



Bridge Structural Knowledge and Pre-trained Language Models for Knowledge Graph Completion

Qiao Qiao^(✉), Yuepei Li, Qing Wang, Kang Zhou, and Qi Li[✉]

Iowa State University, Ames, IA, USA

{qqiao1, liyp0095, qingwang, kangzhou, qli}@iastate.edu

Abstract. Knowledge graph completion (KGC) is a task of inferring missing triples based on existing Knowledge Graphs (KGs). Both structural and semantic information are vital for successful KGC. However, existing methods only use either the structural knowledge from the KG embeddings or the semantic information from pre-trained language models (PLMs), leading to suboptimal model performance. Moreover, since PLMs are not trained on KGs, directly using PLMs to encode triples may be inappropriate. To overcome these limitations, we propose a novel framework called Bridge, which jointly encodes structural and semantic information of KGs. Specifically, we strategically encode entities and relations separately by PLMs to better utilize the semantic knowledge of PLMs and enable structured representation learning via a structural learning principle. Furthermore, to bridge the gap between KGs and PLMs, we employ a self-supervised representation learning method called BYOL to fine-tune PLMs with two different views of a triple. Unlike BYOL, which uses augmentation methods to create two semantically similar views of the same image, potentially altering the semantic information. We strategically separate the triple into two parts to create different views, thus avoiding semantic alteration. Experiments demonstrate that Bridge outperforms the SOTA models on three benchmark datasets.

Keywords: Knowledge representation · Knowledge graph completion

1 Introduction

Knowledge graphs (KGs) are graph-structured databases composed of triples (facts), where each triple (h, r, t) represents a relation r between a head entity h and a tail entity t . KGs such as Wikidata [25] and WordNet [5] have a significant impact on various downstream applications such as named entity recognition [13, 33], relation extraction [28], and entity linking [34]. Nevertheless, the effectiveness of KGs has long been hindered by the challenge of the incompleteness problem.

To address this issue, researchers have proposed a task known as Knowledge Graph Completion (KGC), which aims to predict missing relations and provides a valuable supplement to enhance KG's quality. Most existing KGC methods fall into two main categories: structure-based and pre-trained language model (PLMs)-based methods. Structure-based methods represent entities and relations as low-dimensional continuous embeddings, which effectively preserve their intrinsic structure [2, 4, 6, 12]. While

Q. Qiao and Y. Li—These authors contributed equally to this work.

effective in KG's structure representation learning, these methods overlook the semantic knowledge associated with entities and relations. Recently, PLM-based models have been proposed to leverage the semantic understanding captured by PLMs, adapting KGC tasks to suit the representation formats of PLMs [11, 17, 26, 27, 32].

While these models offer promising potential to enhance KGC performance, there is space to improve: (1) Existing structure-based methods do not explore knowledge provided by PLMs. (2) Existing PLM-based methods aim to convert KGC tasks to fit language model format and learn the relation representation from a semantic perspective using PLMs, overlooking the context of the relation in KGs. Consequently, they lack optimal alignment with structural knowledge. For example, given a triple (*trade name*, *member of domain usage*, *metharbital*)¹, the semantic of the relation *member of domain usage* is ambiguous since “it is not a standard used term in the English²”; hence, PLMs may lack accurate semantic representation. Thus, it becomes imperative to enable the model to leverage the principle of structural learning to grasp structural knowledge and compensate for the limitations of semantic understanding. (3) Existing PLM-based methods utilize PLMs directly, overlooking the disparity between PLMs and triples arising from the lack of triple training during PLMs pre-training. This oversight limits the expressive power of PLMs and their adaption to the KG's domain.

To address the limitations of existing methods, we propose a two-in-one framework named Bridge. To overcome the challenge of lacking structural knowledge in PLMs, we propose a structured triple knowledge learning phase. Specifically, we follow the widely applied principle in traditional structured representation learning for KGs [1, 2, 16, 19, 21], which posits that the relation is a translation from the head entity to the tail entity. We strategically extract the embedding of h , r and t separately from PLMs and employ various structure-based scoring functions to assess the plausibility of a triple. This approach allows us to reconstruct KG's structure in the semantic embedding via the structured learning principle. This principle has been widely applied in traditional structured representation learning for KGs, but there is no previous study that investigates this principle using PLM-based representation.

However, due to the different principles between traditional structured representation learning and PLMs, there is a gap between them since PLMs are not trained on KGs. To bridge the gap between PLMs and KGs, we fine-tune PLMs to integrate structured knowledge from KGs into PLMs. By taking this step, we unify the space of structural and semantic knowledge, making integration of KGs and PLMs more reasonable. In summary, our main contributions are:

1. We propose a general framework, Bridge, that jointly encodes structural and semantic information of KGs and can incorporate various scoring functions.
2. We utilize BYOL innovatively for fine-tuning PLM to bridge the gap between structural knowledge and PLMs.
3. We conduct empirical studies with two widely used structural-based scoring functions on three benchmark datasets. Experiment results show that Bridge consistently and significantly outperforms other baseline methods.

¹ This is a triple from WordNet, and metharbital is an anticonvulsant drug used in the treatment of epilepsy.

² Interpretation from ChatGPT when asking “what does *member of domain usage* mean?”.

2 Related Work

2.1 Structure-Based KGC

Structure-based KGC aims to embed entities and relations into a low-dimensional continuous vector space while preserving their intrinsic structure through the design of different scoring functions. Various knowledge representation learning methods can be divided into the following categories: (1) Translation-based models, which assess the plausibility of a fact by calculating the Euclidean distance between entities and relations [2, 6, 9, 21]; (2) Semantic matching-based models, which determine the plausibility of a fact by calculating the semantic similarity between entities and relations [1, 14, 15, 31]; and (3) Neural network-based models, which employ deep neural networks to fuse the graph network structure and content information of entities and relations [8, 12, 16, 19, 20, 24]. All these structure-based models are limited to using graph structural information from KGs, and they do not leverage the rich contextual semantic information of PLMs to enrich the representation of entities and relations.

2.2 PLM-Based KGC

PLM-based KGC refers to a method for predicting missing relations in KGs using the implicit knowledge of PLMs. KG-BERT [32] is the first work to utilize PLMs for KGC. It treats triples in KGs as textual sequences and leverages BERT [10] to model these triples. MTL-KGC [11] utilizes a multi-task learning strategy to learn more relational properties. This strategy addresses the challenge faced by KG-BERT, where distinguishing lexically similar entities is difficult. To improve the inference efficiency of KG-BERT, StAR [26] partitions each triple into two asymmetric parts and subsequently constructs a bi-encoder to minimize the inference cost. SimKGC [27] proposes to utilize contrastive learning to improve the discriminative capability of the learned representation. Adopting the architecture of SimKGC, GHN [17] develops an innovative self-information-enhanced contrastive learning approach to generate high-quality negative samples. MPIKGC [30] utilizes large language models (LLMs) to enrich the descriptions of entities/relations. In contrast to previous encode-only models, [3, 18] explore the generation-based models that directly generate a target entity. However, all these methods simply involve fine-tuning PLMs directly, disregarding both the absence of structured knowledge in PLMs and the gap between PLMs and KGs.

3 Preliminary

3.1 Bootstrap Your Own Latent (BYOL)

Bootstrap Your Own Latent (BYOL) is an approach to self-supervised image representation learning without using negative samples. It employs two networks, referred to as the *online* and *target* networks, working collaboratively to learn from one another. The *online* network is defined by a set of weights θ , while the *target* network shares the same architecture as the *online* network but utilizes a different set of weights ξ .

Given the image x , BYOL generates two augmented views (v, v') from the image x using different augmentations. These two views (v, v') are separately processed by the *online* and the *target* encoders. The *online* network produces a representation $\mathbf{y}_\theta = f_\theta(v)$ and a projection $\mathbf{z}_\theta = g_\theta(\mathbf{y}_\theta)$, while the *target* network outputs a representation $\mathbf{y}'_\xi = f_\xi(v')$ and a projection $\mathbf{z}'_\xi = g_\xi(\mathbf{y}'_\xi)$. Next, only the *online* network applies a prediction $q_\theta(\mathbf{z}_\theta)$, creating an asymmetric between the *online* and the *target* encoders. Finally, the loss function is defined as the mean squared error between the normalized predictions and target projections:

$$\mathcal{L}_{\theta, \xi} \triangleq \|\bar{q}_\theta(\mathbf{z}_\theta) - \bar{\mathbf{z}}'_\xi\|_2^2 = 2 - 2 \cdot \frac{\langle q_\theta(\mathbf{z}_\theta), \mathbf{z}'_\xi \rangle}{\|q_\theta(\mathbf{z}_\theta)\|_2 \cdot \|\mathbf{z}'_\xi\|_2}, \quad (1)$$

where $\bar{q}_\theta(\mathbf{z}_\theta)$ and $\bar{\mathbf{z}}'_\xi$ are the $l2$ -normalized term of $q_\theta(\mathbf{z}_\theta)$ and \mathbf{z}'_ξ .

To symmetrize the loss $\mathcal{L}_{\theta, \xi}$, BYOL swaps the two augmented views of each network, feeding v' to the *online* network and v to the *target* network to compute $\tilde{\mathcal{L}}_{\theta, \xi}$. During each training step, BYOL performs a stochastic optimization step to minimize $\mathcal{L}_{\theta, \xi}^{BYOL} = \mathcal{L}_{\theta, \xi} + \tilde{\mathcal{L}}_{\theta, \xi}$ with respect to θ only. ξ are updated after each training step using an exponential moving average of the online parameters θ as follows:

$$\xi \leftarrow \tau \xi + (1 - \tau) \theta, \quad (2)$$

where τ is a target decay rate.

3.2 Problem Definition

Knowledge Graph Completion The knowledge graph completion (KGC) task is to either predict the tail/head entity t/h given the head/tail entity h/t and the relation r : $(h, r, ?)$ and $(?, r, t)$, or predict relation r between two entities: $(h, ?, t)$. In this work, we focus on head and tail entity prediction.

4 Methodology

In this section, we present Bridge in detail. We first introduce a structure-aware PLM encoder, which aims to learn structure knowledge by PLMs. Then we introduce two essential modules in Bridge. The first module utilizes a fine-tuning process with BYOL to seamlessly integrate structural knowledge from KGs into PLMs, thereby bridging the gap between the two. The second module aims to learn structure-enhanced triple knowledge with PLMs, allowing PLMs to acquire domain knowledge of KGs. As shown in Fig. 1a, Bridge integrates these two modules by sequentially training two objectives. We take the tail entity prediction task $(h, r, ?)$ as an example to illustrate the procedure, and the procedure for the head entity prediction task $(?, r, t)$ is the same.

4.1 Structure-Aware PLMs Encoder

Existing structure-based and PLM-based methods can lead to suboptimal performance, especially when dealing with ambiguous relations. Hence, it is essential to incorporate

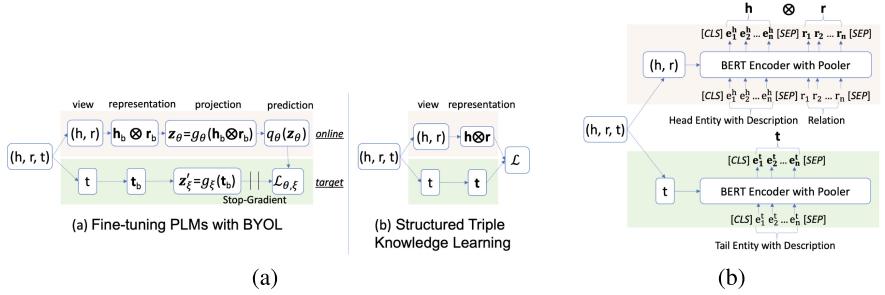


Fig. 1. (a) The framework of Bridge. \otimes represents different interaction strategies between entities and relations, determined by various scoring functions. (b) Structure-Aware PLM Encoder.

structural knowledge with semantic knowledge to achieve a structure-enhanced relation representation.

To facilitate structure representation learning, we use two BERT encoders. Given a triple (h, r, t) , the first encoder takes the textual description of the head entity h and relation r as input, where the textual description of the head entity h is denoted as $(e_1^h, e_2^h, \dots, e_n^h)$, and relation r is denoted as a sequence of tokens (r_1, r_2, \dots, r_n) , the input sequence is: $[CLS] e_1^h e_2^h \dots e_n^h [SEP] r_1 r_2 \dots r_n [SEP]$. The second encoder takes the textual description of the tail entity t as input, where the textual description of the tail entity t is denoted as a sequence of tokens $(e_1^t, e_2^t, \dots, e_n^t)$, the input sequence format is: $[CLS] e_1^t e_2^t \dots e_n^t [SEP]$. The design of these two encoders is illustrated in Fig. 1b. The embedding of h, r, t is computed by taking the mean pooling of the corresponding BERT output:

$$\begin{aligned} \mathbf{h} &= \text{MeanPooling}(\mathbf{e}_1^h, \mathbf{e}_2^h, \dots, \mathbf{e}_n^h), \\ \mathbf{r} &= \text{MeanPooling}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n), \\ \mathbf{t} &= \text{MeanPooling}(\mathbf{e}_1^t, \mathbf{e}_2^t, \dots, \mathbf{e}_n^t). \end{aligned} \quad (3)$$

To reconstruct KG's structure in the semantic embedding, we analyze two widely applied scoring functions in the KGC task, including TransE and RotatE. The corresponding structure scoring functions $\phi(h, r, t)$ are designed as follows:

$$\phi(h, r, t) = \phi(h \otimes r, t)_{TransE} = \cos(\mathbf{h} + \mathbf{r}, \mathbf{t}) = \frac{(\mathbf{h} + \mathbf{r}) \cdot \mathbf{t}}{\|\mathbf{h} + \mathbf{r}\| \|\mathbf{t}\|}. \quad (4)$$

$$\phi(h, r, t) = \phi(h \otimes r, t)_{RotatE} = \cos(\mathbf{h} \circ \mathbf{r}, \mathbf{t}) = \frac{(\mathbf{h} \circ \mathbf{r}) \cdot \mathbf{t}}{\|\mathbf{h} \circ \mathbf{r}\| \|\mathbf{t}\|}. \quad (5)$$

where \circ denotes the Hadamard (element-wise) product, and \otimes represents different interaction strategies between entities and relations. Note that Bridge is flexible enough to be generalized to other existing structure-based scoring functions.

4.2 Fine-Tuning PLMs with BYOL

Previous PLM-based approaches leverage PLMs directly and disregard the gap between structure knowledge and PLMs because PLMs are not trained on triples. Therefore,

strategic fine-tuning PLMs is necessary. Considering the existence of one-to-many, many-to-one, and many-to-many relations in KGs, we exclusively consider positive samples and adopt BYOL [7] as it does not require negative samples. We generate an alternative view of KG by separating a triple into two parts, and leveraging the widely used structural principles to learn KG information.

BYOL generates two augmented views of the same instance, with one view serving as the input for the *online* network and the other view serving as the input for the *target* network. Here, the *online* encoder takes the textual descriptions of the head entity h and relation r as input and produces an *online* representation $\mathbf{h}_b \otimes \mathbf{r}_b$. The *target* encoder takes the textual descriptions of the tail entity t as input and produces a *target* representation \mathbf{t}_b . The design of the encoder is elaborated in Sect. 4.1.

The *online* projection network g_θ takes the *online* representation $\mathbf{h}_b \otimes \mathbf{r}_b$ as input and outputs an *online* projection representation \mathbf{z}_θ :

$$\mathbf{z}_\theta = g_\theta(\mathbf{h}_b \otimes \mathbf{r}_b) = \mathbf{W}_2[\sigma(\mathbf{W}_1[\mathbf{h}_b \otimes \mathbf{r}_b])], \quad (6)$$

where \mathbf{W}_1 and \mathbf{W}_2 are trainable parameters, g_θ is a MLP network with one hidden layer, $\sigma(\cdot)$ is a PReLU function, and \otimes represents different interaction strategies between entities and relations, determined by various scoring functions.

The *target* projection network g_ξ takes the *target* representation \mathbf{t}_b as input and outputs a *target* projection representation \mathbf{z}'_ξ :

$$\mathbf{z}'_\xi = g_\xi(\mathbf{t}_b) = \mathbf{W}_4[\sigma(\mathbf{W}_3\mathbf{t}_b)], \quad (7)$$

where \mathbf{W}_3 and \mathbf{W}_4 are trainable parameters, g_ξ is a MLP network with one hidden layer, and $\sigma(\cdot)$ is a PReLU function.

The prediction network q_θ takes the *online* projection representation \mathbf{z}_θ as input and outputs a representation $q_\theta(\mathbf{z}_\theta)$ which is a prediction of the *target* projection representation \mathbf{z}'_ξ , the goal is to let the *online* network predict the *target* network's representation of another augmented view of the same triple:

$$q_\theta(\mathbf{z}_\theta) \approx \mathbf{z}'_\xi, \quad (8)$$

where q_θ is a MLP network with one hidden layer.

4.3 Structured Triple Knowledge Learning

To reconstruct KG's structures in the semantic embedding, after fine-tuning PLMs with BYOL, we employ the fine-tuned *online* encoder and the *target* encoder to facilitate structure learning. The *online* BERT encoder takes the textual description of the head entity h and the relation r as input. The *target* BERT encoder takes the textual description of the tail entity t as input. The structure scoring function $\phi(h, r, t)$ is utilized to train these two encoders further to incorporate structure knowledge into PLMs.

4.4 Objective and Training Process

During the Fine-tuning PLMs with BYOL phase, we optimize the PLMs for domain adaption in KGs using the loss $\mathcal{L}_{\theta,\xi}$, which is computed according to Eq. (1). The *online*

parameters θ are updated by a stochastic optimization step to make the predictions $q_\theta(\mathbf{z}_\theta)$ closer to \mathbf{z}'_ξ for each triple, while the target parameters ϕ are updated by Eq. (2). To symmetrize this loss, we also swap the input of the *online* and *target* encoder.

In the Structured Triple Knowledge Learning phase, we use contrastive loss with additive margin [27] to simultaneously optimize the structure and PLMs objectives:

$$\mathcal{L} = -\log \frac{e^{(\phi(h, r, t) - \gamma)/\tau}}{e^{(\phi(h, r, t) - \gamma)/\tau} + \sum_{i=1}^{|\mathcal{N}|} e^{(\phi(h, r, t'_i) - \gamma)/\tau}}, \quad (9)$$

where τ denotes the temperature parameter, t'_i denotes the i -th negative tail, $\phi(h, r, t)$ is the score function as in Eq. (4), Eq. (5), and the additive margin $\gamma > 0$ encourages the model to increase the score of the correct triple (h, r, t) .

5 Experimental Study

5.1 Datasets and Evaluation Metrics

We run experiments on three datasets: WN18RR [5], FB15k-237 [23], and Wikidata5M [29]. The statistics are shown in Table 1. We employ two evaluation metrics: Hits@K and mean reciprocal rank. Hits@K indicates the proportion of correct entities ranked in the top k positions, while MRR represents the mean reciprocal rank of correct entities.

Table 1. Statistics of the Datasets. Columns 2–6 represent the number of entities, relations, triples in the training set, validation set, and the test set, respectively.

Dataset	#Ent	#Rel	#Train	#Valid	#Test
WN18RR	40,943	11	86,835	3,034	3,134
FB15k-237	14,541	237	272,115	17,535	20,466
Wikidata5M-Trans	4,594,485	822	20,614,279	5,133	5,163

5.2 Baseline

We compare Bridge with two categories of baselines in Table 2. **Structure-based methods** aim to learn entity and relation embeddings by modeling relational structure in KGs. **PLM-based methods** aim to enrich knowledge representation by leveraging the semantic knowledge of PLMs but ignore the structural knowledge of KGs, and disregard the disparity between PLMs and KGs, as PLMs are not trained on KGs.

5.3 BridgeSetups

We use the bert-base-uncased model as the initialized encoder. In the fine-tuning PLMs with BYOL module, we train Bridge-TransE on WN18RR, FB15k-237, and Wikidata5M datasets for 2, 2, and 1 epoch(s), respectively. For Bridge-RotateE, we conduct training on the WN18RR, FB15k-237, and Wikidata5M datasets for 1, 2, and 1 epoch(s), respectively. The initial learning rates are $4 * 10^{-4}$, $3 * 10^{-5}$, $4 * 10^{-5}$. In the

Table 2. Main results. **Bold** represents the best results and underline denotes the runner-up results, \dagger cites the results from [27], * cites the results from original papers. - indicates that the original papers do not present results related to the corresponding dataset.

Model	WN18RR				FB15k-237				Wikidata5M-Trans			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
Structure-based Methods												
TransE \dagger	24.3	4.3	44.1	53.2	27.9	19.8	37.6	44.1	25.3	17.0	31.1	39.2
DistMult \dagger	44.4	41.2	47.0	50.4	28.1	19.9	30.1	44.6	-	-	-	-
ComplEx \dagger	44.9	40.9	46.9	53.0	27.8	19.4	29.7	45.0	-	-	-	-
RotatE \dagger	47.6	42.8	49.2	57.1	33.8	24.1	37.5	53.3	29.0	23.4	32.2	39.0
TuckER \dagger	47.0	44.3	48.2	52.6	35.8	26.6	39.4	54.4	-	-	-	-
CompGCN*	47.9	44.3	49.4	54.6	35.5	26.4	39.0	53.5	-	-	-	-
BKENE*	48.4	44.5	51.2	58.4	38.1	29.8	<u>42.9</u>	57.0	-	-	-	-
CompoundE*	49.1	45.0	50.8	57.6	35.7	26.4	39.3	54.5	-	-	-	-
SymCL*	49.1	44.8	50.4	57.6	37.1	27.6	41.1	56.6	-	-	-	-
MGTCA*	51.1	47.5	52.5	59.3	<u>39.3</u>	29.1	42.8	58.3	-	-	-	-
PLM-based Methods												
KG-BERT*	-	-	-	52.4	-	-	-	42.0	-	-	-	-
MTL-KGC*	33.1	20.3	38.3	59.7	26.7	17.2	29.8	45.8	-	-	-	-
StAR*	40.1	24.3	49.1	70.9	29.6	20.5	32.2	48.2	-	-	-	-
KGT5*	50.8	48.7	-	54.4	27.6	21.0	-	41.4	-	-	-	-
KG-S2S*	57.4	53.1	59.5	66.1	33.6	25.7	37.3	49.8	-	-	-	-
SimKGC*	67.1	58.5	73.1	81.7	33.3	24.6	36.2	51.0	35.3	30.1	37.4	44.8
SimKGC-SymCL*	65.7	54.6	70.9	79.1	32.4	23.5	35.4	50.4	-	-	-	-
GHN*	67.8	59.6	71.9	82.1	33.9	25.1	36.4	51.8	36.4	31.7	38.0	45.3
MPIKGC-S*	61.5	52.8	66.8	76.9	33.2	24.5	36.3	50.9	-	-	-	-
Bridge-TransE	69.4	<u>59.4</u>	74.7	85.9	38.0	31.6	41.2	57.4	45.4	40.2	47.8	55.6
Bridge-RotateE	67.3	58.3	<u>73.3</u>	<u>83.2</u>	40.3	<u>31.5</u>	43.2	<u>58.1</u>	46.2	41.1	48.3	<u>55.2</u>

structural triple knowledge learning module, we train Bridge-transE for 7, 10, and 1 epoch(s) on the respective datasets and Bridge-RotateE for 8, 10, and 1 epoch(s). The corresponding initial learning rates are $1 * 10^{-4}$, $1 * 10^{-5}$, $3 * 10^{-5}$. The batch size, additive margin γ of contrastive loss, and the temperature τ are consistent across all datasets, set as 1024, 0.02, and 0.05, respectively.

5.4 Overall Evaluation Results and Analysis

The performances of all models on three datasets are reported in Table 2. Compared with the best baseline results, the improvements obtained by Bridge-TransE in terms of MRR, Hits@3, and Hits@10 are 2.4%, 2.2%, 4.6% on WN18RR. Meanwhile, the improvements obtained by Bridge-RotateE remain competitive with GHN. On Wikidata5M-Trans dataset, both Bridge-TransE and Bridge-RotateE demonstrate substantial improvements. Compared to the best baseline, GHN, Bridge-TransE achieves increases of 24.7% in MRR, 26.8% in Hits@1, 25.8% in Hits@3, and 22.7% in Hits@10. Similarly, Bridge-RotateE achieves increases of 26.9% in MRR, 29.7% in Hits@1, 27.1% in Hits@3, and 21.9% in Hits@10, respectively. On FB15k-237, Bridge-RotateE achieves the best results in MRR and Hits@3, while Bridge-TransE exhibits comparable performance to the best baseline results in MGTCA. Considering that FB15k-237 is much denser (average degree is ~ 37 per entity) [27], MGTCA likely holds an advantage in utilizing abundant neighboring information for learning entity embeddings.

Table 3. Ablation study on WN18RR, FB15k-237 and Wikidata5M-Trans.

Model	WN18RR				FB15k-237				Wikidata5M-Trans			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
SimKGC	67.1	58.5	73.1	81.7	33.3	24.6	36.2	51.0	35.3	30.1	37.4	44.8
w/o structural-TransE	58.2	45.2	64.4	79.3	31.0	24.2	31.9	44.7	30.1	27.7	30.0	38.1
w/o BYOL-TransE	67.3	59.0	72.2	80.8	37.2	30.5	40.8	56.4	40.6	33.8	40.2	50.6
Bridge-TransE	69.4	59.4	74.7	85.9	38.0	31.6	41.2	57.4	45.4	40.2	47.8	55.6
w/o structural-RotateE	53.9	43.2	60.1	74.1	31.8	24.1	33.8	46.3	31.4	28.2	29.8	38.4
w/o BYOL-RotateE	65.4	57.2	70.8	79.6	39.6	30.8	42.7	57.3	41.1	34.0	41.5	50.8
Bridge-RotateE	67.3	58.3	73.3	83.2	40.3	31.5	43.2	58.1	46.2	41.1	48.3	55.2

Table 4. Case study on the tail entity prediction $(h, r, ?)$ task using the test set of Wikidata5M-Trans. The **Bold** font represents the true tail entity. Top 3 shows the first three tail entities that SimKGC and Bridge predicted, respectively.

Triple	SimKGC			Bridge		
	Rank	Top 3	Rank	Top 3	Rank	Top 3
<i>(rio pasion, mouth of the watercourse, Usumacinta river)</i>	119	<i>Golfo de Paria, El Golfo de Guayaquil, Yuma River</i>	2	<i>Tabasco River, Usumacinta river, tzala river</i>		
<i>(lewis gerhardt goldsmith, instance of, Human)</i>	11	<i>plant death, dispute, internet hoax</i>	1	Human , <i>Lists of people who disappeared, Strange deaths</i>		
<i>(cross country championships - short race, sport, Athletics)</i>	4	<i>Cross-country running, long distance race, Road run</i>	1	Athletics , <i>Tower running, Athletics at the Commonwealth</i>		

5.5 Ablation Study

To explore the effectiveness of each module, we conduct two variants of Bridge: (1) removing the structural Triple Knowledge Learning module (referred to as “w/o structural-TransE” and “w/o structural-RotateE”). For inference, we use the fine-tuned *online* BERT and *target* BERT to encode (h, r) and t , respectively, and rank the plausibility of each triple based on their cosine similarity (refer to Eq. (4) and Eq. (5)); (2) remove the Fine-tuning PLMs with BYOL module (referred to as “w/o BYOL-TransE” and “w/o BYOL-RotateE”). The results are summarized in Table 3.

Effectiveness of Structured Triple Knowledge Learning: Compared with Bridge-TransE and Bridge-RotateE, the results of “w/o structural-TransE” and “w/o structural-RotateE” reveal that removing the Structured Triple Knowledge Learning module results in notable decreases. This indicates that contrastive loss effectively distinguishes similar yet distinct instances. The objective of BYOL is to utilize a non-negative strategy to acquire a good initialization that can be applied in downstream tasks. Negative samples continue to play a crucial role in maintaining high performance in these downstream tasks [12, 22]. The limitation of relying solely on BYOL arises from the fact that while the non-negative strategy can effectively minimize the gap between representations of distinct views from the same object, it is unable to sufficiently distinguish and disentangle the representations of views originating from similar yet distinct objects.

Effectiveness of Fine-Tuning PLMs with BYOL: Comparing with Bridge-TransE and Bridge-RotateE, the results of “w/o BYOL-TransE” and “w/o BYOL-RotateE” reveal that removing the fine-tuning BERT with BYOL module results in notable decreases across all metrics in Wikidata5M-Trans, and a minor decline on both WN18RR and FB15k-237. This phenomenon illustrates the necessity for fine-tuning PLMs. While PLMs utilize vast, unlabeled corpora during training to construct a comprehensive language

Table 5. Error Analysis on the tail entity prediction $(h, r, ?)$ on WN18RR. The **Bold** represents the true tail entity. Top 3 shows the first three tail entities predicted by Bridge.

Triple	Rank	Top 3
<i>(position, hypernym, location)</i>	3	<i>region, space, location</i>
<i>(take a breather, derivationally related form, breathing time)</i>	1	<i>breathing time, rest, restfulness</i>
<i>(Africa, has part, republic of cameroon)</i>	14	<i>Eritrea, sahara, tanganyika</i>

model that embodies textual content, achieving competitive performance in particular tasks often requires an additional fine-tuning step. The results validate our previous speculation that abundant data is crucial for fine-tuning the model since Wikidata5M-Trans is larger than the other two datasets. Therefore, removing fine-tuning BERT with the BYOL module has a more significant negative impact on Wikidata5M-Trans. Compared with SimKGC, “w/o BYOL-TransE” and “w/o BYOL-RotateE” outperforms on FB15k-237 and Wikidata5M-Trans. On WN18RR, “w/o BYOL-TransE” outperforms SimKGC in Hits@1 and MRR while being comparable in Hits@3 and Hits@10. This illustrates that our structural scoring function can effectively reconstruct KG’s structures in the semantic embedding.

5.6 Case Study

As shown in Table 4, for the first example, the top three tail entities predicted by Bridge-TransE are rivers in Mexico and geographically close to the true tail entity ***Usumacinta river***. However, the top three tail entities SimKGC predicted are rivers in South America. In the second example, the relation *instance of* has ambiguous semantic interpretations. SimKGC cannot capture the semantics of this relation for this triple from the PLMs, resulting in incorrect predictions for the top three tail entities. Bridge-TransE can understand this relation from the structural perspective, allowing for better predictions. These two toy examples show that when the semantics of the relations are ambiguous, integrating structural knowledge can help to learn a better relation representation. In the third example, although Bridge-TransE predicts the true tail entity ***Athletics***, the prediction *Cross-country running* made by SimKGC can be regarded as correct. *Cross-country running* and ***Athletics*** are not mutually exclusive concepts. However, the evaluation metrics consider it an incorrect answer since the triple *(cross country championships - men’s short race, sport, Cross-country running)* is not present in KGs.

5.7 Error Analysis

As shown in Table 5, in the first example, Bridge-TransE ranks the true tail entity ***location*** as the third. However, the first two tail entities are correct based on human observation. In the second example, *rest* can also be a valid tail due to the fact that *rest* and ***breathing time*** are lexically similar concepts. In the third example, Bridge-TransE ranks the true tail entity ***republic of cameroon*** as 14th, attributed to the nature of the relation *has part*, which is a many-to-many relation. The first three tail entities predicted by Bridge-TransE are correct because they are all located in Africa. Drawing

from these observations, some predicted triples might be correct based on human evaluation. However, these triples might not be present in KGs. This false negative issue results in diminished performance.

5.8 Efficiency of Bridge

We run SimKGC³ on WN18RR and conduct an efficiency comparison with Bridge-TransE. Table 6 illustrates the model efficiency of Bridge-TransE and SimKGC on WN18RR with a batch size of 1024. In Bridge-TransE, the Fine-tuning PLMs with BYOL step converges in 2 epochs, and the Structured Triple Knowledge Learning step achieves convergence in 7 epochs (9 epochs in total). The total training time is 3550 s. SimKGC converges in 8 epochs, and the total training time is 3331 s. Consequently, the overall computational cost of Bridge is comparable with SimKGC.

Table 6. Model efficiency of Bridge-TransE and SimKGC on WN18RR.

Model	# Total Training Epoch	# Total Training Time
SimKGC	8	3331 s
Bridge-TransE	9	3550 s

6 Conclusion

In this paper, we introduce Bridge, which integrates PLMs with structure-based models. Since no previous study investigates structural principles using PLM-based representation, we jointly encode structural and semantic information of KGs to enhance knowledge representation. Further, existing work overlooks the gap between KGs and PLMs due to the absence of KG training in PLMs. To address this issue, we utilize BYOL to fine-tune PLMs. Experimental results demonstrate Bridge outperforms most baselines.

Acknowledgement. The work is supported in part by NSF-CAREER 2237831.

References

1. Balazevic, I., et al.: Tucker: tensor factorization for knowledge graph completion. In: EMNLP (2019)
2. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: NeurIPS (2013)
3. Chen, C., Wang, Y., Li, B., Lam, K.Y.: Knowledge is flat: a seq2seq generative framework for various knowledge graph completion. arXiv preprint [arXiv:2209.07299](https://arxiv.org/abs/2209.07299) (2022)
4. Dettmers, T.: Convolutional 2D knowledge graph embeddings. In: AAAI (2018)
5. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press (1998)

³ <https://github.com/intfloat/SimKGC>.

6. Ge, X., Wang, Y.C., Wang, B., Kuo, C.C.J.: Compounding geometric operations for knowledge graph completion. In: ACL (2023)
7. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, et al.: Bootstrap your own latent-a new approach to self-supervised learning. In: NeurIPS (2020)
8. Guan, S., Jin, X., Wang, Y., Cheng, X.: Shared embedding based neural networks for knowledge graph completion. In: CIKM (2018)
9. Ji, G., et al.: Knowledge graph embedding via dynamic mapping matrix. In: IJCNLP (2015)
10. Kenton, J.D.M.W.C., Toutanova, L.K.: Bert: pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT (2019)
11. Kim, B., Hong, T., Ko, Y., Seo, J.: Multi-task learning for knowledge graph completion with pre-trained language models. In: COLING (2020)
12. Kim, J.S., Ahn, S.J., Kim, M.H.: Bootstrapped knowledge graph embedding based on neighbor expansion. In: CIKM (2022)
13. Li, Y., Zhou, K., Qiao, Q., Wang, Q., Li, Q.: Re-examine distantly supervised NER: a new benchmark and a simple approach. In: COLING (2025)
14. Liang, K., et al.: Knowledge graph contrastive learning based on relation-symmetrical structure. IEEE Trans. Knowl. Data Eng. (2023)
15. Nickel, M., Tresp, V., Kriegel, H.P.: A three-way model for collective learning on multi-relational data. In: ICML (2011)
16. Qiao, Q., Li, Y., Zhou, K., Li, Q.: Relation-aware network with attention-based loss for few-shot knowledge graph completion. In: PAKDD. Springer (2023)
17. Qiao, Z., Ye, W., Yu, D., Mo, T., Li, W., Zhang, S.: Improving knowledge graph completion with generative hard negative mining. In: ACL 2023 (2023)
18. Saxena, A., Kochsiek, A., Gemulla, R.: Sequence-to-sequence knowledge graph completion and question answering. arXiv preprint [arXiv:2203.10321](https://arxiv.org/abs/2203.10321) (2022)
19. Shang, B., Zhao, Y., Liu, J., Wang, D.: Mixed geometry message and trainable convolutional attention network for knowledge graph completion. In: AAAI (2024)
20. Shang, C., Tang, Y., Huang, J., Bi, J., He, X., Zhou, B.: End-to-end structure-aware convolutional networks for knowledge base completion. In: AAAI (2019)
21. Sun, Z., Deng, Z.H., Nie, J.Y., Tang, J.: Rotate: knowledge graph embedding by relational rotation in complex space. In: ICLR (2018)
22. Thakoor, S., Tallec, C., Azar, M.G., Munos, R., Veličković, P., Valko, M.: Bootstrapped representation learning on graphs. In: ICLR (2021)
23. Toutanova, K., Chen, D., Pantel, P., Poon, H., Choudhury, P., Gamon, M.: Representing text for joint embedding of text and knowledge bases. In: EMNLP (2015)
24. Vashishth, S., Sanyal, S., Nitin, V., Talukdar, P.: Composition-based multi-relational graph convolutional networks. arXiv preprint [arXiv:1911.03082](https://arxiv.org/abs/1911.03082) (2019)
25. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Commun. ACM **57**(10), 78–85 (2014)
26. Wang, B., Shen, T., Long, G., Zhou, T., Wang, Y., Chang, Y.: Structure-augmented text representation learning for efficient knowledge graph completion. In: WWW (2021)
27. Wang, L., Zhao, W., Wei, Z., Liu, J.: Simkgc: simple contrastive knowledge graph completion with pre-trained language models. In: ACL (2022)
28. Wang, Q., Zhou, K., Qiao, Q., Li, Y., Li, Q.: Improving unsupervised relation extraction by augmenting diverse sentence pairs. In: EMNLP (2023)
29. Wang, X., et al.: Kepler: a unified model for knowledge embedding and pre-trained language representation. TACL (2021)
30. Xu, D., Zhang, Z., Lin, Chen, E.: Multi-perspective improvement of knowledge graph completion with large language models. arXiv preprint [arXiv:2403.01972](https://arxiv.org/abs/2403.01972) (2024)
31. Yang, B., Yih, S.W.T., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: ICLR (2015)

32. Yao, L., Mao, C., Luo, Y.: KG-BERT: BERT for knowledge graph completion (2020)
33. Zhou, K., Li, Y., Li, Q.: Distantly supervised named entity recognition via confidence-based multi-class positive and unlabeled learning. In: ACL (2022)
34. Zhou, K., Li, Y., Wang, Q., Qiao, Q., Li, Q.: Gendecider: integrating “none of the candidates” judgments in zero-shot entity linking re-ranking. In: NAACL (2024)