Semi-Supervised Entity Alignment via Knowledge Graph Embedding with Awareness of Degree Difference

Shichao Pei, Lu Yu, Robert Hoehndorf, Xiangliang Zhang
King Abdullah University of Science and Technology
Thuwal, 23955, SA
{shichao.pei,lu.yu,robert.hoehndorf,xiangliang.zhang}@kaust.edu.sa

ABSTRACT

Entity alignment associates entities in different knowledge graphs if they are semantically same, and has been successfully used in the knowledge graph construction and connection. Most of the recent solutions for entity alignment are based on knowledge graph embedding, which maps knowledge entities in a low-dimension space where entities are connected with the guidance of prior aligned entity pairs. The study in this paper focuses on two important issues that limit the accuracy of current entity alignment solutions: 1) labeled data of priorly aligned entity pairs are difficult and expensive to acquire, whereas abundant of unlabeled data are not used; and 2) knowledge graph embedding is affected by entity's degree difference, which brings challenges to align high frequent and low frequent entities. We propose a **semi-supervised entity alignment** method (SEA) to leverage both labeled entities and the abundant unlabeled entity information for the alignment. Furthermore, we improve the **knowledge graph embedding with awareness of the degree difference** by performing the adversarial training. To evaluate our proposed model, we conduct extensive experiments on real-world datasets. The experimental results show that our model consistently outperforms the state-of-the-art methods with significant improvement on alignment accuracy.

KEYWORDS

Knowledge Graph; Entity Alignment; Semi-supervised Learning

ACM Reference Format:

Shichao Pei, Lu Yu, Robert Hoehndorf, Xiangliang Zhang. 2019. Semi-Supervised Entity Alignment via Knowledge Graph Embedding with Awareness of Degree Difference. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3308558.3313646

1 INTRODUCTION

Knowledge graphs have been constructed and widely applied to organize and represent the knowledge of different domains, including the most popular ones such as Freebase [1], YAGO [27] and DBpedia [17]. Even in the same domain, knowledge graphs (KG) are generated by different methods in different languages. To comprehensively represent the knowledge in one domain, it is thus essential to connect multiple knowledge graphs in same domain. More specifically, entities in different knowledge graphs are aligned if they are semantically same (as known as entity alignment problem).

Given two knowledge graphs with labeled entities that are known to be same in semantics, existing entity alignment approaches need human-involved feature design [20] or extra resources [16, 22, 27], such as information of the entity and the relation, as the training set, to supervise the learning process of entity alignment. Most of these methods require a sufficient number of labeled entities to generalize well in downstream applications. However, getting labeled entities is difficult and expensive. The number of accessible prior entity is usually a small proportion of a whole knowledge graph [5, 29]. ITransE [41], IPTransE [41], BootEA [29] and KDCoE [4] tried to propose new aligned pairs of entities as training data iteratively during training process. However, the focus is on how to acquire more aligned entities, leaving the abundant unaligned entities without consideration. Therefore, this paper targets on designing **semi-supervised entity alignment** model, which learns from both labeled and unlabeled entities. The underlying distribution of abundant unaligned entities will help on mitigating the risk of generalization errors caused by limited amount of aligned entities [42].

Recently, knowledge graph embedding methods, e.g., TransE [3] and PTransE [18] show significant improvement on entity alignment. For example, MtransE [5] encodes entities and relations of each language in a separated embedding space, and learns transitions to map each embedding vector to its cross-lingual counterparts in other spaces. ITransE [41] and IPTransE [41] were proposed to encode both entities and relations of different KGs into a unified low-dimensional space jointly and iteratively. JAPE [28], KDCoE [4], and Graph Convolutional Network-based approach [34] all jointly model structure and attribute information of knowledge graphs. Our work also takes advantage of embedding methods for building a semi-supervised entity alignment model. However, rather than directly using existing embedding methods, we address an important issue in the embedding process, which is caused by the **degree difference of entities** in different knowledge graphs. This phenomenon is demonstrated in Figure 1.

Entities in KGs have different degrees, i.e., popular entities are more connected with other concepts than rare entities. As pointed in recent study of natural language processing [7, 21] and machine translation [24, 25], word embedding methods encode more frequency information than semantic information of words in the resulted low-dimension space. We also investigate and find the same problem in knowledge graph embedding. Our findings verify

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313646

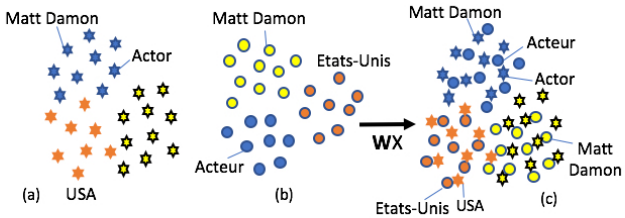


Figure 1: Illustration of the impact of entity's degree difference on embedding and alignment results. (a): embedding of (only entities in) a KG in English, where "Matt Damon" and "Actor" are both popular (with high degree) and thus mapped close in blue. (b): embedding of a KG in French, where "Acteur" is more popular (in blue) than "Matt Damon" (in yellow). (c): alignment of English KG and French KG ("Actor" with "Acteur", and "USA" with "Etats-Unis"), which shows that "Matt Damon" in French is far from the "Matt Damon" in English, even they are actually the same entity.

that entities with similar degree tend to be aggregated into a same region in the embedding space.

This deficiency of embedding methods brings challenges to entity alignment problem. Figure 1 shows a toy example of the problem, when aligning "Matt Damon" in English and French KGs. "Matt Damon" is a very popular name in English KG, while "Matt Damon" in French has fewer records. Embedding results of English KG show "Matt Damon" is close to other popular entities like "Actor" (in color blue in Figure 1 (a)), while the results of French KG (in Figure 1 (b)) have "Matt Damon" in yellow region that is distant to popular entity "Acteur" in blue region. By aligning "Actor" in English and "Acteur" in French, "USA" in English and " Etats-Unis" in French, the embedding space of French KG can be transformed to the embedding space of English KG (e.g., with linear transformation $\mathbf{W}$). The alignment is illustrated in Figure 1 (c)), where we can see that "Matt Damon" in French is far from the "Matt Damon" in English, even they are actually the same entity.

To tackle the above challenges, we propose a **semi-supervised entity alignment** framework called SEA, which takes advantages of both aligned with unaligned entities, which are represented in low-dimension space by knowledge graph embedding with **awareness of the degree difference**. In particular, the impact of degree difference is mitigated by an adversarial training, which prevents entities with similar degree from being aggregated into the same region in the embedding space during training. We also design a cycle-consistency based translation loss using the unaligned entities to reduce the search space when learning the mappings. Thus, the learned mappings can better align the entity in one knowledge graph to the corresponding entity in another knowledge graph.

Our main contributions in this work are summarized as follows:

- We propose to solve entity alignment in a semi-supervised way, not only using the given aligned entity, but also incorporating the unaligned entity to enhance the performance.
- We indicate the impact of entity's degree difference on embedding of knowledge graph, and address the problem under the adversarial training framework.
- We conduct extensive experiments on four real-world datasets to evaluate the proposed SEA model for the task of entity alignment. The results demonstrate its advantages over the state-of-the-art methods, with significant improvement on several datasets.

The rest of the paper is organized as follows. We discuss the related works in Section 2. The proposed method is described in Section 3 and followed by the experimental results in Section 4. Finally Section 5 concludes the whole paper.

2 RELATED WORK

2.1 KG Embedding

Knowledge graph (KG) embedding has became an important tool for knowledge graph analysis and semantic information modeling tasks with the fast growth of large-scale knowledge graphs. The KG embedding approaches can be roughly categorized into two groups: translational distance models and semantic matching models [33].

TransE [3] is the most representative translational distance model. It considers a relation as the translation from its head entity to its tail entity and represents both entities and relations as vectors into a same low dimensional vector space. TransE [3] characterizes a triple (h, r, t) following a common assumption $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$, where $\mathbf{h}$ and $\mathbf{t}$ are the representations of h and t, $\mathbf{r}$ is the representation of r. TransE has shown its feasibility for KG modeling, and it has been improved by many following studies, such as TransH [35], TransR [19], TransD [12], TranSparse [13], and PTransE [18].

Semantic matching models use the similarity-based scoring functions instead of the distance-based scoring functions. RESCAL [23] and Bilinear [11] model each relation as a matrix and associate each entity with a vector. DistMult [36] restricts the relation matrices to diagonal matrices. And HolE combines RESCAL [23] and DistMult [36] to get expressive and efficiency model. ComplEx [31] introduces complex-valued embeddings for better asymmetric relations modeling. Several works in [2, 26] conduct semantic matching using neural network.

The effect of word's frequency difference on embedding results has been recognized in [7, 21] and [24, 25], but the effect of entity's degree difference on embedding results has not been explored. In this paper, we address the effect of entity's degree on the embedding and propose to improve the most widely-used TransE method by mitigating the effect of entity's degree difference. The obtained KG embedding are used for improving entity alignment, especially for the rare and popular entities.

2.2 Entity Alignment

Some forerunners proposed to address the entity alignment problem using crowdsourcing [17, 32] and well-designed hand-crafted features [20]. However, the approaches suffer from the requirement of heavy human efforts, and is thus costly and labor-expensive. Many works leverage the extra resources, such as OWL properties [9], entity descriptions [37], information of entities and relations [16, 22, 27]. Such methods are complex and usually limited by the availability of the extra information about a knowledge graph.

In recent study, KG embedding-based approaches become the most popular solution for entity alignment. MTransE [5] is the representative work which encodes entities and relations of each language in a separated embedding space, and learns transitions to map each embedding vector to its cross-lingual counterparts in the other space. Then ITransE [41] and IPTransE [41] are proposed to encode both entities and relations of different KGs into a unified low-dimensional space jointly by an iterative and parameter sharing method. BootEA [29] addresses the lack of labeled data by bootstrapping strategy and tries to reduce error accumulation during iterations by employing an alignment editing method. It also proposes an improved KG embedding approach with the limit-based loss and truncated uniform negative sampling. BootEA achieves a significant performance improvement. Further, several works, such as JAPE [28] and KDCoE [4], consider to jointly model the structure and attributes information of KGs. Recently, a Graph Convolutional Network-based entity alignment approach [34] uses GCNs to embed entities of each language into a unified vector space combining structural and attributes information.

All the methods above focus on the utilization of aligned entities, including BootEA [29], which tried iteratively enlarge the labeled entity pairs based on the bootstrapping strategy. The abundant unaligned entities that have rich content information are not used in the alignment process. Therefore, we propose a semi-supervised framework to align entities in different KGs based on both labeled and unlabeled data.

3 METHODOLOGY

In this section, we firstly introduce the notation and problem definition, then describe the proposed method in detail.

3.1 Notation and Problem Definition

A knowledge graph can be noted as $G = (E, R, T)$, where E is the set of entities, R is the set of relations, and T is the set of triples, each of which is a triple (h, r, t), including the head entity h, the relation r and the tail entity t. By using KG embedding, each triple can be presented as $(\mathbf{h}, \mathbf{r}, \mathbf{t})$, in which boldfaced $\mathbf{h}$, $\mathbf{r}$, and $\mathbf{t}$ represent the embedding vectors of head h, relation r, and tail t, respectively.

Let $G_1 = (E_1, R_1, T_1)$ and $G_2 = (E_2, R_2, T_2)$ be two KGs in different languages. $AS^L = \{(e_{i_1}, e_{i_2}) | e_{i_1} \in E_1^L, e_{i_2} \in E_2^L\}$ is a set of labeled entity pairs that are same in semantics, e.g., e_{i_1} in G_1 shares same meaning with its counterpart e_{i_2} in G_2. Entity alignment is a task to find and align the remaining semantically same entities $\{e_{i_1} \in E_1^U\}$ and $\{e_{i_2} \in E_2^U\}$ where $E_1^U = E_1 \setminus E_1^L$ and $E_2^U = E_2 \setminus E_2^L$. Unlike the previous study which builds alignment model based on AS^L only, our approach SEA builds the semi-supervised alignment model based on both AS^L and E_1^U and E_2^U.

The framework of our proposed method SEA is shown in Figure 2. SEA has two modules, knowledge graph embedding with awareness of degree difference of entities (called degree-aware KGE for short), and semi-supervised entity alignment. We introduce them in details next.

3.2 Degree-Aware Knowledge Graph Embedding

Like the previous works in [4, 5, 41], we build our degree-aware KG embedding model by following TransE [3], which is the most representative translational distance model. It is worth mentioning that our ideas of mitigating the effect of entity's degree difference on embedding can also be applied to other KG embedding methods, which is not the focus of this work.

When applying TransE on both knowledge graphs G_1 and G_2, entities and relations are projected into the same low-dimensional vector space by encoding the triples (h, r, t), and making $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ when (h, r, t) holds. Specifically, the embeddings of relations can translate the embeddings of head entities to tail entities. The margin-based ranking object function minimized by TransE over a knowledge graph G_i is defined as:

$$L_{G_i}(G_i; \varphi^i, \theta_e^i) = \sum_{(h,r,t) \in T_i} L_t(h, r, t) \tag{1}$$

where φ^i refers to the model parameters for G_i, θ_e^i presents the learned embedding from G_i, and $L_t(h, r, t)$ is the object function for a triple (h, r, t):

$$L_t(h, r, t) = \sum_{(h', r, t') \in T'_{(h,r,t)}} \left[\gamma + E(h, r, t) - E(h', r, t')\right]_+ \tag{2}$$

where $[x]_+ = \max\{0, x\}$ denotes the positive part of x, γ is a margin hyper-parameter which is greater than 0, and $E(h, r, t)$ indicates the energy function:

$$E(h, r, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2 \tag{3}$$

and T' denotes the negative sample set for the triple (h, r, t):

$$T'_{(h,r,t)} = \{(h', r, t) | h' \in E\} \cup \{(h, r, t') | t' \in E\} \tag{4}$$

where (h', r, t) and (h, r, t') are the Bernoulli negative-sampled triples by replacing h or t in (h, r, t). Finally, we can get the object function for both knowledge graphs G_1 and G_2:

$$L_G = L_{G_1} + L_{G_2} \tag{5}$$

Next, we present the degree-aware KGE method to address the impact of entity's degree difference on the knowledge graph embedding. As discussed before, the entities with similar degree values tend to be aggregated into the same regions in the embedding space. However, for entity alignment, entities with similar semantic information are expected to be closer without the impact of entities' degree. Thus, we design the degree-aware KGE model by training the knowledge graph embeddings in an adversarial framework, inspired by Generative Adversarial Network (GAN) [8]. Given a graph G_i, we design two discriminators to classify the entities with different degrees in G_i, and the degree-aware KGE model can be regarded as a generator which produces "Good" embeddings to fool the discriminators. We catalog the degree of entities into three levels, *high* degree, *normal* degree, and *low* degree. One discriminator D_1 categorizes the entities with *high* degree and *normal* degree, while the other discriminator D_2 is in charge of classifying the entities with *low* degree and *normal* degree. The expectation is: the learned knowledge graph embeddings not only minimize the margin-based ranking loss function defined above, but also can fool the two discriminators. The impact of degree is thus removed from the learned embeddings when the two discriminators cannot distinguish entities according to the information of degree.

Let D_1 be the first discriminator with parameters ϕ_1^i, and D_2 be the other discriminator with parameter ϕ_2^i. The inputs of D_1 are the entities from E_{hd}^i and E_{nd}^i, which are entities with *high* degree and *normal* degree in graph G_i. The inputs of D_2 are the entities

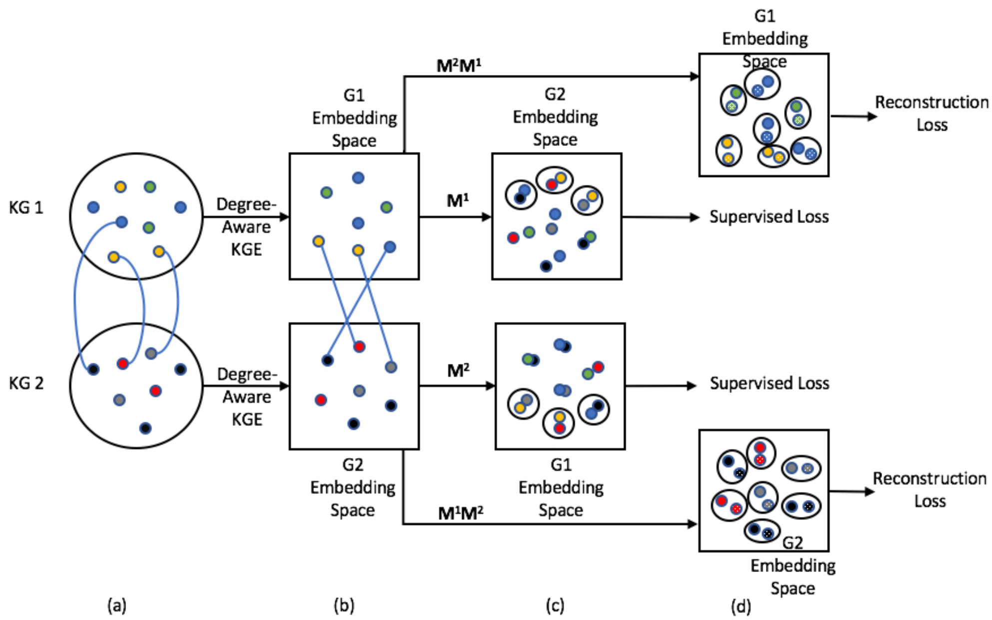


Figure 2: Framework of SEA for semi-supervised entity alignment. (a) shows two knowledge graphs G_1 and G_2 with three pairs of aligned entities. (b) presents the learned embeddings via the degree-aware KG embedding approach. (c) demonstrates the embedding spaces after transferring G_1 to G_2 and G_2 to G_1 using M^1 and M^2 learned by the pairs of aligned entities. The entities within small circles are close due to the supervised alignment guidance. (d) shows the cycle consistency of entities after being transferred back the original embedding spaces. The transferred embedding of each entity should be close to the original embedding of this entity in the original embeddings spaces (as indicated in the small circles).

from E^i_{ld} and E^i_{nd}, which are entities with *low* degree and *normal* degree in graph G_i. Both discriminators are learned to minimize loss functions:

$$L_{D_1} = \frac{1}{\left|E^i_{hd}\right|} \sum_{e \in E^i_{hd}} \log D_1(\theta^i_e, \phi^i_1) + \frac{1}{\left|E^i_{nd}\right|} \sum_{e \in E^i_{nd}} \log(1 - D_1(\theta^i_e, \phi^i_1)) \tag{6}$$

$$L_{D_2} = \frac{1}{\left|E^i_{ld}\right|} \sum_{e \in E^i_{ld}} \log D_2(\theta^i_e, \phi^i_2) + \frac{1}{\left|E^i_{nd}\right|} \sum_{e \in E^i_{nd}} \log(1 - D_2(\theta^i_e, \phi^i_2)) \tag{7}$$

Taking also the objective function Eq. (1), the over all minimax object function for learning embeddings of G_i will be:

$$\begin{aligned} \min_{\varphi^i, \theta^i_e} \max_{\phi^i_1, \phi^i_2} & L_{G_i}(G_i; \varphi^i, \theta^i_e) - \\ & \alpha L_{D_1}(E^i_{hd}, E^i_{nd}; \phi^i_1, \theta^i_e) - \alpha L_{D_2}(E^i_{ld}, E^i_{nd}; \phi^i_2, \theta^i_e) \end{aligned} \tag{8}$$

where α is a tradeoff parameter. In the adversarial training, L_{G_i} is minimized w.r.t. φ^i and θ^i_e to produce "Good" embeddings to fool D_1 and D_2. Meanwhile, D_1 and D_2 are trained w.r.t. ϕ^i_1 and ϕ^i_2 to distinguish entities with different degree levels. Following the iterative training process of GAN [8], we first train the KG embedding model with D_1 and D_2 fixed, and then train the discriminators with KG embedding model fixed. The same process will be applied to G_1 and G_2.

3.3 Semi-Supervised Entity Alignment

After obtaining entity embeddings of graph G_1 and G_2, labeled entities are aligned to minimize

$$\sum_{(e_i, e_j) \in AS^L} \left\| \mathbf{M}^1 \theta^1_{e_i} - \theta^2_{e_j} \right\|_2 + \left\| \mathbf{M}^2 \theta^2_{e_j} - \theta^1_{e_i} \right\|_2 \tag{9}$$

where $\mathbf{M}^1$ and $\mathbf{M}^2$ are the $d \times d$ translation matrices, d is the dimension of embeddings of entities. Note that our model jointly learns the translation of embeddings of entities in both directions for each knowledge graph. That is to say, $\mathbf{M}^1$ is learned to transfer the embeddings of G_1 into the embeddings space of G_2, and $\mathbf{M}^2$ is to transfer the embeddings of G_2 into the embeddings space of G_1. Due to the limited labeled set, the learned $\mathbf{M}^1$ and $\mathbf{M}^2$ cannot be generalized well to all entities in the two knowledge graphs.

To improve generalizability and incorporate unlabeled entities of two knowledge graphs in the alignment process, we define a cycled consistent loss, inspired by the work of CycleGAN [38] in computer vision field, where transitivity is used to regularize structured data [38] to do visual tracking [14, 30], 3D shape matching [10], co-segmentation [39, 40] and depth estimation [6]. Translation $\mathbf{M}^1$ and $\mathbf{M}^2$ should be able to

1) bring e_i in G_1 back after mapping cycle $e_i \rightarrow e_j \rightarrow \hat{e}_i$, i.e., the distance d_{g_1} between e_i and $\hat{e}_i$ should be small;
2) bring bring e_j back after mapping cycle $e_j \rightarrow e_i \rightarrow \hat{e}_j$, i.e., the distance d_{g_2} between e_j and $\hat{e}_j$ should be small.

Formally, the cycle process is:

$$\theta_{e_i}^1 \rightarrow \mathbf{M}^1\theta_{e_i}^1 \rightarrow \mathbf{M}^2\mathbf{M}^1\theta_{e_i}^1 \tag{10}$$

$$\theta_{e_j}^2 \rightarrow \mathbf{M}^2\theta_{e_j}^2 \rightarrow \mathbf{M}^1\mathbf{M}^2\theta_{e_j}^2 \tag{11}$$

Combing the alignment loss function in Eq. (9) and the cycle consistency restriction, we define the loss function of our semi-supervised entity alignment as:

$$\begin{aligned} L_{SEA}(\mathbf{M}^1, \mathbf{M}^2) = \alpha_1 \sum_{(e_i,e_j)\in AS^L} & \left\|\mathbf{M}^1\theta_{e_i}^1 - \theta_{e_j}^2\right\|_2 + \left\|\mathbf{M}^2\theta_{e_j}^2 - \theta_{e_i}^1\right\|_2 \\ & + \left\|\mathbf{M}^2\mathbf{M}^1\theta_{e_i}^1 - \theta_{e_i}^1\right\|_1 + \left\|\mathbf{M}^1\mathbf{M}^2\theta_{e_j}^2 - \theta_{e_j}^2\right\|_1 \\ & + \alpha_2 \sum_{e_i\in E_1^U} \left\|\mathbf{M}^2\mathbf{M}^1\theta_{e_i}^1 - \theta_{e_i}^1\right\|_1 + \alpha_2 \sum_{e_j\in E_2^U} \left\|\mathbf{M}^1\mathbf{M}^2\theta_{e_j}^2 - \theta_{e_j}^2\right\|_1 \end{aligned} \tag{12}$$

where α_1 and α_2 are the tradeoff parameters for balancing the loss between labeled and unlabeled data.

We initialize the embeddings of KGs by drawing from a Gaussian initialization, and initialize the matrices by using orthogonal initialization. We use SGD as our optimizer, and normalize all embeddings by $L2$ norm. The trade-off parameter α, α_1 and α_2 are set by grid search. Once $\mathbf{M}^1$, $\mathbf{M}^2$ are learned, an entity e in G_1 can be aligned by first transferring to G_2 as $\mathbf{M}^1\theta_e$ and then selecting the most similar entity in G_2. Similarly, an entity e in G_2 can be aligned by first transferring to G_1 as $\mathbf{M}^2\theta_e$ and then selecting the most similar entity in G_1.

4 EXPERIMENTS

In this section, we conduct experiments on several real-world datasets with different sizes, and evaluate our proposed method for entity alignment.

4.1 Datasets and Baselines

To comprehensively evaluate the effectiveness of our SEA method, we use two trilingual knowledge graph datasets from WK31 provided in [5]. WK31 datasets consist of English(En), French(Fr), and German(De) knowledge graphs which are extracted from Person domain of DBpedia's with known aligned entities as ground truth. WK31 includes two datasets with different sizes, which are WK31-15K and WK31-120K. The statistics of the datasets are given in Table 2 and Table 3. For datasets WK31-15K and WK31-120K, we extract the aligned entities from aligned triples.

To verify the effectiveness of our proposed method, we include the following methods for performance comparison, including: **MTransE** [5], **ITransE** [41], **JAPE** [28], **GCN-based method**[34], **BootEA**[29], **SEA w/o DA** (a variant of the proposed SEA, by removing the degree-aware KGE part), and **SEA**.

4.2 Evaluation Metrics and Parameter Settings

We adopt two popular metrics, *Hits@k* and *MRR* for evaluating entity alignment results. *Hits@k* measures the proportion of correctly aligned entities ranked in the top k proposed candidates. In our work, we report *Hits*@1, *Hits*@5 and *Hits*@10. Both metrics are preferred to be higher to present better performance.

Table 1: Statistics of the WK31 dataset

dataset	#Triple	#Entity	#Relation
WK31-15K En-Fr	En: 203,502 Fr: 170,605	En: 15,170 Fr: 15,393	En: 2,228 Fr: 2,422
WK31-15K En-De	En: 203,502 De: 145,616	En: 15,127 De: 14,603	En: 1,841 De: 596
WK31-120K En-Fr	En: 1,376,011 Fr: 767,750	En: 119,749 Fr: 118,591	En:3,109 Fr:2,336
WK31-120K En-De	En: 1,376,011 De: 391,108	En: 67,650 De: 61,942	En: 2,393 De: 861

Table 2: Number of aligned entity in different datasets.

Dataset	En-Fr	Fr-En	En-De	De-En
WK31-15K	10,108	10,164	11,594	11,445
WK31-120K	117,947	117,212	55,640	54,287

For all methods compared in 4.3, we set the dimension of knowledge graph embeddings $d = 100$ on all datasets. We find the optimal parameter settings for all baseline methods.

For our SEA method, we set that the *high* degree entities are those with top 20% degree values, while the *low* degree entities are those with bottom 20% degree values, and the rest are the *normal* degree entities. we search the margin γ among $\{0.5, 1, 1.5, 2\}$, and the tradeoff parameter α among $\{0.1, 0.3, 0.5, 0.7, 1.0\}$, α_1 and α_2 among $\{1, 2.5, 5, 7.5\}$ and $\{0.05, 0.15, 0.25, 0.35, 0.45\}$, respectively. The best configuration is $\gamma = 0.5$, $\alpha = 0.5$, $\alpha_1 = 2.5$ and $\alpha_2 = 0.25$. Discriminators are set as two-layers MLPs with 500 hidden units. We use Adam [15] to optimize the object function. Meanwhile, we use $L2$ norm to avoid potential over-fitting.

We randomly sample 30% of the aligned entities as the training set, and the rest aligned entities for testing. Each evaluation is repeated 5 times and we report the averaged *Hits@k* and *MRR*.

4.3 Performance Evaluation Results

The evaluation results are presented in Table 3-4. The best results are shown in bold among the group of methods, along with the percentage of improvement when comparing SEA without degree-aware (SEA w/o DA) and SEA with the best baseline methods. From these evaluation results, we have the following findings:

(1) **Our proposed SEA consistently outperforms the baseline methods on all datasets under different evaluation metrics**. This observation verifies that our proposed model effectively unifies the labeled and unlabeled data for improving entity alignment accuracy. In particular, our method achieves significant improvement when matching the top-1 ranked entity (*Hits*@1 is improved by more than 10%, or even 30-56%). Especially on the largest dataset WK31-120k, our SEA method has improvements on all metrics from 18% to 56%. BootEA is often the second best baseline due to the effective bootstrapping strategy for selecting labeled data. However, on the largest dataset WK31-120k, it sometimes performs worse than MTransE, mainly because it is difficult for bootstrapping to propose effective entities to label in a large dataset. In addition, bootEA need more time to calculate the similarity between each

Table 3: Entity alignment results of different methods on WK31-15K dataset. The best results are in bold, along with the percentage of improvement when comparing SEA w/o DA and SEA with the best baseline methods.

Language	En-Fr				Fr-En			
Metric	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10	MRR
MtransE	16.77	21.64	25.35	0.198	19.85	31.27	38.21	0.261
ITransE	18.21	24.34	27.41	0.214	18.61	33.64	36.28	0.248
JAPE	15.68	23.45	28.69	0.208	16.22	28.93	34.71	0.219
GCN	17.24	27.29	31.16	0.220	17.58	30.82	36.21	0.237
BootEA	29.72	52.92	61.19	0.395	30.77	55.44	63.67	0.428
SEA w/o DA	36.78	54.89	62.37	0.454	38.61	58.69	62.51	0.481
SEA	**37.28**	**55.91**	**63.56**	**0.468**	**39.76**	**59.32**	**66.31**	**0.489**
Improvement %	**23.75 / 25.47**	**3.72 / 5.65**	**1.93 / 3.87**	**14.93 / 18.48**	**25.48 / 29.22**	**5.86/ 7.00**	**-1.82 / 4.14**	**12.38 / 14.25**
Language	En-De				De-En			
Metric	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10	MRR
MtransE	6.170	8.48	10.39	0.078	4.69	6.61	7.74	0.059
ITransE	15.98	28.63	32.71	0.218	13.42	25.63	31.17	0.205
JAPE	16.85	27.32	34.74	0.226	13.92	22.15	29.68	0.189
GCN	18.25	31.30	37.26	0.248	15.70	27.53	33.31	0.217
BootEA	33.13	54.13	61.70	0.435	30.47	45.33	53.52	0.381
SEA w/o DA	37.74	54.81	**65.74**	0.462	**32.86**	46.21	53.53	0.393
SEA	**38.59**	**55.21**	64.06	**0.473**	32.11	**47.53**	**55.82**	**0.402**
Improvement %	**13.91 / 16.48**	**1.26 / 2.01**	**6.55 / 3.82**	**6.21 / 8.74**	**7.84 / 5.38**	**1.94/ 4.85**	**0.19 / 4.31**	**3.15 / 5.51**

Table 4: Entity alignment results of different methods on WK31-120K.

Language	En-Fr				Fr-En			
Metric	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10	MRR
MtransE	21.01	22.24	22.82	0.217	21.11	23.63	25.24	0.227
ITransE	11.54	20.41	23.92	0.176	13.35	21.20	24.18	0.197
JAPE	6.98	16.10	22.74	0.127	8.64	17.85	23.38	0.134
GCN	9.32	18.62	25.48	0.146	10.81	18.22	26.39	0.153
BootEA	17.56	27.41	31.85	0.235	18.46	28.65	31.85	0.241
SEA w/o DA	26.48	34.38	38.57	0.315	27.82	35.99	39.77	0.320
SEA	**28.02**	**35.82**	**39.70**	**0.321**	**28.72**	**36.74**	**41.37**	**0.331**
Improvement %	**26.04 / 33.36**	**25.42 / 30.68**	**21.10 / 24.65**	**34.04 / 36.59**	**31.78 / 36.05**	**25.62/ 28.23**	**24.87 / 29.89**	**32.78 / 37.34**
Language	En-De				De-En			
Metric	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10	MRR
MtransE	5.38	6.53	7.33	0.062	4.97	7.39	9.12	0.066
ITranE	7.62	15.54	19.41	0.112	6.41	12.82	15.27	0.085
JAPE	4.37	12.91	14.49	0.076	5.23	10.46	14.10	0.071
GCN	6.32	15.14	20.77	0.109	5.91	13.85	17.68	0.092
BootEA	11.57	22.08	27.75	0.179	10.32	22.11	26.36	0.169
SEA w/o DA	16.21	25.49	30.73	0.213	14.59	25.66	29.76	0.202
SEA	**17.23**	**27.48**	**32.83**	**0.227**	**16.11**	**27.06**	**32.48**	**0.218**
Improvement %	**40.10 / 48.91**	**15.44 / 24.45**	**10.74 / 18.31**	**19.01 / 26.82**	**41.38 / 56.10**	**16.06/ 22.38**	**12.89 / 23.21**	**19.53 / 28.99**

entity pair in the dataset except the training data. All the results show that the advantage of our semi-supervised method.

(2) **Our proposed degree-aware knowledge graph embedding approach produces improved entity representations, and thus improves entity alignment results**. This conclusion is drawn from the comparison of improvement made by SEA w/o DA and SEA. After mitigating the effect of entity's degree difference on embedding results, SEA can further improve the semi-supervised results obtained by SEA w/o DA, with only few exceptions. It is worth noting here that the aligned entities with *high* and *low* degree is a small proportion of the whole aligned entity set. How much degree-aware KE embedding can help is limited by this small proportion of entities with *high* and *low* degree levels. Therefore, the improvement made from SEA w/o DA to SEA is justifiable.

5 CONCLUSION

Entity alignment is an important research problem in knowledge graph analysis and management. In this work, we propose a semi-supervised method with degree-aware KG embedding to do the entity alignment. We design a cycle-consistency based translation loss to leverage the unaligned entity to enhance the ability of alignment instead of only using labeled aligned entities. In addition, we observe that the entity degree can influence the learned embeddings and degrade the performance of downstream application. We thus adopt adversarial training to alleviate the problem and improve the embedding results. We conduct experiments on four real-world datasets. The experimental results show that our model consistently outperforms the state-of-art methods on the entity alignment task. Based on the success of this first attempt of semi-supervised entity alignment, in future, we will consider the relation information in the graph to enhance the model.

ACKNOWLEDGMENTS

The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST), under award number FCC/1/1976-19-01.

REFERENCES

- [1] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. AcM, 1247–1250.
- [2] Antoine Bordes, Xavier Glorot, Jason Weston, and Yoshua Bengio. 2014. A semantic matching energy function for learning with multi-relational data. *Machine Learning* 94, 2 (2014), 233–259.
- [3] 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.
- [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. *IJCAI* (2018).
- [5] Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. 2016. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. *arXiv preprint arXiv:1611.03954* (2016).
- [6] Clément Godard, Oisin Mac Aodha, and Gabriel J Brostow. 2017. Unsupervised monocular depth estimation with left-right consistency. In *CVPR*, Vol. 2. 7.
- [7] Chengyue Gong, Di He, Xu Tan, Tao Qin, Liwei Wang, and Tie-Yan Liu. 2018. FRAGE: Frequency-Agnostic Word Representation. *Advances in neural information processing systems* (2018).
- [8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
- [9] Wei Hu, Jianfeng Chen, and Yuzhong Qu. 2011. A self-training approach for resolving object coreference on the semantic web. In *Proceedings of the 20th international conference on World wide web*. ACM, 87–96.
- [10] Qi-Xing Huang and Leonidas Guibas. 2013. Consistent shape maps via semidefinite programming. In *Proceedings of the Eleventh Eurographics/ACMSIGGRAPH Symposium on Geometry Processing*. Eurographics Association, 177–186.
- [11] Rodolphe Jenatton, Nicolas L Roux, Antoine Bordes, and Guillaume R Obozinski. 2012. A latent factor model for highly multi-relational data. In *Advances in Neural Information Processing Systems*. 3167–3175.
- [12] 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)*, Vol. 1. 687–696.
- [13] Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. 2016. Knowledge Graph Completion with Adaptive Sparse Transfer Matrix.. In *AAAI*. 985–991.
- [14] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. 2010. Forward-backward error: Automatic detection of tracking failures. In *Pattern recognition (ICPR), 2010 20th international conference on*. IEEE, 2756–2759.
- [15] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [16] Simon Lacoste-Julien, Konstantina Palla, Alex Davies, Gjergji Kasneci, Thore Graepel, and Zoubin Ghahramani. 2013. Sigma: Simple greedy matching for aligning large knowledge bases. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 572–580.
- [17] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. 2015. DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195.
- [18] Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. 2015. Modeling relation paths for representation learning of knowledge bases. *arXiv preprint arXiv:1506.00379* (2015).
- [19] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion.. In *AAAI*, Vol. 15. 2181–2187.
- [20] Farzaneh Mahdisoltani, Joanna Biega, and Fabian M Suchanek. 2013. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*.
- [21] Jiaqi Mu, Suma Bhat, and Pramod Viswanath. 2017. All-but-the-top: Simple and effective postprocessing for word representations. *arXiv preprint arXiv:1702.01417* (2017).
- [22] Thanh Nguyen, Viviane Moreira, Huong Nguyen, Hoa Nguyen, and Juliana Freire. 2011. Multilingual schema matching for Wikipedia infoboxes. *Proceedings of the VLDB Endowment* 5, 2 (2011), 133–144.
- [23] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A Three-Way Model for Collective Learning on Multi-Relational Data.. In *ICML*, Vol. 11. 809–816.
- [24] Myle Ott, Michael Auli, David Granger, and Marc'Aurelio Ranzato. 2018. Analyzing uncertainty in neural machine translation. *arXiv preprint arXiv:1803.00047* (2018).
- [25] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural Machine Translation of Rare Words with Subword Units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 1715–1725.
- [26] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems*. 926–934.
- [27] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 697–706.
- [28] Zequn Sun, Wei Hu, and Chengkai Li. 2017. Cross-lingual entity alignment via joint attribute-preserving embedding. In *International Semantic Web Conference*. Springer, 628–644.
- [29] Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. 2018. Bootstrapping Entity Alignment with Knowledge Graph Embedding.. In *IJCAI*. 4396–4402.
- [30] Narayanan Sundaram, Thomas Brox, and Kurt Keutzer. 2010. Dense point trajectories by GPU-accelerated large displacement optical flow. In *European conference on computer vision*. Springer, 438–451.
- [31] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*. 2071–2080.
- [32] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.
- [33] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering* 29, 12 (2017), 2724–2743.
- [34] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. 2018. Cross-lingual Knowledge Graph Alignment via Graph Convolutional Networks.. In *EMNLP*.
- [35] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes.. In *AAAI*, Vol. 14. 1112–1119.
- [36] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575* (2014).
- [37] Yang Yang, Yizhou Sun, Jie Tang, Bo Ma, and Juanzi Li. 2015. Entity matching across heterogeneous sources. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1395–1404.
- [38] Zili Yi, Hao (Richard) Zhang, Ping Tan, and Minglun Gong. 2017. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation.. In *ICCV*. 2868–2876.
- [39] Tinghui Zhou, Yong Jae Lee, Stella X Yu, and Alyosha A Efros. 2015. Flowweb: Joint image set alignment by weaving consistent, pixel-wise correspondences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1191–1200.
- [40] Tinghui Zhou, Philipp Krahenbuhl, Mathieu Aubry, Qixing Huang, and Alexei A Efros. 2016. Learning dense correspondence via 3d-guided cycle consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 117–126.
- [41] Hao Zhu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. 2017. Iterative entity alignment via joint knowledge embeddings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 4258–4264.
- [42] Xiaojin Zhu and Andrew B Goldberg. 2009. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning* 3, 1 (2009), 1–130.

6 APPENDIX

6.1 Analysis of the Impact of Entity's Degree Difference on Embedding

In this section, we investigate the impact of entity's degree difference on the embeddings learned by knowledge graph embedding approaches. We catalogs the degree of entities into three levels, *high degree*, *normal degree*, and *low degree*. The *high degree* entities are those with top 20% degree values in a KG, while the *low degree* entities are those with bottom 20% degree values, and the rest are the *normal degree* entities. We apply the most popular KG embedding method TransE on the evaluation datasets presented in Section 4, and report only the embedding results from WK31-15 dataset [5], due to the space limits and the similar observations from other datasets.

Figure 3 (a) and (b) visualize the embeddings of entities in WK31-15 English and French, respectively, through applying Singular Value Decomposition (SVD) to reduce the embeddings into 3-dim space. Entities with different degree levels are shown in different colors. We can observe that in both Figure 3 (a) and (b), entities at the same degree level tend to be close in the mapped space. Entities with high degree values and normal degree values are mostly mapped in three dense regions (in red and green), while entities with low degree values relatively more spread out around centers (in blue) deviating from those in red and green.

Figure 3 (c) shows the transferred embeddings of French KG in the English embedding space. The mapping function minimizes the distance between the given aligned entities. Figure 3 (a) and (c) generally have similar distribution patterns. However, careful comparison between (a) and (c) shows that entities with low and high degree values (in blue and red) are less matchable than entities with normal degree values (in green). Thus, Figure 3 visually illustrates the impact of entity's degree difference on the embeddings.

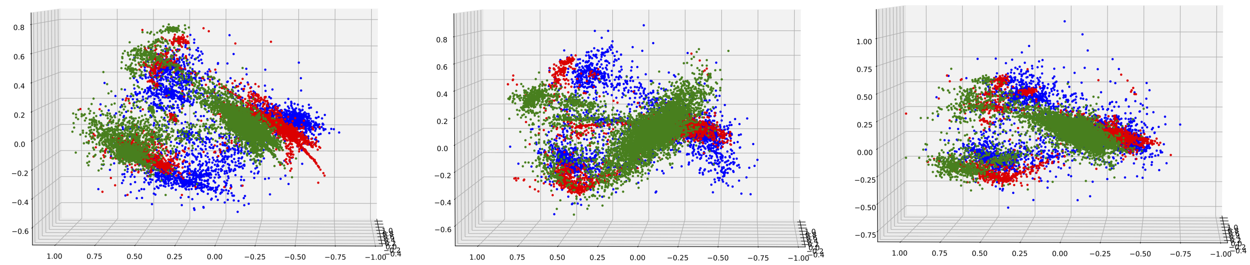
To further investigate the impact, we find the 50 nearest neighbors of each entity in the embedding results, and show the distribution of neighbors at different degree level in Table 5. The first row of Table 5 shows that 91% of the nearest neighbors of entities with *high* degree are the entities with *high* and *normal* degree, only 9% are entities with *low* degree, in both English and French KG embedding results. For the entities with *normal* degree, the second row of Table 5 shows that 74% (76%) of their neighbors are entities also with *normal* degree in the English graph (French graph). Last, the third row of Table 5 shows that the entities with *low* degree have more neighbors with *normal* and *low* degrees than with *high* degrees (11% and 13% in English and French graph, respectively). In summary, a significant observation is that entities with *high* degree are far from the entities with *low* degree in the embedding space. Since entities with *normal* degree make up the largest proportion of whole set of entity, we therefore design a model to treat entities with *normal* degree as anchors and pull entities with *high* and *low* degree close to the anchors in the embedding process. By doing so, the impact of entity's degree difference on embedding can be mitigated. More details of the proposed model are given in Section 3.2. And we present the embedding results here also in Figure 3 for an easy comparison.

Table 5: Distribution of the 50 nearest neighbors of entities at different degree levels in embedding space of TransE (top) and our degree-aware KG embedding model (bottom). Values in columns of %High (%Normal, %Low) are the portion of the 50 nearest neighbors with *high* (*normal, low*) degrees.

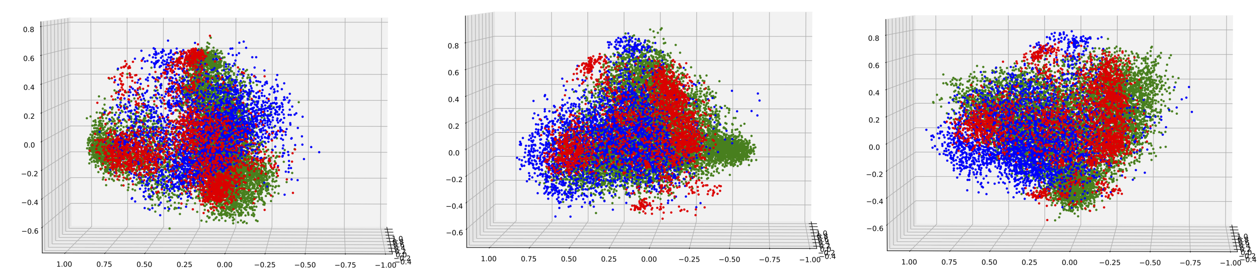
	English			French		
	%High	%Normal	%Low	%High	%Normal	%Low
High	42	49	9	51	40	9
Normal	15	74	11	12	76	12
Low	11	47	42	13	49	38

	English			French		
	%High	%Normal	%Low	%High	%Normal	%Low
High	28	54	18	32	50	18
Normal	17	70	13	15	69	16
Low	18	55	27	26	52	22

Figure 3 (d) and (e) visualize the embeddings learned by our degree-aware KG embedding method proposed in Section 3.2. Compare the same set of entities represented in Figure 3 (a) and (d), and those in Figure 3 (b) and (e), we can observe that the embedding results of our model in Figure 3 (d) and (e) spread more than those in Figure 3 (a) and (b). The influence of entity's degree is less severe. Moreover, we can find that the transferred embeddings of French KG in Figure 3 (f) also show the similar distribution pattern, and entities with different degree values share same regions in the space. Also, we calculate the distribution of the 50 nearest neighbors of entities at different degree levels in our degree-aware KG embedding space. The results are given in the bottom part of Table 5. We find that the proportion of entity with *low* degree of nearest neighbors of the entity with *high* degree become 18% (it is 9% in the top part in Table 1). The overall comparison in Table 1 and Figure 3 demonstrates that our degree-aware KG embedding method can mitigate the effect of the degree difference of entities. Its effectiveness on improving entity alignment will be evaluated and presented in Section 4.



(a) Embeddings of entities in English KG by TransE (b) Embeddings of entities in French KG by TransE (c) Transferred embeddings of entities in French KG



(d) Embeddings of entities in English KG by our model (e) Embeddings of entities in French KG by our model (f) Transferred embeddings of entities in French KG

Figure 3: Demonstration of the impact of entity's degree difference on embedding, obtained by TransE and our proposed model on WK31-15 dataset. (a) and (b) visualizes the embeddings of entities in WK31-15 English and French KG, respectively. It can be observed that entities at the same degree level tend to be close in the mapped space. (c) shows the transferred embeddings of French KG in the English embedding space. (d) and (e) show the embeddings of the same entities in WK31-15 English and French KG obtained by our degree-aware KG embedding model proposed in Section 3.2. (f) shows the transferred embeddings of French KG in the English embedding space. Comparison of (a) and (c) shows that entities with low and high degree values (in blue and red) are less matchable than entities with normal degree values (in green). Results in (d-f) show that the phenomena in (a-c) have been mitigated by our model. Figure best viewed in pdf or colored print.