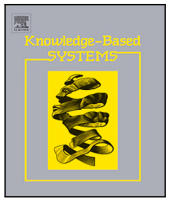
Knowledge-Based Systems 281 (2023) 111070



Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



Short communication



Conflict-aware multilingual knowledge graph completion

Weihang Zhang [a,*], Ovidiu Şerban [a], Jiahao Sun [a,b], Yike Guo [a]

[a] Data Science Institute, Imperial College London, Exhibition Rd, South Kensington, London, SW7 2BX, United Kingdom
[b] Royal Bank of Canada, 100 Bishopsgate, London, EC2M 1GT, United Kingdom

ARTICLE INFO

Keywords:
Knowledge graph
Knowledge graph completion
Knowledge representation
Knowledge transfer
Active learning

ABSTRACT

Knowledge graph completion (KGC), a task that aims at predicting missing links with existing information inside a knowledge graph (KG), has emerged as a popular research area in recent years. While many existing works have demonstrated effectiveness on a single knowledge graph KGC, limited effort has been devoted to exploring the potentially complementary nature of multiple KGs. In this work, we proposed a novel method called CA-MKGC (Conflict-aware Multilingual Knowledge Graph Completion) for multiple knowledge graph completion (MKGC), aiming to alleviate the sparseness of a single knowledge graph by leveraging information from other knowledge graphs. We designed an intra-KG graph convolutional network encoder that regards the seed alignments between KGs as edges for intra-KG message propagation to model all KGs in a unified model while also adopting an iterative mechanism to progressively incorporate newly predicted alignments along with the newly inferred facts into the learning process. Additionally, we employed an active learning mechanism and a greedy approximation to a semi-constrained optimization problem to focus on learning the structural prior knowledge that is difficult to learn in semantic space, limiting the propagation of error in the iterative training process. Experimental results on multilingual KG datasets demonstrated that our method achieved state-of-the-art results.

1. Introduction

With the vast amount of data available on the internet, knowledge graphs (KGs) have become a popular tool for organizing and representing structured knowledge. KGs enable the creation of a semantic network of entities and their relationships, which can be leveraged for various tasks in the artificial intelligence field, such as question-answering and recommendation systems [1]. However, KGs are known to be incomplete [2]. In recent years, there has been a surge of research aimed at addressing the incompleteness problem in KGs through the task of knowledge graph completion (KGC). However, most of the existing works [3–8] focused on the KGC task on single KG, overlooking the fact that existing information inside single KG might not be sufficient to infer all missing edges or relations between entities. In this work, we propose a novel method called CA-MKGC that focuses on the MKGC problem, which aims at leveraging the collaborative and complementary nature of multiple KGs to infer missing edges and relations of each other. Specifically, we first propose an intra-KG GCN encoder to model alignments between entities in different KG as relations for intra-KG message propagation to train all KGs in a unified model while progressively introducing new confident predictions of entity alignments to generate additional supervisions. Similar to existing works [9–12] that address the same problem setting, we identified the important collaborative nature of entity alignment (EA) prediction and KGC task and used this nature to generate incremental supervision for the training process by predicting more alignments between KGs and use them for further training. However, unlike existing works that assume this to be only beneficial, we view this as the source of better supervision and error propagation. Most existing works neglect the transitive nature of alignments across multiple KGs and only consider predictions of alignments between pairs of KGs rather than holistically for the entire set of KGs. In Section 3.4, we show that this could result in the propagation of potentially incorrect facts and alignments when predicted alignments are used for further training. To mitigate the potential propagation of errors, we propose an active learning mechanism and a greedy approximation of a semi-constrained optimization problem to impose a simple yet effective unique alignment restriction between KGs to help refine the prediction of alignments and restrict the error propagation when predictions are used for incremental supervision. We show that imposing this restriction is surprisingly effective in restricting the error propagation of incorrect predictions and results in state-of-the-art results for multiple knowledge graph completion on various datasets. Fig. 1 illustrates an overview of the framework of the proposed method. The main contributions of this work can be summarized as follows: (1) we propose a novel intra-KG

* Corresponding author.
E-mail address: w.zhang21@imperial.ac.uk (W. Zhang).

https://doi.org/10.1016/j.knosys.2023.111070
Received 19 June 2023; Received in revised form 29 September 2023; Accepted 12 October 2023
Available online 14 October 2023
0950-7051/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

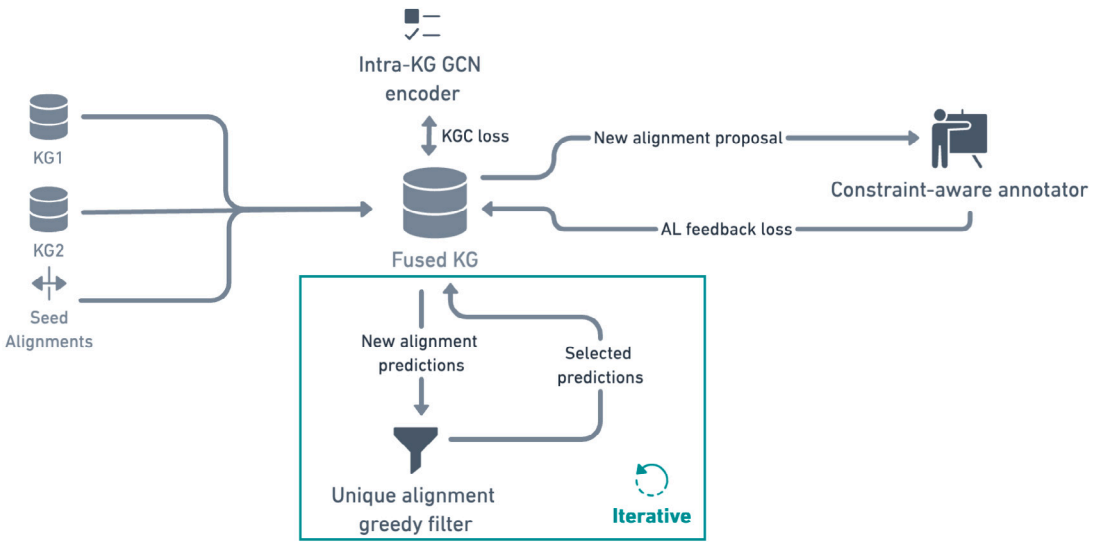


Fig. 1. An overview of the framework of our proposed method on MKGC task.

GCN encoder that models all KGs in a unified model for efficient intra-KG message propagation and knowledge transfer, (2) we examine the source of error propagation in MKGC problem when using predictions to generate additional training samples and propose a first attempt to limit the propagation of errors using a simple yet effective restriction. We describe the details of the proposed method in Section 3, and the detailed experiments and ablation analysis are reported in Section 4.

2. Related work

2.1. Single-knowledge graph completion

Knowledge graph completion (KGC) has been a popular research topic in the past few years, and many embedding-based methods that encode nodes and edges into the low-dimensional vector space have been proposed to tackle the KGC problem. Translation-based models like TransE [3], semantic matching-based models like RESCAL [13] along with their variations [2,5,6,14] have all shown promising results on KGC task. In recent years, many methods [15–17] were proposed to use Graph Convolutional Network (GCN) [18] and its variations as encoders to help better utilize the graph structure to encode the entities and relations in a KG, and achieved even better results on KGC task on single KG. Advancements in knowledge representation learning are not limited to deeper networks and more advanced encoder structures. Many recent works [19,20] have also proposed to use techniques like contrastive learning with graph augmentation to improve the quality of knowledge representation. We believe the development of more advanced knowledge representation methods on a single KG can significantly benefit the methods focused on the MKGC task. This is because most MKGC methods target more on maximizing the "knowledge transfer" between graphs while using the aforementioned knowledge representation techniques as their basis.

2.2. Multi-knowledge graph completion

Compared to the KGC task on a single KG, the knowledge graph completion task on multiple KGs has received less attention over the years. This task was first briefly discussed in [21] but only as a side product of the entity alignment method. Chen et al. [9] was the first to propose multilingual KGC problem setting and generate a dedicated dataset DBP-5L. Chen et al. [9] tackled the MKGC task by generating ensemble results on shared triples between KGs in different languages. Other recent methods like AlignKGC [10], SS-AGA [11], and most recently, JMAC [12], have also been proposed to tackle the same problem setting. AlignKGC and JMAC have also proposed to use only 50% of the seed alignment between KGs to simulate a lower resource setting for this problem. Most existing works in multi-knowledge graph completion allow the prediction of aligned entities between KGs to generate additional supervision for MKGC tasks. However, when producing predictions of aligned entities between KGs, the existing method only considers a pair of KG each time, which did not fully utilize the benefit of having multiple KGs and could result in error-pruned prediction results and generating sub-optimal additional supervisions. In terms of the encoding strategy for nodes and edges in all KGs, most existing works [1,10,12] proposed to model the embeddings separately for each KG while using loss terms to regularize the relationship between aligned entities across KGs. SS-AGA [11] was a notable exception that took a very similar strategy to our work and proposed to model all embeddings under a unified model via message propagation. However, the proposed SS-AGA employed regularization that forces the embeddings of aligned entities to be similar, which intuitively would create confusion for the knowledge graph embeddings training process under a unified model, which aims to distinguish between different entities.

2.3. Entity alignment

Entity alignment (EA) is a task that aims to find equivalent entities in different KGs. Similar to KGC, embedding-based methods [21–24] have also been proven to perform well in this task. In existing multi-knowledge graph completion research, EA is an important step to identify and generate more supervision to transfer knowledge across KGs better. However, in the EA task, one common problem most embedding-based methods have is the hubness problem [25], which means one entity is sometimes predicted to be aligned with multiple modes in the embedding space. As pointed out by [26], this is more severe when the number of entities differs in each KG. Many research works [26–29] have attempted to solve this issue by enforcing global entity alignment, but all in scenarios involving only 2 KGs. In this work, we identify that in the MKGC task, the hubness problem can be an important source of error propagation and demonstrate a first attempt to solve this in the multi-KG setting.

3. Methods

3.1. Preliminaries

As shown in Fig. 3, the problem setting of MKGC involves two or more KGs, without loss of generality, we assume there are a total of

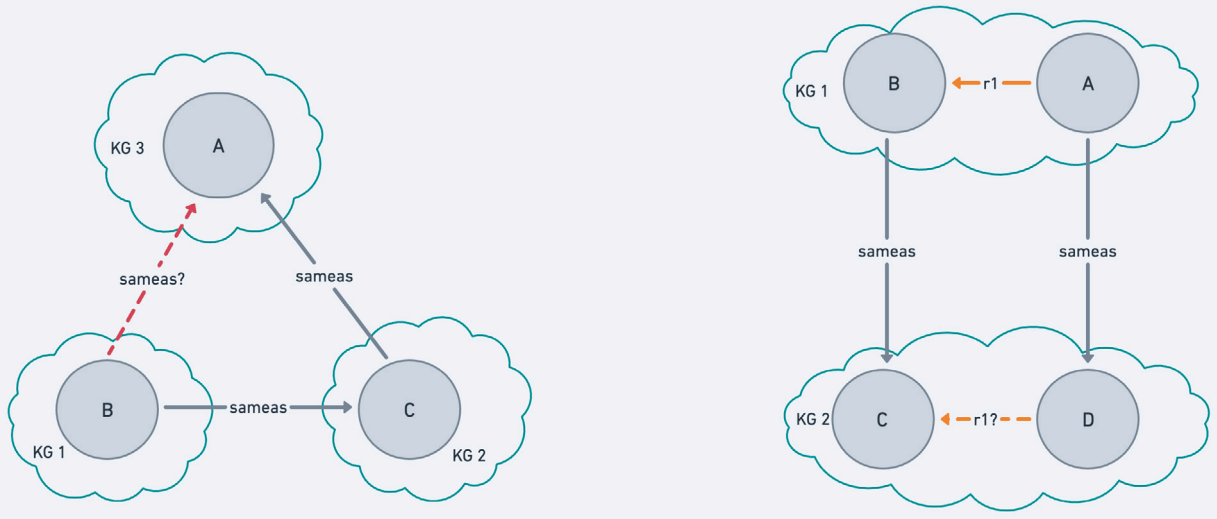


Fig. 2. An illustrative figure on two types of alignment-entailed meta-paths that can be used to infer new alignments and facts.

m KGs. We formalize the ith heterogeneous KG in the task as $KG_i = \{E_i, R_i, T_i\}$, where E_i, R_i, T_i represent the entity set, relation set, and a triple set of knowledge graph KG_i respectively. Seed alignments between each pair of KGs is denoted by $S_{KG_i,KG_j} = \{(e_i, \sim, e_j) : (e_i, e_j) \in E_i \times E_j\}$, where $\sim$ denotes the equivalence relation. The complete set of seed alignments can then be denoted by $S_{align} = \cup_{i=1}^{m} \cup_{j=i+1}^{m} S_{KG_i,KG_j}$.

3.2. Intra-KG encoder

To encode the heterogeneous KGs into low-dimensional vector representations, we view seed alignments as normal relations for message propagation to model all KGs in one unified encoder. In the method, we design a GCN-based intra-KG encoder with non-parametric composition operations [16] to embed the fused KG. The composition operation was first proposed in GCN encoders in the work of CompGCN [16], where the entity-relation composition operator was created to embed both nodes and edges in a multi-relational graph jointly. The update equation of CompGCN node embeddings can be written as:

$$h_v^t = f(\Sigma_{(u,r)\in N(v)} Me(u,r)) \tag{1}$$

$$Me(u,r) = W_{\lambda(r)}\phi(h_u^{t-1}, h_r^{t-1}), \tag{2}$$

where h_v^t denotes the updated embedding for node v at tth layer, $N(v)$ denotes the set of neighboring entities and relations of node v, h_u^{t-1} and h_r^{t-1} denotes the embeddings for node u and relation r at $(t-1)$th layer respectively, ϕ denotes the non-parametric composition operation and $W_{\lambda(r)}$ denotes the direction specific transformation matrix where λ denotes the direction of relation. In our method, we modify the original CompGCN encoder for better intra-KG message propagation for knowledge transfer on the fused KG_f. Specifically, although seed alignments are viewed as regular relations to connect individual KGs into a unified KG, to facilitate the intra-KG message propagation, we remove the composition operator between the seed alignments and instead treat alignment as a special bidirectional relation with its unique transformation matrix. The modified message function in the fused KG can then be written as:

$$Me(u,r) = \begin{cases} W_{align} h_u^{t-1} & \text{if}(u, \sim, v) \in S_{align} \\ W_{\lambda(r)}\phi(h_u^{t-1}, h_r^{t-1}) & \text{otherwise} \end{cases}, \tag{3}$$

where W_{align} denotes the transformation matrix specifically for intra-KG message propagation. Many existing methods [10–12,30] all use a loss regularization to ensure the equivalent entities in each KG have very similar embeddings with or without transformation. However, when modeling all KGs in a unified model, we find directly imposing this regularization creates confusion for the overall embedding training because the intuition of the downstream KGC task is to distinguish between entities with trained embeddings essentially. To solve this, we model alignment relation also as edges in the fused KG, and the "information sharing" between aligned entities is imposed via the intra-KG message passing, where the contextualized node embeddings of entities in each knowledge graph are passed to their counterparts in other KGs during encoding. This ensures the aligned entities between different KGs have similar but distinguishable embeddings, while the differences are the local neighbors they have in each KG. Another benefit of the intra-KG encoder is that it allows for sharing of relation embeddings between all KGs under a unified model, which leads to improved capture of the complex dynamics and relationships between relations in the KGs. With the intra-KG GCN encoder, the contextualized embedding of entities in each KG can be shared across the KGs by the message propagation in the encoding phase and optimized during the training of the KGC task on fact triples.

The encoded entity and relation embeddings are then passed to the decoder, which performs the link prediction task on KG triples and computes the knowledge completion loss. We used a margin-based knowledge representation loss which can be written as:

$$L_c = \Sigma_{t_i \in T_i, t_i' \in T_i'} f(t_i) - f(t_i') + \gamma, \tag{4}$$

where T_i' denotes the negative samples created from corrupting head or tail entity in triple t_i; $f(t_i)$ denotes the scoring function for TransE or similar knowledge embedding model; and γ denotes the margin, which is a hyperparameter describing the ideal distance between positive triples and negative triples.

3.3. Incremental supervision with alignment-entailed meta-paths

In the MKGC task, in addition to using an effective encoder to enhance knowledge transfer in the representation space, unambiguous ways exist to introduce newly inferred alignments and facts in the training process as additional samples. As illustrated in Fig. 2, there exist two different types of alignment-entailed meta-paths: (1) with the composition rule applied on alignments, additional alignments can potentially be inferred via meta-path: $(e_A \overset{\sim}{\longrightarrow} e_B \overset{\sim}{\longrightarrow} e_C)$ implies $(e_A \overset{\sim}{\longrightarrow} e_C)$, (2) with alignment information available, a meta-path of $(e_A \overset{\sim}{\longrightarrow} e_B \overset{r_1}{\longrightarrow} e_C \overset{\sim}{\longrightarrow} e_D)$ implies a new fact triple $(e_A \overset{r_1}{\longrightarrow} e_D)$ (this is sometimes referred as parameter-swapping [31] in entity alignment algorithms). Previous works [11,12] have predominantly focused on applying the second meta-path to generate new facts while neglecting the potential synergy that could be achieved by incorporating both meta-paths. In

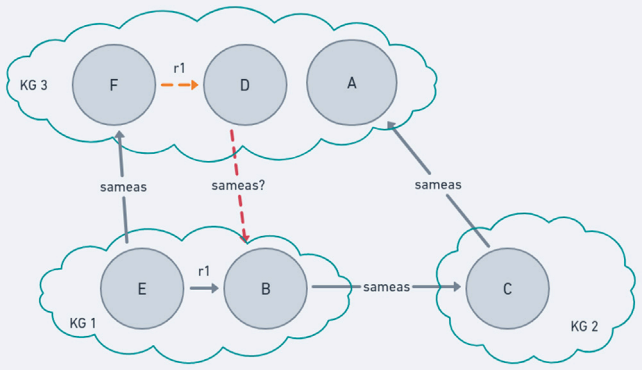


Fig. 3. An illustrative figure on how new facts can be inferred from alignment predictions. The yellow fact (F, r_1, D) is inferred from the predicted alignment (D, sameas, B). However, this fact could potentially be a result of error propagation from the predicted alignments (D, sameas, B), which produces a conflict with the inferred alignments (A, sameas, B).

this paper, we iteratively apply both alignment-entailed meta-paths: we enable the prediction of potential alignments during training, and the predicted alignments t_i with confidence $f(t_i) > \theta$ are leveraged to infer new alignments and fact triples as additional supervisions for training in the next iteration using alignment-entailed meta-paths. θ here denotes a hyperparameter representing the minimum confidence score for selecting the alignment predictions. Let S_{inc} and T_{inc} denote the set of newly produced alignments and fact triples in the iterative process. We can then formalize the loss term for the incremental supervision to be:

$$L_{inc} = \Sigma_{t_i \in T_{inc} \cup S_{inc}, t_i' \in T'_{inc} \cup S'_{inc}} f(t_i) - f(t_i') + \gamma, \tag{5}$$

where T'_{inc} and S'_{inc} here respectively denote the set of negative samples for newly produced fact triples and alignments. The negative samples are created similar to the negative samples for Eq. (4). When the textual information for nodes of KGs is available, textual embeddings can be used as an additional resource to assist with alignment prediction. In those cases, we use multilingual Bert [32] to encode the textual information of nodes: $e_i^{text} = mBERT(e_i)$, and the confidence score of alignment $t_i = (e_i, \sim, e_j)$ then becomes: $f(t_i) + \alpha cos(e_i^{text}, e_j^{text})$, where α denotes the hyperparameter controlling the trade-off between graph and text embeddings, and cos denotes the cosine similarity between the text embeddings.

3.4. Unique alignment constraint

Although incremental supervision with alignment-entailed meta-paths can introduce new training samples iteratively, including predicted alignments in this process may also introduce potential errors that could propagate throughout the model as illustrated in Fig. 3. The error propagation could worsen when more KGs are involved in the learning process. An incorrect alignment prediction can lead to a higher probability of generating erroneous facts through alignment-entailed meta-paths if the entities involved in the incorrect alignment have multiple known alignments across KGs. Under the assumption that entities in high-quality KGs do not usually have duplicates, we impose a prior restriction on the alignment predictions. Concretely, let $CC_{e_i} = \{e_j : (e_i, e_j) \in S_{align}\}$ denotes the alignment connected component set of entity e_i containing its corresponding alignments across different KGs, two restrictions are imposed:

$$|CC_{e_i}| \leq m \tag{6}$$

$$|CC_{e_i}| = |\{KG(e_j), \forall e_j \in CC_{e_i}\}|, \tag{7}$$

where m denotes the total number of KGs and $KG()$ denotes the original KG lookup of an entity. This restriction can essentially be interpreted as each entity should ideally be aligned with no more than one entity in each knowledge graph. While the restriction might seem trivial, this was not fully incorporated into the existing MKGC methods. Most of the existing works alleviate the hubness problem [25] and guarantee the 1-to-1 mapping of entities in low-dimensional space by incorporating CSLS [25] heuristic at prediction time or by converting and solving optimal transport problem between the KGs [26]. However, when making predictions, they only considered alignments between a pair of KGs, which overlooked the important transitive nature of alignments between multiple KGs. When more than two KGs are involved at the prediction time of alignments, consider the same illustrative Fig. 3, if alignment relations $(B, \sim, D)$ and $(A, \sim, C)$ are both predicted, the one-to-one restriction of alignments is not violated when considering only pair of (KG_1, KG_3) or (KG_2, KG_3). However, globally, this violates the above-mentioned unique alignment constraint since both predicted alignments indicate node A is equivalent to node D in KG_3. This conflict could result in the alignment of connected components of considerable size (many alignments across KGs are predicted to be aligned with each other), and the alignment-entailed meta-path could rapidly propagate incorrect facts within the connected component. We designed two components to impose the above unique alignment restriction into the iterative training process: one at the training time and another at the prediction time when generating additional supervision. For convenience, in the rest of the paper, we denote $e_j \in Align(e_i)$ as entity e_j can be predicted to align with entity e_i without violating above mentioned unique alignment restriction.

3.4.1. Active learning based prior restriction loss

When training KG embeddings, we incorporated the unique alignment restriction to produce negative alignment samples, which would help the model to learn to avoid making predictions of alignment that violate the restriction imposed by existing seed alignments. However, directly implementing the restriction would generate a large number of negative samples for each entity, rendering an excessive and overwhelming amount of negative samples. To avoid generating unhelpful negative samples, we utilized an active learning mechanism that allows the model to predict possible alignments while the unique alignment restriction serves as an "annotator" to provide feedback on the proposed alignments. If a proposed alignment violates the restriction, it is labeled as negative and added to negative set A_{neg}. In practice, generating only negative examples would still inevitably discourage the predictions of all alignments. Therefore, we follow a similar idea in [33], using the model to generate pseudo-aligned positive examples for each negative example to serve as "soft" positive examples. Concretely, for each negative sample $(e_i, \sim, e_j) \in A_{neg}$ between KG_i and KG_j, we choose the most confident prediction $(e_i, \sim, e_{j'})$ as a pseudo-aligned example if it fulfills $f(e_i, \sim, e_{j'}) > \theta, e_{j'} \in Align(e_i)$, where θ is the same hyperparameter we use to select alignment predictions when generating incremental supervisions. For simplicity, we denote the set of pseudo-aligned positive examples as A_{pos}. With additional negative and "soft" positive alignment samples generated from the active learning mechanism, the overall loss term for training when active learning is enabled then becomes:

$$L_{al} = \Sigma_{t_i \in T_f \cup A_{pos}, t_i' \in T'_f \cup A_{neg}} f(t_i) - f(t_i') + \gamma \tag{8}$$

3.4.2. Greedy alignment prediction filter

It is noteworthy that imposing the unique alignment restriction during training is insufficient to prevent the violation at prediction time since predicted alignments could collaboratively violate the unique alignment constraint. To address this issue, we can convert this problem into a constraint optimization problem at prediction time. We treat the triple scoring function f as the confidence score for alignment predictions. The overall goal at prediction time can be viewed as a mixed integer programming optimization problem to select a group of predicted alignments such that the total confidence score of selected

Table 1
Dataset statistics.

Dataset	Language	Entities	Relations	Triples	Aligned links
DBP-5L	EL	5,231	111	13,839	9,042
	JA	11,805	128	28,774	16,263
	ES	12,382	144	54,066	16,347
	FR	13,176	178	49,015	16,877
	EN	13,996	831	80,167	16,916
E-PKG	JA	2,642	21	16,703	5,175
	ES	9,595	21	30,163	20,184
	IT	15,670	21	71,292	23,827
	DE	17,223	21	75,870	21,382
	FR	17,068	21	80,015	24,812
	EN	16,544	21	100,531	21,382

alignments is maximized while not violating the unique alignment restriction. The overall goal can then be formalized as:

$$\begin{aligned} \text{maximize} \quad & \sum_{i\in 1..m, j\in i+1..m} M_{ij} S_{ij} \\ \text{s.t.} \quad & [1,1,\ldots,1] M_{ij} \le 1, \\ & [1,1,\ldots,1] M \le m \end{aligned} \tag{9}$$

where S_{ij} denotes the confidence score of corresponding alignments, M_{ij} denotes the alignment indicator matrix between entities in KG_i and KG_j, M denotes the alignment indicator matrix between all entities in the connected KG_f after applying transitivity rule, and $[1,1,\ldots,1]M_{ij}$ denotes the row sum of indicator matrix. The two constraints are sufficient for the unique alignment restriction across all KGs. However, directly solving this optimization is difficult due to the high computation cost, especially because of the second constraint. We, therefore, decide to use a greedy filter on alignment predictions to approximate the solution: all predicted alignments are sorted in descending order based on their confidence scores; then we iteratively include the predictions that comply with the unique alignment restriction while also updating the alignment connected components for each entity dynamically as we go. With the updated alignment of connected components, in the next iteration, the active learning mechanism would naturally provide more useful "feedback" to aid the training of the embedding model. This approach allows us to continually refine and select the alignment predictions, ensuring the best possible accuracy for our results in a greedy way. We find this simple greedy approach very effective in practice.

3.5. Model training

The model is initially trained with iterative optimization of KGC loss L_c and active learning loss L_{al} until stable. The predictions of alignments are made under the greedy filter for the incremental supervision, and subsequent model training becomes the iterative optimization of $L_c + L_{inc}$ and L_{al}. More details about the training of the algorithm can be found in Appendix B.

4. Experiments

4.1. Experiment settings

4.1.1. Dataset and evaluation protocols

We conducted our experiments on two existing multilingual KG datasets *DBP-5L* [9] and *E-PKG* [11]. *DBP-5L* is a dataset sampled from DBpedia [34], containing five KGs from different languages: English (EN), French (FR), Spanish (ES), Japanese (JA) and Greek (EL). *E-PKG* is an industrial multilingual E-commerce product KG dataset, which describes phone-related products across six different languages: English (EN), German (DE), French(FR), Japanese (JA), Spanish (ES) and Italian (IT). The detailed dataset statistics are reported in Table 1. We follow the same evaluation protocols in previous works [9–12]: the performance of models are evaluated on tail prediction task (h, r, ?) for all KGs with Hit@1, Hit@10 and mean reciprocal rank (MRR) metrics. All experiments are conducted on a server running Ubuntu Linux, 128 GB of RAM, and an NVIDIA A100 GPU (80 GB).

4.1.2. Selected baseline models

In the experiments, we have selected four popular state-of-the-art models discussed in Section 2 on the MKGC task to compare with our proposed method. KenS [9] was the first work to propose the MKGC problem setting, and it uses ensemble prediction results from different KGs to generate the final prediction. AlignKGC [10] is a multi-task model that leverages the collaborative nature between alignments and link prediction tasks and trains them jointly. SS-AGA [11] and JMAC [12] are the two most recent works that both use alignment predictions to generate additional training signals and produce state-of-the-art performances for the MKGC task.

4.1.3. Model hyperparameters

Our model is implemented in PyTorch. For the fairness of comparison, we set the embedding dimension to be 100 in our CA-MKGC method, which is strictly less than or equal to the selected baseline models. We used TransE [3] as our KG embedding module, which is consistent with the existing models we compare with. Our method can also be easily adapted to use other KG embedding modules. For training, our CA-MKGC method is trained for a total of 90 epochs, where active learning is enabled after 30 epochs and alignment prediction is enabled after 60 epochs. Grid search is used for all hyperparameters: the margin of KG embedding training $\gamma \in \{1, 5, 9, 13\}$, alignment confidence threshold $\theta \in \{0.02, 0.04, 0.06, 0.08\}$ when text embeddings are not used to assist with alignment prediction, and $\theta \in \{0.92, 0.94, 0.96, 0.98, 1.0\}$ and text embedding weight $\alpha \in \{0.5, 1, 2\}$ for when text embeddings are used. The best checkpoints are all selected based on the best MRR performance on the validation set.

4.2. Results

Table 2 reports the results of our model compared to other existing state-of-the-art models on *DBP-5L* dataset. In the table, we follow the evaluation in [12] and compare the models under two separate settings: *w/ SI* and *w/o SI*, denoting experiments with and without surface information (text information of nodes) respectively. In both settings, our method CA-MKGC achieved state-of-the-art performances in almost all metrics. In the *w/ SI* setting, where textual information can be used as an additional resource for alignment prediction, our method achieved an average 3.4% absolute improvement on Hit@1 and 1.1% absolute improvement on MRR while providing slightly worse performance on Hit@10 compared to the recent state-of-the-art JMAC. In the *w/o SI* setting, where alignments are predicted purely based on structural information of graphs, our method achieved significant improvements across all KGs compared with state-of-the-art works. The comparison with state-of-the-art models verifies that prior knowledge, like unique alignment restriction, is challenging to learn purely from semantics, and imposing such prior knowledge is very effective for the MKGC task. Compared to existing works that all train one model for each individual KG for knowledge transfer, our unified encoder model achieved better results while requiring fewer total trainable parameters in both settings.

As a recently curated industrial E-commerce product KG dataset, *E-PKG* dataset exhibits lower data quality compared to the *DBP-5L*, which was generated from sampling the widely adopted and verified DBpedia [34]. Specifically, when we connect all entities with seed alignment annotations to form the aforementioned "alignment connected components", we discovered several connected components of size over 1000 in such a graph on the E-PKG dataset. This suggests two possible scenarios: either there is a significant number of duplicate entities within each KG, or the quality of the seed alignment annotations is relatively low.

Table 2
Results on DBP-5L dataset. Performances of existing state-of-the-art models were directly reported from [12].

	Align.	EL H@1/H@10/MRR	JA H@1/H@10/MRR	ES H@1/H@10/MRR	FR H@1/H@10/MRR	EN H@1/H@10/MRR
KenS	100%	28.1/56.9/–	32.1/65.3/–	23.6/60.1/–	25.5/62.9/–	15.1/39.8/–
SS-AGA	100%	30.8/58.6/35.3	34.6/66.9/42.9	25.5/61.9/36.6	27.1/65.5/38.4	16.3/41.3/23.1
AlignKGC w/o SI	50%	55.1/84.2/65.5	46.7/74.4/56.6	44.0/71.4/53.4	44.5/74.0/54.9	28.5/54.9/37.5
JMAC w/o SI	50%	46.8/90.3/62.3	48.1/**85.2**/61.3	35.6/76.7/50.3	43.8/**83.8**/58.0	24.5/65.3/38.3
CA-MKGC w/o SI	50%	**58.2**/**90.7**/**70.2**	**51.5**/84.0/**63.3**	**46.4**/**80.6**/**58.9**	**48.4**/83.6/**61.2**	**33.3**/**68.9**/**45.6**
AlignKGC w/ SI	50%	58.2/88.6/69.4	49.3/78.7/60.1	48.0/76.6/58.0	48.4/79.4/59.5	31.7/59.8/41.3
JMAC w/ SI	50%	55.2/**97.5**/71.7	53.3/**91.4**/**66.8**	45.4/**88.2**/61.0	49.3/**91.3**/64.5	29.5/72.7/44.6
CA-MKGC w/ SI	50%	**59.6**/93.8/**72.5**	**54.6**/87.5/**66.8**	**49.1**/85.9/**62.2**	**51.3**/87.9/**64.6**	**34.9**/**72.8**/**48.0**

Table 3
Results on E-PKG dataset. We directly report the benchmarking results from [11] for the first two rows, while we produce the performances of the JMAC model since the E-PKG dataset was not available when Tong et al. [12] proposed JMAC.

	Align.	EN H@1/H@10/MRR	DE H@1/H@10/MRR	FR H@1/H@10/MRR	JA H@1/H@10/MRR	ES H@1/H@10/MRR	IT H@1/H@10/MRR
KenS	100%	26.2/69.5/–	24.3/65.8/–	25.4/68.2/–	33.5/73.6/–	21.3/59.5/–	25.1/64.6/–
SS-AGA	100%	26.7/69.8/41.5	24.6/66.3/39.4	25.9/68.7/40.2	33.9/74.1/48.3	21.0/60.1/36.3	24.9/63.8/38.4
JMAC w/o SI	50%	46.1/77.5/58.6	38.2/73.3/52.0	38.1/74.9/52.3	60.8/82.3/69.0	33.0/66.3/46.0	39.7/72.8/53.0
CA-MKGC w/o SI	50%	**51.7**/**82.0**/**63.7**	**42.4**/**76.6**/**55.4**	**41.5**/**78.5**/**55.7**	**63.3**/**84.1**/**71.1**	**37.0**/**69.3**/**49.4**	**43.8**/**76.0**/**56.4**

Table 4
Ablation study results of Intra-KG GCN encoder on DBP-5L.

	Align.	Metric	EL	JA	ES	FR	EN
KGC-I	100%	H@1	**64.8**	55.2	48.3	51.7	34.5
		H@10	94.8	88.9	83.7	87.1	71.8
		MRR	**76.4**	67.3	61.1	64.8	47.6
Intra-KG	100%	H@1	62.3	**57.4**	**51.7**	**53.3**	**35.8**
		H@10	**96.1**	**89.6**	**85.8**	**89.5**	**73.0**
		MRR	75.1	**69.2**	**64.2**	**66.5**	**48.8**
KGC-I	50%	H@1	50.6	47.3	38.1	40.0	27.3
		H@10	84.7	77.8	72.6	77.3	64.5
		MRR	63.3	58.1	50.4	53.3	40.4
Intra-KG	50%	H@1	**52.8**	**47.5**	**40.1**	**41.8**	**29.1**
		H@10	**86.9**	**80.6**	**75.0**	**79.7**	**65.9**
		MRR	**65.5**	**59.2**	**52.5**	**55.1**	**42.0**

This particular issue within the dataset can lead to an excessive amount of inferred information when utilizing alignment-entailed meta-paths to infer new alignments and facts from seed alignments. As a result, it hampers the convergence of existing methods, including our own. To solve this issue, we produce a cleaned version of *E-PKG* dataset that removes the seed alignment annotations resulting in alignment connected components of size 20. We then compare our model on this dataset with the state-of-the-art models reported in [11] and the most recent model, JMAC, which is the top performer on *DBP-5L* dataset. The comparison was made in *w/o SI* scenario because the names of entities in *E-PKG* dataset were not readily available, and the results are reported in Table 3. Our model outperforms all other models by a significant margin across all metrics, averaging 4% absolute increase compared to the state-of-the-art JMAC model. This indicates the effectiveness of the proposed CA-MKGC under a sparse and noisy setting.

4.3. Ablation study

4.3.1. Intra-KG GCN encoder

In Table 4, we report the results of the ablation study for our intra-KG GCN encoder. In this table, we created a static setting by only using 50% and 100% seed alignments to train the models while disabling the alignment prediction component that generates incremental supervision and the active learning component that generates additional feedback for embedding training. Under the static setting, we compare and showcase the effectiveness of our intra-KG GCN encoder (represented by *Intra-KG* in table) by comparing to vanilla CompGCN models

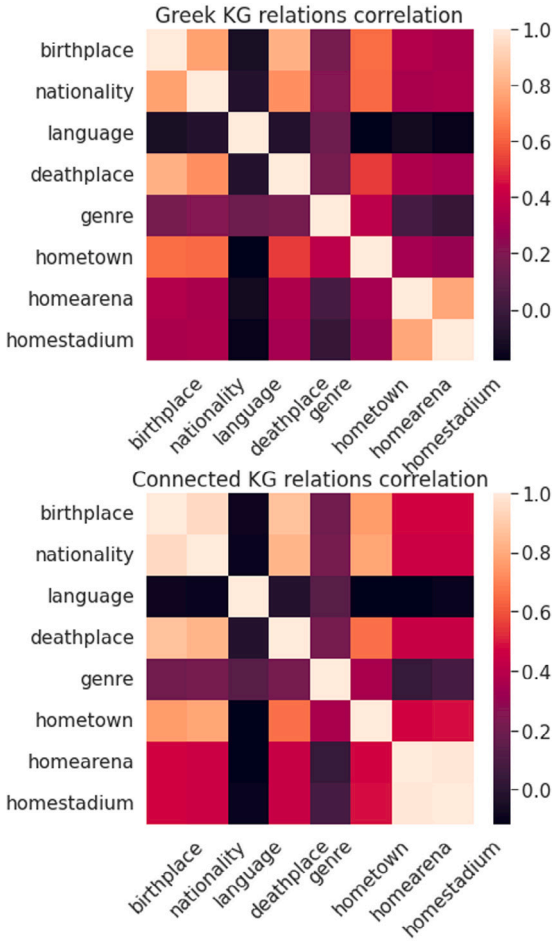


Fig. 4. Relation correlation heatmap of Greek KG vs. fused KG.

(represented by *KGC-I* in table) trained on individual KGs. To highlight the contribution provided by only intra-KG message propagation and parameter sharing, we infer additional facts from seed alignments using alignment-entailed meta paths for both models to ensure they use the same set of training facts. It can be observed that in both scenarios,

Table 5
Ablation study results on DBP-5L dataset w/o surface information.

	EL H@1/H@10/MRR	JA H@1/H@10/MRR	ES H@1/H@10/MRR	FR H@1/H@10/MRR	EN H@1/H@10/MRR
CA-MKGC	**58.2/90.7/70.2**	**51.5**/84.0/**63.3**	**46.4/80.6/58.9**	**48.4/83.6/61.2**	**33.3/68.9/45.6**
CA-MKGC w/o greedy filter	55.9/87.5/67.9	48.2/81.5/59.9	42.4/76.5/54.3	44.0/79.8/56.7	30.1/65.9/42.6
CA-MKGC w/o AL	57.6/89.6/69.89	50.5/83.6/62.8	45.4/80.0/57.8	46.9/82.7/59.8	31.5/67.7/44.2
CA-MKGC w/o alignment prediction	55.8/87.4/67.7	47.8/80.0/59.6	42.1/77.8/54.8	44.6/81.1/57.7	30.4/66.7/43.0

Table 6
Results of quality of alignment predictions made by our proposed CA-MKGC.

	Precision	Recall
CA-MKGC w/o SI	79.7%	59.6%
CA-MKGC w SI	93.8%	87.3%

the intra-KG GCN encoder model generates superior results compared to the vanilla CompGCN model. This verifies the effectiveness of the intra-KG GCN encoder by modeling all KGs under one unified encoder and enabling parameter sharing and message propagation across KGs.

To further understand the benefit of the intra-KG GCN encoder on fused KG, we compared the correlations of learnt relation embeddings. In Fig. 4, we visualize the correlations of relation embeddings learnt on Greek KG and the fused KG on the DBP-5L dataset, with vanilla CompGCN encoder and the intra-KG GCN encoder, respectively. The correlations of learnt KG relations in the Greek KG differ from those discovered in the fused KG with the unified model. Specifically, pairs of relatable connections like (“birthplace”, “nationality”), (“birthplace”, “deathplace”), and (“homearena”, “homestadium”), all exhibit a higher correlation in the fused KG compared to the individual Greek KG. This implies the unified model can capture more accurate dynamics between relations with shared network parameters and relation embeddings. This also helps prevent the low-resource Greek KG from converging to a suboptimal region due to the lack of complete information.

4.3.2. Unique alignment restriction

In Table 5, we report the results of the ablation study for our designed components to impose the unique alignment restriction. In the table, the performances of the following variants of our model are reported: (1) *w/o greedy filter* is the model without greedy filter applied at prediction time of alignments when generating additional supervisions; (2) *w/o AL* is variant with active-learning based loss removed during training; (3) *w/o alignment prediction* is the model without incremental supervision. For consistency of comparison, all models in the table are reported on **DBP-5L** dataset with 50% seed alignment, trained without surface information. The table shows that the model performance decreased by about 8% across the board when incremental supervision was turned off. This indicates the importance of the iterative inclusion of confident alignment predictions. We also observe a noticeable decrease in all metrics when active learning loss is excluded from the training process. This shows that the active learning feedback mechanism helped the model learn to avoid potentially harmful predictions. Interestingly, when the greedy filter is turned off,[1] the model performance is equally inferior to the variant without any incremental supervision. This shows that: (1) the errors in predictions can propagate across the model and ultimately offset most of the benefits introduced by the incremental supervision inferred via alignment-entailed meta-paths; (2) the greedy filter is very effective in preventing large clusters (connected components) of aligned entities and therefore effectively limiting the propagation of errors.

[1] To run the w/o greedy filter variant, we had to set θ to be relatively high to avoid the excessive run time when generating additional supervision caused by large-size alignment connected components

4.4. Predicted alignments quality

Another benefit of our proposed model is its independence from the information of unaligned entities. Concretely, we discovered that the previous state-of-the-art models like JMAC implicitly rely on knowledge of the population of the remaining 50% of unaligned entities when only 50% of the seed alignments are used for training. We want to point out that this is impractical and unrealistic in real-world settings. Therefore, we choose not to use this information in CA-MKGC. In our proposed model, we do not limit the search space of alignment predictions, allowing the model to make predictions between any entities. To provide insights into the quality of the alignment predictions made by our model, in Table 6, we report the precision and recall rates on the 50% of seed alignments that are not available to the model during training in *DBP-5L* dataset. Unsurprisingly, surface information appears to be very useful in predicting alignments: both the precision and recall rate improves substantially when surface information is included. The improvement in alignment prediction quality is also the main driver of superior performances of models with surface information on the MKGC tasks reported in Table 2. However, we must note that surface information is not always guaranteed to be available in the real-world dataset, e.g. *E-PKG* dataset. In those cases, as shown in Table 6, our model still produces decent-quality alignment predictions despite limited information.

4.5. Reliance on seed alignments

In the experiment, one of the most important assumptions is the existence of seed alignments. Given the size and scale of KGs nowadays, it is difficult to obtain high-quality annotations of seed alignments on a large scale. Therefore, we find it very important to study our model’s reliance on seed alignments. In addition to the 100% and 50% seed alignments we tested in previous tables, we additionally evaluate our model on the DBP-5L dataset with varying percentages (0%–100%) of seed alignments available during training. Fig. 5 shows the plot of the performance (MRR) of our model against varying percentages of seed alignments. It is not surprising to observe that the MRR for each language decreases as the percentage of seed alignment decreases. However, with the same interval size in percentages, the performance drop from 40% to 0% is significantly larger than other intervals. This shows that the proposed model could leverage the context from all KGs to learn a reasonably good representation with $\geq$40% of the provided seed alignments, which translates to roughly 19% of total entities. While the model shows decent performance when fewer seed alignments are provided, we want to note that the ablation study also reveals the model’s reliance on the availabilities of seed alignments to a certain degree, which is expected since the availability of seed alignments was one of the assumptions for the study.

When no seed alignments are visible to the model during training, the initial prediction of alignment for *w/o SI* scenario would be extremely difficult. We therefore additionally experimented with 0% seed alignment under *w/ SI* scenario, in which initial entity alignment predictions are made based purely on multilingual text embeddings. The results are also plotted in Fig. 5. It can be observed that when no seed alignment annotations are available to the model, multilingual text embeddings could also provide decent signals to predict alignments between KGs, and the model outperforms *w/o SI* scenario by a large

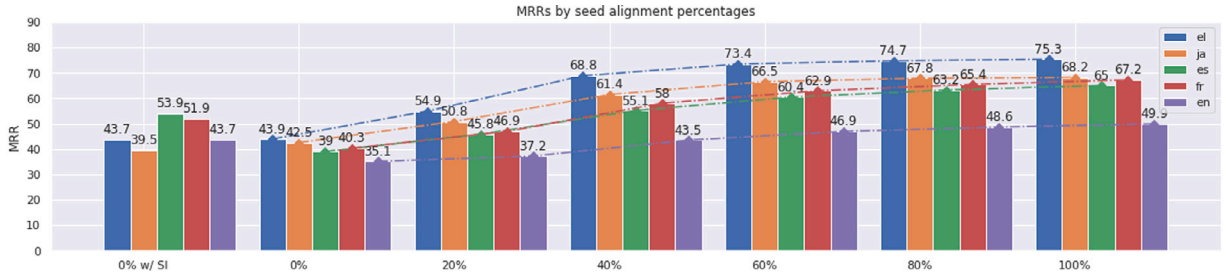


Fig. 5. MRRs of DBP-5L dataset with varying seed alignment percentages.

margin on Spanish, French and English KG. However, on smaller KGs (Greek and Japanese) that are more sensitive to noises introduced by alignment predictions, we observed performance degradation.

5. Conclusion & limitations

In this paper, we proposed a novel method that focuses on the MKGC task called CA-MKGC. The proposed method uses an intra-KG GCN encoder for message propagation across multiple KGs and encodes them in a unified model. In addition, we examined how incorrect facts propagate in the MKGC problem when confident predictions are used to generate additional supervision for the training. Based on our findings, we proposed an active learning mechanism and a greedy approximation of a semi-constrained optimization problem to help impose an effective unique alignment restriction between KGs while iteratively adding confident alignment predictions for incremental supervision. We demonstrate that the proposed intra-KG GCN encoder effectively leverages the collective knowledge of multiple KGs, and the simple restriction between KGs is very effective in restricting the propagation of errors when predictions of potential alignments are used for incremental training. Extensive experiments on multilingual datasets show that our method CA-MKGC achieves state-of-the-art results on multiple datasets compared with existing methods.

There also exist limitations in our proposed method. Although we proposed a first attempt in this paper to solve the error propagation issue in MKGC problem, the proposed solution only serves as a reasonable approximation to the optimization problem. To obtain a fully optimized solution, more research effort is needed to design a more computationally efficient method. Additionally, more comprehensive metrics can be used in the optimization to determine the model's confidence in the alignment predictions. We have also planned future works around these limitations to improve current systems on multi-knowledge graph completion.

CRediT authorship contribution statement

Weihang Zhang: Software, Methodology, Data curation, Conceptualization. **Ovidiu Şerban:** Writing – review & editing, Validation, Methodology, Conceptualization. **Jiahao Sun:** Writing – review & editing, Funding acquisition, Conceptualization. **Yike Guo:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Funding

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

Appendix A. Discussion of computational efficiency and scalability

With the ever-growing size of modern KGs, computational efficiency and scalability have become important factors when applying a method for real-world problems. The proposed CA-MKGC method in this paper generally scales well, and it has a few advantages compared to existing methods. One of the unique properties of the proposed model is that it considers the implication of alignments between all KGs globally, and models all KGs in the dataset together as one unified model. In doing so, we can avoid the common approach of existing methods that train one model for each KG in the dataset. This saves time on both training time and hyperparameter tuning. This makes the model very scalable as the number of KGs increases. Admittedly, modeling all KG in a unified model has its own disadvantage: compared to other methods, the memory footprint of our model is larger. The intra-KG message propagation mechanism requires storing all node and edge embeddings as well as the sparse adjacency matrices of the unified graph. Luckily, the memory consumption of them all scales linearly with the number of nodes, which is standard for all methods for this problem. On the very large datasets where the memory footprint on GPU might become a concern, we can potentially consider distributed training to alleviate the problem. At the same time, we would still like to note that the advantages of training a unified model that considers all KGs globally have unique advantages that justify the additional burden on memory footprint.

Appendix B. Training process of CA-MKGC

Algorithm 1 Pseudocode of the training process of CA-MKGC.

▷ Stage 1: Train the model M_f on fused KG KG_f with the KGC loss and AL loss.
while M_f not converged **do**
  $M_f \leftarrow$ Update w.r.t L_t
  $M_f \leftarrow$ Update w.r.t L_{al}
end while
▷ Stage 2: Iteratively include additional supervisions inferred from new alignment predictions and further train the model.
while not converged **do**
  $T_{inc} \leftarrow$ predict potential alignments with M_f
  $T_{inc}^{filtered} \leftarrow$ select alignments with greedy filter
  for epoch $\in 1..k$ **do**
    $M_f \leftarrow$ Update w.r.t $L_t + L_{inc}^{filtered}$
    $M_f \leftarrow$ Update w.r.t L_{al}
  end for
end while

References

[1] Z. Chen, Y. Wang, B. Zhao, J. Cheng, X. Zhao, Z. Duan, Knowledge graph completion: A review, IEEE Access 8 (2020) 192435–192456, http://dx.doi.org/10.1109/ACCESS.2020.3030076, Conference Name: IEEE Access.

[2] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, W. Zhang, Knowledge vault: a web-scale approach to probabilistic knowledge fusion, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2014-08-24, pp. 601–610, http://dx.doi.org/10.1145/2623330.2623623.

[3] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, O. Yakhnenko, Translating embeddings for modeling multi-relational data, Adv. Neural Inf. Process. Syst. 26 (2013).

[4] Z. Wang, J. Zhang, J. Feng, Z. Chen, Knowledge graph embedding by translating on hyperplanes, in: Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014.

[5] Y. Lin, Z. Liu, M. Sun, Y. Liu, X. Zhu, Learning entity and relation embeddings for knowledge graph completion, in: Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

[6] B. Yang, W.-t. Yih, X. He, J. Gao, L. Deng, Embedding entities and relations for learn- ing and inference in knowledge bases, 2015, p. 13.

[7] X. Lv, Y. Lin, Y. Cao, L. Hou, J. Li, Z. Liu, P. Li, J. Zhou, Do pre-trained models benefit knowledge graph completion? A reliable evaluation and a reasonable approach, in: Findings of the Association for Computational Linguistics: ACL 2022, Association for Computational Linguistics, 2022, pp. 3570–3581, http://dx.doi.org/10.18653/v1/2022.findings-acl.282.

[8] W. Cui, X. Chen, Instance-based learning for knowledge base completion, 2022.

[9] X. Chen, M. Chen, C. Fan, A. Uppunda, Y. Sun, C. Zaniolo, Multilingual knowledge graph completion via ensemble knowledge transfer, in: Findings of the Association for Computational Linguistics: EMNLP 2020, Association for Computational Linguistics, Online, 2020, pp. 3227–3238, http://dx.doi.org/10.18653/v1/2020.findings-emnlp.290.

[10] H. Singh, P. Jain, M. Mausam, S. Chakrabarti, Multilingual knowledge graph completion with joint relation and entity alignment, 2021.

[11] Z. Huang, Z. Li, H. Jiang, T. Cao, H. Lu, B. Yin, K. Subbian, Y. Sun, W. Wang, Multilingual knowledge graph completion with self-supervised adaptive graph alignment, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 474–485, http://dx.doi.org/10.18653/v1/2022.acl-long.36.

[12] V. Tong, D.Q. Nguyen, H.T. Trung, T.T. Nguyen, Q.V.H. Nguyen, N. Mathias, Joint multilingual knowledge graph completion and alignment, in: Findings of the Association for Computational Linguistics: EMNLP 2022, 2022.

[13] M. Nickel, V. Tresp, H.-P. Kriegel, A three-way model for collective learning on multi-relational data, 2011, p. 8.

[14] S. Ji, S. Pan, E. Cambria, P. Marttinen, P.S. Yu, A survey on knowledge graphs: Representation, acquisition and applications, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–21, http://dx.doi.org/10.1109/TNNLS.2021.3070843, arXiv:2002.00388.

[15] M. Schlichtkrull, T.N. Kipf, P. Bloem, R. van den Berg, I. Titov, M. Welling, Modeling relational data with graph convolutional networks, in: A. Gangemi, R. Navigli, M.-E. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, M. Alam (Eds.), The Semantic Web, in: Lecture Notes in Computer Science, Springer International Publishing, Cham, 2018, pp. 593–607, http://dx.doi.org/10.1007/978-3-319-93417-4_38.

[16] S. Vashishth, S. Sanyal, V. Nitin, P. Talukdar, Composition-based multi-relational graph convolutional networks, 2020, p. 16.

[17] D. Yu, Y. Yang, R. Zhang, Y. Wu, Knowledge embedding based graph convolutional network, in: Proceedings of the Web Conference 2021, WWW '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 1619–1628, http://dx.doi.org/10.1145/3442381.3449925.

[18] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2017, arXiv:1609.02907 [cs, stat].

[19] Y. Yang, C. Huang, L. Xia, C. Li, Knowledge graph contrastive learning for recommendation, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 1434–1443, http://dx.doi.org/10.1145/3477495.3532009.

[20] Z. Tan, Z. Chen, S. Feng, Q. Zhang, Q. Zheng, J. Li, M. Luo, KRACL: Contrastive learning with graph context modeling for sparse knowledge graph completion, in: Proceedings of the ACM Web Conference 2023, WWW '23, Association for Computing Machinery, 2023, pp. 2548–2559, http://dx.doi.org/10.1145/3543507.3583412.

[21] M. Chen, Y. Tian, M. Yang, C. Zaniolo, Multilingual knowledge graph embeddings for cross-lingual knowledge alignment, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization, Melbourne, Australia, 2017, pp. 1511–1517, http://dx.doi.org/10.24963/ijcai.2017/209.

[22] Z. Sun, W. Hu, Q. Zhang, Y. Qu, Bootstrapping entity alignment with knowledge graph embedding, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, 2018, pp. 4396–4402, http://dx.doi.org/10.24963/ijcai.2018/611.

[23] X. Zhao, W. Zeng, J. Tang, W. Wang, F. Suchanek, An experimental study of state-of-the-art entity alignment approaches, IEEE Trans. Knowl. Data Eng. (2020) 1, http://dx.doi.org/10.1109/TKDE.2020.3018741, Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[24] K. Zeng, C. Li, L. Hou, J. Li, L. Feng, A comprehensive survey of entity alignment for knowledge graphs, AI Open 2 (2021) 1–13, http://dx.doi.org/10.1016/j.aiopen.2021.02.002.

[25] G. Lample, A. Conneau, M. Ranzato, L. Denoyer, H. Jégou, Word translation without parallel data, in: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, OpenReview.net, 2018.

[26] S. Luo, P. Cheng, S. Yu, Semi-constraint optimal transport for entity alignment with dangling cases, 2022, http://dx.doi.org/10.48550/ARXIV.2203.05744, Publisher: arXiv Version Number: 3.

[27] W. Zeng, X. Zhao, J. Tang, X. Lin, Collective entity alignment via adaptive features, 2020, http://dx.doi.org/10.1109/ICDE48307.2020.00191.

[28] R. Zhu, M. Ma, P. Wang, RAGA: Relation-aware graph attention networks for global entity alignment, in: K. Karlapalem, H. Cheng, N. Ramakrishnan, R.K. Agrawal, P.K. Reddy, J. Srivastava, T. Chakraborty (Eds.), Advances in Knowledge Discovery and Data Mining, in: Lecture Notes in Computer Science, Springer International Publishing, Cham, 2021, pp. 501–513, http://dx.doi.org/10.1007/978-3-030-75762-5_40.

[29] X. Lin, E. Haihong, W. Song, H. Luo, EchoEA: Echo information between entities and relations for entity alignment, ArXiv (2021).

[30] H. Wang, S. Li, R. Pan, An adversarial transfer network for knowledge representation learning, in: Proceedings of the Web Conference 2021, ACM, Ljubljana Slovenia, 2021, pp. 1749–1760, http://dx.doi.org/10.1145/3442381.3450064.

[31] Z. Sun, Q. Zhang, W. Hu, C. Wang, M. Chen, F. Akrami, C. Li, A benchmarking study of embedding-based entity alignment for knowledge graphs, Proc. VLDB Endow. 13 (12) (2020) 2326–2340, http://dx.doi.org/10.14778/3407790.3407828.

[32] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186, http://dx.doi.org/10.18653/v1/N19-1423.

[33] K. Zeng, Z. Dong, L. Hou, Y. Cao, M. Hu, J. Yu, X. Lv, L. Cao, X. Wang, H. Liu, Y. Huang, J. Feng, J. Wan, J. Li, L. Feng, Interactive contrastive learning for self-supervised entity alignment, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, CIKM '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 2465–2475, http://dx.doi.org/10.1145/3511808.3557364.

[34] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, DBpedia: A nucleus for a web of open data, in: K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, P. Cudré-Mauroux (Eds.), The Semantic Web, in: Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2007, pp. 722–735, http://dx.doi.org/10.1007/978-3-540-76298-0_52.