Hits@1 compared to the current leading model.

- Additionally, we delve into the performance of LLMs on link prediction, aiming to reveal both the opportunities and challenges.

2. Related Work

Structure-based KGE deduces representation of entities and relations from structural interaction. TransE (Bordes et al., 2013) embeds entities and relations as $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$ and measures the plausibility of a triple (h, r, t) by $\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$. RotatE (Sun et al., 2019) views relation r as a rotation from h to t in a complex space, in which entites and relations are embed, i.e. $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^d$. Semantic matching models define score function based on the similarity of latent semantics in a triple (h, r, t). RESCAL (Nickel et al., 2011) represents r as a matrix M_r and expect bilinear function $h^\top M_r t$ to approach X. Distmult (Yang et al., 2015) simplifies M_r to be diagonal matrix to lighten the parameters. However, the utilization of structure-based in inductive KGC needs retraining, which is not cost-effective.

Description-based KGE that incorporates descriptive information has gained more attention than structure-based methods with the rapid development of natural language processing (NLP) technique. The first model to utilize exogenous text information is DKRL (Xie et al., 2016), which encodes entities' descriptions via convolution neural network. KEPLER (Wang et al., 2021b) uses PLMs to jointly optimize KGE and Masked Language Model(MLM) objectives. BLP (Daza et al., 2021) harnesses PLMs for learning representations of entities via a link prediction objective. StAR (Wang et al., 2021a) optimize the distance-based objectives while utilizing the description and thus structural information at the same time. RAILD (Gesese et al., 2022) proposes a novel GNN-based approach to generate features for relations.

Contrastive Learning is a popular technique in the field of unsupervised representation learning, which has shown promising results in various computer vision (Grill et al., 2020) and natural language processing tasks (Gunel et al., 2021). The principal idea behind contrastive learning is contrasting the representation between positives and negatives in a latent space. In contrastive learning, positive examples are pairs of similar data points that are pulled closer together in the latent space, while negative examples are pairs of dissimilar data points that are pushed farther apart. It is typically achieved through a contrastive loss function such as InfoNCE (van den Oord et al., 2018), which encourages the model to learn representations that are more discriminative.

3. Methodology

In this section, we offer a comprehensive introduction to the language model employed and how PReSA leverages in-batch negatives and schema-guided negatives to facilitate efficient contrastive

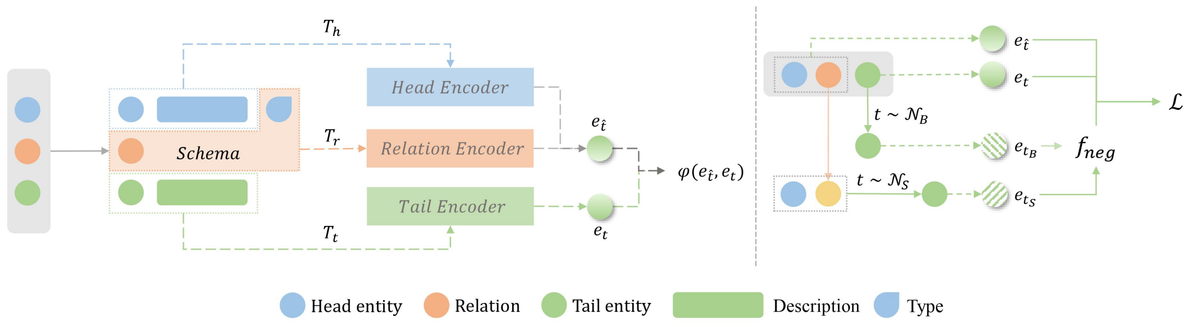


Figure 2: Overview of PReSA. The model employs three PLM encoders to capture information pertaining to the head entity, tail entity, and relational schema. Within the contrastive learning framework, two distinct categories of negative samples are introduced: i) Other tail entities in the same mini-batch, denoted as $\mathcal{N}_B$, and ii) Head entities and similar schemas are passed through the MLP module, as demonstrated in Eq.5, to produce $\mathcal{N}_S$.

learning, with the aim of obtaining high-quality embeddings for entities and upper schemas.

3.1. Language Models

As in previous work, we utilize PLMs to encode entities and relations. PLMs have exhibited significant progress in recent years (Peters et al., 2018). They undergo extensive training on large text corpora to acquire knowledge and have proven to be highly valuable in a wide range of downstream NLP tasks.
BERT (Devlin et al., 2019) is a variation of the Transformer Encoder architecture that utilizes masked language modeling (MLM) to predict concealed words, enabling it to capture contextual representations. We provide textual data related to entities and relations as sequential input to BERT, leveraging the state of the final layer as a learned vector, awaiting further operations and ultimately using it for KGC.

3.2. Model Architecture

As depicted in Figure 2, for a given triple (h, r, t), we incorporate external descriptions d_h, d_t, along with the types associated with the head entity h, denoted as $S^h = (s_1^h, s_2^h, \cdots, s_{n_h}^h)$, as auxiliary input. The components of a triple undergo transformation into sequences of tokens, which are subsequently ingested by PLMs. To be more specific, the head entity h is presented as a sequence that encompasses its name, description, and a selection of names from neighboring entities. We assume that the linked entities of the central entity e are denoted as N_e. The sentence-like prompts for h and t are illustrated in Eq.1 and Eq.2.

$$T^h = (h, d_h, concat(e_i)), e_i \in N_h \tag{1}$$

$$T^t = (t, d_t, concat(e_i)), e_i \in N_t \tag{2}$$

Instead of directly using name of relation, we model schema which comprises the types of head entity h and the corresponding relation's name r in the specific triple. Its prompt is illustrated in Eq.3.

$$T^r = (concat(s_i^h, [SEP]), r), s_i^h \in S^h \tag{3}$$

Given the sequence representations, we employ three encoders to compute the embeddings of T_h, T_r, T_t. As depicted in Eq.4, following the acquisition of the last-layer hidden states, we utilize a pooling layer, as suggested for enhanced performance (Gao et al., 2021).

$$\begin{aligned} e_h &= Pooling(BERT_1(T^h)) \\ e_r &= Pooling(BERT_2(T^r)) \\ e_t &= Pooling(BERT_3(T^t)) \end{aligned} \tag{4}$$

To effectively incorporate the complex structural interaction between h and r, we exploit several structural actions and a two-layer MLP to compute the embedding of the anticipated tail entity $e_{\hat{t}}$:

$$e_{\hat{t}} = MLP([e_h; e_r; e_h - e_r; e_h \cdot e_r]) \tag{5}$$

Finally, we calculate the cosine similarity, which we use to quantify the difference between $e_{\hat{t}}$ and e_t, as follows:

$$\varphi(e_{\hat{t}}, e_t) = \frac{e_{\hat{t}} \cdot e_t}{\|e_{\hat{t}}\| \|e_t\|} \tag{6}$$

3.3. Optimization

During the training process, we employ a contrastive learning framework to fine-tune Pretrained Language Models (PLMs). Furthermore, we exploit two negative sampling methods, in order to balance the efficiency of negative sample generation and the exploration of hard samples (Wu and Wang, 2021; Wang et al., 2022c).
We firstly propose in-batch negative samples, denoted as $\mathcal{N}_B$, which consist of other entities within

the same mini-batch. As discussed in previous work (Chen et al., 2021), a larger in-batch negative size tends to enhance the effectiveness of contrastive learning. However, simply introducing more negative tail entities results in increased computational overhead. However, since the schema is rich in semantic information, it can be used to mine hard samples within a mini-batch while expanding the pool of negatives.

Specifically, we leverage the most similar schema to generate schema-guided negatives. We firstly compute the cosine similarity matrix W of schemas within the mini-batch, as denoted in Eq.7.

$$W = \phi(E_r)\phi(E_r)^T \tag{7}$$

$\phi(\cdot)$ in Eq.7 means normalization. As demonstrated in Eq.8, we then select the most similar one for each schema, excluding itself.

$$V_r = \{j|argmax(DM(W)[i,:]) = j, i = 0, 1, \cdots\} \tag{8}$$

DM operation means diagonal masking. Following the reorganization of V_r, we derive the similar schema matrix denoted as $E_r^{'} = E_r[V_r,:]$. Subsequently, we utilize E_r and E_h as inputs for the MLP outlined in Eq.5 to acquire the embeddings representing schema-guided negatives $\mathcal{N}_S$. We propose that, when subjected to analogous guidance or behavior, the head entity and the relation may correspond to tail entities that exhibit proximity within the vector space but remain conceptually distinct.

To accommodate the negative samples presented above, we design loss function based on InfoNCE (van den Oord et al., 2018):

$$f_{neg} = \sum_{i=1}^{|\mathcal{N}_B|} e^{\varphi(e_{\hat{t}}, e_{t_B^i})/\tau} + \sum_{i=1}^{|\mathcal{N}_S|} \lambda e^{\varphi(e_{\hat{t}}, e_{t_S^i})/\tau} \tag{9}$$

$$\mathcal{L} = -log \frac{e^{(\varphi(e_{\hat{t}}, e_t)-\gamma)/\tau}}{e^{(\varphi(e_{\hat{t}}, e_t)-\gamma)/\tau} + f_{neg}} \tag{10}$$

The parameter λ in Eq.9 is used to adjust the loss weight of schema-guided negatives. Additive positive γ in Eq.10 encourages factual triple to obtain a higher score. Here, e_{t_B} and e_{t_S} represent embeddings for in-batch negatives and schema-guided negatives, respectively.

4. Experiments

4.1. Experimental Setup

Datasets - To evaluate the performance of PReSA, we have conducted experiments within

Dataset	#entity	#relation	#train	#valid	#test
HumanWiki-Trans	37262	221	92821	6045	6044
HumanWiki-Ind	37262	221	101195	1705	2010
WN18RR	40943	11	69585	11381	12087
FB15k-237	14541	237	215082	42164	52870

Table 1: Statistics of four used for evaluation.

a semi-inductive setting using the WN18RR and FB15k-237 datasets, both generously provided by BLP (Daza et al., 2021). Additionally, we have curated two additional datasets: HumanWiki-Trans and HumanWiki-Ind for a more comprehensive evaluation of the model. These datasets are derived from the original HumanWiki dataset (Rosso et al., 2021), which was initially designed for instance completion tasks. The HumanWiki dataset, sourced from Wikidata[1], comprises triples, with each triple featuring an entity classified as *human*. We obtain type and description information conveniently via the Wikidata open API. Further details regarding the four datasets are provided in Table 1.

Additionally, we conduct an evaluation on the classical triple classification task, for which we exploit the WN11 and FB13 datasets (Socher et al., 2013).

Baselines - In the evaluation of the semi-inductive setting across three datasets, we have selected several description-based approaches for comparison. These approaches include BE-BOW, BE-DKRL (Xie et al., 2016), BLP (Daza et al., 2021), StAR (Wang et al., 2021a), RAILD (Gesese et al., 2022), SimKGC (Wang et al., 2022a), and kNN-KGE (Wang et al., 2022b). For transductive link prediction on the HumanWiki-Trans, we have extended our baseline comparison to include structure-based models such as TransE (Bordes et al., 2013), RotatE (Sun et al., 2019), and HAKE (Zhang et al., 2020).

Metrics - In the context of link prediction, the most widely used metrics for evaluation include Mean Reciprocal Rank (MRR) and Hits@n (where n takes values from the set $\{1, 3, 10\}$). In this paper, we adhere to the established evaluation protocol. In test mode, for a given factual test triple (h, r, t), the model is tasked with predicting the missing t in $(h, r, ?)$ and h in $(?, r, t)$, respectively, by providing a sorted plausibility ranking of candidate entities. MRR calculates the average reciprocal rank across all test triples, providing a measure of overall prediction accuracy. Hits@n assesses the proportion of the correct target entity appearing within the top n positions of the ranking. Importantly, our evaluation is conducted under the filtered setting (Bordes et al., 2013). This entails excluding correct matches that already exist in the training set for each query in the test set.

[1]https://www.wikidata.org/wiki/Wikidata:Main_Page

For triple classification, we report the model's accuracy in correctly determining whether triples are factual or not.

Implementation Details - We inherit the choice of most previous work to use *bert-base-uncased* as the initial encoder. We perform a grid search on the following hyperparameters: the learning rate of the PLM from the set $\{1e-5, 3e-5, 5e-5\}$, and the batch size from the set $\{256, 512, 1024\}$. The temperature coefficient τ is empirically set to 0.05. As for optimization, we utilize AdamW and implement a linear learning rate decay strategy. The selection of the saved model for inference is determined based on Hits@1 performance on the validation dataset. The maximum number of training epochs for HumanWiki (in both settings), WN18RR, and FB15k-237 are set to 5, 50, and 15, respectively. All experiments are executed on a system equipped with $3\times$ A40 GPUs.

4.2. Main Results

Table 2 presents a comparison between PReSA with some other recent state-of-the-art models in transductive and inductive settings. The table demonstrates that PReSA consistently attains either the best or second-best results on the two HumanWiki datasets. Significantly, on the HumanWiki-Ind dataset, PReSA performs relatively 10.2% and 20.5% improvement in MRR and Hits@1, respectively. Moreover, Table 3 provides a detailed depiction of PReSA's performance in comparison to recent baseline models in the inductive setting, on the FB15k-237 and WN18RR datasets. Remarkably, PReSA achieves the best or second-best results across all metrics.

We continue to present the transductive link prediction performance of PReSA on FB15K-237 and WN18RR, and traditional triple classification on WN11 and FB13, to further portray the universal superiority of PReSA as a KGE approach in Table 4 and 5 respectively. In the triple classification task, we utilize the similarity function described in Eq.5 to evaluate the plausibility of each triple and establish a score threshold accordingly.

The results presented in Table 4 highlight PReSA's strong performance on the WN18RR dataset but also reveal a performance gap when compared to state-of-the-art models on FB15k-237. We attribute this difference in performance to the varying structural complexities inherent in these two datasets. Specifically, the average node degree within the heterogeneous graph formed by the WN18RR dataset stands at 4.59, whereas FB15k-237 exhibits a considerably higher average node degree at 42.65. Consequently, reasoning on FB15k-237 necessitates a more comprehensive consideration of structural information due to its higher complexity.

It's worth noting that PReSA demonstrates outstanding performance across all triple classification tasks. This achievement can be attributed to the fine-tuning of the PLM on specific knowledge graph triple contexts, enhancing its overall performance.

5. Analysis

5.1. Ablation on Batch Size

In Figure 3, we perform an evaluation to gauge the influence of varying batch sizes on contrastive learning, a process inherently involving different quantities of negative samples. In Figure 3, we perform an evaluation to gauge the influence of varying batch sizes on contrastive learning, a process inherently involving different quantities of negative samples. Our experiments consistently demonstrate that increasing the batch size has a beneficial effect on the model's performance. However, it is essential to highlight that this performance improvement gradually wanes as the batch size becomes larger, ultimately approaching an upper threshold. It is noteworthy that excessively large batch sizes can significantly impede the model's convergence, while simultaneously imposing greater demands on graphics memory resources.

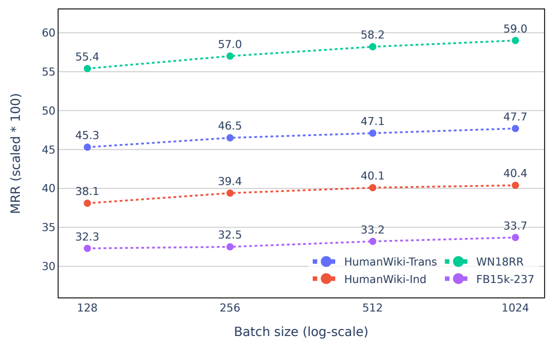


Figure 3: MRR on the four datasets w.r.t the number of negatives including $\mathcal{N}_B$ and $\mathcal{N}_S$.

Moreover, Figure 3 underscores the advantages of integrating schema-guided negatives $\mathcal{N}_S$. This approach proves to be beneficial as it doubles the pool of negative sample embeddings, all while incurring minimal computational overhead. Importantly, this augmentation of negative samples does not compromise the overall effectiveness of the model.

5.2. Ablation on Proposed Modules

Table 6 presents the results of our ablation study concerning type-augmented schemas and schema-guided negatives. PReSA (w/o type)

Method	HumanWiki-Trans				HumanWiki-Ind			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TransE (Bordes et al., 2013)	0.392	0.323	0.432	0.509	-	-	-	-
RotatE (Sun et al., 2019)	0.403	0.334	0.441	0.523	-	-	-	-
HAKE (Zhang et al., 2020)	0.394	0.322	0.435	0.521	-	-	-	-
BLP-TransE (Daza et al., 2021)	0.386	0.319	0.414	0.513	0.365	0.213	0.443	**0.664**
BLP-ComplEx (Daza et al., 2021)	0.368	0.314	0.398	0.491	0.351	0.201	0.430	0.627
BLP-SimplE (Daza et al., 2021)	0.366	0.301	0.396	0.493	0.359	0.208	0.442	<u>0.650</u>
RAILD-ComplEx (Gesese et al., 2022)	0.371	0.321	0.401	0.498	<u>0.372</u>	<u>0.244</u>	<u>0.448</u>	0.626
SimKGC (Wang et al., 2022a)	**0.478**	**0.396**	<u>0.510</u>	<u>0.636</u>	0.294	0.230	0.323	0.416
PReSA	<u>0.475</u>	<u>0.390</u>	**0.517**	**0.650**	**0.410**	**0.294**	**0.460**	<u>0.650</u>

Table 2: Link prediction results on HumanWiki-Trans and HumanWiki-Ind. The results of baselines come from our implementation. '-' indicates that these approaches cannot be directly leveraged for inductive setting. Best results are in **bold** and the seconds are <u>underlined</u>.

Method	FB15k-237				WN18RR			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
BE-BOW†	0.172	0.099	0.188	0.316	0.180	0.045	0.244	0.450
BE-DKRL†	0.144	0.084	0.151	0.263	0.139	0.048	0.169	0.320
BLP-TransE (Daza et al., 2021)†	0.195	0.113	0.213	0.363	0.285	0.135	0.361	0.580
BLP-ComplEx (Daza et al., 2021)†	0.148	0.081	0.154	0.283	0.261	0.156	0.297	0.472
BLP-SimplE (Daza et al., 2021)†	0.144	0.077	0.152	0.274	0.239	0.144	0.265	0.435
StAR (Wang et al., 2021a)*	0.169	0.101	0.180	0.312	0.307	0.184	0.364	0.566
RAILD-ComplEx (Gesese et al., 2022)‡	0.197	0.117	0.212	<u>0.364</u>	<u>0.320</u>	0.177	**0.390**	**0.609**
kNN-KGE (Wang et al., 2022b)**	**0.198**	**0.146**	<u>0.214</u>	0.293	0.294	**0.223**	0.320	0.431
PReSA	**0.198**	<u>0.121</u>	**0.215**	**0.367**	**0.323**	<u>0.193</u>	<u>0.370</u>	<u>0.582</u>

Table 3: Inductive link prediction results on WN18RR and FB15k-237. † Results from (Daza et al., 2021). ‡ Results from (Gesese et al., 2022). * Results from our implementation. ** Results from (Wang et al., 2022b). Best results are in **bold** and the seconds are <u>underlined</u>.

Method	WN18RR		FB15k-237	
	Hits@1	Hits@10	Hits@1	Hits@10
TransE (Bordes et al., 2013)	0.043	0.532	0.198	0.441
ComplEx (Trouillon et al., 2016)	0.410	0.510	0.158	0.428
RotatE (Sun et al., 2019)	0.428	0.571	0.241	<u>0.533</u>
HAKE (Zhang et al., 2020)	0.452	0.582	**0.250**	**0.542**
MEM-KGC (Choi et al., 2021)	<u>0.473</u>	0.636	<u>0.249</u>	0.522
StAR (Wang et al., 2021a)	0.243	0.709	0.205	0.482
$\text{LASS}_{\text{BERT-base}}$ (Shen et al., 2022)†	0.459	0.720	0.217	0.473
KGLM (Youn and Tagkopoulos, 2022)	0.330	0.741	0.200	0.468
LP-BERT (Li et al., 2022)	0.343	<u>0.752</u>	0.223	0.490
PReSA	**0.498**	**0.760**	0.221	0.512

Table 4: Transductive link prediction results on WN18RR and FB15k-237. † Results from our implementation due to the ungiven Hits@1. Other results are taken from original papers.

Method	WN11	FB13	Avg
TransE (Bordes et al., 2013)	75.9	81.5	78.7
TransD (Bordes et al., 2013)	86.4	89.1	87.8
TransR (Lin et al., 2015)	85.9	82.5	84.2
DistMult (Yang et al., 2015)	87.1	86.2	86.7
ConvKB (Nguyen et al., 2018)	87.6	88.8	88.2
KG-BERT (Yao et al., 2019)	**93.5**	90.4	92.0
R-MeN (Nguyen et al., 2020)	90.5	88.9	89.7
$\text{LASS}_{\text{BERT-base}}$ (Shen et al., 2022)	<u>93.3</u>	<u>91.2</u>	<u>92.3</u>
PReSA	<u>93.3</u>	**91.7**	**92.5**

Table 5: Accuracy of triple classification on WN11 and FB13. Results of baselines are taken from Nguyen et al. (2020), Shen et al. (2022), and Nassiri et al. (2022).

model variant which leverages only relations' name is similar to BLP (Daza et al., 2021). PReSA (w/o sgn) confines its use to in-batch negatives while ensuring the total number of negative samples aligns with other variants. The combination of these two strategies leads to improvements across most evaluation metrics compared to their individual utilization. Nonetheless, it's worth noting that PReSA displays minimal distinguishable differences from the two aforementioned counterparts.

We hypothesize that this observation stems from the fact that entities are visible in the transductive setting, which inherently provides a good fit for ordinary relations. Nonetheless, the remarkable improvements in MRR and Hits@n from PReSA(w/o type) to PReSA underscore that expanding relations based purely on lexical context remains equally beneficial.

5.3. Fine-grained Analysis

Following widely accepted conventions (Bordes et al., 2013; Wang et al., 2014), relations are categorized into four distinct types: 1-1, 1-n, n-1, and n-n. Previous research has consistently highlighted the relative difficulty in modeling relations of the 1-n and n-n types. This complexity arises from the model's limited capacity to generalize effectively when the target entity has multiple potential associations. Additionally, some accurate triples may not be present within the KGs. We address that PReSA alleviates the dilemma by em-

	HumanWiki-Trans				HumanWiki-Ind			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
PReSA	<u>0.475</u>	**0.390**	0.517	**0.650**	**0.410**	**0.294**	<u>0.460</u>	**0.650**
w/o type	**0.479**	0.384	<u>0.519</u>	0.636	0.397	0.281	**0.462**	0.639
w/o sgn	0.474	<u>0.388</u>	**0.520**	<u>0.644</u>	<u>0.402</u>	<u>0.287</u>	0.455	<u>0.645</u>
	FB15k-237				WN18RR			
PReSA	**0.198**	<u>0.121</u>	<u>0.215</u>	**0.367**	<u>0.323</u>	**0.193**	**0.370**	**0.582**
w/o type	0.191	0.117	0.206	0.354	0.317	0.184	0.361	0.569
w/o sgn	<u>0.197</u>	**0.124**	**0.218**	<u>0.364</u>	**0.325**	<u>0.190</u>	<u>0.362</u>	<u>0.575</u>

Table 6: Ablation studies w.r.t entity type information and proposed schema-guided negative samples.

Method	1-1	1-n	n-1	n-n
BLP-TransE (Daza et al., 2021)	0.606	0.367	0.644	0.434
SimKGC (Wang et al., 2022a)	0.586	0.196	0.502	0.333
PReSA	0.630	0.486	0.693	0.565

Table 7: MRR grouped by four types of relations on the HumanWiki-Ind dataset (tail-batch).

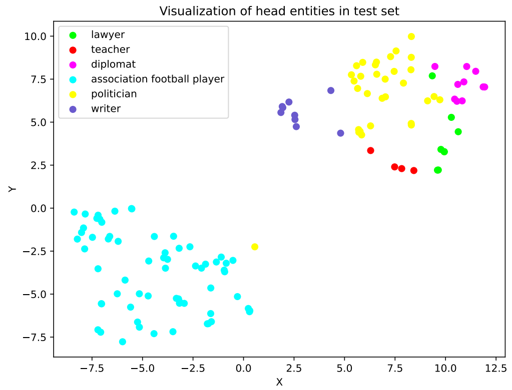
ploying schema-level learning which guides unfamiliar entities toward actions under familiar populations. The results, as presented in Table 7, clearly illustrate a significant performance disparity within the 1-n and n-n relations. In these categories, SimKGC and BLP-TransE exhibit marked inferiority compared to PReSA.

To elucidate this point further, we offer a specific example from HumanWiki-Ind dataset. Let's consider the query (*Wolfgang Schäuble*, *participant_in*, ?). Notably, PReSA yields a precise prediction, correctly identifying *coalition talks between the CDU/CSU and SPD in 2013*. In stark contrast, SimKGC and BLP-TransE provide predictions of *Paris Peace Conference* and *Nuremberg trials*, respectively. This discrepancy can be traced back to the acquired schema, namely *politician_lawyer_participant_in*. The schema augmented by entity types effectively directs the model's attention toward activities of a political and legal nature. Despite the availability of descriptive information, SimKGC and BLP-TransE generate responses that lack relevance.

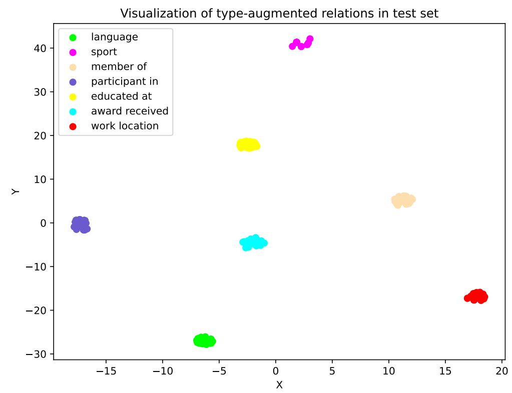
5.4. Visualization

We provide a 2-D visualization of schemas and entities on test set to demonstrate the generalization and rationality. Specifically, we randomly select 30 embeddings of schemas derived from each of the seven most frequently occurring relations. Furthermore, we visualize entities from six randomly selected categories. In Figure 4(a), it is obvious that the utilization of different types of information does not lead to significant deviations. This phenomenon can be attributed to the effectiveness of our contrastive learning approach.

Nevertheless, we observe a degree of overlap among the types ”diplomat”, ”politician”, and ”lawyer”. This overlap can be attributed to the presence of entities in the dataset that exhibit behaviors associated with all three occupations. This observation highlights the inherent limitations of embedding learning.



(a) Visualization of inductive entity embeddings.



(b) Visualization of relation embeddings.

Figure 4: 2-D visualization of the head entities and schemas in test set of HumanWiki-Ind dataset using t-SNE (van der Maaten and Hinton, 2008) downscaling method.

Figure 4(b) illustrates that the relation embeddings, even after generalization through entity type information, still maintain a clustered distribution as expected, without significant deviations.

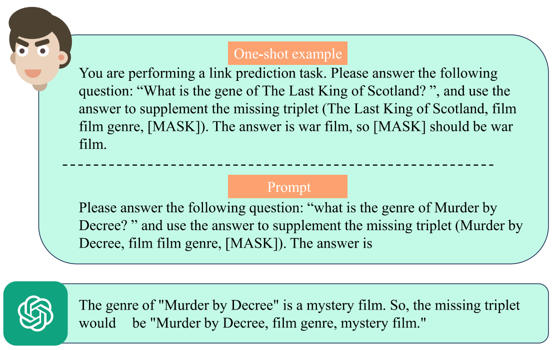


Figure 5: Examples of our evaluation. In the zero-shot setting, we only serve the content under the dotted line.

6. Discussion with LLMs

In recent times, the rapid advancement of LLMs has garnered significant attention. Notable LLMs like ChatGPT and GPT-4 (OpenAI, 2023) have emerged as exceptional language understanding systems. In this section, we delve into an assessment of various LLMs' performance in link prediction tasks, with the aim of exploring their potential in knowledge graph reasoning.

Given the extensive amount of entity information stored within these large models, it is impractical to anticipate a ranked list of candidates for every query. Thus, we employ exact-match accuracy as a metric, aligning it with the Hits@1 metric commonly used in link prediction tasks. We not only present the results in a zero-shot setting but also delve into a comprehensive exploration of how the one-shot setting can augment the reasoning capabilities of LLMs. Refer to Figure 5 for a visual representation of the prompt.

Model	FB15k-237	WN18RR	HumanWiki
	Zero-shot		
GPT-3.5-Turbo	0.220	0.320	0.380
GPT-4	0.280	0.440	0.480
Claude-2	0.220	0.420	0.420
	One-shot		
GPT-3.5-Turbo	0.260	0.400	0.400
GPT-4	**0.340**	**0.500**	**0.520**
Claude-2	0.240	0.480	0.440

Table 8: Exact-match accuracy results on GPT-3.5-Turbo, GPT-4 and Claude-2. Best results are in **bold** and the seconds are underlined.

As depicted in Table 8, even when employing LLMs, achieving optimal accuracy on FB15k-237 remains a challenging task, whereas more satisfactory results are obtained on WN18RR and HumanWiki. We attribute this disparity in performance to several factors. Firstly, FB15k-237 represents real-world knowledge, which can be intricate. In contrast, WN18RR primarily deals with logical relations between words, which tends to reduce ambiguity and result in more favorable performance for LLMs. HumanWiki, reliant on the acquired corpus, exhibits distinctive behaviors among the models. For instance, we observe that GPT-3.5-turbo occasionally requests additional contextual information to aid in decision-making instead of providing a direct answer. This behavior may stem from the limitations in the scope of stored knowledge within the model. Furthermore, Claude-2 exhibits a similar approach to CoT (Wei et al., 2022). It initially presents relevant knowledge, then seeks to understand the query's intent, and finally predicts the target entity.

The results clearly demonstrate that the one-shot setting significantly enhances the model's reasoning capabilities. In both settings, GPT-4 stands out as the top performer, largely owing to its substantial parameter count and superior training data. However, it's worth noting that the impact of the one-shot setting appears to be more limited when applied to FB15k-237 and HumanWiki. In these cases, the models seem to rely more on their intrinsic knowledge rather than discerning underlying logic from the provided relations.

Given the extensive training data used for LLMs, it is reasonable to consider KG reasoning on them as a form of transductive inference. Our results, presented in Figure 4, demonstrate that the PReSA model performs only slightly less effectively than GPT-4 in the one-shot setting on the WN18RR dataset, achieving scores of 0.498 and 0.500, respectively. This performance is indicative of high competitiveness. Furthermore, our model exhibits similar performance to gpt-3.5-turbo and Claude-2 in zero-shot scenarios on FB15k-237, with scores of 0.221, 0.220, and 0.220, and on HumanWiki, with scores of 0.390, 0.380, and 0.420.

7. Conclusion

This paper focuses on investigating the generalization properties of entity and relation representations using PLMs and a contrastive learning framework. It addresses the oversight in previous research concerning relation representations. To comprehensively characterize contrastive learning and account for computational cost, we introduce shema-guided negative sampling. Extensive experiments conducted on the FB15k-237, WN18RR, and HumanWiki datasets validate the superior performance of PReSA in the inductive setting. Additionally, PReSA also achieves outstanding results in transductive link prediction and traditional triple classification tasks.

Regarding future research directions, we emphasize the importance of improving the effectiveness of head-batch techniques. This is crucial because PLMs still struggle to fully comprehend the nuances of inverse relations. Furthermore, at the level of contrastive learning, considering the inherent incompleteness of knowledge graphs, we propose designing more robust loss functions to enhance model resilience, as discussed in RINCE (Chuang et al., 2022).

8. Ethical Considerations

We believe that all authors highly value and strictly adhere to a code of ethics. In our approach, we use a language model and make use of person-related data from the publicly available Wikidata. However, we do not expect the model to generate harmful conclusions based on the dataset. We solely focus on the efficient application of the language model in KGC tasks.

References

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*, volume 12922 of *Lecture Notes in Computer Science*, pages 74–92. Springer.

Akash Anil, Víctor Gutiérrez-Basulto, Yazmín Ibáñez-García, and Steven Schockaert. 2023. Inductive knowledge graph completion with gnns and rules: An analysis. *CoRR*, abs/2308.07942.

Kurt D. Bollacker, Colin Evans, Praveen K. Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008*, pages 1247–1250. ACM.

Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795.

Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam Hruschka, and Tom Mitchell. 2010. Toward an architecture for never-ending language learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 24, pages 1306–1313.

Junya Chen, Zhe Gan, Xuan Li, Qing Guo, Liqun Chen, Shuyang Gao, Tagyoung Chung, Yi Xu, Belinda Zeng, Wenlian Lu, Fan Li, Lawrence Carin, and Chenyang Tao. 2021. Simpler, faster, stronger: Breaking the log-k curse on contrastive learners with flatnce. *CoRR*, abs/2107.01152.

Bonggeun Choi, Daesik Jang, and Youngjoong Ko. 2021. MEM-KGC: masked entity model for knowledge graph completion with pre-trained language model. *IEEE Access*, 9:132025–132032.

Ching-Yao Chuang, R. Devon Hjelm, Xin Wang, Vibhav Vineet, Neel Joshi, Antonio Torralba, Stefanie Jegelka, and Yale Song. 2022. Robust contrastive learning against noisy views. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 16649–16660. IEEE.

Daniel Daza, Michael Cochez, and Paul Groth. 2021. Inductive entity representations from text via link prediction. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, pages 798–808. ACM / IW3C2.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. Simcse: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 6894–6910. Association for Computational Linguistics.

Genet Asefa Gesese, Harald Sack, and Mehwish Alam. 2022. RAILD: towards leveraging relation features for inductive link prediction in knowledge graphs. In *Proceedings of the 11th International Joint Conference on Knowledge Graphs,* *IJCKG 2022, Hangzhou, China, October 27-28, 2022*, pages 82–90. ACM.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Ávila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. 2020. Bootstrap your own latent - A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Beliz Gunel, Jingfei Du, Alexis Conneau, and Veselin Stoyanov. 2021. Supervised contrastive learning for pre-trained language model fine-tuning. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Da Li, Sen Yang, Kele Xu, Ming Yi, Yukai He, and Huaimin Wang. 2022. Multi-task pre-training language model for semantic network completion.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*, pages 2181–2187. AAAI Press.

Lihui Liu, Boxin Du, Jiejun Xu, Yinglong Xia, and Hanghang Tong. 2022. Joint knowledge graph completion and question answering. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1098–1108.

George A. Miller. 1994. WORDNET: A lexical database for english. In *Human Language Technology, Proceedings of a Workshop held at Plainsboro, New Jerey, USA, March 8-11, 1994*. Morgan Kaufmann.

Armita Khajeh Nassiri, Nathalie Pernelle, Fatiha Saïs, and Gianluca Quercini. 2022. Knowledge graph refinement based on triplet bert-networks. *CoRR*, abs/2211.10460.

Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen, and Dinh Q. Phung. 2018. A novel embedding model for knowledge base completion based on convolutional neural network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers)*, pages 327–333. Association for Computational Linguistics.

Dai Quoc Nguyen, Tuan Nguyen, and Dinh Phung. 2020. A relational memory-based embedding model for triple classification and search personalization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 3429–3435. Association for Computational Linguistics.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, pages 809–816. Omnipress.

OpenAI. 2023. GPT-4 technical report. *CoRR*, abs/2303.08774.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 2227–2237. Association for Computational Linguistics.

Paolo Rosso, Dingqi Yang, Natalia Ostapuk, and Philippe Cudré-Mauroux. 2021. RETA: A schema-aware, end-to-end solution for instance completion in knowledge graphs. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, pages 845–856. ACM / IW3C2.

Jianhao Shen, Chenguang Wang, Linyuan Gong, and Dawn Song. 2022. Joint language semantic and structure embedding for knowledge graph completion. In *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*, pages 1965–1978. International Committee on Computational Linguistics.

Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 926–934.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Thomas Pellissier Tanon, Denny Vrandecic, Sebastian Schaffert, Thomas Steiner, and Lydia Pintscher. 2016. From freebase to wikidata: The great migration. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, pages 1419–1428. ACM.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 2071–2080. JMLR.org.

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *CoRR*, abs/1807.03748.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605.

Bo Wang, Tao Shen, Guodong Long, Tianyi Zhou, Ying Wang, and Yi Chang. 2021a. Structure-augmented text representation learning for efficient knowledge graph completion. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, pages 1737–1748. ACM / IW3C2.

Liang Wang, Wei Zhao, Zhuoyu Wei, and Jingming Liu. 2022a. Simkgc: Simple contrastive knowledge graph completion with pre-trained language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 4281–4294. Association for Computational Linguistics.

Peng Wang, Xin Xie, Xiaohan Wang, and Ningyu Zhang. 2022b. Reasoning through memorization: Nearest neighbor knowledge graph embeddings.

Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, and Jian Tang. 2021b. KEPLER: A unified model for knowledge embedding and pre-trained language representation. *Trans. Assoc. Comput. Linguistics*, 9:176–194.

Xintao Wang, Qianyu He, Jiaqing Liang, and Yanghua Xiao. 2022c. Language models as knowledge embeddings. *arXiv preprint arXiv:2206.12617*.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, pages 1112–1119. AAAI Press.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837.

Haiping Wu and Xiaolong Wang. 2021. Contrastive learning of image representations with cross-video cycle-consistency. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10149–10159.

Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, and Maosong Sun. 2016. Representation learning of knowledge graphs with entity descriptions. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 2659–2665. AAAI Press.

Chenyan Xiong, Russell Power, and Jamie Callan. 2017. Explicit semantic ranking for academic search via knowledge graph embedding. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 1271–1279. ACM.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Yuhao Yang, Chao Huang, Lianghao Xia, and Chenliang Li. 2022. Knowledge graph contrastive learning for recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1434–1443.

Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. KG-BERT: BERT for knowledge graph completion. *CoRR*, abs/1909.03193.

Jason Youn and Ilias Tagkopoulos. 2022. KGLM: integrating knowledge graph structure in language models for link prediction. *CoRR*, abs/2211.02744.

Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. 2020. Learning hierarchy-aware knowledge graph embeddings for link prediction. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 3065–3072. AAAI Press.