Cross-Knowledge Graph Relation Completion for Non-isomorphic Cross-lingual Entity Alignment

Yuhong Zhang[1,2], Dan Lu[1,3], Chenyang Bu[1,3], Kui Yu[1,3], Xindong Wu[1,3]
[1]Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Hefei 230009, China
[2]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, Anhui, China
[3]School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China
Corresponding: Yuhong Zhang(zhangyh@hfut.edu.cn)

Abstract—**The Cross-Lingual Entity Alignment (CLEA) aims to find the aligned entities that refer to the same identity from two Knowledge Graphs (KGs) in different languages. In real-world applications, the neighborhood structures of the same entities in different KGs tend to be non-isomorphic, which makes the entity representation contain diverse semantics information and poses a great challenge for CLEA. In this paper, we address this challenge from two perspectives. On the one hand, cross-KG relation completion rules are designed with the alignment constraint of entities and relations to improve the isomorphism of two KGs. On the other hand, a representation method combining isomorphic weights is designed to include more isomorphic semantics for counterpart entities, which will benefit CLEA. Experimental results show that our model can improve the isomorphism of two KGs and the alignment performance, especially for two non-isomorphic KGs.**

Keywords: Knowledge Graphs, Cross-Lingual Entity Alignment, Non-isomorphism, Relation Completion

I. Introduction

Knowledge Graphs (KGs) play an important role in NLP field and data mining-related fields, such as question answering [4], industrial and academic settings [12]. But the construction of KGs is very hard that needs substantial resources. Due to the scarcity of available resources, it is difficult to build KGs for under-resourced languages, such as Greek, Arabic, etc. To address this problem, recent research has proposed Cross-Lingual Entity Alignment (CLEA) to enhance KG of under-resourced language using well-resourced language [4].

CLEA is to identify the aligned entity pairs referring to the same objects from two KGs in different languages. To this end, CLEA methods try to map the entities and relations in two KGs into a shared space, in which, the embeddings of the same objects in two KGs are as close as possible. Existing CLEA methods are classified into TransE-based methods [11] and Graph Neural Network-based (GNN-based) methods [22]. TransE-based methods assume that two KGs in different languages have a similar structure, so the embeddings of aligned entity pairs should have relative similar positions in the vector spaces. Recently, GNN-based methods have gained a lot of attention due to their great performance. GNN-based methods first learn the entity embeddings by aggregating the neighboring entities and then evaluate the similarity between entities based on their embeddings. The entities with the nearest geometric distance are regarded as a pair of aligned entities. These methods have proven their effectiveness for the isomorphic KGs.

However, owing to imbalanced resources and different cultures, two KGs in different languages are non-isomorphic generally. Particularly, the ratio of non-isomorphic neighbors is more than 85% for two KGs [17]. The non-isomorphism means that the counterpart entities in two KGs tend to contain heterogeneous neighboring entities, the different numbers of neighbors and relations. As shown in Fig. 1, Given two non-isomorphic KGs and some aligned entities as supervised seeds, (represented by the same shape in yellow), we aim to find more new aligned pairs, such as “林肯” and “Lincon” (red double dashed line). However, The neighborhoods of “林肯” and that of “Lincon” are heterogeneous. They have different neighbors. However, GNN-based methods aggregate these heterogeneous neighbors, which will lead to monolingual embeddings containing different semantics. Thus, the non-isomorphism will hold back CLEA and pose a great challenge.

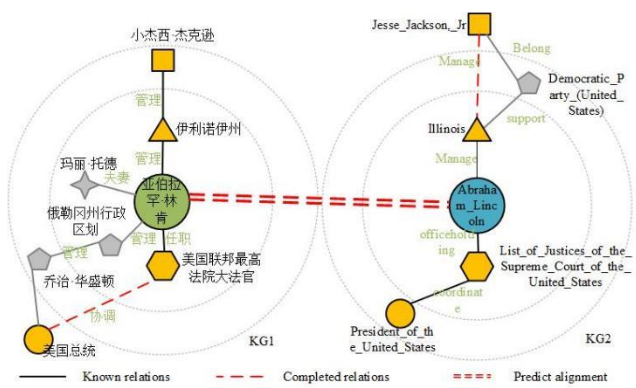

Figure 1. The illustration of non-isomorphism of KGs and our idea of cross-KG relation completion. The neighborhoods of “林肯” and “Lincon” are heterogeneous. Our idea is to change the topology of two neighborhoods by completing the relations (red dashed lines).

A few studies have focused on non-isomorphic CLEA. Their common idea is to expand the neighboring scopes or filter noisy neighbors. Alinet [17] thinks that distant neighbors may include more homogeneous entities and expands neighborhoods to cover more neighbors. DAEA [15] identifies the useful neighborhood with the importance of relations, and then the embedding will include similar neighbors and exclude noisy ones. However, the above methods try to find available information from a single KG, which has two limitations.

DOI reference number: 10.18293/DMSVIVA23-051

1) Although expanding the neighbors' scope can cover more homogeneous neighbors, it inevitably covers some noisy neighboring entities. Moreover, with the increasing scope, the cost of representation methods also increases exponentially. In addition, the expansion of neighbor scope cannot improve the topology isomorphism of two KGs.

2) Existing methods enrich the representation by paying more attention for closer or more similar neighbors. If these closer neighbors are non-isomorphic, the representation of entities will focus more on heterogeneous neighbors and lead to semantic difference.

To address these problems, we propose a method named cross-KGs relation completion for non-isomorphic CLEA. Our method addresses the non-isomorphism in two views. Firstly, with the assumption that counterpart entities should have isomorphic neighborhoods, cross-KGs relation completion is designed to change the topology of KGs and improve the isomorphism of two KGs, as shown with the red dashed lines in Fig.1. Secondly, the isomorphic weights are introduced into representation learning to make entity representation focus more on the isomorphic neighbors, which will benefit non-isomorphic CLEA. Our contributions are summarized as:

1) To address the non-isomorphic CLEA, we propose to improve the topology isomorphism of KGs by cross-KG relation completion. To our best knowledge, there is little work focusing on cross-KG completion owing to the unavailable connection KGs [25]. In this paper, we complete the relations with some supervised aligned information. With our cross-KG completion, both completeness and topology isomorphism can be improved. And KG representation will cover more isomorphic information on a smaller neighborhood.

2) To reduce the semantic discrepancy of counterpart entities, the isomorphic weights for two neighborhoods, not the similarity or importance to the central entity, are introduced into representation learning. The isomorphic weights will make the embedding include more isomorphic semantics and exclude non-isomorphic semantics, making CLEA more easily.

II. Related Work

A. Methods for CLEA

Existing CLEA methods are classified into TransE-based methods [11] and Graph Neural Network-based (GNN-based) methods [22]. And recent studies have shown that GNN-based methods can achieve outperformance. Gnn-based methods can be divided into two types: 1) **GNN with entity attention.** As an expansion of GNN, Graph Convolutional Network (GCN) can learn the node-level representation. GCNAlign [22] is the first study using GCN to learn the representation in low-dimensional space, and then measures the distance of entities to find new alignments. Some works [5] put two KGs into one GCN to learn a shared representation space, in which, it uses aligned pairs to make entities closer to each other. To represent the entities with more semantics, Graph Attention network (GAT) is used to make the representation focus more on the important or similar neighbors [20]. 2) **GCN with relation attention.** To measure the importance of neighbors accurately, some studies [13] combine the relations with neighbors to find useful neighbors. By giving more attention to those useful neighbors, the representation is enhanced further. Other studies use specific relation to update the attention for neighbors, such as the node attributes [14] and relation types [19][24].

In sum, GNN-based methods have proven their superiority for CLEA. However, they only achieve a good performance for similar knowledge graphs [17].

B. Methods for Non-isomorphic CLEA

Non-isomorphism is common in applications. That means the counterpart entities have non-isomorphic neighbors. It is a huge challenge for CLEA. The studies focus on non-isomorphic CLEA can be divided into two categories. 1) **Using additional information**. KDCoE [10] uses both entity description and multilingual literal description as additional information to co-train the embeddings of entities. N-gram [18] uses the attribute triples to generate the embeddings for attribute characters. Other works [1][21] also merge additional configuration information for entities by entities' attributes. 2) **Changing the range of neighborhood**. AliNet [17] is the first work for non-isomorphic CLEA. It expands the scope to cover more distant neighbors to increase the overlapping KGs. And it uses attention to reduce the noisy neighbors and emphasis the useful neighbors. KE-GCN [23] selects the right relations and their corresponding neighbors from all neighborhoods using translated method [11]. DAEA [15] uses the relation and level attention to filter useless and distant neighbors respectively.

Non-isomorphism has attracted much attention. However, existing methods find available information in one KG while neglecting the information from cross-KGs.

III. Our Proposed Method

Formally, KG is defined as $KG=(E, R, T)$. It consists of a set of entities E and a set of relations R, and the knowledge facts are stored in a collection T in form of triples (h, r, t), where h, t $\in E$ and r $\in R$. Given two non-isomorphic KGs, denoted as $KG1=(E1, R1, T1)$ and $KG2=(E2, R2, T2)$, the task of our CLEA is to find new aligned entity pairs using some supervised entity pairs $EP=(e1_i, e2_j) \mid (e1_i \in E1; e2_j \in E2)$.

The framework of our method is shown in Fig.2, which includes three steps. The first step is cross-KG relation completion. With the aligned entities as supervised information, the relations alignment is treated as a constraint to predict the potential relations. This step changes the topology of each KG and then improves the topology isomorphism of KGs. The second step is augmented representation. Isomorphic weights are introduced into GAT to make KG representation focus more on the isomorphic neighbors and then the semantic discrepancy of counterpart entities is reduced. The third step is the alignment with a loss function. The distance is used to search for the nearest entity in the whole space and then to find more alignment pairs.

A. Cross-KGs completion

In this paper, non-isomorphic CLEA is addressed differently. We introduce the cross-KG relation completion to change the structure of original KGs and make them more isomorphic.

It can be easily accepted that the same object in different KGs should have homogeneous neighborhoods. If they do not, there may be some relations missing. Motivated by this, we propose to complete these missing relations according to the aligned entities and relations. The supervised aligned entities are known as seeds, and then we need to align the relations. Owing to that the number of relation types is smaller than that of entities. Thus the alignment of relations is easier. We construct the set of aligned relations RP=($r1_i$, $r2_j$) | ($r1_i \in R1$; $r2_j \in R2$) as a constraint to the completion, which will reduce the noise in the cross-KG relation completion.

The cross-KG completion searches the potential relations in global KGs, and it includes 2 steps, relation alignment and relation completion. We will describe the two steps in detail.

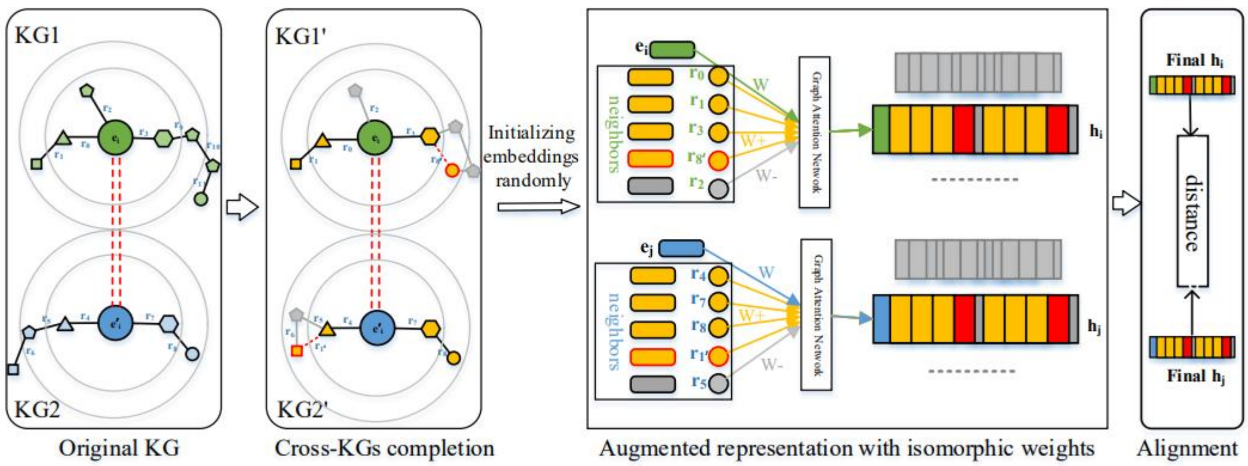

Figure 2. The framework of our method. In cross-KG relation completion step, isomorphic entities in yellow are treated as supervised pairs, and the red dashed lines are the completed relations. In augmented representation step, the yellow W^+ and gray W^- mean isomorphic and non-isomorphic weights respectively.

Relation Alignment Rule: Relations can be aligned based on whether their related entities are aligned. For two triples *(h1, r1, t1)* and *(h2, r2, t2)* from *KG1* and *KG2*, respectively, if the entities $<h1,h2>$ and $<t1,t2>$ are both aligned, we can infer that *r1* and *r2* should be aligned. Based on this observation, we design the first rule for relation alignment, which is formally expressed as follows.

$$\text{IF } ((h1, r1, t1) \in KG1 \text{ and } (h2, r2, t2) \in KG2 \text{ and } < h1, h2 > \in EP \text{ and } < t1, t2 > \in EP\,) \\ \text{THEN } \; RP += < r1, r2 >$$

Relation Completion Rule: Relations can be completed when two entities are connected in one KG but their aligned entities are not connected in another KG. With the aligned pairs $<h1,h2>$, $<t1,t2>$, and $<r1,r2>$ as constraints, if *(h1,r1,t1)* exists in *KG1* but *(h2,r2,t2)* is not included in *KG2*, we will complete the triple *(h2,r2,t2)* in *KG2*. We formally write this as the second rule for relation completion:

$$\text{IF } (< h1, h2 > \in EP \text{ and } < t1, t2 > \in EP \text{ and } < r1, r2 > \in RP \text{ and } (h1, r1, t1) \in KG1 \\ \text{THEN } KG2 += (h2, r2, t2)$$

The relation alignment rule is used to align relations in two KGs, and the aligned relations serve as constraints for relation completion. The relation completion rule is used to complete potential relations, and these completed relations provide more triples for relation alignment. The two rules are run iteratively. Finally, with the completed relations, the neighborhoods of entities change, and the non-isomorphism is reduced.

It is worth noting that the relation-aligned constraint is important for relation completion. Firstly, the aligned relations are introduced as additional information besides neighbors, which is helpful for CLEA. Secondly, the aligned relation constraint connects entities and relations into triples as an aligned unit, ensuring that the completed related neighbors are unambiguous and can reduce noise. If the aligned relations are ignored, we can only connect entities but cannot distinguish their neighbors by relation awareness. Therefore, the relation constraint is necessary.

B. The isomorphic weights for augmented representation

Although the isomorphism of KGs has been improved after completion, it cannot ensure the complete isomorphism of KGs. There still are some heterogeneous neighbors. Thus, we propose isomorphic weights to focus more on homogeneous neighbors and ignore heterogeneous ones in representation. And then the entity embedding will be more suitable for non-isomorphic CLEA. In this subsection, we first set different weights for isomorphic neighbors and non-isomorphic neighbors, and then learn the representation KGs. We take *KG1* as an example to show the weighted aggregation representation, and the representation of *KG2* is similar.

Isomorphic Weight Setting: For two counterpart entities, if their neighbors are known as aligned seeds, they are called isomorphic neighbors. It can be defined as follows.

Isomorphic Neighbors: Given $e_i \in KG1$ and $e_j \in KG2$, and $N_{ei} = n_i^{ei}$ and $N_{ej} = n_j^{ej}$ denote the neighbors of e_i and e_j. If $< n_i^{ei}, n_j^{ej} > \in EP$, they are called isomorphic neighbors. Otherwise, they are non-isomorphic neighbors. The isomorphic value is set as $I(n_i^{ei}, n_j^{ej})$. When $I(\cdot) = 1$, it means n_i^{ei} and n_j^{ej} are isomorphic, and when $I(\cdot) = -1$, it means they are heterogeneous.

$$I(n_i^{ei}, n_j^{ej}) = \begin{cases} 1 & < n_i^{ei}, n_j^{ej} > \in EP \\ -1 & < n_i^{ei}, n_j^{ej} > \notin EP \end{cases} \quad (1)$$

The isomorphic weights of neighbors are set according to whether they are isomorphic. Especially, the weights are initialized equally, and then the isomorphic neighbors will make the weight $W_{n_i^{ei}}$ larger owing to $I(\cdot) = 1$, and the non-isomorphic ones will make the weight smaller with $I(\cdot) = -1$. $W_{n_i^{ei}}$, the weight for *i*-th neighbor of entity e_i, is calculated as following.

$$W_{n_i^{ei}} = \frac{W_{initial}}{Z_{|N_{ei}|}} exp(I(n_i^{ei}, n_j^{ej})) \quad (2)$$

where $W_{initial}$ is the initial weight of neighbors, and it is set to one out of the number of neighbors equally. $Z_{N_{ei}}$ is the normalized factor used to normalize the weight value, and it is calculated as:

$$Z_{N_{ei}} = \sum_{i=1}^{|N_{ei}|} W_{initial} exp(I(n_i^{ei}, n_j^{ej})) \quad (3)$$

Weighted Augmented Representation: Isomorphic weights are combined with GAT to learn KG representation. Firstly, embeddings of entities and relations are initialized randomly, denoted as $h_{e_1}, h_{e_2}, \ldots, h_{e_k}$ and $h_{r_1}, h_{r_2}, \ldots, h_{r_k}$. Then the isomorphic weights are set to all neighbors and relations to augment those isomorphic neighbors. The weighted aggregation of relations and neighbors is shown in Formula 4 and 5.

$$h_{e_i,r} = \frac{1}{|N_{ei}|} (\sum_{r_k} W_{n_i^r} h_{r_k}) \quad (4)$$

$$h_{e_i,N} = \frac{1}{|N_{ei}|+1} (\sum_{e_i \in N_{ei}} W_{n_i^{ei}} h_{n_k^{ei}} + W_{initial} h_{ei}) \quad (5)$$

where h_{r_k} refers to the embeddings of relations associated with entity e_i, and $h_{n_k^{ei}}$ refers to the embeddings of neighbors that belong to entity e_i. Combining both relations and neighbors enables the representation of the entity e_i.

$$h_{e_i} = [h_{e_{i,r}} || h_{e_{i,N}}] \quad (6)$$

Secondly, the weighted embeddings h_{ei} are used as the input of GAT to learn the final representation of *KG*. Formula 7 shows the learning representation process of GAT.

$$h_{e_i}^{final} = ReLU(\frac{1}{Z} \sum_{z=1}^{1} [\sum_{e_i \in N_{ei}} \alpha_{i,j}^z h_{ei}]) \quad (7)$$

where Z is the number of head attention, $\alpha_{i,j}^z$ is the attention. Both Z and $\alpha_{i,j}^z$ are computed as MRAEA [8] does.

C. Entity Alignment

With the representation KGs, we find new entity pairs by searching the nearest entity to each other globally [22]. In this process, the distance between entities can be computed by the Manhattan distance.

$$dis(e_1, e_2) = |\hat{h}_{e_1}^{out} - \hat{h}_{e_2}^{out}| \quad (8)$$

where $\hat{h}_{e_1}^{out}$ and $\hat{h}_{e_2}^{out}$ represent the embeddings of e_1 and e_2.

To bring similar entities closer to each other in a uniform space, we shorten the distance by minimizing the following loss.

$$L = \sum_{<e_1,e_2>\epsilon EP} ReLU(dis(e_1, e_2) - dis(e'_1, e_2) - dis(e_1, e'_2) + \lambda \quad (9)$$

where λ represents the margin hyper-parameter. The entities e'_1 and e'_2 are considered as negative entities. We randomly select negative pairs from *E1* and *E2*, similar to MRAEA [8]. As shown in Formula 9, our calculations enhance positive samples and weaken negative samples in order to narrow the alignment entity distance.

IV. EXPERIMENT

A. Datasets and Baselines

The performance of our method is evaluated on three large cross-lingual datasets from DBP15K, which are used commonly in many studies. For this dataset, the ratio of overlap coefficient (OC) is used to show the isomorphism of KGs. The OC is proposed in [17], and it is computed by the ratio of aligned neighbors to all neighborhoods in one-hop neighboring range. Higher the OC value, the more isomorphic two KGs [17]. For example, the OC value of ZH-EN is 11.7%, which means that only 11.7% of neighborhoods in Chinese and English KGs are homogeneous. It can be seen that the two KGs are non-isomorphic. the OC value of JA-EN is 11.6% and the OC value of FR-EN is 13.1%. Baselines

To validate the effectiveness of our method, fifteen baselines are compared with our method. These baselines fall into 3 categories. TransE-based baselines include MTransE [11], IPTransE [26], and NAEA [27]. GNN-based baselines include GCN-Align [22], MuGCN [2], GAT [20], R-GCN [13], MuGCN [2], MRAEA [8], RREA [9], Dual-AMN [7] , PSR [6], Sparse[3] and RpAlign [16]. There are also some baselines focusing on the non-isomorphic CLEA, including of AliNet [17], KE-GCN [23] and DAEA [15].

It is worthy to note that a few recent works [1][3][28] have achieved remarkable performances. [1][28] use some additional information, such as entities' attribute information and entities' description information. [3][28] initialize the representation with Glove embedding. In this paper, we compare our methods with its variants igoring the additional information for fair comparision.

B. Experimental Setting

DBP15K consists of three cross-lingual tasks, namely $DBP_{ZH\text{-}EN}$, $DBP_{JA\text{-}EN}$, and $DBP_{ER\text{-}EN}$. For a fair comparison, we use 30% of alignments data as training and the other 70% as testing, as other methods did [8]. In addition, there are some common parameters, which all are set to the same values as the previous works. The embeddings' dimensions of entities and relations *d=100*, attention head number *k=2*, the depth of GNN is set to 2, the dropout rate is 0.3 and the learning rate of Adam is 0.005. The margin-based loss function integrates some negative entities. The aggregation range, dropout rate, and learning ratio are set to 2, 0.3, and 0.005 respectively. In this paper, Hits@k and Mean Reciprocal Rank (MRR) are used to measure performance.

C. *Main Results*

Table I shows the performance of our method and baselines. Experimental results of all baselines are obtained from their original papers. Some conclusions can be drawn from Table I.

1) GCN-based methods outperform TransE-based methods, which is consistent with the conclusion of other works [19].

2) As for GCN-based methods, methods with GAT [7], [20] perform better than those with GCN. It is because that the similar or closer entities are given more attention to enrich the representation of KGs. In addition, the GNN-based methods focusing on both relations and entities [9], [13] perform better than those only focusing on the entities.

3) The methods for non-isomorphic CLEA, including AliNet [17], KE-GCN [23], DAEA [15] and ours, perform better than the GCN-based methods on average owing to that they address the non-isomorphism of KGs. It shows that the non-isomorphism does exist commonly in two KGs and addressing it will benefit the CLEA.

4) Our method has an obvious improvement than other non-isomorphism baselines, including AliNet [17], KE-GCN [23], DAEA [15], KE-GCN [23] and RpAlign[16]. H@1 of our method is improved by 14.5%, 13.2%, and 14.9% averagely on three datasets. Compared with AliNet [17], our method not only covers more neighbors but also changes the topology of the neighbors using cross-KG relation completion. By improving the isomorphism of KGs, our method achieves an improvement. And compared with KE-GCN, our method enriches the entity embedding by supplementing missing homogeneous neighbors rather than deleting heterogeneous neighbors, which includes more related and similar semantic information. Compared with RpAlign[16], Our completion rule depends on non-isomorphic relation and assign isomorphic weights to make the representation include more isomorphic information.

TABLE I. OVERALL PERFORMANCE OF ALL METHODS ON DBP15K DATASET

Types	Models	$\mathrm{DBP_{ZH-EN}}$			$\mathrm{DBP_{JA-EN}}$			$\mathrm{DBP_{FR-EN}}$		
		H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR
TransE-based CLEA	**MTransE [11]**	30.8	61.4	36.4	27.9	57.5	34.9	24.4	55.6	33.5
	IPTransE [26]	40.6	73.5	51.6	36.7	69.3	47.4	33.3	68.5	45.1
	NAEA [27]	65.0	86.7	72.0	64.1	87.2	71.8	67.3	89.4	75.2
GCN-based CLEA	**GCN-Align [22]**	41.3	74.4	54.9	39.9	74.5	54.6	37.3	74.5	53.2
	GAT [20]	41.8	66.7	50.8	44.6	69.5	53.7	44.2	73.1	54.6
	R-GCN [13]	46.3	73.4	56.4	47.1	75.4	57.1	46.9	75.8	57.0
	MuGCN [2]	49.4	84.4	61.1	50.1	85.7	62.1	49.5	87.0	62.1
	MRAEA [8]	65.7	89.5	74.4	72.7	92.3	79.8	73.9	93.8	81.0
	RREA [9]	71.5	**92.9**	79.0	71.3	**93.3**	79.3	73.9	94.6	81.6
	Dual-AMN [7]	73.1	92.3	79.9	72.6	92.7	79.9	75.6	**94.8**	**82.7**
	PSR [6]	70.2	92.4	78.1	69.8	93.0	78.2	73.1	94.1	80.7
	Sparse [3] (L=0)	58.5	78.0	-	59.1	79.1	-	76.0	91.5	-
non-isomorphism CLEA	**AliNet [17]**	53.9	82.6	62.8	54.9	83.1	64.5	55.2	85.2	65.7
	DAEA [15]	56.8	88.3	67.7	57.6	89.2	68.3	58.0	91.2	69.5
	KE-GCN [23]	56.2	84.2	66.4	57.0	85.2	67.0	57.2	88.5	68.3
	RpAlign [16]	74.7	88.8	79.4	72.9	89.0	78.2	75.2	89.9	80.1
	Ours	**74.9**	92.2	**80.6**	**73.8**	92.5	**83.0**	**76.3**	93.4	81.7

TABLE II. ABLATION OF OUR METHOD ON DBP15K DATASET

Models	$\mathrm{DBP_{ZH-EN}}$			$\mathrm{DBP_{JA-EN}}$			$\mathrm{DBP_{FR-EN}}$		
	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR
Baseline(AliNet)	53.9	82.6	62.8	54.9	83.1	64.5	55.2	85.2	65.7
Baseline(MRAEA)	65.7	89.5	74.4	72.7	92.3	79.8	73.9	93.8	81.0
W/O isomorphic weights	72.9	90.5	79.7	73.3	92.3	81.8	74.4	93.0	81.2
W/O rel Completion	73.8	91.5	83.0	73.1	92.1	79.9	74.5	93.1	81.3
Ours	74.9	92.2	80.6	73.8	92.5	83.0	76.3	93.4	81.7

D. *Ablation Studies*

Ablation is conducted in Table II. w/o rel completion means ignoring the cross-KG relation completion and finds new aligned pairs from the original KGs. And w/o isomorphic weights means ignoring isomorphic weights and learns the representation only with GAT.

The effectiveness of cross-KGs relation completion: Compared with MRAEA, w/o isomorphic weights improves H@1 by 2.7% average. Compared with w/o rel completion, our method also improves H@1. It reveals that cross-KG completion can improve the completeness and isomorphism of KGs.

The effectiveness of isomorphic weights: Compared with w/o isomorphic weights, our method improves H@1 by 3.56% average. This reveals the augmented representation for isomorphic neighborhoods can enhance KG representation and is helpful for non-isomorphic CLEA.

E. *Analysis*

1) ***Cross-KG relation completion can improve the completeness and isomorphism of KGs.***

Cross-KG relation completion results are shown in Table III.

a) After completion, the number of triples increases by 32455 and 6173 for KG_{ZH} and KG_{EN} respectively. It means 32455 and 6173 relations are completed and the completeness of KGs is improved. With the increase of the number of isomorphic edges, the isomorphism of the graph is enhanced, so that the representation of aligned entities is closer, and the training results of the graph neural network are more accurate.

b) After completion, OC values are improved by 8.8%, 9.3%, and 9.5%, which shows that isomorphism of KGs is improved.

TABLE III. THE RESULT OF CROSS-KG RELATION COMPLETION

Datasets	Indicators	Original	After Completed	Increase
DBP_{ZH-EN}	$Triples_{ZH}$	70,414	102,869	32,455
	$Triples_{EN}$	95,142	101,317	6,175
	OC	11.7%	20.5%	8.8%
	H@1	67.0%	74.9%	7.9%
DBP_{JA-EN}	$Triples_{JA}$	77,214	89,804	12,590
	$Triples_{EN}$	93,484	120,489	27,005
	OC	11.6%	20.9%	9.3%
	H@1	55.2%	73.8%	18.6%
DBP_{FR-EN}	$Triples_{FR}$	105,998	206,658	100,660
	$Triples_{EN}$	115,722	155,477	39,755
	OC	13.1%	22.6%	9.5%
	H@1	55.2%	76.3%	21.1%

c) With the improvement of completeness and isomorphism, H@1 is improved by 7.9%, 18.6%, and 21.1%. This shows the effectiveness of the cross-KG relation completion.

2) Our method can achieve an identical performance only covering the least neighboring scopes.

Some baseline methods, such as AliNet [17], expand the neighborhood scope to improve the isomorphism between two KGs. We compared the performance of AliNet and our method with varying ranges, as shown in Fig.Figure 3. .

a) When the neighborhood range changes from 1-hop to 2-hops, the performance of AliNet improves significantly, indicating that expanding the scope covers more homogeneous neighbors. However, when the scope changes to 3-hops and 4-hops, the performance of AliNet decreases sharply, demonstrating that a larger scope covers more heterogeneous and noisy neighbors, which hinders CLEA.

b) As the neighborhood scope increases, the performance of both our method and our w/o weight remains relatively stable. This is because our method changes the topology of all entities through cross-KG completion. After completion, the isomorphism between two KGs will not change when the neighborhood scope varies. Additionally, as the neighborhood scope expands, the isomorphic weights weaken distant and heterogeneous neighbors. This implies that our method does not require aggregating too many neighbors in the representation learning.

3) ***The robustness for non-isomorphism of KGs.***

Figure 4. shows the performance of our method with different OC. The larger the OC value is, the stronger isomorphism of knowledge graph is. We randomly drop out some homogeneous neighbors from the original KGs to get several datasets with different OC. We delete [5%-30%] isomorphic neighbors and OC value will decrease from 11.1% to 8.1% for ZH-EN, 11.0% to 8.1% for JA-EN and 12.4% to 9.2% for FR-EN.

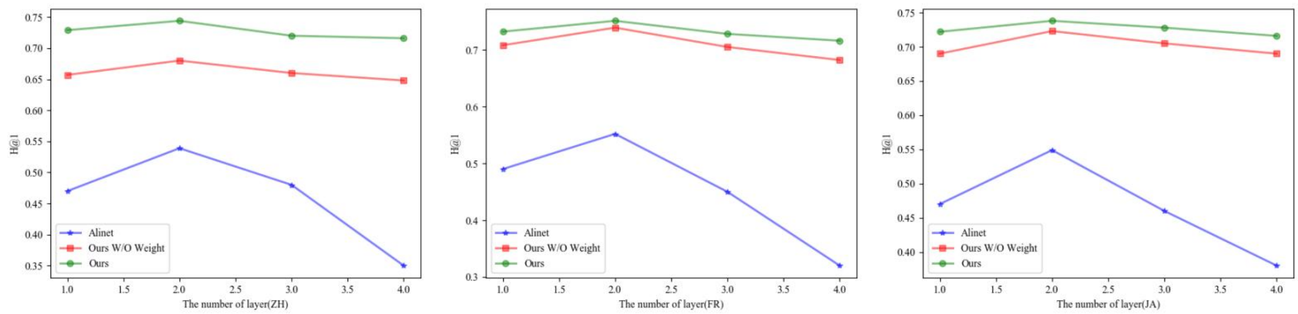

Figure 3. The influence for performances of different ranges of neighborhoods on ZH-EN, JA-EN, FR-EN.

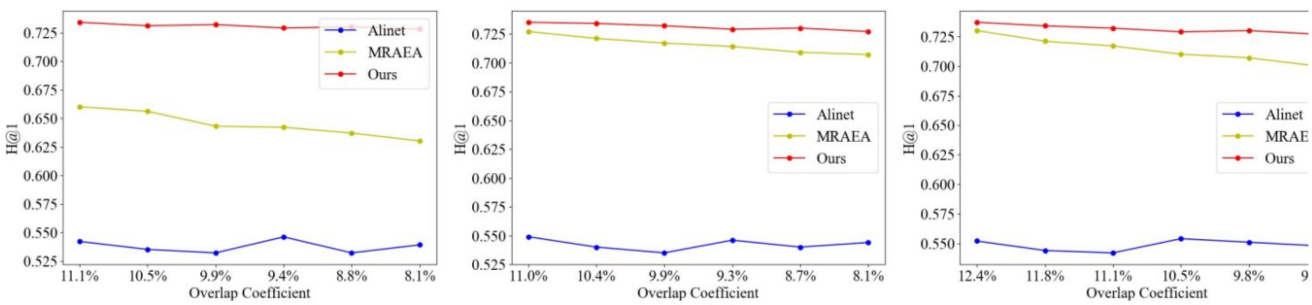

Figure 4. The performance varies with different OC values on three datasets.

a) With the decreasing of OC, the performances of our method and MRAEA [8] decrease obviously. It shows that non-isomorphism will hold back CLEA. While the performance of Alinet fluctuates over a range because it randomly selects distant neighbors and does not rely on direct neighbors.

b) Compared with MRAEA[8], our method degrades more slowly, and performs relatively higher and more stably. This reveals that our method is robust, especially for non-isomorphic CLEA.

4) ***The robustness for available alignment seeds.***

All baselines in this paper are supervised methods. The number of available seeds will influence the alignment performance. In our method, the available seeds will influence both the supervised learning and the cross-KG relation completion. The results of Hits@1 and the OC value varying with the size of aligned entity pairs are shown in Figure 5. and Figure 6.

In Figure 5, the alignment accuracy increases with the number of pre-aligned seed entities increasing. In Figure 6, OC value increases more obviously with the increasing of the seeds number. For Zh-EN, when there are only 1500 seeds available, the OC increases by 1.84%, while 4500 seeds are available, the OC increases by nearly 9%. It shows that more aligned seeds will be conducive to our entity alignment.

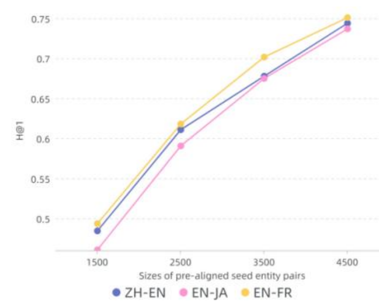

Figure 5. The performance changes with the size of available aligned entity pairs.

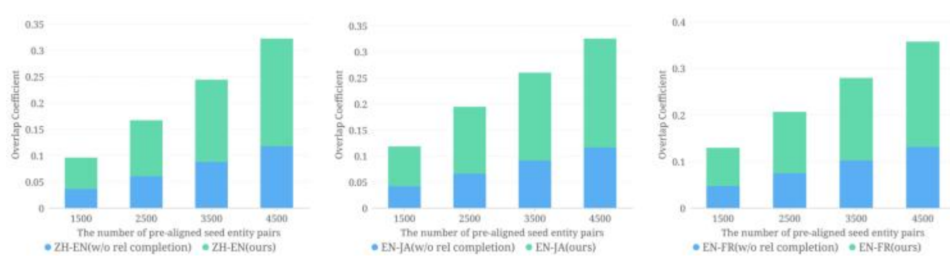

Figure 6. The OC changes with the size of available aligned entity pairs.

V. CONCLUSIONS

This paper focuses on non-isomorphic CLEA. To address the non-isomorphism, cross-KG relation completion is proposed to complete the missing relations and improve the completeness and isomorphism. And then, the isomorphic weights, not the importance of central entities in one KG, are designed to learn a representation more suitable for CLEA. In near future, we will explore more suitable method to mearue the isomorphism of two KGs.

ACKNOWLEDGMENTS

This work is supported by the National Key Research and Development Program of China (under grant 2020AAA0106100), the National Natural Science Foundation of China (under grant 61976077), the Natural Science Foundation of Anhui Province (under grant 2208085MF170) and the University Synergy Innovation Program of Anhui Province (GXXT-2022-040).

References

[1] W. Cai, Y. Wang, S. Mao, J. Zhan, and Y. Jiang, "Multi-heterogeneous neighborhood-aware for knowledge graphs alignment." Info. Proc. & Mana., vol. 59, pp. 529-551, 2022.

[2] Y. Cao, Z. Liu, C. Li, Z. Liu, J. Li, and T. Chua, "Multi-channel graph neural network for entity alignment." Proc. 57th Conf. Asso. Comp. Ling., Florence, Italy, vol. 1. Long Papers, pp. 1452–1461, 2019.

[3] M. Fey, J. E. Lenssen, C. Morris, J. Masci, and N. M. Kriege, "Deep graph matching consensus." 8th Inte. Conf. Lear. Repr. Ethiopia, April 26-30, 2020.

[4] H. Cui, T. Peng, F. Xiao, J. Han, R. Han,L. Liu, "Incorporating anticipation embedding into reinforcement learning framework for multi-hop knowledge graph question answering." Info. Scie., vol. 619, pp. 745-76, 2023.

[5] X. Kun, W. Liwei, Y. Mo, F. Yansong, S. Yan, W. Zhiguo, and Y. Dong, "Cross-lingual knowledge graph alignment via graph matching neural network." Proc. 57th Conf. Asso. Comp. Ling., Florence, Italy, vol. 1. Long Papers, pp. 3156–3161, July 28- August 2, 2019.

[6] X. Mao, W. Wang, Y. Wu, and M. Lan, "Are negative samples necessary in entity alignment?: An approach with high performance, scalability and robustness." 30th ACM Inte. Conf. Info. Know. Mana., Queensland, Australia, pp. 1263-1273, November 1 - 5, 2021,.

[7] X. Mao, W. Wang, Y. Wu, and M. Lan, "Boosting the speed of entity alignment 10 ×: Dual attention matching network with normalized hard sample mining." WWW' 21: Web Conf., Ljubljana, Slovenia, pp. 821-832, April 19-23, 2021.

[8] X. Mao, W. Wang, H. Xu, M. Lan, and Y. Wu, "Mraea: an efficient and robust entity alignment approach for crosslingual knowledge graph." Proc. 13th Inte. Conf. Web Sear. Data Mini., pp. 420-428, 2020.

[9] X. Mao, W. Wang, H. Xu, Y. Wu, and M. Lan, "Relational reflection entity alignment." CIKM' 20: 29th ACM Inte. Conf. on Info. Know. Mana., Ireland, pp. 1095-1104, October 19-23, 2020.

[10] C. Muhao, T. Yingtao, C. Kai-Wei, S. Steven, and Z. Carlo, "Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment." Proc. Twen. Inte. Join. Conf. Arti. Inte., Stockholm, Sweden, pp. 3998-4004, July 13-19, 2018.

[11] C. Muhao, T. Yingtao, Y. Mohan, and Z. Carlo, "Multilingual knowledge graph embeddings for cross-lingual knowledge alignment." Proc. 26th Inte. Join. Conf. Arti. Inte., Melbourne, Australia, pp. 1511-1517, August 19-25, 2017.

[12] A. Rossi, D. Barbosa, D. Firmani, A. Matinata, and P. Merialdo, "Knowledge graph embedding for link prediction: A comparative analysis." ACM Tran. Know. Disc. Data, vol. 15, pp. 1-49, 2021.

[13] M. Schlichtkrull, T. N Kipf, P. Bloem, R. V. D. Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks." Euro. Sema. Web Conf., pp. 593-607, 2018.

[14] C. Shang, Y. Tang, J. Huang, J. Bi, X. He, and B. Zhou. "End-to-end structure-aware convolutional networks for knowledge base completion." Proc. AAAI Conf. Arti. Inte., pp. 3060-3067, 2019.

[15] J. Sun, Y. Zhou, and C. Zong, "Dual attention network for cross-lingual entity alignment." Proc. 28th Inte. Conf. Comp. Ling., pp. 3190-3201, 2020.

[16] H. Huang, C. Li, X. Peng, L. He, S. Guo, H. Peng, et al, "Cross-knowledge-graph entity alignment via relation prediction." Know. Syst. vol. 240, pp. 107813, 2022.

[17] Z. Sun, C. Wang, W. Hu, M. Chen, J. Dai, W. Zhang, and Y. Qu, "Knowledge graph alignment network with gated multi-hop neighbor-hood aggregation." Proc. AAAI Conf. Arti. Inte., pp. 222-229, 2020.

[18] B. D. Trisedya, J. Qi, and R. Zhang, "Entity alignment between knowledge graphs using attribute embeddings." Proc. AAAI Conf. Arti. Inte., pp. 297-304, 2019.

[19] S. Vashishth, S. Sanyal, V. Nitin, and P. Talukdar, "Composition-based multi-relational graph convolutional networks." 8th Inte. Conf. Lear. Repr., Addis Ababa, Ethiopia, April 26-30, 2020.

[20] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Li`o, and Y. Bengio, "Graph attention networks." 6th Inte. Conf. Lear. Repr., Vancouver, BC, Canada, April 30 - May 3, 2018.

[21] C. Wang, B. Han, S. Pan, J. Jiang, G. Niu, and G. Long, "Cross-graph: Robust and unsupervised embedding for attributed graphs with corrupted structure." 2020 IEEE Inte. Conf. Data Mini., pp. 571-580, 2020.

[22] Z. Wang, Q. Lv, X. Lan, and Y. Zhang, "Crosslingual knowledge graph alignment via graph convolutional networks." Proc. 2018 Conf. Empi. Meth. Natu. Lang. Proc., pp. 349-357, 2018.

[23] D. Yu, Y. Yang, R. Zhang, and Y. Wu, "Knowledge embedding based graph convolutional network." Proc. Web Conf., pp. 1619-1628, 2021.

[24] Z. Zhang, F. Zhuang, H. Zhu, Z. Shi, H. Xiong, and Q. He, "Relational graph neural network with hierarchical attention for knowledge graph completion." Proc. AAAI Conf. Arti. Inte., pp. 9612-9619, 2020.

[25] Y. Zheng, "Methodologies for cross-domain data fusion: An overview." IEEE Tran. Big Data, 1(1):16-34, 2015.

[26] H. Zhu, R. Xie, Z. Liu, and M. Sun, "Iterative entity alignment via knowledge embeddings." Proc. Inte. Join. Conf. Arti. Inte., pp. 4258-4264, 2017.

[27] Q. Zhu, X. Zhou, J. Wu, J. Tan, and L. Guo, "Neighborhood-aware attentional representation for multilingual knowledge graphs." Proc. 28th Inte. Join. Conf. Arti. Inte., Macao, China, pp. 1943-1949, August 10-16, 2019.

[28] R. Zhu, M. Ma, and P. Wang, "RAGA: relation-aware graph attention networks for global entity alignment." Adva. Know. Disc. Data Mini. 25th Paci. Conf., pp. 501-513, May 11-14, 2021