MRAEA: An Efficient and Robust Entity Alignment Approach for Cross-lingual Knowledge Graph

Xin Mao[1,2], Wenting Wang[2], Huimin Xu[1,2], Man Lan[1*], Yuanbin Wu[1*]
{xmao,hmxu}@stu.ecnu.edu.cn,{sunian.mx,xinjiu.xhm}@alibaba-inc.com
wenting.wang@lazada.com,{mlan,ybwu}@cs.ecnu.edu.cn
[1]East China Normal University, [2]Alibaba Group

ABSTRACT

Entity alignment to find equivalent entities in cross-lingual Knowledge Graphs (KGs) plays a vital role in automatically integrating multiple KGs. Existing translation-based entity alignment methods jointly model the cross-lingual knowledge and monolingual knowledge into one unified optimization problem. On the other hand, the Graph Neural Network (GNN) based methods either ignore the node differentiations, or represent relation through entity or triple instances. They all fail to model the meta semantics embedded in relation nor complex relations such as n-to-n and multi-graphs. To tackle these challenges, we propose a novel Meta Relation Aware Entity Alignment (MRAEA) to directly model cross-lingual entity embeddings by attending over the node's incoming and outgoing neighbors and its connected relations' meta semantics. In addition, we also propose a simple and effective bi-directional iterative strategy to add new aligned seeds during training. Our experiments on all three benchmark entity alignment datasets show that our approach consistently outperforms the state-of-the-art methods, exceeding by 15%-58% on *Hit@1*. Through an extensive ablation study, we validate that the proposed meta relation aware representations, relation aware self-attention and bi-directional iterative strategy of new seed selection all make contributions to significant performance improvement. The code is available at https://github.com/MaoXinn/MRAEA.

CCS CONCEPTS

• **Computing methodologies** → **Knowledge representation and reasoning**.

KEYWORDS

Knowledge Graph, Entity Alignment, Graph Neural Network, Cross-lingual

WSDM '20, February 3–7, 2020, Houston, TX, USA

ACM ISBN 978-1-4503-6822-3/20/02...$15.00
https://doi.org/10.1145/3336191.3371804

ACM Reference Format:
Xin Mao[1,2], Wenting Wang[2], Huimin Xu[1,2], Man Lan[1*], Yuanbin Wu[1*]. 2020. MRAEA: An Efficient and Robust Entity Alignment Approach for Cross-lingual Knowledge Graph. In *The Thirteenth ACM International Conference on Web Search and Data Mining (WSDM '20), February 3–7, 2020, Houston, TX, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3336191.3371804

1 INTRODUCTION

Recently, Knowledge Graphs (KGs) have been widely applied in areas of artificial intelligence and natural language processing. Besides open source multilingual KGs such as DBpedia [1], YAGO [16] and ConceptNet [15], many companies including Google and Amazon have constructed their own KGs to improve the results in search and solve many business problems. With more multi-lingual KGs emerging, it becomes necessary and beneficial to integrate cross-lingual KGs. Two main motivations are: (1) Different KGs are constructed independently from different data sources, so they contain complementary facts. (2) Equivalent entities in multi-lingual KGs play an important role to bridge the language gap and support downstream multi-lingual applications. Therefore, the task of aligning entities in cross-lingual KGs attracts increasing attentions. The existing methods can be divided into two main groups:

Translation-based approach. These approaches assume that two KGs with respective language have a similar structure, so the embeddings of aligned entity pairs should have relative similar positions in the vector space. Thus, these methods use TransE [2] or its extensions to obtain the representations of entities and relations in each KG, and then map these representations into a unified vector space through joint learning [5, 6, 13, 17, 18, 25]. Such extensions of TransE have been successfully applied in cross-lingual entity alignment. However, they still suffer from two fundamental issues: (1) TransE is originally designed to learn structure embeddings in a single KG. When moving to cross KG alignment, these variants attempt to learn one representation jointly for both cross-lingual and monolingual knowledge. Therefore, during unified optimization, the losses have to be balanced very carefully between entity alignment and link prediction. In addition, these methods usually have many hyper-parameters to tune with, e.g., five in the loss function of state-of-the-art system NAEA [25], which may weaken the robustness of the system across different datasets. (2) TransE assumes $\boldsymbol{h}+\boldsymbol{r}\approx\boldsymbol{t}$ holds for each relational triple (h,r,t). This enables the effective modeling of 1-to-1 relations, but fails to

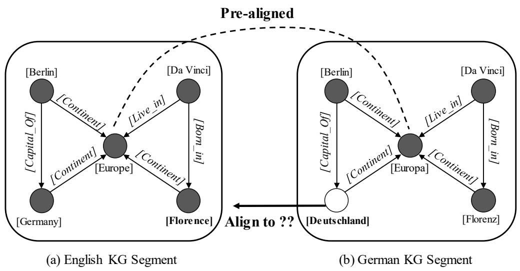

Figure 1: Examples of graph structure segments.

represent more complicated relations such as 1-to-n, n-to-1 relations (e.g., both *Batman* and *Superman* are *Members* of *JusticeLeague*), and multiple relations among one entity pair in multi-graphs (e.g., *Marie wasBornIn* and *DiedIn Madrid*). **GNN-based approach**. Different from the above translation-based methods, Graph Neural Network(GNN) generates node-level embeddings by aggregating information from the neighboring nodes. Thus GNN is able to learn aligment oriented embeddings directly and model more complex relations beyond 1-to-1. However, existing GNN-based methods still fail to effectively model the relational information hidden in the triple. For example, some research studies [3, 20, 23] ignore the type of relations and treat all relations as the same edge with no context. Others [22, 25] apply self-attention to include relation type information into entity feature, but only leverage the transformed triples to represent the relation. Intuitively, we believe both relation type context and its pointing direction matter a lot in entity alignment. For example, as shown in Figure 1, in order to find the right alignment of [*Deutschland*] in (a), both relation type and direction information are mandatory to eliminate the three false candidates [*Berlin*], [*Florence*] and [*Da Vinci*]. Without relation type, [*Florence*] and [*Da Vinci*] could not be removed from the candidate set; while without relation direction, [*Berlin*] could not be removed.

Inspired by the above observations, in this paper, we propose a new entity alignment approach called *Meta Relation Aware Entity Alignment* (MRAEA). MRAEA is a novel alignment approach that operates on cross-lingual KGs, leveraging meta relation-aware embedding and relation-aware self-attention to address the shortcomings of prior methods. Specifically, our approach generates representations of each entity by attending over not only its neighbors but also the meta relation semantics, then utilizes this neighborhood-and-relation-aware embeddings to directly align cross-lingual KGs by a direct alignment loss function. We further adopt an iterative training process, by selecting predicted alignments and adding them back to the training set for next iteration. Our proposed bi-directional iterative selecting strategy based on the asymmetric nature of cross-KG alignment is simple but proved to be effective. The experimental results show that our model outperforms previous works by a statistically significant and large margin on $Hit@1$ across all datasets. We summarize the main contributions of this paper as follows:

- **Model.** We propose a new entity alignment method, which effectively models the structure and meta relation information in KGs. With this simpler but effective architecture, we have addressed several key issues from prior methods.
- **Learning.** Our bi-directional iterative strategy improve alignment results while lessening the gap between two alignment directions. In addition, we can greatly reduce the number of hyper-parameters which needs manual adjustment in training, and thus foster the generalization and robustness across different datasets.
- **Experiments.** We conduct extensive experiments on three most common public cross-lingual datasets to demonstrate the efficacy of our model. Our proposed iterative model significantly outperforms all the state-of-the-art approaches with a 15%-58% increase on $Hits@1$. Moreover, we carry out a thorough empirical analysis, including the impact of each new component, hyper-parameter tuning, size of pre-aligned pairs and the space-time cost. All experimental results validate that our model is ranked consistently as the best across datasets, and thus is very robust. In addition, by further incorporating machine translation of entity names to build pre-aligned seeds, we propose an unsupervised approach based on our MRAEA model. The experiment results show that our unsupervised approach even beats previous supervised systems by at least 4.72% on $Hits@1$.

2 RELATED WORK

2.1 KG Embeddings

In recent years, there has been an increasing interest in learning embeddings of KG. TransE [2], a translation-based method, interprets a relation as the translation from the head entity to its tail entity. It learns embeddings by making $\boldsymbol{h} + \boldsymbol{r} \approx \boldsymbol{t}$ when a triple (h, r, t) is hold, a concise and reasonable hypothesis, which proved to be effective in subsequent studies. Later, several translation-based models have been proposed to solve the problem of modeling 1-to-n, 1-to-1, n-to-n relations, such as TransH, TransR and TransD. TransH [21] maps entities to different hyperplanes for different relations. TransR [10] maps entities and relations into two embedding spaces respectively and sets up a transformation matrix for each relation. TransD [7] uses dynamic matrix to transfer entities and relations rather than a fixed one.

2.2 Graph Neural Networks

Benefitting from the ability to model non-Euclidean space, GNN has become increasingly popular in many areas, including social networks, KGs, recommendation systems and even life sciences. Graph Convolutional Network (GCN) [9] is an extension of GNN, which generates node-level embeddings by aggregating information from the nodes' neighborhoods. Veličković et al. [19] introduces a new graph embedding model Graph Attention Network (GAT) which calculates the hidden representations of each entity by attending over its neighbors, via a self-attention strategy. Recently, many studies focus on extending GNN to modeling KGs. Relational GCN (RGCN) [14] further incorporates relation type information by setting a transformation matrix for each relation. Nathani et al. [12] and Zhu et al. [25] modify the attention mechanism of GAT, assign different weights to each neighboring node according to the relation between them.

2.3 KGs Alignment

Existing studies on KGs alignment are generally based on embeddings learned by TransE and its variant algorithms. MTransE [5] employs TransE to embed different KGs into independent embedding spaces and learns the transitions between KGs via five different alignment methods. JAPE [17] further enhances the performance by leveraging attributes and attribute correlations to refine entity embedding. BootEA [18] has made improvements based on JAPE by a new sampling method, a new loss function and a bootstrapping strategy. Guo et al. [6] proposes RSNs which integrates recurrent neural networks (RNNs) with residual learning to efficiently capture the long-term relational dependencies within and between KGs. The above approaches are translation-based KGs embedding model, and they all suffer from the two problems as listed in the Section 1.

In order to solve these problems, Wang et al. [20] innovatively uses GCN to model entity information for KGs alignment and achieves performance comparable to JAPE. MuGNN [3] learns alignment-oriented KG embeddings by robustly encoding two KGs via multiple channels. Wu et al. [22] proposes a novel Relation-aware Dual-Graph Convolutional Network (RDGCN) to incorporate relation information via attentive interactions between the KG and its dual relation counterpart.

3 PROBLEM FORMULATION

In KG, facts are stated in relational triples, in the form of $\langle enetity_1, relation, entity_2\rangle$, to describe the relations between two entities. Formally, we define a KG as $G = (E, R, T)$, where E and R represent the sets of entities and relations respectively, $T \subset E \times R \times E$ is the set of relational triples.

Considering two KGs in different languages, there are some entity pairs refering to the same real world object, which are called aligned entities. Formally, G_1 and G_2 are two KGs in different languages, $P = \{(e_{i_1}, e_{i_2}) | e_{i_1} \in E_1, e_{i_2} \in E_2\}_{i=1}^{p}$ is the set of pre-aligned entity pairs between G_1 and G_2. The task of cross-lingual entity alignment aims to find new aligned entity pairs based on the pre-aligned seeds P.

4 THE PROPOSED APPROACH

Figure 2 depicts the overall framework of our proposed approach. The main idea is to build an entity alignment model which leverages meta relation knowledge to embed entities from two different languages into one unified vector space, where aligned entities are expected to be close to each other. In addition, training is conducted iteratively. That is, in each iteration, newly found entity pairs which are mutually nearest aligned are added into the training set of next iteration. Once no such new pairs emerge, the training ends. Finally, the alignment for each source entity can be found by searching the nearest cross-lingual neighbors in the learned unified vector space.

4.1 Meta Relation Aware Representation

As shown by the examples in Figure 1, relation level information is vital to align entities in KGs. Previous studies attempt to represent a relation by either replacing it with transformed triples [25] or treat it as a transformation matrix attached to the entity [14]. However, in either way, relation is binded too closely to entity instances in KGs.

Intuitively, relation expresses three categories of crucial meta semantics:

(1) **Relation Type:** Independent of real triple instances, relation type already conveys rich information by itself. For example, $[Capital\_Of]$ derives the same semantics no matter what triple and entity instances are currently presented in the KGs.
(2) **Relation Direction:** Relation is usually not reflective, thus its pointing directions actually imposes extra but delicate constraints on the head and tail entity individually. For example, given relation $[Capital\_Of]$, its head entity should be a city while its tail entity should be a country.
(3) **Inverse Relation:** Integration of relation direction raises an interesting issue which, however, has been neglected by all existing graph attention mechanisms. The directed edges force the neighboring information to be accumulated only through the flow-in direction. For instance, in triple ($Berlin$, $Capital\_Of$, $Germany$), only node $[Germany]$ is attended to neighbor $[Berlin]$ through inward $[Capital\_Of]$ edge, but not vice versa. We believe it would be beneficial if flow-out information is also accumulated. Furthermore, directed and typed edges are extremely sparse in real world KGs. Thus, explicitly allowing information to flow in and out would facilitate more information to be bridged and propagated in such sparse graph.

Motivated by the above characteristics of meta semantics in relation, we define a *Meta Relation Aware Representation* to fuse the meta semantics of relations with the structure information of entities in KG by the following steps.

First, we create an ***inverse*** for each relation, i.e., extending the set of relations R to have its inverse version.

$$R = \{r_1, r_2, \cdots, r_{|R|}\} \rightarrow \{r_1^+, r_1^-, \cdots, r_{|R|}^+, r_{|R|}^-\} \quad (1)$$

Here r^+ and r^- represent the outward and inward relation respectively. For example, $\langle e_x, r_y, e_z\rangle$ will be extended to $\langle e_x, r_{y_+}, e_z\rangle$ and $\langle e_z, r_{y_-}, e_x\rangle$. Correspondingly, the sizes of relations and triples, i.e., $|R|$ and $|T|$, will be doubled.

Next, in order to force the input features to also reflect hidden semantics of relations (i.e., ***relation type*** and ***direction***), we average the embeddings of its neighboring entities and relations respectively, then concatenate them together

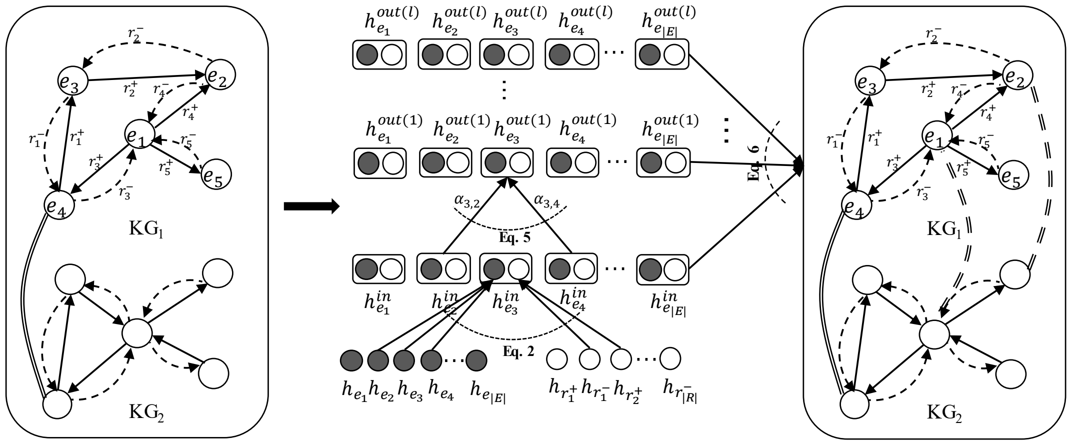

Figure 2: The framework of MRAEA. The solid arrow between entities represents the existent relations in KGs, while the dashed arrows represent the inverse ones we create. Solid double lines represent the pre-aligned pairs and dashed double lines represent the predicted new aligned pairs.

as the initial features:

$$h_{e_i}^{in} = \left[\frac{1}{|\mathcal{N}_i^e|+1} \sum_{e_j \in \mathcal{N}_i^e \cup \{e_i\}} h_{e_j} \,\Big|\Big|\, \frac{1}{|\mathcal{N}_i^r|} \sum_{r_k \in \mathcal{N}_i^r} h_{r_k} \right] \quad (2)$$

We randomly initialize the embedding of each entity and relation, represented by $\{h_{e_1}, ..., h_{e_{|E|}}\}$ and $\{h_{r_1^+}, h_{r_1^-}, \cdots, h_{r_{|R|}^+}, h_{r_{|R|^-}}\}$. $|\mathcal{N}_i^e|$ and $|\mathcal{N}_i^r|$ represent the sets of entities and relations which are outward from e_i, respectively.

4.2 Relation Aware Self-Attention

Unlike [19] whose approach generates representations for each entity by only attending over its neighbors, we believe that relations, including their types and directions, also play an important role to model the structures. In order to achieve this, for each entity e_i, we augment the original self-attention mechanism to include relation features as follows:

$$s_{i,j} = v^T \left[h_{e_i}^{in} \Big|\Big| h_{e_j}^{in} \Big|\Big| \frac{1}{|M_{i,j}|} \sum_{r_k \in M_{i,j}} h_{r_k} \right] \quad (3)$$

$s_{i,j}$ indicates the weighted importance of neighboring entity e_j and connected relations r_k to entity e_i. Theoretically, M is a tensor of order 3 to model the connections between e_i, e_j and r_k. But to simplify the implementation and reduce tensor computation complexity, we define the entry $M_{i,j} = \{r_k | (e_i, r_k, e_j) \in T\}$ to be a sparse vector representing the set of linked relation in the direction from entity e_i to entity e_j. $||$ represents concatenate operation, while v is a shared attention weight vector.

It is worth noting that here is no weight matrix W for either h_{e_i} or h_{e_j}. This is comply with the implementation in Wang et al. [20]. Some other recent works [11] also adopt this implementation.

By applying non-linear activation function and normalization, the coefficients $\alpha_{i,j}$ computed by the attention mechanism is as below:

$$\alpha_{ij} = \frac{exp(LeakyReLU(s_{i,j}))}{\sum_{t \in \mathcal{N}_i^e} exp(LeakyReLU(s_{i,t}))} \quad (4)$$

The output features $h_{e_i}^{out}$ for entity e_i are obtained by applying non-linear activation function to the linear combination of attended neighbors' features.

$$h_{e_i}^{out} = ReLU\left(\frac{1}{Z} \sum_{z=1}^{Z} \left[\sum_{e_j \in \mathcal{N}_i^e} \alpha_{i,j}^z h_{e_j}^{in} \right] \right) \quad (5)$$

Here we employ multi-head attention in the same way as GAT to stabilize learning process. Z is the number of heads, $\alpha_{i,j}^z$ represents attention coefficient in the z-th head.

In previous research [19, 24], GNN is able to expand to multi-hop neighboring level information by the addition of more layers, thus creates a more global-aware representation of the graph. Let $h_{e_i}^{out(0)}, ..., h_{e_i}^{out(l)}$ be the output features of e_i from 0-th(input features) to l-th layer. We concatenate them together to obtain the final output features $\widehat{h}_{e_i}^{out}$ of entity e_i:

$$\widehat{h}_{e_i}^{out} = [\ h_{e_i}^{out(0)}\ ||\ ...\ ||\ h_{e_i}^{out(l)}\ ] \quad (6)$$

4.3 Entity Alignment Model

Entity Alignment Prediction. Entity alignment can be performed by simply measuring the distance between two entities. Following Wang et al. [20], we also use Manhattan Distance to be the distance metric. For $e_i \in G_1$ and $e_j \in G_2$, the distance between entity pair (e_i, e_j) is calculated by the formula:

$$dis\,(e_i, e_j) = |\widehat{h}_{e_i}^{out} - \widehat{h}_{e_j}^{out}| \quad (7)$$

Thus, in order to find G_1's entity e_i's alignment, we calculate its distance to all entities in G_2 and choose the nearest one to

Datasets	DBP_{ZH}	DBP_{EN}	DBP_{JA}	DBP_{EN}	DBP_{FR}	DBP_{EN}	SRP_{EN}	SRP_{FR}	SRP_{EN}	SRP_{DE}	WK_{EN}	WK_{FR}	WK_{EN}	WK_{DE}
Relation	2,830	2,317	2,043	2,096	1,379	2,209	221	177	222	120	458	277	458	172
Entity	19,388	19,572	19,814	19,780	19,661	19,993	15,000	15,000	15,000	15,000	64,539	45,255	64,539	43,503
Triple	70,414	95,142	77,214	93,484	105,998	115,722	36,508	33,532	38,363	37,377	569,393	258,337	569,393	244,647

Table 1: Statistics of the Datasets. The statistics of DBP are different from that of previous papers. The reason is that all previous papers actually use a subset of DBP rather than what they reported, this Statistics are consistent with the real situation.

be e_i's counterpart entity in G_2. Here the order of e_i and e_j in the distance function matters because in most cases, the nearest relation is asymmetric. That is to say, although e_j in G_2 is the most similar entity to e_i, we may find another entity e_t in G_1 is closer to e_j. We will report alignment performance of our model in two alignment directions in Section 5.4.

Model Training. In order to make the equivalent cross-lingual entities close to each other in the unified vector space, we define the following margin-based *loss* function:

$$L = \sum_{(e_i, e_j) \in P} ReLU\left(dis\left(e_i, e_j\right) - dis\left(e_i', e_j\right) + \lambda\right) + ReLU\left(dis\left(e_i, e_j\right) - dis\left(e_i, e_j'\right) + \lambda\right) \tag{8}$$

where $ReLU$ is linear rectifier function to ensure non-negative output while λ is a hyper-paremeter of margin. e_i' and e_j' represent negative entities of e_i and e_j respectively.

In this work, the entities in G_1 and G_2 are randomly selected as negative entities by the uniform probability. Adam [8] is adopted to minimize the loss function. Our design of loss function fosters the robustness of our model, since there is only 1 hyper-parameter that needs to be tuned, instead of tuning 5 hyper-parameter in recent studies [25]. In Section 5.5 , we further empirically prove that our model performs stably with different values of λ.

4.4 Bi-directional Iterative Strategy For Newly-Aligned Seeds Selection

In practice, the aligned seeds are often inadequate due to the high cost of manual annotations and the huge size of KG. To expand training data, KDCoE [4] and BootEA [18] adopt iterative co-training and bootstrapping respectively and achieve great performance improvement. However, their strategies are either based on a heuristic threshold or complicated selection criteria, which inevitably introduces a set of hyper-parameters and often lead to unstability.

Considering entities in KG are corresponding one by one, we propose a bi-directional iterative strategy based on the asymmetric nature of alignment directions. In current iteration, if and only if the entities e_i and e_j are mutually nearest neighbors of each other, then the pair (e_i, e_j) is considered as newly aligned entities and will be added into the training set of next iteration. The detailed iterative procedure is given in Algorithm 1.

Algorithm 1 Bi-directional Iterative strategy

Input: Graphs G_1, G_2, pre-aligned seed entity pairs P. $E_1' \subseteq E_1, E_2' \subseteq E_2$ represent entity sets that does not exist in P respectively.
Output: parameters θ for model

```
repeat
    Reinitialize model;
    Train model on G1, G2, P until the loss of development set does not decrease;
    for e ∈ E1' do
        e' ← NN(e, E2')  /*NN(e, E) represents the entity nearest to e in E.*/
        if NN(e', E1') = e then
            P ← {(e, e')} ∪ P
            E1' ← E1' − e
            E2' ← E2' − e'
        end if
    end for
until no more aligned entities are added to P
```

5 EXPERIMENT

5.1 Datasets

In order to verify the effectiveness, robustness, scalability and performance stability of our model, we experiment on three most commonly used public datasets for entity alignment:

- **DBP-15K**: This [17] includes three cross-lingual subsets built from DBpedia [1]: DBP_{ZH-EN}[1], DBP_{JA-EN} and DBP_{FR-EN}.
- **WK31-60K** [4] consists of two subsets built from Wikidata with two different language pairs ($WK31_{FR-EN}$ and $WK31_{DE-EN}$) for size $60K$.
- $\mathbf{SRP_{Normal}}$[2] [6] contains two cross-lingual datasets built from DBpedia and Wikidata: SRP_{FR-EN} and SRP_{DE-EN} with normal distribution.

Table 1 lists the statistics of all datasets. Each of the datasets has its own special characteristics and thus demonstrates different challenges in entity alignment: (1) DBP-15K is the most widely used dataset in entity alignment task and contains the richest relations. (2) WK31-60K, as the largest dataset of the three, brings challenges on space and time complexity. Moreover, the scale of WK31-60K varies greatly among different languages, e.g., the numbers of triples and relations in French are just half of those in English, which reflects a real world phenomenon facing by low-resource languages. (3) SRP_{Normal} has much fewer relations and more sparse triples,

[1]ZH, EN, JA, FR and DE represent Chinese, English, Japanese, French and German, respectively.
[2]We define this name "SRP" according to the entity sampling method: segment-based random PageRank sampling method.

Models		$\mathrm{DBP_{ZH-EN}}$			$\mathrm{DBP_{JA-EN}}$			$\mathrm{DBP_{FR-EN}}$		
		Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
NonIter	MTransE	30.83	61.41	0.364	27.86	57.45	0.349	24.41	55.55	0.335
	JAPE	41.18	74.46	0.490	36.25	68.50	0.476	32.39	66.68	0.430
	GCN	41.25	74.38	0.549	39.91	74.46	0.546	37.29	74.49	0.532
	MuGNN*	49.40	84.40	0.611	50.10	85.70	0.621	49.50	87.00	0.621
	$\mathrm{MRAEA_{EN\to\mathbb{L}}}$	63.49	88.21	0.729	63.61	88.73	0.731	**66.64**	**91.18**	0.764
	$\mathrm{MRAEA_{\mathbb{L}\to EN}}$	**63.76**	**88.64**	**0.736**	**64.56**	**89.14**	**0.735**	66.39	91.11	**0.765**
	Improv. %	**29.07**	**5.23**	**20.46**	**28.86**	**4.01**	**18.36**	**34.63**	**4.80**	**23.19**
Iter	BootEA	62.94	84.75	0.703	62.23	85.39	0.701	65.30	87.44	0.731
	NAEA*	65.01	86.73	0.720	64.14	87.27	0.718	67.32	89.43	0.752
	$\mathrm{MRAEA_{EN\to\mathbb{L}}}$	75.28	92.31	0.824	75.34	93.30	0.825	**78.09**	94.70	0.843
	$\mathrm{MRAEA_{\mathbb{L}\to EN}}$	**75.70**	**92.98**	**0.827**	**75.78**	**93.38**	**0.826**	78.04	**94.81**	**0.849**
	Improv. %	**15.80**	**6.43**	**14.44**	**17.46**	**6.91**	**14.90**	**15.11**	**5.89**	**12.10**

Models		$\mathrm{WK_{FR-EN}}$			$\mathrm{WK_{DE-EN}}$			$\mathrm{SRP_{FR-EN}}$			$\mathrm{SRP_{DE-EN}}$		
		Hits@1	Hits@5	MRR	Hits@1	Hits@5	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
NonIter	MTransE	13.95	20.25	0.177	3.37	10.07	0.072	25.10	55.10	0.350	31.20	58.60	0.400
	JAPE	16.85	35.41	0.271	14.71	23.86	0.192	25.60	56.20	0.360	32.00	59.90	0.410
	GCN	21.47	37.81	0.293	13.80	24.55	0.190	15.50	34.50	0.220	25.30	46.40	0.330
	RSNs*	-	-	-	-	-	-	34.70	63.10	0.440	48.70	72.00	0.570
	$\mathrm{MRAEA_{EN\to\mathbb{L}}}$	42.79	66.54	0.544	**37.64**	57.34	**0.474**	**40.93**	**74.24**	**0.524**	54.75	**79.24**	0.634
	$\mathrm{MRAEA_{\mathbb{L}\to EN}}$	**45.14**	**69.24**	**0.561**	37.46	**58.55**	0.472	40.15	73.54	0.515	**54.88**	79.03	**0.636**
	Improv. %	**110.2**	**83.13**	**91.52**	**155.9**	**138.5**	**146.9**	**17.95**	**16.07**	**19.09**	**12.69**	**10.06**	**11.58**
Iter	BootEA	33.31	51.14	0.425	23.28	39.29	0.316	31.30	62.90	0.420	44.20	70.10	0.530
	OTEA*	36.07	54.08	0.447	26.97	43.97	0.352	-	-	-	-	-	-
	$\mathrm{MRAEA_{EN\to\mathbb{L}}}$	53.74	75.26	0.644	42.74	62.16	0.523	**46.02**	**76.80**	**0.559**	59.43	81.52	0.664
	$\mathrm{MRAEA_{\mathbb{L}\to EN}}$	**54.34**	**76.03**	**0.646**	**42.76**	**62.77**	**0.524**	46.01	76.67	0.558	**59.55**	**81.81**	**0.666**
	Improv. %	**50.65**	**40.58**	**43.17**	**58.54**	**42.75**	**47.72**	**32.59**	**21.71**	**27.04**	**22.20**	**13.63**	**15.79**

Table 2: Comparative results of our methods against strong baselines on selected datasets. The table is divided into two parts: iterative and non-iterative methods. We report its performance on two directions. $\mathbb{L}$ represents the non-English language. * represents the state-of-the-art system published.

which demands the model to have a stronger reasoning ability when aligning entities from limited information.

5.2 Baselines

To comprehensively evaluate the effectiveness of our proposed architecture, we compare to the following advanced systems:

- **Strong baselines** are MtransE [5], JAPE [17], GC-N [20], BootEA [18]. We show their performance on all datasets according to the results reported in their papers.
- **SOTA systems** are MuGNN [3], NAEA [25], RSNs [6], OTEA [13]. MuGNN and NAEA are the state-of-the art non-iterative and iterative system on DBP-15K, respectively. RSNs is the state-of-the-art non-iterative system on $\mathrm{SRP_{Normal}}$. OTEA is the state-of-the-art iterative system on WK31-60K.

5.3 Experiment Setting

In order to make a fair comparison with the previous works, we adopt the same data split method and evaluation metrics. Specifically, for DBP and WK, we random split 30% of the pre-aligned entity pairs as training and development data while 70% for testing; for SRP, we use the fixed set provided by the original author which is also 30% for training and 70% for testing. $Hits@k$ and *Mean Reciprocal Rank*(MRR) are used as the evaluation metrics. We report the averaged performance of 5 independent runs from our model as the final performance.

For all experiments if otherwise explicitly stated, the same default setting is adopted for our model: dimensions of entities and relations $d=100$, margin $\lambda=3$, attention head number $K=2$, GNN's depth $l=2$, the dropout rate is 0.3 and the learning rate of Adam is 0.005.

5.4 Main Results

Table 2 shows the performance of all compared approaches on the evaluation datasets. Across all datasets, our model is consistently ranked as the best over all competing approaches in both iterative and non-iterative groups respectively (i.e. relatively improvements of $12-155\%$ and $15-58\%$ on $Hist@1$ for iterative and non-iterative groups respectively).

In addition, our non-iterative version is already comparable to the state-of-the-art iterative methods. This suggests that our MRAEA model does capture rich and subtle subgraph information which is the core for entity alignment task. We have statistically significant improvement for on $Hits@1$ and

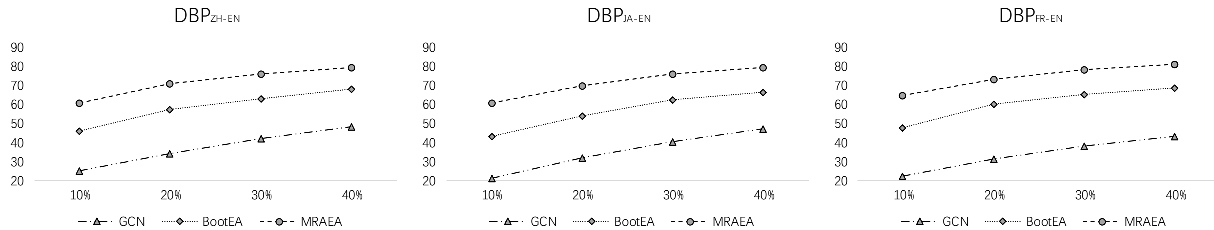

Figure 3: *Hits@1* on entity alignment w.r.t. different sizes of pre-aligned seed entity pairs on cross-lingual datasets.

Models	$\text{SRP}_{\text{FR-EN}}$			$\text{SRP}_{\text{DE-EN}}$		
	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
baseline (GCNs)	15.50	37.81	0.220	25.30	46.40	0.330
+RA-SA	25.44	57.55	0.363	42.72	67.56	0.512
+Inverse Rel.	30.98	62.15	0.417	46.51	70.29	0.553
+Rel. Aware Emb.	40.93	74.24	0.524	54.75	79.24	0.634

Table 3: Ablation experiment result of non-iterative model on SRP dataset.

	$\text{SRP}_{\text{FR-EN}}$			$\text{SRP}_{\text{DE-EN}}$		
	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
$\lambda=1$	44.53	75.63	0.542	58.21	80.71	0.645
$\lambda=2$	45.46	76.47	0.556	59.01	81.53	0.659
$\underline{\lambda=3}$	46.02	76.80	0.559	59.55	81.81	0.666
$\lambda=4$	46.15	76.92	0.561	59.41	81.72	0.664
range	**1.63**	**1.29**	**0.019**	**1.34**	**1.10**	**0.021**
$l=1$	45.80	75.91	0.554	59.37	81.59	0.663
$\underline{l=2}$	46.02	76.80	0.559	59.55	81.81	0.666
$l=3$	46.57	77.03	0.562	59.58	81.87	0.667
$l=4$	45.65	76.17	0.553	58.78	80.92	0.658
range	**0.92**	**1.12**	**0.009**	**0.80**	**0.95**	**0.009**
$K=1$	45.70	76.27	0.556	59.21	81.54	0.662
$\underline{K=2}$	46.02	76.80	0.559	59.55	81.81	0.666
$K=3$	46.01	76.74	0.560	59.13	81.37	0.659
$K=4$	46.08	76.83	0.561	59.21	81.46	0.661
range	**0.38**	**0.56**	**0.005**	**0.42**	**0.44**	**0.007**

Table 4: Hyper-parameter control experiment results on SRP dataset. Underline represents the default setting.

MRR for all datasets. Specifically, when testing in the large-scale dataset (i.e., $\text{WK}_{\text{FR-EN}}$ and $\text{WK}_{\text{DE-EN}}$), our model improves all metrics by 40% to 58% relatively over the state-of-the-art system.

Therefore, we are confident to conclude that our proposed model pushes the state-of-the-art performance of cross-lingual entity alignment to an obvious new level on all datasets.

5.5 Ablation Studies

Effect of New Designs in Our Non-Iterative Model. In our non-iterative model, we introduce three innovative designs to incorporate structural and relational information for entity alignment. In order to study the effectiveness of each design, we start from a basic GCN model [20] and gradually add each new component in the following order:

- **Replace GCN layer with *Relation Aware Self-Attention***: To attend the entity to not only its neighboring entities but also the relations it connects to (i.e., following Eq. 3).
- **Create an *inverse* for each relation**: Differentiate relation's representation between inward and outward directions (i.e., following Eq. 1).
- **Use *Meta Relation Aware Representation* to capture relation type and direction**: To fuse entity and relation level features as the entity embedding (i.e., following Eq. 2).

The results of each gradually improved model on $\text{SRP}_{\text{Normal}}$ datasets are shown in Table 3. We observe that adding every component brings to a significant performance improvement (i.e., *Hits*@1 increases by +9.94%, +5.54%, +9.95% gradually). This validates that each of the new designs is able to model some unique information, thus all contribute significantly to the performance improvement. Combining all the components together enables our non-iterative model to outperform the state-of-the-art (RSNs) by 12%. This also proves that our *Meta Relation Aware Representation* is more effective for KGs.

Effect of Our Iterative Strategy. In order to shed light on the impact of iterative strategy, we report our models' alignment performances in both alignment directions as shown in Table 2. Noticeably, our proposed bi-directional iterative strategy significantly improves the performance against non-iterative version, with at least an absolute increase of 9% on *Hits*@1. In addition, when testing on different alignment directions, the performance gap between alignment directions becomes neglectable when we add iterative strategy to our model, i.e. dropping from ±2.35% to ±0.60% on $\text{WK}_{\text{FR-EN}}$. We believe the main reason is our design of the bi-directional iterative strategy fully considers the asymetric nature of cross-lingual alignment direction, thus reduces the error propagation problem brought by adding falsely aligned pairs into the training set of next epoch.

5.6 Robustness Analysis

Sensitivity to Size of Pre-Aligned Entity Pairs. It is costly to annotate pre-aligned entity pairs, especially for the larger scale KGs. To investigate the robustness of our iterative model to pre-aligned seed size, following [18], we

Datasets	$\|E\|+\|T\|$	Mem(MB)	MB/ent.	Time(S/epoch)	S/ent.
SRP_{FR-EN}	100,438	1,561	1.55×10^{-3}	0.149	1.49×10^{-6}
SRP_{DE-EN}	106,082	1,561	1.47×10^{-3}	0.156	1.48×10^{-6}
DBP_{ZH-EN}	209,663	2,405	1.14×10^{-3}	0.241	1.15×10^{-6}
DBP_{JA-EN}	214,431	2,549	1.19×10^{-3}	0.251	1.17×10^{-6}
DBP_{FR-EN}	264,962	2,879	1.09×10^{-3}	0.298	1.12×10^{-6}
WK_{DE-EN}	773,696	7,495	0.96×10^{-3}	0.892	1.15×10^{-6}
WK_{FR-EN}	803,246	8,265	1.02×10^{-3}	0.847	1.06×10^{-6}

Table 5: Time and space costs on all datasets.

	Models	DBP_{ZH-EN}		DBP_{JA-EN}		DBP_{FR-EN}	
		Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10
sup.	JAPE	73.09	90.43	82.84	94.65	-	-
	GMNN	67.93	78.48	73.97	87.15	89.38	95.24
	RDGCN	70.75	84.55	76.74	89.54	88.64	95.72
unsup.	MT	54.79	70.48	71.02	85.12	76.91	89.64
	MRAEA (s)	74.54	**93.47**	84.03	**96.85**	86.88	**98.43**
	MRAEA (s+text)	**77.81**	83.19	**88.89**	92.71	**95.04**	97.04

Table 6: Unsupervised experiment result on SRP. '*s*' represents structure embedding and '*text*' represents using translated entity names.

test GCN, BootEA and our model on DBP15K dataset with pre-aligned proportion from 10% to 40% with step size of 10%. Figure 3 depicts the changes of $Hits@1$ with respect to different proportions. The proposed model is not only significantly superior to other methods in all proportions, but also has a more gradual slope curve. This demonstrates that our model is less dependent on additional training data and it is promising to have good capability of generalization.

Sensitivity to Hyper-parameter. For neural network model, tuning hyper-parameters is usually tedious but essential. Thus the fewer number of hyper-parameters, the more robust of the model. In order to study the tuning impact of the newly introduced and most important hyper-parameters in our model, we design a quick control experiment on SRP dataset. Starting with the default setting listed in Section 5.3, then each time we only change the value of one hyper-parameter in the set of margin λ, network depth l and attention head number K. Table 4 shows the results for each test setting. We observe that our proposed approach constantly maintain a stable performance even the hyper-parameters change values. This experiment proves that our method is robust to hyper-parameter tuning. This is also the reason why our model outperforms state-of-the-art systems on all datasets with the same setting.

Time and Space Complexity. To examine whether our model increases any complexity compared with GCN, we examine the memory and time cost of our model on different dataset with default setting. Detailed results are listed in the Table 5. From the results, we observe that the costs of time and space are proportional to the sum of entities and relations, which are exactly the same as GCN. In average, our system needs four or five iterations to achieve optimum performance, where each iteration takes about $1,000$ to $2,000$ epoches. This means that even on the largest data set like WK_{FR-EN}, with only one GPU, our model is still able to output results in two hours. Thus, this proves that our model outperforms previous studies without increasing time and space complexity.

5.7 Unsupervised Entity Alignment Based on Machine Translation

In the above sections, we discuss entity alignment models built purely based on structural and relational information. But several recent studies [17, 22, 23] show great success by just adding string feature machine translation. For datasets where the aligned entities are related mostly by string names, these approaches are effective although too dependent on machine translation quality. Inspired by their studies, we further propose an unsupervised entity alignment method by extending our MRAEA models with machine translated seeds, then compare it with recent studies which also utilize machine translation.

Unsupervised Pre-aligned Seeds. Sun et al. [17] designs a machine translation based approach that (1) employs Google Translate to translate the entity names in one KG; (2) computes Levenshtein distance between the translations and the entity names in the other KG. We augment the above approach to obtain the pre-aligned seeds and ensure only mutually nearest neighbors are included, i.e., both e_i's translation is the closest to e_j and e_j's translation is the closest to e_i. By testing on DBP-15K dataset, we obtain $7,623$, $9,648$ and $10,843$ seeds on DBP_{ZH-EN}, DBP_{JA-EN} and DBP_{FR-EN} respectively, with the accuracy of 95.97%, 98.58% and 99.30% respectively. Therefore, in this way, we can obtain very high quality pre-aligned seeds automatically.

Unsupervised Experimental Results. After getting unsupervised pre-aligned seeds, we leverage our iterative MRAEA model to align entities and obtain a rank based on structure similarity among entities. Then to combine structure and text information, for each candidate aligned entity pair, we take the worse rank between the structure rank and the Levenshtein rank as their final rank. Then we conduct experiments on DBP-15k datasets to compare against recent studies [17, 22, 23]. For a fair comparison, we use the same translations provided by Xu et al. [23]. The experimental results in Table 6 show that our unsupervised method outperforms all previous studies by a large margin even with no supervised data. We also observe an interesting fact of our model that adding text information improves $Hits@1$ performance, but $Hits@10$ performance is degraded. The reason is that when machine translation errors occur, the Levenshtein distance between candidate pair may have larger gap. However, the sole structure-based method is not affected by translation errors and demonstrates the stable performance across datasets.

6 CONCLUSION AND FUTURE WORK

In this paper, we propose a novel *Meta Relation Aware Entity Alignment* (MRAEA) model for cross-lingual entity alignment, and also present a simple but effective iterative training strategy based on the asymmetric nature of alignments. Our approach outperforms the state-of-the-art models by a large margin across all datasets. This validates that our MRAEA model is able to effectively capture the hidden semantics in relations and complex graphs. Furthermore, we build an unsupervised entity alignment framework on top of MRAEA model by introducing translated entity names as seeds, which also achieves the best performance compare to previous machine translation based methods. In the future, we intend to extend our method to other KG applications, such as link prediction, relation extraction, entity classification and so on, to verify the generalization of our method. We will also continue to investigate the inherent reason for performance degradation when introducing weight W in self attention.

ACKNOWLEDGEMENTS

This work was supported by Alibaba Group through Alibaba Innovative Research (AIR) Program and East China Normal University fund for international conferences.

REFERENCES

[1] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 2009. DBpedia - A crystallization point for the Web of Data. *J. Web Semant.* 7, 3 (2009), 154–165. https://doi.org/10.1016/j.websem.2009.07.002

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*. 2787–2795.

[3] Yixin Cao, Zhiyuan Liu, Chengjiang Li, Zhiyuan Liu, Juanzi Li, and Tat-Seng Chua. 2019. Multi-Channel Graph Neural Network for Entity Alignment. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*. 1452–1461. https://www.aclweb.org/anthology/P19-1140/

[4] Muhao Chen, Yingtao Tian, Kai-Wei Chang, Steven Skiena, and Carlo Zaniolo. 2018. Co-training Embeddings of Knowledge Graphs and Entity Descriptions for Cross-lingual Entity Alignment. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.* 3998–4004. https://doi.org/10.24963/ijcai.2018/556

[5] Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. 2017. Multilingual Knowledge Graph Embeddings for Cross-lingual Knowledge Alignment. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*. 1511–1517. https://doi.org/10.24963/ijcai.2017/209

[6] Lingbing Guo, Zequn Sun, and Wei Hu. [n.d.]. Learning to Exploit Long-term Relational Dependencies in Knowledge Graphs. ([n. d.]).

[7] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 687–696.

[8] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. http://arxiv.org/abs/1412.6980

[9] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. https://openreview.net/forum?id=SJU4ayYgl

[10] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Twenty-ninth AAAI conference on artificial intelligence*.

[11] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*. 2181–2187. http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9571

[12] Deepak Nathani, Jatin Chauhan, Charu Sharma, and Manohar Kaul. [n.d.]. Learning Attention-based Embeddings for Relation Prediction in Knowledge Graphs. ([n. d.]).

[13] Shichao Pei, Lu Yu, and Xiangliang Zhang. 2019. Improving Cross-lingual Entity Alignment via Optimal Transport. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 3231–3237. https://doi.org/10.24963/ijcai.2019/448

[14] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*. Springer, 593–607.

[15] Robert Speer, Joshua Chin, and Catherine Havasi. 2017. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*. 4444–4451. http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14972

[16] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2008. YAGO: A Large Ontology from Wikipedia and WordNet. *J. Web Semant.* 6, 3 (2008), 203–217. https://doi.org/10.1016/j.websem.2008.06.001

[17] Zequn Sun, Wei Hu, and Chengkai Li. 2017. Cross-Lingual Entity Alignment via Joint Attribute-Preserving Embedding. In *The Semantic Web - ISWC 2017 - 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part I*. 628–644. https://doi.org/10.1007/978-3-319-68288-4_37

[18] Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. 2018. Bootstrapping Entity Alignment with Knowledge Graph Embedding.. In *IJCAI*. 4396–4402.

[19] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[20] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. 2018. Cross-lingual Knowledge Graph Alignment via Graph Convolutional Networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*. 349–357. https://aclanthology.info/papers/D18-1032/d18-1032

[21] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Twenty-Eighth AAAI conference on artificial intelligence*.

[22] Yuting Wu, Xiao Liu, Yansong Feng, Zheng Wang, Rui Yan, and Dongyan Zhao. 2019. Relation-aware entity alignment for heterogeneous knowledge graphs. *arXiv preprint arXiv:1908.08210* (2019).

[23] Kun Xu, Liwei Wang, Mo Yu, Yansong Feng, Yan Song, Zhiguo Wang, and Dong Yu. 2019. Cross-lingual Knowledge Graph Alignment via Graph Matching Neural Network. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*. 3156–3161. https://www.aclweb.org/anthology/P19-1304/

[24] Kun Xu, Liwei Wang, Mo Yu, Yansong Feng, Yan Song, Zhiguo Wang, and Dong Yu. 2019. Cross-lingual Knowledge Graph Alignment via Graph Matching Neural Network. *arXiv preprint arXiv:1905.11605* (2019).

[25] Qiannan Zhu, Xiaofei Zhou, Jia Wu, Jianlong Tan, and Li Guo. 2019. Neighborhood-aware attentional representation for multilingual knowledge graphs. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, 1943–1949.