Multilingual Knowledge Graph Completion with Self-Supervised Adaptive Graph Alignment

Zijie Huang[1][*], **Zheng Li**[2][†]**, Haoming Jiang**[2]**, Tianyu Cao**[2]**, Hanqing Lu**[2],

Bing Yin[2]**, Karthik Subbian**[2]**, Yizhou Sun**[1]**, Wei Wang**[1]

[1]University of California, Los Angeles, CA, USA [2]Amazon.com Inc, CA, USA
{zijiehuang,yzsun,weiwang}@cs.ucla.edu
{amzzhe,jhaoming,caoty,luhanqin,alexbyin,ksubbian}@amazon.com

Abstract

Predicting missing facts in a knowledge graph (KG) is crucial as modern KGs are far from complete. Due to labor-intensive human labeling, this phenomenon deteriorates when handling knowledge represented in various languages. In this paper, we explore multilingual KG completion, which leverages limited seed alignment as a bridge, to embrace the collective knowledge from multiple languages. However, language alignment used in prior works is still not fully exploited: (1) alignment pairs are treated equally to maximally push parallel entities to be close, which ignores KG capacity inconsistency; (2) seed alignment is scarce and new alignment identification is usually in a noisily unsupervised manner. To tackle these issues, we propose a novel self-supervised adaptive graph alignment (SS-AGA) method. Specifically, SS-AGA fuses all KGs as a whole graph by regarding alignment as a new edge type. As such, information propagation and noise influence across KGs can be adaptively controlled via relation-aware attention weights. Meanwhile, SS-AGA features a new pair generator that dynamically captures potential alignment pairs in a self-supervised paradigm. Extensive experiments on both the public multilingual DBPedia KG and newly-created industrial multilingual E-commerce KG empirically demonstrate the effectiveness of SS-AGA[1].

1 Introduction

Knowledge graphs (KGs) like Freebase (Bollacker et al., 2008) and DBPedia (Lehmann et al., 2015) are essential for various knowledge-driven applications such as question answering (Yasunaga et al., 2021) and commonsense reasoning (Lin et al., 2021). A KG contains structured and semantic information among entities and relations, where prior
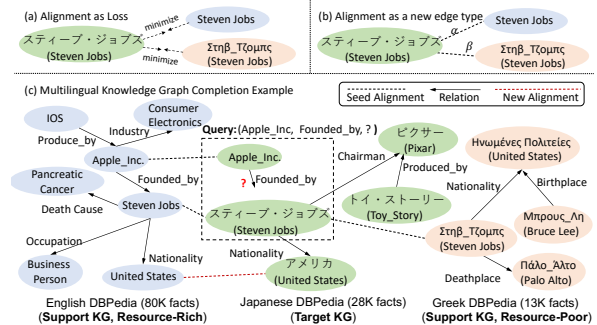

Figure 1: (a) Existing methods treat alignment pairs equally as a loss, which maximally ensures the same entity from different languages to be as similar as possible. (b) Our method differentiates alignment pairs as a new type edge with dynamic attention weights such as α and β, which control the influence and information propagation from other support KGs. (c) An example of MKGC task answering the query in the Japanese KG.

knowledge can be instantiated as factual triples (head entity, relation, tail entity), e.g., (*Apple Inc.*, *Founded by*, *Steven Jobs*). As new facts are continually emerging, modern KGs are still far from being complete due to the high cost of human annotation, which spurs on the Knowledge Graph Completion (KGC) task to automatically predict missing triples to complete the knowledge graph.

The KG incompletion circumstance is exacerbated in the multilingual setting, as human annotations are rare and difficult to gather, especially for low-resource languages. Unfortunately, most efforts for KGC have been devoted to learning each monolingual KG separately (Peng et al., 2021; Xu et al., 2021; Liang et al., 2021; Cao et al., 2021; Lovelace et al., 2021), which usually underperform in low-resource language KGs that suffer from the sparseness (Chen et al., 2017, 2020; Sun et al., 2020). In contrast, KGs from multiple languages are not naturally isolated, which usually share some real-world entities and relations. The transferable knowledge can be treated as a bridge to align different KGs, which not only facilitates the knowledge propagation to low-resource KGs but also alleviates

---

costly manual labeling for all languages.

In this paper, we explore multilingual KG completion (MKGC) (Chen et al., 2020) with limited seed alignment across languages. To mitigate language gaps, some efforts have been initiated on multilingual KG embedding methods, which leverage a KG embedding module (e.g., TransE (Bordes et al., 2013)) to encode each language-specific KG independently and then employ an alignment loss to force pairs of aligned entities to be close maximally (Chen et al., 2020; Zhang et al., 2019; Sun et al., 2020). However, such approaches mainly involve two limitations: (1) the KG inconsistency issue among different languages is neglected due to the equal treatment for parallel entities; (2) the scarcity of seed alignment hinders the efficient knowledge transfer across languages.

Concretely, prior methods treat all alignment pairs equally by forcing all parallel entities to be maximally close to each other (Chen et al., 2018; Sun et al., 2018; Chen et al., 2017). This ignores potentially negative effects from the KG inconsistency due to the language diversity. For example, as shown in Figure 1, the support English KG in DBP-5L (Chen et al., 2020) has much more enriched knowledge (80K facts) than the Greek one (13K facts). In order to complete the query (*Apple Inc.*, *Founded by*, ?) in the resource-poor Japanese KG (28K facts), we can transfer more knowledge from resource-rich English KG through the alignment link of *Steven Jobs* than that of the low-data Greek. However, if roughly pushing *Steven Jobs* to be equally close to that English KG and Greek KG, the learned embeddings for *Steven Jobs* will be similar even though they have different structures, KG capacity, coverage and quality. As such, it will bring in irrelevant information regarding this query and may cause the model to get the wrong answer. Thus, we encourage the model to automatically distinguish the underlying inconsistency and transfer knowledge from suitable support KGs[2] for better language-specific KGC performance.

One the other hand, seed alignment is critical for cross-lingual transfer (Chen et al., 2020; Sun et al., 2020), while acquisition of such parallel entities across languages is costly and often noisy. To mitigate such issue, some recent works (Chen et al., 2018, 2020) propose to generate new alignment pairs based on the entity embedding similarity dur-

ing the training process. The generated new pairs can increase the inter-connectivity between KGs to facilitate knowledge transfer. However, simple usage of correlations between entities without any supervision may increase the noise during training, and inhibit the effectiveness of realistic language alignment in KGs (Sun et al., 2020).

Motivated by these observations, we propose a **S**elf-**S**upervised **A**daptive **G**raph **A**lignment (**SS-AGA**) framework for MKGC. To tackle the knowledge inconsistency issue, SS-AGA regards alignment as a new edge type between parallel entities instead of a loss constrain, which fuses KGs from different languages as a whole graph. Based on such unified modeling, we propose a novel GNN encoder with a relation-aware attention mechanism, which aggregates local neighborhood information with learnable attention weights and differs the influence received from multiple alignment pairs for the same entity as shown in Figure 1(b). To alleviate the scarcity of seed alignment, SS-AGA exploits a new pair generator that iteratively identifies new alignment pairs in a self-supervised manner. This is achieved by masking some seed alignment in the fused KG before GNN encoding and teaching the generation module to recover them. Empirically, SS-AGA outperforms popular baselines in both public and industrial datasets. For the public dataset, we use the multilingual DBPedia KG (Chen et al., 2020) and for the industrial dataset, we create a multilingual E-commerce Product KG called E-PKG.

Our contributions are as follows: (1) We handle the knowledge inconsistency issue for MKGC by treating entity alignment as a new edge type and introducing a relation-aware attention mechanism to control the knowledge propagation; (2) We propose a new alignment pair generation mechanism with self-supervision to alleviate the scarcity of seed alignment; (3) We constructed a new industrial-level multilingual E-commerce KG dataset; (4) Extensive experiments verify the effectiveness of SS-AGA in both public and industrial datasets.

2 Preliminaries

2.1 Knowledge Graph Completion

A knowledge graph $G = (\mathcal{E}, \mathcal{R}, \mathcal{T})$ consists of a set of entities $\mathcal{E}$, relations $\mathcal{R}$, and relational facts $\mathcal{T} = \{(e_h, r, e_t)\}$, where $e_h, e_t \in \mathcal{E}$ are head and tail entities, and $r \in \mathcal{R}$ is a relation. Entities and relations are represented by their text descriptions. The

---

[2] We regard the remaining KGs as the support KGs when conducting the KGC task in the target one.
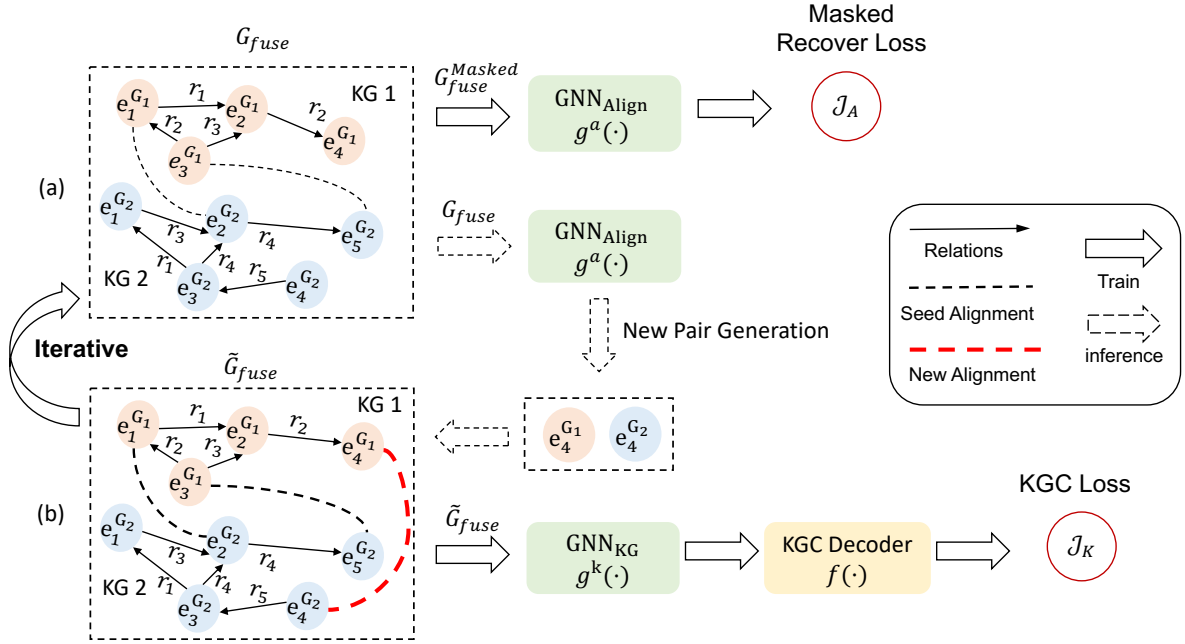
Figure 2: The overall framework of the Self-Supervised Adaptive Graph Alignment (SS-AGA).

KG completion task seeks to impute the missing head or tail entity of a triple given the relation and the other entity. Without loss of generality, we hereafter discuss the case of predicting missing tails, which we also refer to as a query $q = (e_h, r, ?e_t)$. **Multilingual KG completion (MKGC)** utilizes KGs across multiple languages to achieve more accurate KG completion task on each individual KG (Chen et al., 2020). Formally, we are given M different language-specific KGs as $G_1, G_2, \cdots, G_M$, and only limited entity alignment pairs $\Gamma_{G_i \leftrightarrow G_j} \subseteq \{(e_i, e_j) : e_i \in \mathcal{E}_i, e_j \in \mathcal{E}_j\}$ between G_i and G_j. We also call $\Gamma_{G_i \leftrightarrow G_j}$ the *seed* alignment pairs to distinguish it from the new or pseudo alignment. Each KG G_i has their own relation set $\mathcal{R}_i$. We denote the union of relation sets from all KGs as a unified relation set $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \cdots \mathcal{R}_M$. MKGC is related to but different from the entity alignment (EA) task (Cao et al., 2019; Sun et al., 2020). In MKGC, seed alignment is not direct supervision while the auxiliary input features, all used in the training stage for cross-lingual transfer to boost the KGC results.

2.2 KG Embedding Models

KG embedding models aim to learn latent low-dimensional representations for entities $\{e\}_{e \in \mathcal{E}}$ and relations $\{r\}_{r \in \mathcal{R}}$. A naive implementation is an embedding lookup table (Bordes et al., 2013; Sun et al., 2019). Recently, Graph Neural Networks (GNN) have been explored to aggregate neighbor-

hood information in KGs, where each triple is no longer considered independent of each other (Hao et al., 2019). Mathematically, these methods employ a GNN-based encoder g that embeds entities considering the neighborhood information,

$$\{e\}_{e \in \mathcal{E}} = g(G).$$

Then, the plausibility of a relational fact (e_h, r, e_t) can be measured by the triple score:

$$f(e_h, r, e_t),$$

where f can be any scoring function such as TransE (Bordes et al., 2013), RotatE (Sun et al., 2019). We also refer it to as the KGC decoder.

3 Method

We introduce SS-AGA for MKGC, consisting of two alternating training components (a) and (b) in Figure 2: (a) A new alignment pair generation module for alleviating the limited seed alignment in G_{fuse}. Specifically, we mask some seed alignment in the fuse KG to obtain $G_{\text{fuse}}^{\text{Masked}}$ and train the generator $g^a(\cdot)$ to recover them. Then, the trained generator will propose new edges based on the learned entity embeddings, which will be incorporated to G_{fuse} as $\widetilde{G}_{\text{fuse}}$ for MKG embedding model $g^k(\cdot)$ in the next iteration; (b) A novel relation-aware MKG embedding model $g^k(\cdot)$ for addressing the knowledge inconsistency across multilingual KGs. Specifically, we fuse different KGs as a

whole graph G_{fuse} by treating alignment as a new edge type. Then $g^k(\cdot)$ computes the contextualized embeddings for each node with learnable relation-aware attention weights that differ the influence received from multiple alignment pairs. Finally, a KGC decoder $f(\cdot)$ computes the triple scores.

3.1 Relation-aware MKG Embedding

As mentioned before, the knowledge transfer is inefficient in existing MKGC methods, as they encode each KG separately and transfer knowledge by forcing aligned entities to share the same embedding. To handle the knowledge inconsistency, we first fuse all KGs as a whole, which relaxes the entity alignment to relational facts. We then design an attention-based relation-aware GNN to learn the contextualized MKG embeddings for entities, which can differ the influence from multiple alignment sources with learnable attention weights. Afterwards, we apply a KGC decoder on the contextualized embedding to get the triple scores for relational facts.

More specifically, we create the fused KG by preserving triples within each KG and converting each cross-KG alignment pair (e_i, e_j) to two relational facts $(e_i, r_{\text{align}}, e_j)$ and $(e_j, r_{\text{align}}, e_i)$ with the alignment edge as a newly introduced relation r_{align}. In this way, we enable direct message passing among entities from different KGs, where the attention weight can be learned automatically from data to differ the influence from multiple alignment pairs. We denote the fused knowledge graph as $G_{\text{fuse}} = (\mathcal{E}_{\text{fuse}}, \mathcal{R}_{\text{fuse}}, \mathcal{T}_{\text{fuse}})$, where $\mathcal{E}_{\text{fuse}} = \bigcup_{i=1}^{M} \mathcal{E}_i$, $\mathcal{R}_{\text{fuse}}! = (\bigcup_{i=1}^{M} \mathcal{R}_i) \cup \{r_{\text{align}}\}$ and $\mathcal{T}_{\text{fuse}} = (\bigcup_{i=1}^{M} \mathcal{T}_i) \cup (\bigcup_{i,j}\{(e_h, r_{\text{align}}, e_t) : (e_h, e_t)$ or $(e_t, e_h) \in \Gamma_{G_i \leftrightarrow G_j}\})$.

Given the fused KG G_{fuse}, we propose an attention-based relation-aware GNN encoder $g^k(\cdot)$ to learn contextualized embeddings for entities following a multi-layer message passing architecture.

At the l-th layer of GNN, we first compute the relation-aware message delivered by the entity e_i in a relational fact (e_i, r, e_j) as follows:

$$\boldsymbol{h}_{i(r)}^l = \text{Msg}\left(\boldsymbol{h}_i^l, r\right) := \boldsymbol{W}_v^l \text{Concat}(\boldsymbol{h}_i^l, \boldsymbol{r}),$$

where $\boldsymbol{h}_i^l$ is the latent representation of e_i at the l-th layer, $\text{Concat}(\cdot, \cdot)$ is the vector concatenation function, and $\boldsymbol{W}_v^l$ is a transformation matrix. Then, we propose a relation-aware scaled dot product attention mechanism to characterize the importance of each entity's neighbor e_i to itself e_j, which is computed as follows:

$$\text{Att}\left(\boldsymbol{h}_{i(r)}^l, \boldsymbol{h}_j^l\right) = \frac{\exp(\alpha_{ij}^r)}{\sum_{(e_{i'}, r) \in \mathcal{N}(e_j)} \exp\left(\alpha_{i'j}^r\right)}$$

$$\alpha_{ij}^r = \left(\boldsymbol{W}_k^l \boldsymbol{h}_{i(r)}^l\right)^T \cdot \left(\boldsymbol{W}_q^l \boldsymbol{h}_j^l\right) \cdot \frac{1}{\sqrt{d}} \cdot \beta_r, \quad (1)$$

where d is the dimension of the entity embeddings, $\boldsymbol{W}_k^l, \boldsymbol{W}_q^l$ are two transformation matrices, and β_r is a learnable relation factor. Different from the traditional attention mechanism (Veličković et al., 2018; Bai et al., 2019), we introduce β_r to characterize the general significance of each relation r. It is essential as not all the relationships contribute equally to the query entity. We also remark that the neighborhood is bidirectional, i.e. $\mathcal{N}(e_j) := \{(e_{i'}, r) : (e_{i'}, r, e_j) \in \mathcal{T}_{\text{fuse}} \text{ or } (e_j, r, e_{i'}) \in \mathcal{T}_{\text{fuse}}\}$ as the tail entity will also influence the head entity.

We then update the hidden representation of entities by aggregating the message from their neighborhoods based on the attention score:

$$\boldsymbol{h}_j^{l+1} = \boldsymbol{h}_j^l + \sigma\left(\sum_{(e_{i'}, r) \in \mathcal{N}(e_j)} \text{Att}\left(\boldsymbol{h}_{i'(r)}^l, \boldsymbol{h}_j^l\right) \cdot \boldsymbol{h}_{i'(r)}^l\right),$$

where $\sigma(\cdot)$ is a non-linear activation function, and the residual connection is used to improve the stability of GNN (He et al., 2015).

Finally, we stack L layers to aggregate information from multi-hop neighbors and obtain the contextualized embedding for each entity e_j as: $\boldsymbol{e}_j = \boldsymbol{h}_j^L$. Given the contextualized entity embeddings, the KGC decoder computes the triple score for each relational fact: $f(\boldsymbol{e}_h, \boldsymbol{r}, \boldsymbol{e}_t)$. The learning object is to minimize the following hinge loss:

$$\mathcal{J}_K = \sum_{\substack{(e_h, r, e_t) \in \mathcal{T}_m \\ (e_{h'}, r, e_{t'}) \notin \mathcal{T}_m \\ m=1,\ldots,M}} \left[f\left(\boldsymbol{e}_h{}', \boldsymbol{r}, \boldsymbol{e}_t'\right) - f\left(\boldsymbol{e}_h, \boldsymbol{r}, \boldsymbol{e}_t\right) + \gamma\right]_+, \quad (2)$$

where $\gamma > 0$ is a positive margin, f is the KGC decoder, $(e_{h'}, r, e_{t'})$ is a negative sampled triple obtained by replacing either head or tail entity of the true triple (e_h, r, e_t) randomly by other entities in the same language-specific KG.

Remark 1. Our method views cross-KG alignment as a relation r_{align} in the fused KG. The knowledge transfer cross KGs is essentially conducted via the learnable attention weight $\alpha_{ij}^{r_{\text{align}}}$, where e_i and e_j are connected through the relation r_{align}. Thanks to the power of GNN, $\alpha_{ij}^{r_{\text{align}}}$ differs the influence from multiple alignment sources, as

opposed to some existing models that simply force pairs of entities to be close to each other through a pre-defined alignment loss. In this way, we properly conduct knowledge transfer among KGs with aware of their knowledge inconsistency.

Scalability issue. Since we fuse all the M KGs as a whole, and duplicate edges for head entities, the scale of the graph G_{fuse} would become very large. We therefore employ a k-hop graph sampler that samples the k-hop neighbors for each node and compute their contextualized embeddings.

3.2 Self-supervised New Pair Generation

In multilingual KGs, we are only provided with limited seed alignment pairs to facilitate knowledge transfer, as they are expensive to obtain and even sometimes noisy (Sun et al., 2020). To tackle such challenge, we propose a self-supervised new alignment pair generator. In each iteration, the generator identifies new alignment pairs which will be fed into the GNN encoder $g^k(\cdot)$ to produce the contextualized entity embeddings in the next iteration. The training of the generator is conducted in a self-supervised manner, where the generator is required to recover masked alignment pairs.

New Pair Generation (NPG) relies on two sets of entity embeddings: the structural embeddings and the textual embeddings. The structural embeddings are obtained by another GNN encoder g^a: $\{e^a\}_{e \in \mathcal{E}_{\text{fuse}}} = g^a(G_{\text{fuse}})$, which shares the same architecture with $g^k(\cdot)$ in the relation-aware MKG Embedding model (Section 3.1). The reason we employ two GNN encoders is that the set of embeddings that generate the best alignment results may differ from those that can best achieve the KG completion task.

The textual embeddings are obtained by entities' text description and mBERT: $e^{\text{text}} = \text{mBERT}(e)$. mBERT is a multilingual pre-trained language model (Devlin et al., 2019) and is particularly attractive to the new alignment pair generation due to the following merits: (1) it captures rich semantic information of the text; (2) the pre-trained BERT embeddings are also aligned across different languages (Devlin et al., 2019; Sun et al., 2020).

We then model the pairwise similarity score between entity e_i and e_j as the maximum of the cosine similarities of their structural embeddings and textual embeddings:

$$\text{sim}(e_i, e_j) = \max\left(\cos\left(e_i^a, e_j^a\right), \cos\left(e_i^{\text{text}}, e_j^{\text{text}}\right)\right).$$

Then we introduce new alignment pairs if a pair of unaligned entities in two KGs are mutual nearest neighbors according to the cross-domain similarity local scaling (CSLS) measure (Conneau et al., 2018) as shown below,

$$\text{CSLS}(e_i, e_j) = 2\text{sim}(e_i, e_j) - s(e_i) - s(e_j)$$

$$\text{subject to } s(e_i) = \frac{1}{K} \sum_{e_{i'} \in \mathcal{N}(e_i)} \text{sim}(e_i, e_{i'}),$$

where K is the number of each node's k-nearest neighbors. CSLS is able to capture the sturctural similarity between pairs of entities. The generated pairs are then utilized to update the graph structure of G_{fuse} to $\widetilde{G}_{\text{fuse}}$ in the next iteration, to alleviate the challenge of limited seed alignment.

Self-Supervised Learning (SSL) Similar to many existing works (Chen et al., 2020; Sun et al., 2020), the aforementioned NPG paradigm is unsupervised and may bring in unexpected noises. Inspired by masked language modeling (Devlin et al., 2019) which captures contextual dependencies between tokens, we propose a self-supervised learning procedure to guide and denoise the new pair generation. Specifically, we randomly mask out some alignment relational facts, $\mathcal{T}_{\text{masked}} \subseteq \{(e_h, r, e_t) \in \mathcal{T}_{\text{fuse}} : r = r_{\text{align}}\}$, and let the generator to recover them. Such masked alignment recovery in KGs can automatically identify the underlying correlations for alignment neighbors and encourage the NPG to generate high-quality alignment pairs that are real existences but hide due to the limited seed alignment.

Given the fused KG with masked alignment $G_{\text{fuse}}^{\text{Masked}} = \{\mathcal{E}_{\text{fuse}}, \mathcal{R}_{\text{fuse}}, \mathcal{T}_{\text{fuse}}/\mathcal{T}_{\text{masked}}\}$, the GNN encoder g^a embeds the entities as

$$\{\widetilde{e}\}_{e \in \mathcal{E}_{\text{fuse}}} = g^a(G_{\text{fuse}}^{\text{Masked}}).$$

The GNN g^a is then trained via minimizing the following hinge loss $\mathcal{J}_A$,

$$\mathcal{J}_A^{G_i \leftrightarrow G_j} = \sum_{\substack{(e_h, e_t) \in \Gamma_{ij}^p \\ (e_{h'}, e_{t'}) \in \Gamma_{ij}^n}} \left[\|\widetilde{e}_h^a - \widetilde{e}_t^a\|_2 - \|\widetilde{e}_{h'}^a - \widetilde{e}_{t'}^a\|_2 + \gamma_a\right]_+$$

$$\mathcal{J}_A = \sum_{1 \leq i < j \leq M} \mathcal{J}_A^{G_i \leftrightarrow G_j}, \qquad (3)$$

where $\Gamma_{ij}^p = \{(e_h \in \mathcal{E}_i, e_t \in \mathcal{E}_j) : (e_h, r_{\text{align}}, e_t) \in \mathcal{T}_{\text{masked}}\}$ is the masked alignment set, $\Gamma_{ij}^n = \{(e_h \in \mathcal{E}_i, e_t \in \mathcal{E}_j) : (e_h, e_t) \notin \Gamma_{G_i \leftrightarrow G_j}\}$ is the unaligned entity pair set, and $\gamma_a > 0$ is a positive margin. $(e_{h'}, e_{t'})$ is randomly sampled by replacing one of the entities in the positive entity pairs.

3.3 Training

The overall loss function is the combination of the KG completion loss Eq. (2) and the self-supervised alignment loss Eq. (3) as shown below

$$\mathcal{J} = \mathcal{J}_K + \lambda\mathcal{J}_A, \qquad (4)$$

where $\lambda > 0$ is a positive hyperparameter to balance between the two losses. We summarize the training process in Algorithm 1 of the Appendix.

4 Experiments

4.1 Dataset

We conduct experiments over two real-world datasets. (i) **DBP-5L** (Chen et al., 2020) contains five language-specific KGs from DBpedia (Lehmann et al., 2015), i.e., English (EN), French (FR), Spanish (ES), Japanese (JA), Greek (EL). As the original dataset only contains structural information, we additionally crawled the text information for these entities and relations based on the given URLs. (ii) **E-PKG** is a new industrial multilingual E-commerce product KG dataset, which describes phone-related product information from an E-commerce platform across six different languages: English (EN), German (DE), French (FR), Japanese (JA), Spanish (ES), Italian (IT). The statistics are shown in Table 1. The # Aligned Links for a specific KG G_i denotes the number of alignment pairs where one of the aligned entities belong to that KG. It is possible for an entity to have multiple alignment pairs across different KG sources. For both datasets, we randomly split the facts in each KG into three parts: 60% for training, 30% for validation, and 10% for testing. Please refer to *Appendix A* for the details of E-PKG construction.

4.2 Evaluation Protocol

In the testing phase, given each query $(e_h, r, ?e_t)$, we compute the plausibility scores $f(e_h, r, \widetilde{e}_t)$ for triples formed by each possible tail entity $\widetilde{e}_t$ in the test candidate set and rank them. We report the mean reciprocal ranks (MRR), accuracy (Hits@1) and the proportion of correct answers ranked within the top 10 (Hits@10) for testing. We also adopt the filtered setting following previous works based on the premise that the candidate space has excluded the triples that have been seen in the training set (Wang et al., 2014a; Yang et al., 2015a).

Dataset	#Entity	#Relation	#Triple	#Aligned Links
Multilingual Academic KG (*DBP-5L*)				
EN	13,996	831	80,167	16,916
FR	13,176	178	49,015	16,877
ES	12,382	144	54,066	16,347
JA	11,805	128	28,774	16,263
EL	5,231	111	13,839	9,042
Multilingual Industrial KG (*E-PKG*)				
EN	16,544	21	100,531	21,382
DE	17,223	21	75,870	24,696
FR	17,068	21	80,015	24,812
JA	2,642	21	16,703	5,175
ES	9,595	21	30,163	20,184
IT	15,670	21	71,292	23,827

Table 1: Statistics of DBP-5L and E-PKG datasets. #Aligned Links denotes the number of alignment pairs where one of the aligned entities belongs to that KG.

4.3 Baselines

• **Monolingual Baselines.** (i) **TransE** (Bordes et al., 2013) models relations as translations in the Euclidean space; (ii) **RotatE** (Sun et al., 2019) models relations as rotations in the complex space; (iii) **DisMult** (Yang et al., 2015b) uses a simple bilinear formulation; (iv) **KG-BERT** (Yao et al., 2020) employs pre-trained language models for knowledge graph completion based on text information of relations and entities.

• **Multilingual Baselines.** (i) **KEnS** (Chen et al., 2020) embeds all KGs in a unified space and exploits an ensemble technique to conduct knowledge transfer; (ii) **CG-MuA** (Zhu et al., 2020) is a GNN-based KG alignment model with collective aggregation. We revise its loss function to conduct MKGC. (iii) **AlignKGC** (Singh et al., 2021) jointly trains the KGC loss with entity and relation alignment losses. For fair comparison, we use mBERT (Devlin et al., 2019) to obtain initial embeddings of entities and relations from their text for all methods. We do not employ any pretrained tasks such as EA to obtain these initial text embeddings as in (Singh et al., 2021).

4.4 Main Results

The main results are shown in Table 2 and Table 3. Firstly, by comparing multilingual and monolingual KG models, we can observe that multilingual methods can achieve better performance. This indicates that the intuition behind utilizing multiple KG sources to conduct KG completion is indeed beneficial, compared with inferring each KG independently. Notably, multilingual models tend

Method	Metric	EL	JA	ES	FR	EN
		Monolingual Baselines				
TransE	H@1	13.1	21.1	13.5	17.5	7.3
	H@10	43.7	48.5	45.0	48.8	29.3
	MRR	24.3	25.3	24.4	27.6	16.9
RotatE	H@1	14.5	26.4	21.2	23.2	12.3
	H@10	36.2	60.2	53.9	55.5	30.4
	MRR	26.2	39.8	33.8	35.1	20.7
DisMult	H@1	8.9	9.3	7.4	6.1	8.8
	H@10	11.3	27.5	22.4	23.8	30.0
	MRR	9.8	15.8	13.2	14.5	18.3
KG-BERT	H@1	17.3	26.9	21.9	23.5	12.9
	H@10	40.1	59.8	54.1	55.9	31.9
	MRR	27.3	38.7	34.0	35.4	21.0
		Multilingual Baselines				
KenS	H@1	28.1	32.1	23.6	25.5	15.1
	H@10	56.9	65.3	60.1	62.9	39.8
	MRR	-	-	-	-	-
CG-MuA	H@1	21.5	27.3	22.3	24.2	13.1
	H@10	44.8	61.1	55.4	57.1	33.5
	MRR	32.8	40.1	34.3	36.1	22.2
AlignKGC	H@1	27.6	31.6	24.2	24.1	15.5
	H@10	56.3	64.3	60.9	62.3	39.2
	MRR	33.8	41.6	35.1	37.4	22.3
SS-AGA	H@1	**30.8**	**34.6**	**25.5**	**27.1**	**16.3**
	H@10	**58.6**	**66.9**	**61.9**	**65.5**	**41.3**
	MRR	**35.3**	**42.9**	**36.6**	**38.4**	**23.1**

Table 2: Main results on DBP-5L.

Method	Metric	EN	DE	FR	JA	ES	IT
		Monolingual Baselines					
TransE	H@1	23.2	21.2	20.8	25.1	17.2	22.0
	H@10	67.5	65.5	66.9	72.7	58.4	63.8
	MRR	39.4	37.4	37.5	43.6	33.0	37.8
RotatE	H@1	24.2	22.3	22.1	26.3	18.3	22.5
	H@10	66.8	64.3	67.1	71.9	58.9	64.0
	MRR	40.0	38.2	38.0	41.8	33.7	38.1
DisMult	H@1	23.8	21.4	20.7	25.9	17.9	22.8
	H@10	60.1	54.5	53.5	62.6	46.2	51.8
	MRR	37.2	35.4	35.1	38.0	30.9	34.8
KG-BERT	H@1	24.3	21.8	22.3	26.9	18.7	22.9
	H@10	66.4	64.7	67.2	72.4	58.8	63.7
	MRR	39.6	38.4	38.3	44.1	33.2	37.2
		Multilingual Baselines					
KenS	H@1	26.2	24.3	25.4	33.5	**21.3**	**25.1**
	H@10	69.5	65.8	68.2	73.6	59.5	**64.6**
	MRR	-	-	-	-	-	-
CG-MuA	H@1	24.8	22.9	23.0	30.4	19.2	23.9
	H@10	67.9	64.9	67.5	72.9	58.8	63.8
	MRR	40.2	38.7	39.1	45.9	33.8	37.6
AlignKGC	H@1	25.6	22.1	22.8	31.2	19.4	24.2
	H@10	68.3	65.1	67.2	72.3	59.1	63.4
	MRR	40.5	38.5	38.8	46.2	34.2	37.3
SS-AGA	H@1	**26.7**	**24.6**	**25.9**	**33.9**	21.0	24.9
	H@10	**69.8**	**66.3**	**68.7**	**74.1**	**60.1**	63.8
	MRR	**41.5**	**39.4**	**40.2**	**48.3**	**36.3**	**38.4**

Table 3: Main results on E-PKG.

to bring larger performance gains for those low-resource KGs such as Greek in DBP-5L, which is expected as low-resource KGs are far from complete and efficient external knowledge transfer can bring in potential benefits. Among multilingual models, our proposed method SS-AGA can achieve better performance in most cases across different metrics, languages, and datasets, which verifies the effectiveness of SS-AGA.

4.5 Ablation Study

To evaluate the effectiveness of our model design, we conduct ablation study by proposing the following model variants: (i) **GNN** applies the GNN encoder without relation modeling to each KG independently, and directly forces all alignment pairs to be close to each other as in prior works (Chen et al., 2020; Zhu et al., 2020); (ii) **R-GNN** is the proposed relation-aware MKG embedding model (Section 3.1), which utilizes all seed alignment to construct G_{fused} and differs the influence from other KGs by the relation-aware attention mechanism; (iii) **R-GNN + NPG** conducts additional new pair generation for R-GNN; (iv) **R-GNN + NPG + SSL** is our proposed full model SS-AGA, which leverages SSL to guide the NPG process. We also investigate the effect of whether to share or not share the

encoders $g^a(\cdot)$, $g^k(\cdot)$ that generate the embeddings for the SSL and KGC loss, respectively.

We report the average Hits@1, Hits@10 and MRR over DBP-5L as shown in Table 4. As we can see, applying a GNN encoder to each KG independently would cause the performance drop as all aligned entities are being equally forced to be close to each other. Removing the new pair generation process would also cause a performance degradation due to the sparsity of seed alignment, which shows that iteratively proposing new alignment is indeed helpful. If the generation process is further equipped with supervision, the performance would be enhanced, which verifies the effectiveness of the self-supervised alignment loss. Finally, sharing the parameters of two GNN encoders would harm the performance. Though MKGC and entity alignment are two close-related tasks that can potentially benefit each other, the set of embeddings that produce the best alignment result do not necessarily yield the best performance on the MKGC task.

4.6 Impact of Seed Alignment

We next study the effect of seed alignment number as depicted in Figure 3. Firstly, we can observe that SS-AGA consistently outperforms other multilingual models on varying alignment ratios. Secondly,
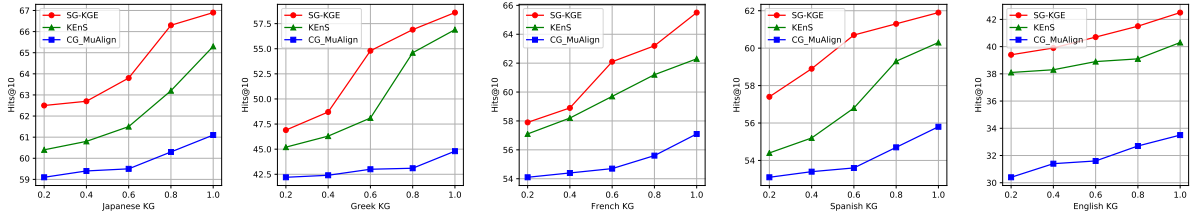
480

Figure 3: Hits@10 with respect to different sampling ratio of seed alignment pairs.

Method	Avg H@1	Avg H@10	Avg MRR
GNN	24.1	56.3	33.2
R-GNN	25.7	57.9	34.4
R-GNN + NPG	26.2	58.3	34.9
R-GNN + NPG + SSL (SS-AGA)			
- encoder (shared)	25.8	57.7	34.1
- encoder (no shared)	**26.9**	**58.7**	**35.3**

Table 4: Ablation results on DBP-5L.

for low-resources KGs such as Japanese and Greek KGs, we can observe a sharp performance drop when decreasing the alignment ratio compared with those popular KGs such as English KG. This indicates that the knowledge transfer among different KGs is especially beneficial for those low-resources KGs, as popular KGs already contain relatively rich knowledge. However, such transfer process is heavily dependent on the seed alignment, which yields the necessity of new alignment generation process.

4.7 Case Study

To interpret the knowledge transfer across different KGs, we visualize the normalized average attention weight for each KG w.r.t. the attention score computed in Eq. (1) from different KG sources. We can see that for those popular KGs, they will receive the highest attention score from themselves such as English and French KGs. Although Japanese KG is low-resource, from the main results table 2, we can see that the gap improvement brought by multilingual methods is relatively small compared to another low-resource Greek KG. This indicates that Japanese KG may contain more reliable facts to facilitate missing triple predictions. However, for Greek KG, we can observe that the attention weights from other languages take the majority, which means that the performance boost in Greek KG is largely attributed to the efficient knowledge transfer from other KG sources.

5 Related Work

5.1 Monolingual KG Embeddings

Knowledge graph embeddings (Bordes et al., 2013; Sun et al., 2019; Con, 2018) achieve the state-of-

the-art performance for KGC, which learn the latent low-dimensional representations of entities and relations. They measure triple plausibility based on varying score functions such as translation-based TransE (Bordes et al., 2013), TransH (Wang et al., 2014b); rotation-based RotatE (Sun et al., 2019) and language-model-based KG-BERT (Yao et al., 2020). Recently, GNN-based methods (Li et al., 2019; Zhang et al., 2020; Javari et al., 2020) have been proposed to capture node neighborhood information for the KGC tasks. GNN is a class of neural networks that operate on graph-structured data by passing local messages (Kipf and Welling, 2017; Veličković et al., 2018; Xu et al., 2019; Bai et al., 2019; Huang et al., 2020, 2021; Wang et al., 2021). Specifically, they use GNN as an encoder to generate contextualized representation of entities by passing local messages (Kipf and Welling, 2017; Veličković et al., 2018; Xu et al., 2019; Bai et al., 2019; Huang et al., 2020, 2021). Then, existing score functions are employed to generate triple scores which outperform the aforementioned methods that treat each triple independently only with the scoring function.

5.2 Multilingual KG Embeddings

Multilingual KG embeddings are extensions of monolingual KG embeddings that consider knowledge transfer across KGs with the use of limited seed alignment (Sun et al., 2020; Singh et al., 2021). Earlier work proposes different ways to reconcile KG embeddings for the **entity alignment (EA)** task: MTransE (Chen et al., 2017) learns a transformation matrix between pairs of KGs. MuGNN (Cao et al., 2019) reconciles structural differences via rule grounding. CG-MuA utilizes collective aggregation of confident neighborhood (Zhu et al., 2020). Others incorporate attribute information such as entity text (Zhang et al., 2019; Chen et al., 2018). To tackle the sparsity of seed alignment, BootEA (Sun et al., 2018) iteratively proposes new aligned pairs via bootstrapping. Zhu et al. (2017) utilizes parameter sharing to improve alignment performance. While they focus
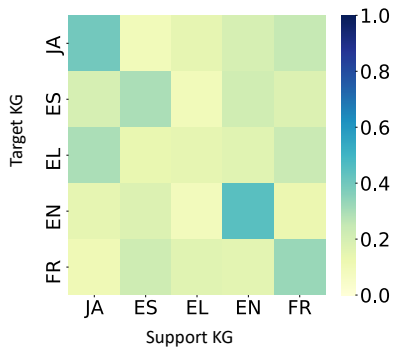
Figure 4: Average attention weight learned in DBP-5L.

on the EA task rather than the MKGC task that we tackle here, such techniques can be leveraged to conduct knowledge transfer among KGs. Recently, Chen et al. (2020) propose an ensemble-based approach for the MKGC task. In this paper, we view alignment as a new edge type and employ a relation-aware GNN to get the contextualized representation of entities. As such, the influence of the aligned entities is captured by the learnable attention weight, instead of assuming each alignment pair to have the same impact. We also propose a self-supervised learning task to propose new alignment pairs during each training epoch to overcome the sparsity issue of seed alignment pairs.

6 Discussion and Conclusion

In this paper, we propose SS-AGA for multilingual knowledge graph completion (MKGC). It addresses the knowledge inconsistency issue by fusing all KGs and utilizing a GNN encoder to learn entity embeddings with learnable attention weights that differs the influence from multiple alignment sources. It features a new pair generation conducted in a self-supervised learning manner to tackle the limited seed alignment issue. Extensive results on two real-world datasets including a newly-created E-commerce dataset verified the effectiveness of SS-AGA. Our current approach may fail to fully exploit the benefit of entity and relation texts. In the future, we plan to study more effective ways to combine text data with graph data for better model performance. We are also interested in studying MKGC where there no alignment pairs are given, which is a very practical setting and our current model is not able to deal with.

7 Ethical Impact

Our paper proposed SS-AGA, a novel multilingual knowledge graph completion model for predicting missing triples in KGs considering their knowledge transfer. SS-AGA neither introduces any social/ethical bias to the model nor amplifies any bias in the data. We the created multilingual E-commerce product KG dataset by masking all customers'/sellers' identity and privacy. We only collect information related to products without any personal information leakage. Our model is built upon public libraries in Pytorch. We do not foresee any direct social consequences or ethical issues.

References

2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the 32th AAAI Conference on Artificial Intelligence*, pages 1811–1818.

Yunsheng Bai, Hao Ding, Song Bian, Ting Chen, Yizhou Sun, and Wei Wang. 2019. Simgnn: A neural network approach to fast graph similarity computation. In *International Conference on Web Search and Data Mining*.

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*, SIGMOD '08, page 1247–1250.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, page 2787–2795.

Yixin Cao, Xiang Ji, Xin Lv, Juanzi Li, Yonggang Wen, and Hanwang Zhang. 2021. Are missing links predictable? an inferential benchmark for knowledge graph completion. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6855–6865, Online.

Yixin Cao, Zhiyuan Liu, Chengjiang Li, Zhiyuan Liu, Juanzi Li, and Tat-Seng Chua. 2019. Multi-channel graph neural network for entity alignment. In *ACL*.

Muhao Chen, Yingtao Tian, Kai-Wei Chang, Steven Skiena, and Zaniolo Carlo. 2018. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3998–4004.

Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. 2017. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*.

Xuelu Chen, Muhao Chen, Changjun Fan, Ankith Uppunda, Yizhou Sun, and Carlo Zaniolo. 2020. Multilingual knowledge graph completion via ensemble knowledge transfer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 3227–3238.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Herve J´egou. 2018. Word translation without parallel data. In *In International Conference on Learning Representations*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Junheng Hao, Muhao Chen, Wenchao Yu, Yizhou Sun, and Wei Wang. 2019. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 1709–1719.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*.

Zijie Huang, Yizhou Sun, and Wei Wang. 2020. Learning continuous system dynamics from irregularly-sampled partial observations. In *Advances in Neural Information Processing Systems*.

Zijie Huang, Yizhou Sun, and Wei Wang. 2021. Coupled graph ode for learning interacting system dynamics. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*.

Amin Javari, Zhankui He, Zijie Huang, Raj Jeetu, and Kevin Chen-Chuan Chang. 2020. Weakly supervised attention for hashtag recommendation using graph data. In *Proceedings of The Web Conference 2020*, WWW '20.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, and Soren Auer. 2015. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. In *Semantic Web*, pages 167–195.

Chengjiang Li, Yixin Cao, Lei Hou, Jiaxin Shi, Juanzi Li, and Tat-Seng Chua. 2019. Semi-supervised entity alignment via joint knowledge embedding model and cross-graph model. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2723–2732.

Zongwei Liang, Junan Yang, Hui Liu, and Keju Huang. 2021. A semantic filter based on relations for knowledge graph completion. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7920–7929.

Bill Yuchen Lin, Haitian Sun, Bhuwan Dhingra, Manzil Zaheer, Xiang Ren, and William Cohen. 2021. Differentiable open-ended commonsense reasoning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4611–4625.

Justin Lovelace, Denis Newman-Griffis, Shikhar Vashishth, Jill Fain Lehman, and Carolyn Rosé. 2021. Robust knowledge graph completion with stacked convolutions and a student re-ranking network. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1016–1029.

Xutan Peng, Guanyi Chen, Chenghua Lin, and Mark Stevenson. 2021. Highly efficient knowledge graph embedding learning with Orthogonal Procrustes Analysis. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2364–2375.

Harkanwar Singh, Prachi Jain, Soumen Chakrabarti, et al. 2021. Multilingual knowledge graph completion with joint relation and entity alignment. *arXiv preprint arXiv:2104.08804*.

Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. 2018. Bootstrapping entity alignment with knowledge graph embedding. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, page 4396–4402.

Zequn Sun, Qingheng Zhang, Wei Hu, Chengming Wang, Muhao Chen, Farahnaz Akrami, and Chengkai Li. 2020. A benchmarking study of embedding-based entity alignment for knowledge graphs. *Proc. VLDB Endow.*, page 2326–2340.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio.

2018. Graph Attention Networks. *International Conference on Learning Representations*.

Ruijie Wang, Zijie Huang, Shengzhong Liu, Huajie Shao, Dongxin Liu, Jinyang Li, Tianshi Wang, Dachun Sun, Shuochao Yao, and Tarek Abdelzaher. 2021. Dydiff-vae: A dynamic variational framework for information diffusion prediction. In *SIGIR'21*.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014a. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, page 1112–1119.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014b. Knowledge graph embedding by translating on hyperplanes. *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 28.

Jingwen Xu, Jing Zhang, Xirui Ke, Yuxiao Dong, Hong Chen, Cuiping Li, and Yongbin Liu. 2021. P-INT: A path-based interaction model for few-shot knowledge graph completion. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 385–394.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How powerful are graph neural networks? In *International Conference on Learning Representations*.

Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, , and Li Deng. 2015a. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the 3th International Conference on Learning Representations (ICLR)*.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015b. Embedding entities and relations for learning and inference in knowledge bases. In *International Conference on Learning Representations (ICLR)*.

Liang Yao, Chengsheng Mao, and Yuan Luo. 2020. Kg-bert: Bert for knowledge graph completion. *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*.

Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. 2021. Qa-gnn: Reasoning with language models and knowledge graphs for question answering. In *North American Chapter of the Association for Computational Linguistics (NAACL)*.

Qingheng Zhang, Zequn Sun, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. 2019. Multi-view knowledge graph embedding for entity alignment. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5429–5435.

Zhao Zhang, Fuzhen Zhuang, Hengshu Zhu, Zhiping Shi, Hui Xiong, and Qing He. 2020. Relational graph neural network with hierarchical attention for knowledge graph completion. *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9612–9619.

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 (IJCAI)*, page 4258–4264.

Qi Zhu, Hao Wei, Bunyamin Sisman, Da Zheng, Christos Faloutsos, Xin Luna Dong, and Jiawei Han. 2020. Collective multi-type entity alignment between knowledge graphs. In *Proceedings of The Web Conference 2020*.

A Data Construction

We introduce the generation process of the multilingual E-commerce KG dataset (E-PKG). E-PKG is a phone-related multilingual product KG across six different languages: English (EN), German (DE), French (FR), Japanese (JA), Spanish (ES), Italian (IT). The statistics are shown in Table 5.

	EN	DE	FR	JA	ES	IT
#Triple_between	90,318	65,077	69,451	14,814	23,671	60,998
#Triple_attributes	5,013	7,345	6,017	946	5,396	6,016
#Triple_products	5,220	3,448	4,547	943	1,096	4,278
#Triples	100,531	75,870	80,015	16,703	30,163	71,292
#Aligned Pairs	21,382	24,696	24,812	5,175	20,184	23,827
#Entities	16,544	17,223	17,068	2,642	9,595	15,670
#Relations	21	21	21	21	21	21

Table 5: Statistics of E-PKG.

Specifically, each KG consists of two types of entities, which are products such as iPhone 12 and attributes such as style and brand. There are three types of triples grouped by their relation types: 1.) The triples that describe relations between a product and an attribute (Triple_between), such as product-belong-to-brand; 2.) The triples that denote relations between a product and a product, such as product-co-buy-with-product (Triple_products); 3.) The triples that refer to relations between an attribute and an attribute, such as manufacturer-has-brand (Triple_attributes). All relations are described in English and entities are in their own languages. The entity type distributions and seed alignment pairs distributions are illustrated in Figure 5 and Figure 6, respectively.
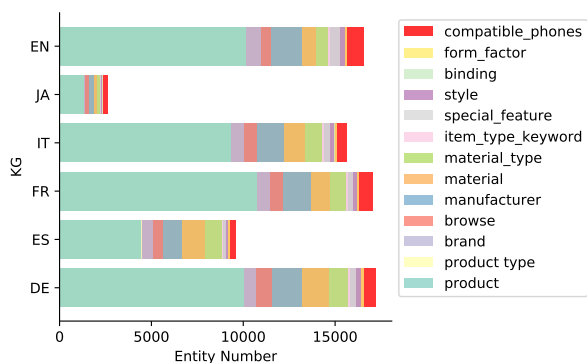

Figure 5: Entity distribution for E-PKG.

B Implementation Details

We use Adam (Kingma and Ba, 2014) as the optimizer to train our model and use TransE (Bordes et al., 2013) as the KG decoder whose margin γ is set to be 0.3. For the two GNN encoders $g^k(\cdot)$ and $g^a(\cdot)$, we set the latent dimension as 256 with
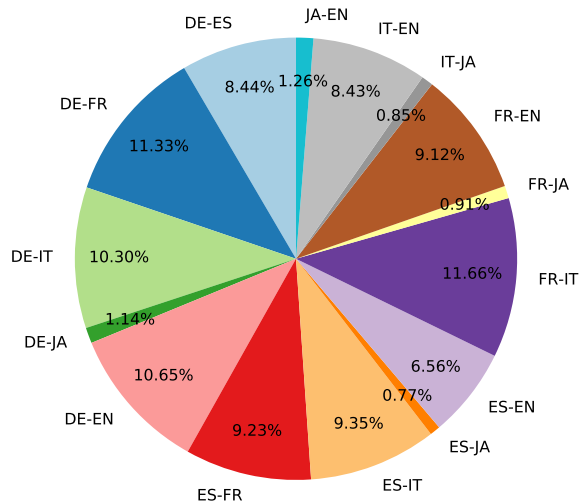

Figure 6: Seed alignment distribution for E-PKG

2 layers, and the dimensions of entity and relation embeddings are also set as 256. We use batch size of 512 and learning rate $lr = 0.005$ during training. The detailed training procedure is illustrated in Algo 1. Instead of directly opmizing $\mathcal{J}$ as in Eqn 4, we alternately update $\mathcal{J}_K$ and $\mathcal{J}_A$ with different learning rate. Specifically, in our implementation, we optimize with $\theta_{new} \leftarrow \theta_{old} - \eta \nabla \mathcal{J}_K$, $\theta_{new} \leftarrow \theta_{old} - (\lambda\eta) \nabla \mathcal{J}_A$ in consecutive steps within one epoch, where θ_{new} denotes our model parameters and ∇ is the training step.

Algorithm 1: SS-AGA training procedure.

Input: KGs $G_1, G_2 \cdots G_M$; Seed
Alignment $\Gamma_{G_i \leftrightarrow G_j} (1 \le i < j < M)$.
Output: Model parameters θ.

1 **while** *model not converged* **do**
2   *//For the masked alignment pairs*:
3   Optimize with the masked recover loss in Eqn 3:
4     $\theta_{new} \leftarrow \theta_{old} - (\lambda\eta) \nabla \mathcal{J}_A$
5   *//For new pair generation*:
6   Propose new pairs with all alignment info using CSLS (Conneau et al., 2018)
7   *//For KG Completion*:
8   Optimize with the KG completion loss in Eqn 2:
9     $\theta_{new} \leftarrow \theta_{old} - \eta \nabla \mathcal{J}_K$
10 **end**