



Article

IPPT4KRL: Iterative Post-Processing Transfer for Knowledge Representation Learning

Weihang Zhang ^{1,*} , Ovidiu Șerban ¹ , Jiahao Sun ^{1,2} and Yike Guo ¹ ¹ Data Science Institute, Imperial College London, London SW7 2BX, UK² Royal Bank of Canada, London EC2M 1GT, UK

* Correspondence: w.zhang21@imperial.ac.uk

Abstract: Knowledge Graphs (KGs), a structural way to model human knowledge, have been a critical component of many artificial intelligence applications. Many KG-based tasks are built using knowledge representation learning, which embeds KG entities and relations into a low-dimensional semantic space. However, the quality of representation learning is often limited by the heterogeneity and sparsity of real-world KGs. Multi-KG representation learning, which utilizes KGs from different sources collaboratively, presents one promising solution. In this paper, we propose a simple, but effective iterative method that post-processes pre-trained knowledge graph embedding (IPPT4KRL) on individual KGs to maximize the knowledge transfer from another KG when a small portion of alignment information is introduced. Specifically, additional triples are iteratively included in the post-processing based on their adjacencies to the cross-KG alignments to refine the pre-trained embedding space of individual KGs. We also provide the benchmarking results of existing multi-KG representation learning methods on several generated and well-known datasets. The empirical results of the link prediction task on these datasets show that the proposed IPPT4KRL method achieved comparable and even superior results when compared against more complex methods in multi-KG representation learning.

Keywords: knowledge graph; knowledge representation; knowledge graph embedding; knowledge transfer; transfer learning



Citation: Zhang, W.; Șerban, O.; Sun, J.; Guo, Y. IPPT4KRL: Iterative Post-Processing Transfer for Knowledge Representation Learning. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 43–58. <https://doi.org/10.3390/make5010004>

Academic Editors: Alberto Simões and Pedro Rangel Henriques

Received: 12 December 2022

Revised: 28 December 2022

Accepted: 2 January 2023

Published: 6 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Knowledge Graphs (KGs) are often viewed as large-scale semantic networks that store facts as triples in the form of (*subject entity, relation, object entity*) or (*subject entity, attribute, value*). Typically, KGs are built on top of different existing data sources to connect data. In the industry settings, the downstream tasks and applications of KGs include, but are not limited to, knowledge acquisition, natural knowledge understanding, recommendation systems, question answering, etc. While KGs provide a sensible and effective way of managing knowledge, almost all of them suffer from heterogeneity and sparsity [1], including the most widely used large-scale KGs, such as DBpedia [2], Freebase [3], and YAGO [4]. Most, if not all, of the KG downstream tasks and applications rely on knowledge representation learning and operate under the Closed World Assumption (CWA) [5], which assumes that unobserved facts are false. Under the CWA, downstream tasks and applications usually perform worse on sparser KGs. Fortunately, real-world KGs constructed from different sources are sometimes complementary. Therefore, in this paper, we focused on the multi-KG representation learning problem and propose an intuitive post-processing method that facilitates knowledge transfer to a pre-trained KG to maximize the expressiveness of knowledge graph embeddings.

Link Prediction (LP), a task that predicts new links in a KG based on existing knowledge, has drawn increasing attention in recent years [6]. Recent advances in Knowledge graph Representation Learning (KRL) are showing very promising results

on this task, with the most-representative techniques being triple-based [7–10] and neighbour-based embedding methods [11,12]. While these methods perform well on the LP task, most work only on individual KGs and cannot harness the power of multiple KGs with potentially complementary information. In domains such as finance, where different data vendors are usually specialized in providing certain data types, performing link prediction and knowledge representation learning on multiple overlapping KGs can be very useful. To address this issue, several recent studies [13–16] have all focused on representation learning over multiple KGs. However, most of the existing works exhibit a relatively complex design in the knowledge transfer module, while they have been validated only on limited datasets without horizontal comparison with similar methods. This paper proposes a relatively simple and lightweight post-processing method for representation learning over multiple KGs while also providing benchmarking results against existing works on four different datasets.

This paper proposes an iterative post-processing method called IPPT4KRL to refine the pre-trained knowledge graph embedding on individual KGs with a small number of overlapping labels and extra triples from another KG. The overlapping labels between KGs are usually known as *seed alignments*, representing a small portion of human-labelled equivalent entity pairs between the KGs. Our method divides the multi-KG representation learning into pre-training and post-processing stages. Knowledge graph embeddings are trained during the pre-training stage on individual KGs to best capture “domain-specific” knowledge. At the same time, additional relevant information from another KG is iteratively introduced in the post-processing stage to maximize knowledge transfer. During the post-processing stage, seed alignments are utilized to regularize the distance between equivalent entities in the embedding space. The pre-trained embedding space is post-processed with extra information introduced by the new KG. In contrast, an extra loss term controls the embedding space deviation from the original pre-trained embedding space. The extra information (triples) from a separate KG is introduced iteratively based on the adjacency to the seed alignments. As illustrated in Figure 1, entity *Virginia* and entity *Washington D.C.* are the two entities in KG_2 with the equivalent counterparts in KG_1 . Therefore, in the first iteration of the post-processing stage, triples (*Washington D.C.*, *AdjacentTo*, *Virginia*) and (*White House*, *LocateIn*, *Washington D.C.*) are included. However, triple (*Joe Biden*, *WorksAt*, *White House*) is not included in the training until the second iteration because of the adjacency to aligned entities. To validate and assess the knowledge transfer induced by our proposed method in a life-like scenario, we generated a new dataset based on entity alignment benchmarking dataset D-W-15K [17] with the sampling method introduced in the work of Sun et al. [18], to create dangling entities (entities without alignment across KGs) within KGs. Triples in the sampled KGs are further split into training, testing, and validation sets for the embedding evaluation with the LP task. We also identified several existing methods of multi-KG representation learning and benchmarked the results against our proposed method across multiple datasets. To the best of our knowledge, this is the first work that provides benchmarking results with multiple relevant works on multi-KG representation learning under the same problem set. Experimental results on the generated dataset and three other datasets showed that our relatively lightweight IPPT4KRL method can achieve comparable and superior results against state-of-the-art models. A more detailed description of our IPPT4KRL method is provided in Section 3, and the experimental results on the empirical datasets and the ablation studies are shown in Section 4.

In summary, the main contributions of this paper are the following:

- A novel iterative post-processing method (IPPT4KRL) for knowledge graph representation learning and knowledge transfer over multiple KGs;
- A benchmark of state-of-the-art models in multi-KG representation learning;
- A newly generated empirical dataset with an existing sampling technique for the multi-KG KRL task that simulates “in the wild” conditions with lower overlaps between KGs.

Extensive ablation studies were conducted to analyse the importance and effects of each component in the proposed method.

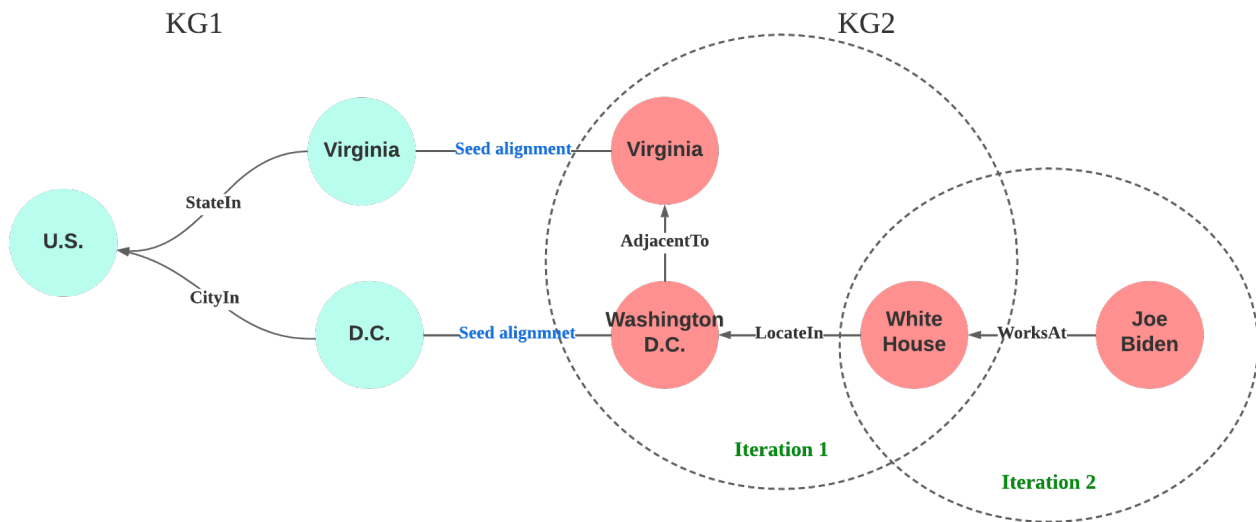


Figure 1. An illustrative figure on how triples from KG_2 are iteratively included in the post-processing.

2. Related Work

2.1. Knowledge Representation Learning

Knowledge Representation Learning (KRL) for generating knowledge graph embedding has recently gained tremendous attention. This task aims to encode the entities and relations of KGs into low-dimensional vectors. The popular techniques involved in KRL can be broadly categorized into two groups, *triple-based embedding* and *neighbour-based embedding*.

Triple-based embedding exploits the local structural information within triples and captures the semantics of relation triples. Triple-based embeddings can be further categorized into translation-based distance models and semantic matching models. Translation-based distance models view the relation as translation in the vector space, while semantic matching models use the scoring function based on semantic similarity to mine the potential semantic association between entities and relationships.

The most widely used translation-based embedding is *TransE* [7], in which both entities and relations are represented in the same embedding space and relations are represented as translations in the embedding space. Specifically, given a triple (h, r, t) , where h , r , and t denote head, relation, and tail, respectively, TransE assumes that the sum of head embedding and relation embedding should be close to the tail embedding, i.e., $\vec{h} + \vec{r} \approx \vec{t}$. Inspired by TransE while trying to improve its flexibility, many translation-based knowledge embedding methods have been proposed, including but not limited to TransH [19], TransR [8], TransD [9], etc. Recently, some translation-based embedding methods such as RotatE [20] and HAKE [21] have been proposed on a polar coordinate system, where relations are represented as rotations from the source entities to the target entities in the complex vector space.

Another popular and widely used triple-based embedding technique is called the semantic matching model. Nickel et al. [22] was one of the first to propose a semantic-matching-based approach, RESCAL, for embedding KGs. RESCAL uses the tensor factorization method to model the inherent structure of KGs in the model. Using ideas from RESCAL as a foundation, DistMult [10] and HolE [23] were then proposed to improve the performance and scalability of semantic matching models. Many other embedding models such as ANALOGY [24], ComplEx [25], TuckER [26], and QuatE [27] also fall into the category of semantic matching models and have shown promising results in embedding-based tasks such as link prediction and node classification.

Neighbour-based embedding employs the graph structure to propagate information between adjacent entities and encode them into embeddings. The learnt entity embeddings are able to capture additional structural information about their neighbouring entities and relations. R-GCN [12], as a variant of the Graph Convolutional Network (GCN) [11], is one of the most widely used graph neural network models that encodes relational knowledge graphs into embeddings. R-GCN [12] employs an encoder–decoder structure for the link prediction task. In the encoder, entity embeddings are generated with the message propagation mechanism of the R-GCN. At the same time, in the decoder, knowledge assumptions such as TransE and DistMult are applied to generate relation embeddings and perform link prediction tasks. The analysis of neighbour-based embedding techniques across multiple KGs is beyond the scope of this work due to the extra complexity of message passing in the encoding phase. However, this could be an important future direction as we explore representation learning across multiple KGs.

2.2. Multi-KG Knowledge Representation Learning

Compared to KRL on individual KGs, KRL across multiple KGs is a relatively under-explored area. Several works have been proposed under similar problem settings to tackle this problem.

Trivedi et al. [13] proposed LinkNBed, a deep relational learning framework that can perform representation learning across multiple KGs. Basic vector representations are created for entities, relations, types, and attributes in a KG and then aggregated by a set of aggregator functions to incorporate this contextual information into a deep representation. In the training process, LinkNBed combines deep representation learning across KGs with the entity alignment task, jointly optimizing both tasks.

Sourty et al. [14] proposed KD-MKB, a model that tackles the multi-graph representation learning from the knowledge distillation perspective. Concretely, KD-MKB mutually and jointly trains the knowledge graph embedding for multiple KGs, where each KG learns embeddings with its own context. A mimicry knowledge distillation process between KGs achieves the knowledge exchange. Unlike traditional one-way knowledge distillation, where the role of student and teacher is fixed, in KD-MKB, each KGs takes turns becoming a teacher and student in the distillation process. However, KD-MKB assumes both entity alignment and relation alignments are available between KGs when performing mutual knowledge distillations, which limits the method's applicability to some extent.

Additionally, Wang et al. [15] proposed ATransN, an Adversarial embedding Transfer Network that aims to facilitate the knowledge transfer from a pre-trained embedding of a teacher KG to a student KG with a set of seed alignments. Like most other methods with similar problem settings, ATransN aims to facilitate knowledge transfer between KGs with multi-task training of knowledge representation and entity alignment. In ATransN, an adversarial adaption module is employed to handle the potential distribution difference between the teacher KG and student KG. The discriminator of the adversarial module is used to assign a consistency score for a pair of seed alignments. This score is later used to weigh the contribution of alignment constraints imposed by seed alignments.

Most of the existing methods in multi-KG KRL exhibit a fairly complicated design in the knowledge transfer module, and the training can sometimes become very time-consuming. We hope to demonstrate a relatively simple method that does not require a complex design, but uses pre-training and post-processing strategies and knowledge representation learning to achieve the same knowledge transfer between KGs.

3. Iterative Post-Processing

3.1. Preliminaries

The framework of multi-KG knowledge representation learning tasks involves two or more KGs. Without loss of generality, we considered the multi-KG KRL task between two KGs. We formalized two heterogeneous KGs as $(KG_1 = \{E_1, R_1, T_1\}, KG_2 = \{E_2, R_2, T_2\})$,

where E_i, R_i, T_i , respectively, represent the entity set, relation set, and fact triple set of the KG. A small set of seed alignments between KGs, known before training, is denoted by $S_{KG_1, KG_2} = \{(e_1, e_2) : (e_1, e_2) \in E_1 \times E_2, e_1 \sim e_2\}$, where \sim denotes the equivalence relation. Let $\{h_{e_i}, h_{r_i}\}$ denote the embeddings for entity e_i and relations r_i , respectively. In our problem setting, we pre-trained the entity and relation embeddings of KG_1 under an arbitrary triple-based embedding model, such as TransE [7]. The post-processing stage aims to facilitate the knowledge transfer between KGs and improve the knowledge graph embedding of KG_1 when triple facts T_2 and seed alignment S_{KG_1, KG_2} are introduced. The improvement of knowledge graph embedding can be evaluated with a link prediction task, discussed in Section 4.

3.2. Methods

3.2.1. The Iterative Inclusion of New Triples

We pre-trained knowledge graph embeddings for KG_1 as our base for post-processing. The pre-trained embeddings on KG_1 were further processed by including triples T_2 from KG_2 in an iterative manner based on the adjacency to seed alignments, as demonstrated in Figure 1. Formally, for a node $e_i \in E_2$, we considered $N_K(e_i) = \{e_j \in E_2 : d(e_i, e_j) = K\}$ as the K_{th} -hop neighbouring entities around node $e_i \in E_2$, where $d : |E_2| \times |E_2| \rightarrow \mathbb{N}$ denotes the shortest path distance between nodes in a KG. We then define $T_{N_K}(e_i) = \{(h, r, t) : t \in N_{K-1}(e_i), r \in R, h \in N_K(e_i)\}$ as the set of triples involved in connecting $(K-1)_{th}$ -hop neighbours and K_{th} -hop neighbours of entity e_i .

Let S_{KG_1} and S_{KG_2} denote the set of KG_1 and KG_2 entities in seed alignment S_{KG_1, KG_2} and T_f denote the set of triples used for multi-KG embedding training. Therefore, T_f was set to be T_1 during the pre-training process of KG_1 . During post-processing, we first set $T_f = T_1 \cup T_{N_1}(S_{KG_2})$, where $T_{N_1}(S_{KG_2}) = \{T_{N_1}(e_i) : e_i \in S_{KG_2}\}$ denotes the triples that connect aligned entities S_{KG_2} and their one-hop neighbours. The embeddings are then trained under the same triple-based embedding model assumption until convergence. After that, we set $T_f = T_1 \cup T_{N_1}(S_{KG_2}) \cup T_{N_2}(S_{KG_2})$ to include the triples in KG_2 that connect 1-hop neighbours and 2-hop neighbours of aligned entities S_{KG_2} . Then, the same training process on the embedding space is repeated. In practice, the same process is repeated k times, where k is a hyper-parameter.

3.2.2. Loss Functions for Embedding Training

Three different losses are optimised in the iterative post-processing of the original embedding space.

Triple-based margin loss: During the post-processing, the triple-based margin loss is used for optimizing the knowledge representation task. The scoring function we used in the post-processing stage is consistent with the triple-based loss function used to pre-train embeddings for KG_1 . The loss function is written as:

$$L_T = \sum_{t_i \in T_f, t'_i \in T'_f} E(t_i) - E(t'_i) + \gamma, \quad (1)$$

where T'_f denotes the negative samples created from corrupting the head or tail entity in triple t_i ; $E(t_i)$ denotes the energy function for TransE or a similar triple-based embedding model for triple t_i ; γ denotes the margin, which is a hyper-parameter describing the ideal distance between positive triples and negative triples. To better facilitate knowledge transfer, we also enabled the option to include extra training triples in the post-processing stage using *parameter swapping*: for triple (e_1, r, e'_1) , if $(e_1, e_2) \in S_{KG_1, KG_2}$, then triple (e_2, r, e'_1) will be included in set T_f for training knowledge representation.

Alignment loss: The alignment loss is used to ensure the learnt embeddings of the same entity in different KGs are close. Alignment loss can also be viewed as a regularization

technique when training knowledge representation across KGs. The alignment in the post-processing can be written as:

$$L_A = \sum_{(e_1, e_2) \in S_{KG_1, KG_2}} d(h_{e_1}, h_{e_2}), \quad (2)$$

where $d(h_{e_1}, h_{e_2})$ denotes the distance function between embeddings of entities e_1 and e_2 . The distance function can be either directly computed by using a typical distance function (i.e., Manhattan, Euclidean, etc.) or derived from cosine similarity using $d(h_{e_1}, h_{e_2}) = 1 - \cos(h_{e_1}, h_{e_2})$.

Embedding space preservation loss: One of the important assumptions we make in this post-processing training is that the original pre-trained embedding space of KG_1 already captures the local semantic information in KG_1 relatively well. Therefore, when including new fact triples and alignments from KG_2 , we introduced this embedding space preservation loss to prevent the post-processed embedding space from deviating too much from the originally trained embedding space. The embedding space preservation loss can be written as:

$$L_P = \sum_{e_i \in E_1} d(h_{e_i}, h_{e'_i}), \quad (3)$$

where e'_i denotes the pre-trained embedding for entity e_i in KG_1 .

3.2.3. Training

The overall loss term optimized during the training is the combination of the three loss terms discussed above:

$$L = L_T + \alpha L_P + \beta L_A, \quad (4)$$

where α and β are the hyper-parameters controlling the contributions of embedding space preservation loss and alignment loss to the overall loss during training.

At each iteration of inclusion of new triples $T_{N_k}(S_{KG_2})$ into the post-processing training set T_f , the embedding post-processing is conducted until convergence. In practice, the convergence is determined by the decreased link prediction performance of embeddings on the validation set. Moreover, the embedding space we produced at the end of the post-processing iteration is then used as a base to further post-process with newly included triples in the next iteration.

4. Experiments

4.1. Datasets and Experiment Settings

Environment: We carried out the experiments on a desktop computer running Ubuntu Linux with an AMD Ryzen 3700x 3.6Ghz CPU, 32GB of RAM memory, and an NVIDIA GeForce GTX 1070 GPU (8GB).

Datasets: Our experiments were conducted on four different datasets: two mono-lingual datasets, D-W-15K-LP(generated) and DBP-FB [28]; and two multi-lingual datasets, CN31 (EN-DE) and WK31-15k (EN-FR) [29]. All of the datasets we chose were originally designed for benchmarking entity alignment algorithms; thus, seed alignments between KGs are readily available. The detailed statistics about the datasets can be found in Table 1. In D-W-15K-LP and DBP-FB, 70% of the triples are reserved for the training set, 10% for the validation set, and 20% for the test set. In CN31 (EN-DE) and WK31-15k (EN-FR), the percentages are 60%, 20%, and 20% for the training, validation, and test sets, respectively. Relation alignments are available in the two multilingual datasets, but not in the two mono-lingual datasets.

Table 1. Dataset statistics; alignment ratio denotes the ratio of seed alignments against the total number of alignments.

Dataset Name	KG1	KG2	Entities	Relations	Triples	Alignment Ratio	Shared Relations
D-W-15K-LP	DBpedia	Wikidata	12,750	160	55,038	24.71%	✗
		Wikidata	12,750	118	61,806	24.71%	✗
DBP-FB	DBpedia	Freebase	29,861	406	96,414	51.32%	✗
		Freebase	25,542	882	111,974	59.99%	✗
CN3I	ConceptNet (DE)	ConceptNet (EN)	4302	7	12,780	90.84%	✓
		ConceptNet (EN)	4316	43	32,528	93.67%	✓
WK3I-15k	Wikipedia (FR)	Wikipedia (EN)	15,393	2422	170,605	15.97%	✓
		Wikipedia (EN)	15,169	2228	203,502	16.45%	✓

D-W-15K-LP generation: D-W-15K-LP is a dataset generated from the entity alignment benchmarking dataset D-W-15K [17]. The sources of the two KGs are DBpedia and WikiData, with 15,000 entities in each KG. However, because D-W-15K was generated for the entity alignment task, every entity has an alignment across KGs. We argue that such a scenario is extremely rare in the real world because KGs from different sources rarely exhibit a large proportion of overlaps, and seed alignment annotations between KGs are very expensive, given the size and scale of real-world KGs. Therefore, we decided to employ the sampling strategy proposed in the work of Sun et al. [18] to create dangling entities (entities without alignment across KGs) within KGs. In the sampling process, triples containing removed entities are excluded by removing part of the alignments from KGs. This results in a more sparse dataset with dangling entities in KGs. Concretely, we created 30% dangling entities in our base dataset D-W-15K; only 30% of the remaining 70% aligned entities (3150 entities) were used as seed alignments to create a more life-like scenario for the experiment.

Settings: The problem set of the experiment is consistent with what we discussed in Section 3: for knowledge graphs KG_1 and KG_2 , the goal is to improve the knowledge graph embedding and its performance on LP tasks for KG_1 with information from KG_2 and seed alignments between them. We chose to compare our IPPT4KRL method against ATransN [15] and MD-MKB [14]. To the best of our knowledge, these two open-source methods achieve state-of-the-art performance in multi-KG KRL settings. In addition, two baseline models, *Individual* and *Connected*, were also included in the experiment. For the baseline *Individual*, we simply trained the knowledge graph embeddings on KG_1 and evaluated the embeddings with their performance on the LP task with test triples in KG_1 . The baseline *Individual* was also used as the pre-trained embeddings for the post-processing stage of our IPPT4KRL method. The second baseline *Connected* was generated by connecting two KGs with seed alignments. Concretely, if there exists entity seed alignment (e_1, e_2) , where $e_1 \in E_1$ and $e_2 \in E_2$, we then merge two entities and replace all entity occurrences e_2 in triples T_2 with entity e_1 . Knowledge graph embeddings are then trained on the connected KG and evaluated on the LP task with test triples in KG_1 . We used TransE as the knowledge graph embedding model for all methods in the experiments for the fairness of comparison. However, all the multi-KG KRL methods in the experiments can be extended to incorporate other triple-based embedding methods for knowledge representation. The pre-training and baseline experiments were conducted using the OpenKE framework [30] with uniform negative sampling.

In the D-W-15K-LP dataset, the DBpedia KG is KG_1 , and the Wikidata KG is added into the post-processing following an iterative manner. In DBP-FB, we chose the DBpedia KG to be KG_1 and the Freebase KG to be KG_2 . The German and French KGs were selected to be

KG_1 for CN3l (EN-DE) and WK3l-15k (EN-FR), respectively, while the English KGs of each dataset were included to provide knowledge transfer in the IPPT4KRL. Because ATransN employs a teacher–student setting in the training, naturally, KG_1 in our setting was regarded as the “student” KG in ATransN. KD-MKB treats each KG equally. Hence, no special configuration was needed in our benchmarking experiments.

Hyper-parameters: For the fairness of comparison, we set the embedding dimension to be the same across all models for a dataset. The embedding size dimension n was set to be 100 for D-W-15K-LP, 200 for DBP-FB, and 200 for CN3l (EN-DE) and Wk3l-15k (EN-FR) (CN3l(EN-DE) and Wk3l-15k (EN-FR) were also reported in the experiments in the ATransN paper. Thus, we set the embedding dimension to be 200 for the ease of reproducing their best model performance on these two datasets). lr_{ke} denotes the learning rate for the overall knowledge representation. The best learning rate for D-W-15K-LP (DBP-FB, WK3l-15K, CN3l) is $1e^{-3}(1e^{-2}1e^{-5}, 1e^{-3})$. For D-W-15K-LP, the best-performing IPPT4KRL model employed hyper-parameters $\gamma = 5.0$, $\alpha = 1$, $\beta = 0.1$, and $k = 2$; for DBP-FB, $\gamma = 12.0$, $\alpha = 2$, $\beta = 1$, and $k = 2$; for WK3l-15K, $\gamma = 8.0$, $\alpha = 1$, $\beta = 0.1$, and $k = 2$; and for CN3l, $\gamma = 16.0$, $\alpha = 10$, $\beta = 0.1$, and $k = 1$. In the iterative inclusion of the k -hop neighbours of KG_2 , the triple-based margin loss would consider more entities as k increases. As a result, triple-based margin loss increases the most every time the neighbour size is increased. We tried to assign different α for each hop, to account for the change in the trade-off between loss terms as k iteratively increases. However, in our experiment, assigning different α for each hop only provided marginal gains in performance.

Evaluation: We used a link prediction task to evaluate and compare the performance of the trained embeddings. The Entity Ranking (ER) protocol was employed for this evaluation: for a test triple (h, r, t) , the ER protocol uses the embedding model to rank all possible answers to link prediction queries $(?, r, t)$ and $(h, r, ?)$ and employs the rank of the correct answer for embedding evaluation. The standard filtered Hit@ m ($m = 1, 3, 10$), Mean Ranks (MRs), and Mean Reciprocal Rank (MRR) metrics are reported in the result tables. The reported results were averaged across multiple runs of the fine-tuned models. (Although we used the two datasets from the ATransN paper in our evaluation, we do not report the results from the ATransN paper. This is because we found a small issue in the open-sourced ATransN code for computing the filtered metrics on the link prediction tasks. Therefore, we followed the hyper-parameters provided in the ATransN paper to generate the embeddings and evaluate them against the corrected, filtered metrics.) The best entries for each metric are highlighted in bold in each of the table.

4.2. Results and Analysis

From the result Tables 2 and 3, we can observe that our IPPT4KRL model achieved comparable and even superior results against the best performers in each dataset. Table 2 shows the experimental results on the generated D-W-15K-LP and DBP-FB. On these two monolingual datasets with no relation alignment available, we extended the original KD-MKB and created KD-MKB*, in which we share the relation knowledge between embedding models of each KG. A more detailed description of the modification can be found in the Appendix B. On D-W-15K-LP, IPPT4KRL outperformed all the baseline models across all metrics. Compared to other multi-KG KRL methods, IPPT4KRL achieved the best performances on the MRR, Hit@1, and Hit@3 while producing a very similar performance against the top performer KD-MKB* on the MR and Hit@10. An important aspect is that IPPT4KRL achieved this comparable performance while requiring less than $\frac{1}{10}$ of the training time of KD-MKB*. Concretely, one complete run of the training process for KD-MKB* on this dataset took 12 hrs on our workstation, while the training for IPPT4KRL only took 1.5 hr, including the pre-training of individual embeddings on KG_1 . D-W-15K-LP is a scenario that was deliberately generated to mimic real-life mono-lingual multi-KG learning. The significant margin that IPPT4KRL produced against the baselines indicated that our model could facilitate positive knowledge transfer even when the seed alignment ratio is low. At the same time, ATransN’s performance on this dataset was unsatisfactory. One

possible reason is that ATransN usually performed well when the teacher KG in their model holds richer information than their student KG, which is not the case for our generated dataset D-W-15K-LP. On DBP-FB, we observed similar performance trends: IPPT4KRL achieved the best performance on the MRR, MR Hit@1, and Hit@3 metrics, while having comparable performance against the top performers on Hit@10. An interesting observation is that, although the alignment ratio was larger in DBP-FB than in D-W-15K-LP, the margins gained by the multi-KG KRL methods were smaller and inconsistent. This indicates that is is “harder” to transfer knowledge between KGs for the dataset DBP-FB. We believe the main driver behind this observation is that DBpedia and Freebase, two KGs, are constructed by two isolated parties, while for D-W-15K-LP, DBpedia and Wikidata practically come from very similar sources. This observation provided a meaningful indicator for our plans for the next steps, which we will discuss more in Section 6.

Table 2. Results on the D-W-15K-LP and DBP-FB datasets, with KD-MKB* representing the modified version of KD-MKB on the datasets without relation alignments.

	D-W-15K-LP					DBP-FB				
	MRR	MR	Hit@1	Hit@3	Hit@10	MRR	MR	Hit@1	Hit@3	Hit@10
Individual	0.3417	577	23.88%	40.54%	52.15%	0.3509	523	25.04%	40.93%	52.49%
Connected	0.3500	750	24.39%	41.42%	53.82%	0.3508	497	24.50%	41.46%	53.36%
KD-MKB*	0.3748	309	27.46%	43.36%	55.19%	0.3439	646	21.65%	42.86%	56.35%
ATransN	0.3361	321	23.29%	39.45%	52.26%	0.3368	584	24.85%	38.49%	49.07%
IPPT4KRL	0.3764	335	27.64%	43.71%	55.16%	0.3782	397	27.58%	43.85%	55.67%

Table 3. Results on the WK3l-15K and CN3l datasets.

	WK3l-15K(EN-FR)					CN3l (EN-DE)				
	MRR	MR	Hit@1	Hit@3	Hit@10	MRR	MR	Hit@1	Hit@3	Hit@10
Individual	0.3819	449	28.17%	42.87%	58.10%	0.1796	708	6.28%	25.84%	37.25%
Connected	0.3793	518	27.92%	42.44%	58.04%	0.2014	670	6.75%	30.14%	40.57%
KD-MKB	0.3856	554	27.48%	44.34%	59.50%	0.2093	494	6.70%	30.58%	45.56%
ATransN	0.3734	417	27.78%	41.37%	56.54%	0.1840	507	1.86%	31.14%	44.76%
IPPT4KRL	0.3823	446	28.22%	42.86%	58.10%	0.2245	473	10.31%	30.44%	42.78%

Table 3 shows the experiment results on CN3l (EN-DE) and WK3l-15K (EN-FR). On CN3l (EN-DE), we can observe that IPPT4KRL outperformed the rest of the models on the MRR, MR, and Hit@1 metrics, while also achieving similar performance on the other metrics compared to the top performers. CN3l (EN-DE) has a relatively large alignment ratio compared to the other three datasets, which usually implies a smaller difference between KGs. IPPT4KRL was also able to match and even outperform the top performers in this case as well. WK3l-15K is one of the harder datasets for multi-KG KRL in the work of Wang et al. [15], mainly because of (1) the lack of alignments and (2) the relatively dense and rich information already existing in the French KG. From Table 3, we can still observe a similar trend that IPPT4KRL achieved a similar performance level compared to other multi-KG KRL methods. However, on WK3l-15K (EN-FR), the margins between all multi-KG KRL methods and the baselines were minimal, which is fairly consistent with the findings in the ATransN paper. The fact that WK3l-15K contains many more relations than the other two datasets might also contribute to this result.

An interesting observation from the performance of the baseline *Connected* is that it outperformed the baseline *Individual* by a significant margin on the CN3l dataset, but the same trend was not observed on the generated D-W-15K-LP. Compared to the CN3l dataset, D-W-15K-LP has a relatively smaller alignment ratio and several aligned entities not visible to the models during training. We view the difference in the baseline performances as

evidence showing that the D-W-15K-LP dataset provides a more “life-like” scenario and is better suited for benchmarking multi-KG representation learning.

To test the generality of IPPT4KRL, we also experimented with our method on the D-W-15K-LP dataset with the roles of KG_1 and KG_2 flipped. From Table 4, we can see that after flipping the KGs, our method still consistently outperformed all of the baselines, showing consistent success in facilitating positive knowledge transfer across KGs in both directions.

Table 4. Results on the D-W-15K-LP dataset, reversing the order of KG_1 and KG_2 .

	MRR	MR	Hit@1	Hit@3	Hit@10
Individual	0.3215	598	23.29%	37.46%	47.51%
Connected	0.3263	649	23.48%	38.10%	48.33%
IPPT4KRL	0.3453	410	25.98%	39.35%	49.50%

5. Ablation Studies

5.1. Contribution of Each Component

In Table 5, we report the results from ablation studies to better understand each component’s contribution to the proposed framework. We conducted ablation studies on datasets D-W-15K-LP and CN3L(EN-DE), where our method showed substantial learning from the baseline models. In the table, *no-preservation* refers to when the embedding space preservation loss was removed from the proposed framework; *non-iterative* refers to when the iterative inclusion of new triples based on neighbouring order was removed and instead included all new information at once; and *no-pretraining* refers to when the pre-training stage of KG_1 was omitted. Both KGs were trained from scratch with uniform negative sampling. Overall, the proposed post-processing method outperformed all ablation models across all metrics with exceptions on the Hit@3 and Hit@10 metrics on CN3L(EN-DE). It can be observed that the embedding space preservation loss provided a more considerable gain in performance when the alignment ratio between KGs was high (CN3L). In contrast, iterative inclusion provided a larger margin in performance when the alignment ratio was relatively low (D-W-15K-LP). In practice, a higher alignment ratio implies more new entities and triples being introduced to the pre-trained embedding space in the first iteration of post-processing. In a high-alignment-ratio scenario, preservation loss showed effectiveness in preventing the embedding space from being “overwhelmed” by a large amount of new information, while in a low-alignment-ratio scenario, iterative inclusion showed effectiveness in restricting the model to learn from the relatively more important information first. From the results, we can also observe that the inclusion of pre-training also had a substantial gain over the performance. We believe this rationale is very similar to our iterative inclusion strategy. In the pre-training stage, the model could focus on individual KGs before utilizing extra information, which helped the convergence to a better local minimum.

Table 5. Ablation study results.

	D-W-15K-LP					CN3L (EN-DE)				
	MRR	MR	Hit@1	Hit@3	Hit@10	MRR	MR	Hit@1	Hit@3	Hit@10
IPPT4KRL	0.3764	335	27.64%	43.71%	55.16%	0.2245	473	10.31%	30.44%	42.78%
No-preservation	0.3716	350	27.16%	43.12%	54.90%	0.1971	493	5.30%	29.85%	43.31%
Non-iterative	0.3694	421	26.96%	43.09%	54.48%	0.2106	490	8.53%	29.57%	41.75%
No-pretraining	0.3303	442	22.28%	39.93%	51.69%	0.2110	565	8.06%	30.89%	41.86%

5.2. Varying Seed Alignment Ratio

We experimented with different seed alignment ratios to better understand the nature of knowledge transfer between KGs. Figure 2 compares the best performances between IPPT4KRL and the baseline *Connected* with varying alignment ratios on the D-W-15K-LP

dataset. In a similar study, ref. [15] used ATransN, where they framed the baseline *Connected* as the upper bound of improvement. This is because they assumed that *seed alignments* used by the models include all possible alignments between KGs. We argue this is extremely unlikely in reality. Therefore, in our study, we kept the total number of aligned entities between KGs constant while varying the number of seed alignments available to the model during training. The best performance for IPPT4KRL was obtained with negative sampling turned off for scenarios with less than a 50% alignment ratio and negative sampling turned on for scenarios with more than a 50% alignment ratio. It can be observed that, as the alignment ratio increased, the performance of both models increased across all metrics. At the same time, IPPT4KRL consistently outperformed the baseline *Combined* in all metrics, except Hit@10, on which the two models achieved very similar performances.

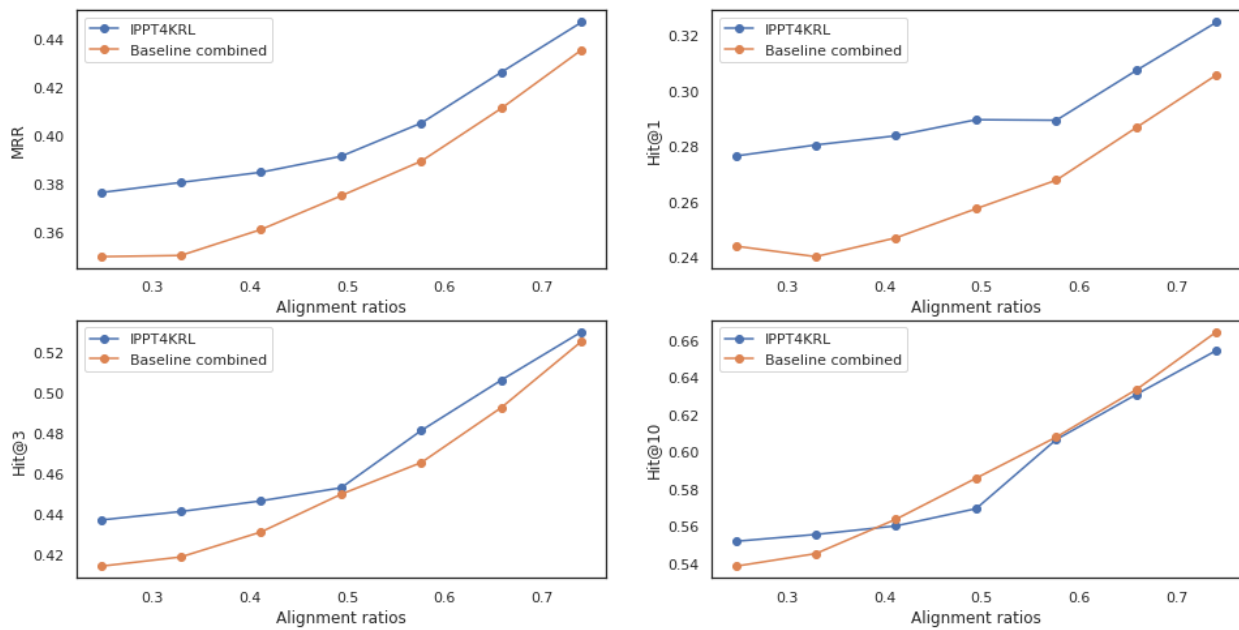


Figure 2. Performance comparison between IPPT4KRL and the *Connected* baseline with varying visible seed alignments on dataset D-W-15K-LP.

5.3. Effect of Negative Sampling

Negative sampling is one of the most-effective techniques in knowledge graph representation learning. In our IPPT4KRL framework, we employed uniform negative sampling for the pre-training stage of KG_1 . In the post-processing stage, we observed some inconsistencies in the effects of including negative sampling: including negative sampling in the post-processing stage did not always provide knowledge gains to the representation learning of KG_1 . In Table 6, we present the experiment results demonstrating the effect of negative sampling during the post-processing stage. From the results, it can be observed that negative sampling tended to be more beneficial for knowledge transfer for datasets with higher seed alignment ratios (CN3I, DBP-FB), while on the datasets with lower seed alignment ratios (D-W-15K-LP, Wk3I), additional negative sampling during the post-processing stage limited the amount of knowledge transfer between the KGs. We further validated this observation on the D-W-15K-LP dataset with varying alignment ratios, comparing model performance with the inclusion and exclusion of negative sampling in the post-processing stage. From Figure 3, it can be observed that, as the alignment ratio increased, the model with negative sampling enabled during the post-processing stage gradually outperformed the model without negative sampling enabled in all metrics. We believe this could be a very good starting point for future research on characterizing positive and negative knowledge transfer between KGs.

Table 6. Model performance comparison including vs. excluding negative sampling in the post-processing stage.

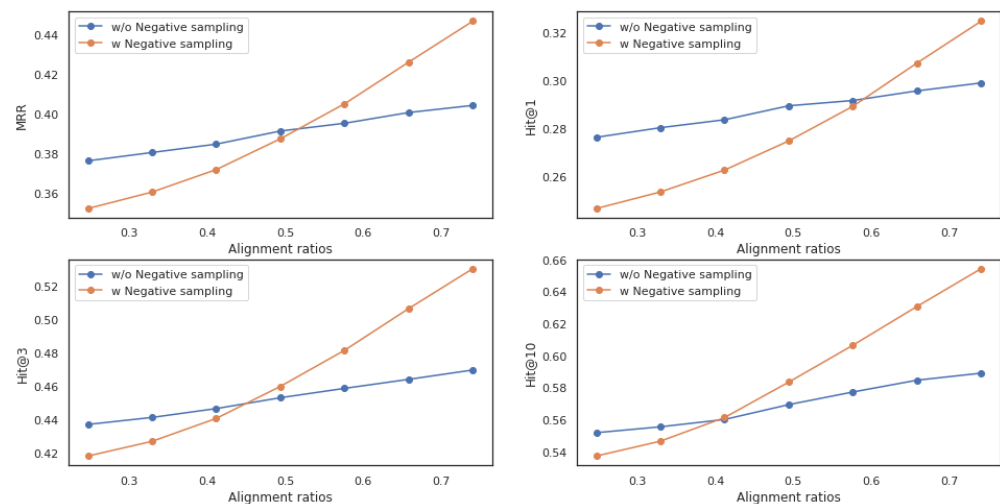
Dataset	Negative Sampling	MRR	MR	Hit@1	Hit@3	Hit@10
D-W-15K-LP	✗	0.3764	335	27.64%	43.71%	55.16%
	✓	0.3486	539	24.33%	41.30%	53.38%
CN3l (EN-DE)	✗	0.2136	501	11.03%	28.17%	38.40%
	✓	0.2238	479	10.17%	30.61%	42.68%
WK3l-15K(EN-FR)	✗	0.3823	446	28.22%	42.86%	58.10%
	✓	0.3817	447	28.17%	42.81%	58.08%
DBP-FB	✗	0.3591	428	25.81%	41.82%	53.44%
	✓	0.3782	397	27.58%	43.85%	55.67%

5.4. Representation Learning on KG2

In IPPT4KRL, most of the focus is on improving the embedding training of KG_1 , while the training of KG_2 embeddings is overlooked. In Table 7, we report the link prediction results on KG_2 embeddings obtained by pre-training/post-processing KG_1 and pre-training/post-processing KG_2 . In the model where pre-training and post-processing were conducted on KG_1 , to obtain the best possible results for KG_2 , we modified the convergence criteria of the method to stop the iterations once we observed a drop in the validations set of KG_2 . The results indicated that, under this framework, the knowledge transfer tends to be more directional towards the initial KG being pre-trained and post-processed, and better embeddings of KG_2 can be obtained if we switch the role and instead apply IPPT4KRL on KG_2 . Directional knowledge transfer is one of the unsatisfactory characteristics of the proposed framework. However, in the scenarios where bi-directional knowledge transfer is desired, the pre-training and post-processing stages of both KGs can potentially be parallelized in practice.

Table 7. Comparison of KG_2 embedding obtained by treating different KGs as the initial KG to pre-train and post-process.

Dataset	Initial KG	MRR	MR	Hit@1	Hit@3	Hit@10
D-W-15K-LP	KG_1	0.3045	374	22.88%	34.15%	44.25%
	KG_2	0.3453	410	25.98%	39.35%	49.50%
CN3l (EN-DE)	KG_1	0.2192	540	13.29%	27.15%	35.93%
	KG_2	0.2199	497	12.19%	28.35%	37.93%

**Figure 3.** Performance comparison between IPPT4KRL w/ and w/o negative sampling enabled in the post-processing stage with varying visible seed alignments on dataset D-W-15K-LP.

6. Conclusions and Future Work

In this paper, we proposed an iterative post-processing method (IPPT4KRL) for multi-KG knowledge representation learning. Furthermore, we provided benchmarking results against other state-of-the-art multi-KG representation learning methods on existing and generated empirical datasets. Compared to the previously proposed multi-KG KRL methods, our method splits the multi-KG representation learning into the pre-training and post-processing stages, focusing on representation learning on individual KGs and knowledge transfer between KGs, respectively, in each stage. As shown in Tables 2 and 3, IPPT4KRL achieved a comparable and superior knowledge transfer while being flexible and relatively lightweight. Apart from the performance of our model, we believe this idea could be very valuable in real life as it could potentially serve as an inexpensive process for assessing the knowledge gained in the industrial application of fused KGs.

We also generated a new mono-lingual dataset for the multi-KG representation learning following a previously proposed sampling technique [18]. We believe this dataset simulates “in the wild” conditions for multi-KG KRL well with the creation of dangling entities and low overlaps between KGs. More datasets with a different number of seed alignments in KGs can be easily generated with the same technique.

As shown in the ablation studies, although effective, the performances of the proposed method still rely on the existence of high-quality seed alignments, which are usually difficult to obtain for large-scale KGs. In the future, we plan to explore the possibility of incorporating the predictions of entity alignments into the multi-KG KRL process and, therefore, alleviate the burden of the manual annotation of alignments. Moreover, with the rapid development in graph neural networks for knowledge representation, existing GNN-related work [16] has already shown some success in multi-graph KRL, and we plan to extend the research work to explore more models with GNN-based encoders under a multi-KG KRL setting.

Author Contributions: Conceptualization, W.Z. and O.Ş.; methodology, W.Z. and O.Ş.; validation, W.Z., O.Ş. and J.S.; experiments, W.Z.; writing—original draft preparation, W.Z.; writing—review and editing, O.Ş., J.S. and Y.G.; visualization, O.Ş. and J.S.; supervision, O.Ş. and Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Royal Bank of Canada’s Wealth Management Digital & Data Solutions group.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The data presented in this study are openly available.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Additional Ablation Study Results

We also conducted experiments with different seed alignment ratios, comparing the proposed IPPT4KRL against the several ablation models we used in the ablation study. Figure A1 shows the plots of the performance of the proposed IPPT4KRL, ablation model *no-preservation*, and ablation model *non-iterative* with different seed alignment ratios on dataset D-W-15K-LP under the same negative sampling setting. Overall, the proposed IPPT4KRL consistently outperformed the two ablation models across all metrics under all seed alignment ratios.

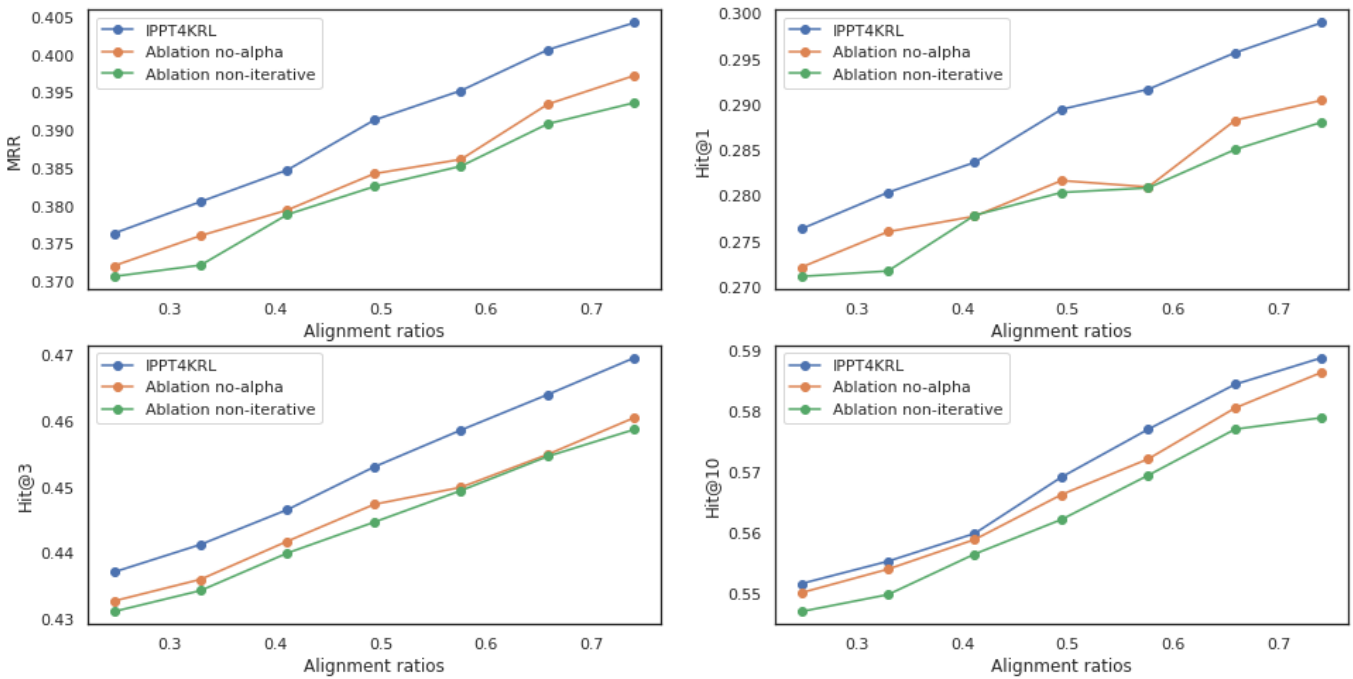


Figure A1. Performance comparison between IPPT4KRL and ablation models with varying visible seed alignments on D-W-15K-LP.

Appendix B. Modification of KD-MKB

In the KD-MKB paper, the relation distillation loss was formalized as

$$\sum_{(e_x^j, r, e_y^j): (e_x^j, e_y^j), (e_y^j, e_x^j) \in I_e(i, j)} D(P(r_{(e_x^j, e_y^j)} \theta^j), P(r_{(e_y^j, e_x^j)} \theta^i)),$$

where $I_e(i, j)$ denotes the entity seed alignments, D denotes the distillation function such as the KL divergence, $r_{(e_x^j, e_y^j)}$ denotes a categorical variable with relation alignment $I_r(i, j)$ values, and $P(r_{(e_x^j, e_y^j)} \theta^j)$ denotes the categorical distribution generated from the embedding model. The loss function for entity distillation follows the same pattern while swapping the role of aligned entities $I_e(i, j)$ and aligned relations $I_r(i, j)$. The main idea behind KD-MKB is to distil knowledge on mutually “shared” information between KGs. However, as indicated in the formula, this requires both KGs to have aligned entities $I_e(i, j)$ and aligned relations $I_r(i, j)$. To adapt KD-MKB to datasets without relation alignments, we created modified KD-MKB*, where $I_r(i, j)$ are changed to be the union of relations in two KGs instead of relation alignments. This implicitly requires the embedding models of each KG to learn representations for links unique to other KGs. These representations are trained indirectly via the knowledge distillation process. This empirically results in substantial knowledge gain compared to the baselines, and we decided to report the results of KD-MKB* on the datasets without aligned relations.

References

1. Paulheim, H. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semant. Web* **2017**, *8*, 489–508. [\[CrossRef\]](#)
2. Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web*; Aberer, K., Choi, K.S., Noy, N., Allemang, D., Lee, K.I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., et al., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2007; pp. 722–735. [\[CrossRef\]](#)
3. Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; Taylor, J. Freebase: A collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD international conference on Management of Data, SIGMOD '08, Vancouver, BC, Canada, 9–12 June 2008; Association for Computing Machinery: New York, NY, USA, 2008; pp. 1247–1250. [\[CrossRef\]](#)

4. Suchanek, F.M.; Kasneci, G.; Weikum, G. YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia. In Proceedings of the 16th International World Wide Web Conference, San Francisco, CA, USA, 8–12 May 2007; pp. 697–706.
5. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Yu, P.S. A Survey on Knowledge Graphs: Representation, Acquisition and Applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 494–514. [[CrossRef](#)] [[PubMed](#)]
6. Mutlu, E.C.; Oghaz, T.; Rajabi, A.; Garibay, I. Review on Learning and Extracting Graph Features for Link Prediction. *Mach. Learn. Knowl. Extr.* **2020**, *2*, 672–704. [[CrossRef](#)]
7. Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating Embeddings for Modeling Multi-relational Data. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 1–9.
8. Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; Zhu, X. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 23–25 January 2015.
9. Ji, G.; He, S.; Xu, L.; Liu, K.; Zhao, J. Knowledge Graph Embedding via Dynamic Mapping Matrix. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); Association for Computational Linguistics: Beijing, China, 2015; pp. 687–696. [[CrossRef](#)]
10. Yang, B.; Yih, W.T.; He, X.; Gao, J.; Deng, L. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
11. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2017**, arXiv:1609.02907.
12. Schlichtkrull, M.; Kipf, T.N.; Bloem, P.; van den Berg, R.; Titov, I.; Welling, M. Modeling Relational Data with Graph Convolutional Networks. In *The Semantic Web*; Gangemi, A., Navigli, R., Vidal, M.E., Hitzler, P., Troncy, R., Hollink, L., Tordai, A., Alam, M., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Lecture Notes in Computer Science, pp. 593–607. [[CrossRef](#)]
13. Trivedi, R.; Sisman, B.; Dong, X.L.; Faloutsos, C.; Ma, J.; Zha, H. LinkNBed: Multi-Graph Representation Learning with Entity Linkage. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; pp. 252–262. [[CrossRef](#)]
14. Sourty, R.; Moreno, J.G.; Servant, F.P.; Tamine-Lechani, L. Knowledge Base Embedding By Cooperative Knowledge Distillation. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020; International Committee on Computational Linguistics: Barcelona, Spain, 2020; pp. 5579–5590. [[CrossRef](#)]
15. Wang, H.; Li, S.; Pan, R. An Adversarial Transfer Network for Knowledge Representation Learning. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021; ACM: New York, NY, USA, 2021; pp. 1749–1760. [[CrossRef](#)]
16. Jiang, M. Cross-Network Learning with Partially Aligned Graph Convolutional Networks. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 14–18 August 2021; ACM: New York, NY, USA, 2021; pp. 746–755. [[CrossRef](#)]
17. Sun, Z.; Zhang, Q.; Hu, W.; Wang, C.; Chen, M.; Akrami, F.; Li, C. A benchmarking study of embedding-based entity alignment for knowledge graphs. *Proc. VLDB Endow.* **2020**, *13*, 2326–2340. [[CrossRef](#)]
18. Sun, Z.; Chen, M.; Hu, W. Knowing the No-match: Entity Alignment with Dangling Cases. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 1–6 August 2021; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 3582–3593. [[CrossRef](#)]
19. Wang, Z.; Zhang, J.; Feng, J.; Chen, Z. Knowledge Graph Embedding by Translating on Hyperplanes. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014.
20. Sun, Z.; Deng, Z.H.; Nie, J.Y.; Tang, J. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. *arXiv* **2018**, arXiv:1902.10197.
21. Zhang, Z.; Cai, J.; Zhang, Y.; Wang, J. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction. *AAAI* **2020**, *34*, 3065–3072. [[CrossRef](#)]
22. Nickel, M.; Tresp, V.; Krieger, H.P. A three-way model for collective learning on multi-relational data. In Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11, Bellevue, WA, USA, 28 June–2 July 2011; pp. 809–816.
23. Nickel, M.; Rosasco, L.; Poggio, T. Holographic Embeddings of Knowledge Graphs. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 28 June–2 July 2016; Volume 30. [[CrossRef](#)]
24. Allen, C.; Hospedales, T. Analogies Explained: Towards Understanding Word Embeddings. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019.
25. Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, E.; Bouchard, G. Complex Embeddings for Simple Link Prediction. In Proceedings of the 33rd International Conference on Machine Learning, PMLR, New York, NY, USA, 19–24 June 2016; pp. 2071–2080.
26. Balazevic, I.; Allen, C.; Hospedales, T. TuckER: Tensor Factorization for Knowledge Graph Completion. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 3–7 November 2019; Association for Computational Linguistics: Hong Kong, China, 2019; pp. 5184–5193. [[CrossRef](#)]
27. ZHANG, S.; Tay, Y.; Yao, L.; Liu, Q. Quaternion Knowledge Graph Embeddings. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32.

28. Zhao, X.; Zeng, W.; Tang, J.; Wang, W.; Suchanek, F.M. An Experimental Study of State-of-the-Art Entity Alignment Approaches. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 2610–2625. [[CrossRef](#)]
29. Chen, M.; Tian, Y.; Yang, M.; Zaniolo, C. Multilingual Knowledge Graph Embeddings for Cross-lingual Knowledge Alignment. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 1511–1517. . [[CrossRef](#)]
30. Han, X.; Cao, S.; Xin, L.; Lin, Y.; Liu, Z.; Sun, M.; Li, J. OpenKE: An Open Toolkit for Knowledge Embedding. In Proceedings of the EMNLP, Brussels, Belgium, 31 October–4 November 2018.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.