A Fused Gromov-Wasserstein Framework for Unsupervised Knowledge Graph Entity Alignment

Jianheng Tang[1,2*], **Kangfei Zhao**[3,4], **Jia Li**[1†]

[1]Hong Kong University of Science and Technology (Guangzhou) [2]Hong Kong University of Science and Technology, [3]Tencent AI Lab, [4]Beijing Institute of Technology
sqrt3tjh@gmail.com, zkf1105@gmail.com, jialee@ust.hk

Abstract

Entity alignment is the task of identifying corresponding entities across different knowledge graphs (KGs). Although recent embedding-based entity alignment methods have shown significant advancements, they still struggle to fully utilize KG structural information. In this paper, we introduce FGWEA, an unsupervised entity alignment framework that leverages the Fused Gromov-Wasserstein (FGW) distance, allowing for a comprehensive comparison of entity semantics and KG structures within a joint optimization framework. To address the computational challenges associated with optimizing FGW, we devise a three-stage progressive optimization algorithm. It starts with a basic semantic embedding matching, proceeds to approximate cross-KG structural and relational similarity matching based on iterative updates of high-confidence entity links, and ultimately culminates in a global structural comparison between KGs. We perform extensive experiments on four entity alignment datasets covering 14 distinct KGs across five languages. Without any supervision or hyper-parameter tuning, FGWEA surpasses 21 competitive baselines, including cutting-edge supervised entity alignment methods. Our code is available at https://github.com/squareRoot3/FusedGW-Entity-Alignment.

1 Introduction

Knowledge Graph (KG) is one of structured data representations that characterizes real-world concepts (also known as entities) with their relationships and attributes. Recent years have witnessed the proliferation of KGs in various areas, ranging from the general ones such as DBpedia (Auer et al., 2007) and ConceptNet (Speer et al., 2017), to those in specific domains such as healthcare (Rotmensch et al., 2017), education (Chen et al., 2018), and e-commerce (Dong, 2018). As the information
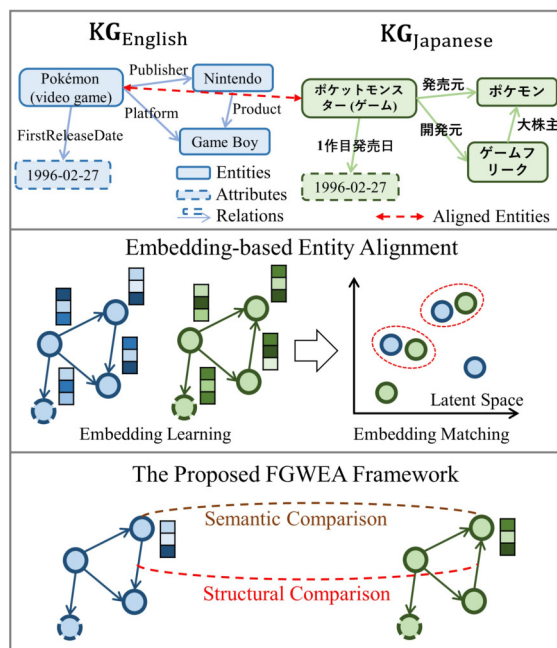

Figure 1: Top: A toy example of cross-lingual entity alignment. Middle and bottom: Comparison between embedding-based EA and our proposed FGWEA.

contained in each individual KG is limited and biased, *entity alignment* (EA) is proposed for linking equivalent entities across two KGs from different sources or languages, and integrating them into a new holistic-view KG. EA task has received a lot of attentions in the computational linguistics community, due to its ability to improve the completeness and fairness of KGs, and enhance a wide range of knowledge-driven downstream applications like question-answering (Saxena et al., 2020; Chen et al., 2021) and dialogue systems (Liu et al., 2021; Xu et al., 2019c). Figure 1 illustrates a toy example of cross-lingual EA between an English KG and a Japanese KG. The main challenge of this task is to leverage the variety of information in KG, such as entity semantics and relations.

In the deep learning era, embedding-based approaches have become the mainstream for ad-

---

dressing the EA task, which primarily follows the "embedding-learning-and-matching" paradigm. As shown in the middle of Figure 1, the embedding module encodes entities from two KGs into a shared latent space. The matching module then infers equivalent entities from the embeddings. The basic principle behind embedding-based EA is that equivalent entities in different KGs share similar neighborhood information. Graph neural networks (Chang et al., 2023; Tang et al., 2022) have been widely adopted as KG encoders, which are usually trained by margin-based losses that encourage equivalent entities to have similar embeddings.

However, the design of the matching module has been overlooked in embedding-based EA. Many existing methods use a greedy strategy that matches entity embeddings to their closest counterparts in another KG, which only relies on the embedding module to incorporate structural information. Unfortunately, even the most powerful KG embedding models and graph neural networks fail to fully preserve structural information. Although some recent methods have attempted to improve the matching module by treating it as a global assignment problem (Mao et al., 2021) or an optimal transport problem (Luo and Yu, 2022), they still fall into the scope of embedding alignment and have limitations in utilizing KG structural information.

To overcome the above issue, we propose FGWEA, an unsupervised EA framework based on the Fused Gromov-Wasserstein (FGW) distance (Titouan et al., 2019), which fuses entity embedding alignment (via the Wasserstein distance) and KG structure alignment (via the Gromov-Wasserstein distance) into a joint optimization framework. As shown in Figure 1, instead of only comparing entity embeddings as most embedding-based EA methods did in the literature, the proposed FGWEA jointly incorporates both KG semantics and structure information. In fact, FGWEA considers cross-KG structural and relational consistencies in optimization objectives to better exploit structural information, rather than implicitly encoding it into embeddings. Moreover, after shifting the inclusion of structural information to the matching module and relieving the workload of embedding module, FGWEA is more compatible with pre-trained language models, which only acts as a main tool for encoding semantic information.

As directly optimizing FGW leads to inefficiency and inferior performance, FGWEA executes a three-stage progressive optimization algorithm, which begins with a relatively simple semantic comparison and then moves on to a more challenging structural comparison. We further develop a fast approximation algorithm and an iterative multi-view OT alignment module to efficiently compare the various KG information. Experiments on four cross-lingual and cross-source EA datasets demonstrate that FGWEA outperforms 21 existing EA methods, including both supervised and unsupervised state-of-the-art approaches.

2 Preliminaries

2.1 Task Definition

Knowledge Graph (KG). Let $\mathcal{E} = \{e_i\}_{i=1}^{|\mathcal{E}|}$, $\mathcal{R} = \{r_i\}_{i=1}^{|\mathcal{R}|}$, $\mathcal{A} = \{a_i\}_{i=1}^{|\mathcal{A}|}$, $\mathcal{L} = \{l_i\}_{i=1}^{|\mathcal{L}|}$ be the set of entities, relations, attributes and literals, respectively. Following Qi et al. (2021), a KG contains a set of relation triples $\mathcal{T}_r = \{(e_i, r_j, e_k)\}$ and attribute triples $\mathcal{T}_a = \{(e_i, a_j, l_k)\}$, denoted as $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{A}, \mathcal{L}, \mathcal{T}_r, \mathcal{T}_a)$. Instances of both types of triples are ⟨*Pokémon, Publisher, Nintendo*⟩ and ⟨*Pokémon, FirstReleaseDate, 1996-02-27*⟩ in Figure 1. While attribute triples are an essential component in KG, some EA datasets simplify them by only considering the relation triples, i.e., $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}_r)$. Besides, we denote the adjacency matrix of $\mathcal{G}$ as A, where $A_{ij} = 1$ if e_i and e_j connected by at least one relation, and 0 otherwise.

Entity Alignment (EA). Given two KGs $\mathcal{G}$ and $\mathcal{G}'$, the EA task is to discover the set of equivalent entity pairs between $\mathcal{G}$ and $\mathcal{G}'$, denoted as $\mathcal{M} = \{(e, e')|e \equiv e', e \in \mathcal{E}, e' \in \mathcal{E}'\}$, where $e \equiv e'$ means an equivalence relation between e and e'. In the unsupervised setting, the EA model predicts $\mathcal{M}$ without observing any pre-aligned entities.

2.2 Optimal Transport (OT)

The core concept of OT is to find a transportation plan (i.e., the coupling matrix) between two distributions that minimize the overall transportation cost. Let $|\mathcal{E}| = m$ and $|\mathcal{E}'| = n$; we denote μ and ν as two discrete distributions on $\mathcal{E}$ and $\mathcal{E}'$, respectively. For simplicity, we assume that μ and ν follow the uniform distribution. That is, $\mu = \frac{1}{m} \sum_{i=1}^{m} \delta_{e_i}$ and $\nu = \frac{1}{n} \sum_{j=1}^{n} \delta_{e'_j}$, where δ_{e_i} and $\delta_{e'_j}$ are the Dirac measure in e_i and e'_j, respectively. We use $\Pi(\mu, \nu)$ to denote the set of all the joint distributions with marginals μ and ν:

$$\Pi(\mu, \nu) = \{\pi \geq 0 : \pi 1_m = \mu, \pi^T 1_n = \nu\}, \quad (1)$$

3321

where π_{ij} signifies the amount of mass transferred from e_i in $\mathcal{G}$ to e'_j in $\mathcal{G}'$, 1_m denotes an m-dimensional all-one vector, and $\pi 1_m$ is the sum of each row in π. The coupling matrix π describes a probabilistic matching of entities between two KGs. A larger value of π_{ij} indicates e_i and e'_j are more likely to be aligned. It is worth noting that when $m = n$ and μ, ν follow a uniform distribution, (1) corresponds to the "assignment polytope", whose vertices correspond to the permutation matrices.

Wasserstein Distance (WD).   WD is used for directly comparing two distributions, such as two sets of entity embeddings. The Wasserstein distance between μ and ν is defined as:

$$\text{WD}(C, \mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \sum_{i,j} C_{ij} \pi_{ij}, \quad (2)$$

where C_{ij} represents the transportation cost between e_i and e'_j, e.g., the cosine distance between entity embeddings. We denote the objective in WD as $f_{\text{WD}}(C, \pi) = \sum_{i,j} C_{ij} \pi_{ij} := \langle C, \pi \rangle$.

2.3   Gromov-Wasserstein Distance

The Gromov-Wasserstein Distance (GWD) (Peyré et al., 2016) is an extension of the classic OT problem, enabling the alignment of two graphs by solely comparing structures within each graph. Consider A and A' are adjacency matrices of $\mathcal{G}$ and $\mathcal{G}'$, GWD is defined as:

$$\text{GWD}(A, A') = \min_{\pi \in \Pi} \sum_{i,j,k,l} |A_{ij} - A'_{kl}|^2 \pi_{ik} \pi_{jl}$$

$$= \min_{\pi \in \Pi} f_{\text{GWD}}(A, A', \pi). \quad (3)$$

In this equation, if π_{ik} and π_{jl} have large values, it suggests that (e_i, e'_k) and (e_j, e'_l) are likely to be two entity pairs. Consequently, the corresponding intra-KG pairs (e_i, e_j) and $(e'k, e'l)$ should exhibit similar structures, i.e., $|Aij - A'kl| \rightarrow 0$. If two KGs possess identical structures and π represents the perfect mapping between them, then $\text{GWD}(A, A') = 0$.

Fused Gromov-Wasserstein Distance (FGW). Neither WD nor GWD is able to depict the full landscape of KGs. Therefore, FGW (Titouan et al., 2019) is introduced, whose objective is a linear combination of f_{WD} and f_{GWD}:

$$f_{\text{FGW}} = \alpha f_{\text{WD}}(C, \pi) + (1 - \alpha) f_{\text{GWD}}(A, A', \pi), \quad (4)$$

where $\alpha \in [0, 1]$ is a trade-off parameter.

However, several challenges emerge when applying FGW to the EA task. First, GWD assumes that both A and A' are homogeneous graphs, whereas KGs are heterogeneous graphs containing relational information. Second, KG entities possess various forms of side information, such as names and attributes, complicating the accurate measurement of entity similarity and the computation of the cost matrix C in WD. Third, although Titouan et al. (2019) invokes the Frank-Wolfe method for optimizing FGW, its effectiveness has only been confirmed on small graphs with hundreds of nodes. We observe that directly applying this method to large-scale sparse KGs results in unstable performance and reduced efficiency. To tackle these issues, we propose a novel EA approach based on FGW in the following section.

3   The Proposed Method

We present an unsupervised EA framework, FGWEA, that performs entity matching based on the FGW distance. As shown in Figure 2, it comprises a semantic embedding module and a three-stage entity matching module. To address the aforementioned challenges, we propose a three-step progressive optimization algorithm. First, FGWEA performs the straightforward semantic embedding matching to obtain high-confidence aligned entity pairs as anchors (Section 3.1). Building on these anchors, FGWEA employs a fast approximation of GWD to compute cross-KG structural and relational similarities, which are then used for iterative multi-view OT alignment (Section 3.2). Upon achieving a better initial point for the coupling matrix, FGWEA proceeds to comparing the global structures of KGs by optimizing GWD, the most challenging component in FGW (Section 3.3).

3.1   Semantic Embedding and Comparison

The embedding module in FGWEA is responsible for encoding entity semantic information, primarily derived from entity names and attributes. Given the remarkable success of pre-trained language models, we employ LaBSE (Feng et al., 2022) for embedding multilingual KGs and SimCSE (Gao et al., 2021) for embedding monolingual KGs, both of which are variations of BERT-base (Devlin et al., 2019) and are tailored for semantic similarity modeling. It is important to note that our embedding module does not necessitate fine-tuning, and any pre-trained sentence Transformers can be used as
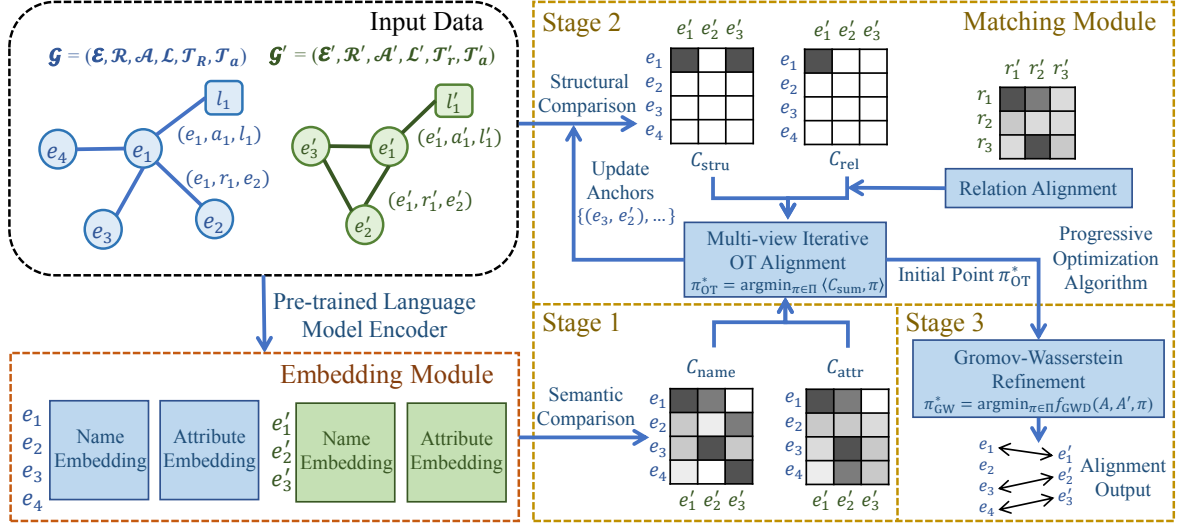
Figure 2: Framework Overview. The embedding module calculates name and attribute embeddings for each entity in KGs. The matching module consists of three stages: semantic comparison (Section 3.1), multi-view iterative OT alignment (Section 3.2), and Gromov-Wasserstein refinement (Section 3.3).

a substitute, such as those presented by (Reimers and Gurevych, 2019).

We represent the entity name of e_i as ne_i and concatenated all attribute triples related to e_i into a single string denoted as ae_i (in the form of $a_1 l_1 a_2 l_2 \cdots$). The order of the triples depends on the attribute frequency in the KG. Let $\text{enc}(\cdot)$ be the encoder function, we calculate the name similarity-based cost matrix C_{name} and attribute similarity-based cost matrix C_{attr} between two KGs as follows:

$$C_{\text{name}} = 1 - \cos(\text{enc}(ne_i), \text{enc}(ne_j')),$$
$$C_{\text{attr}} = 1 - \cos(\text{enc}(ae_i), \text{enc}(ae_j')). \quad (5)$$

In the first matching stage, we use the sum of two semantic similarity matrices as the cost in WD and calculate the initial coupling matrix π^0 by:

$$\pi^0 = \arg\min_{\pi \in \Pi} \langle C_{\text{name}} + C_{\text{attr}}, \pi \rangle. \quad (6)$$

Specifically, we use the Sinkhorn algorithm (Cuturi, 2013) to tackle this problem, and collect high confidence entity pairs in π^0 as *anchors* to facilitate the subsequent matching process. Let $\mathcal{M}_a^0$ denote the initial anchor set and $c = 1/\max(m, n)$ be the maximum potential value of π. We have $\mathcal{M}_a^0 = \{(e_i, e_j') | \pi_{ij}^0 > c - \epsilon\}$, where ϵ is a small threshold satisfying $\epsilon < c/2$ to ensure one-to-one alignment.

3.2 Approximated GWD for Multi-view Iterative OT Alignment

In the second stage, our goal is to incorporate KG structural and relational information into the matching process. Instead of directly optimizing the GWD or FGW objective, we develop an approximate alternative for the sake of efficiency.

Relation-aware GWD We extebd the structural comparison $|A_{ij} - A_{kl}|^2$ in (3) to the relation comparison:

$$\sum_{i,j,k,l} (1 - \text{sim}(r_{i,j}, r_{k,l}')) \pi_{ik} \pi_{jl}, \quad (7)$$

where $r_{i,j}$ represents the relation between e_i and e_j. The relation similarity $\text{sim}(r_{i,j}, r_{k,l}') = 1$ if $A_{ij} = A_{kl}' = 1$ and $r_{i,j} \equiv r_{k,l}'$, otherwise 0. As the relation set in different KGs is also unaligned, we align these relations based on relation name similarity, using the same process in Section 3.1.

Approximation However, optimizing (7) is even more challenging than optimizing GWD. We simplify it by approximating π_{ik} in equation (7) with a sparse coupling matrix $\hat{\pi}$ based on the anchor set $\mathcal{M}_a^0$. Specifically, $\hat{\pi}_{ik} = c$ if $(e_i, e_k') \in \mathcal{M}_a$, and $\hat{\pi}_{ik} = 0$ otherwise. Note that when $\mathcal{M}_a$ is closer to the ground truth alignment, the approximation of GWD is more accurate. Afterward, (7) is converted to a WD objective:

$$\sum_{j,l} (1 - c \sum_{(e_i, e_k') \in \mathcal{M}_a} \text{sim}(r_{i,j}, r_{k,l}')) \pi_{jl} = \langle 1 - cS^{\text{rel}}, \pi \rangle,$$
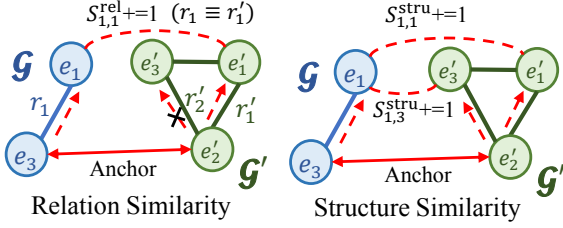$$(8)$$

3323

Figure 3: Illustration of how anchor links contribute to S^{stru} and S^{rel}.

Name	Lang.	$|\mathcal{E}|$	$|\mathcal{R}|$	$|\mathcal{T}_r|$
DBP15K$_{\text{ZH\_EN}}$	ZH	19,388	1,701	70,414
	EN	19,572	1,323	95,142
DBP15K$_{\text{JA\_EN}}$	JA	19,814	1,299	77,214
	EN	19,780	1,153	93,484
DBP15K$_{\text{FR\_EN}}$	FR	19,661	903	105,998
	EN	19,993	1,208	115,722
SRPRS$_{\text{EN\_FR}}$	EN	15,000	177	33,532
	FR	15,000	221	36,508
SRPRS$_{\text{EN\_DE}}$	EN	15,000	120	37,377
	DE	15,000	222	38,363
D-W-15K-V2	EN	15,000	167	73,983
	EN	15,000	121	83,365
Med-BBK-9K	ZH	9,162	32	158,357
	ZH	9,162	20	50,307

Table 1: Dataset statistics. $|\mathcal{E}|$, $|\mathcal{R}|$ and $|\mathcal{T}_r|$ represent the number of entities, relation types and relation triplets in each KG, respectively.

where $S^{\text{rel}}_{j,l}$ reflects the relation similarity between e_j and e'_l. It is calculated by counting the number of anchors $(e_i, e'_k) \in \mathcal{M}_a$ in which e_i is a neighbor of e_j, e'_k is a neighbor of e'_l, and $r_{i,j} \equiv r'_{k,l}$. S^{rel} can be efficiently computed by iterating through all anchor pairs and comparing their corresponding neighbor node pairs. Figure 3 illustrates the computation process. If (e_3, e'_2) is an anchor and e_1, e'_1 are corresponding neighbors with equivalent relations $r_1 \equiv r'_1$, then (e_3, e'_2) contributes to the relation similarity $S^{\text{rel}}_{1,1}$. In the right of Figure 3, we repeat this process to calculate the relation-agnostic structure similarity matrix S^{stru}, which can be regarded as an approximation of GWD that only compares between anchor entity pairs and other pairs.

Multi-view OT Alignment To perform a joint comparison of structures and semantics between KGs, we rescale $1 - cS^{\text{rel}}$ and $1 - cS^{\text{stru}}$ to a range of [0,1] and obtain the corresponding cost matrices C_{rel} and C_{stru}. The multi-view OT combines all four cost matrices that represent discrepancies between KGs from different perspectives:

$$\pi_1^* = \arg\min_{\pi \in \Pi} \langle C_{\text{sum}}, \pi \rangle, \tag{9}$$

where $C_{\text{sum}} = C_{\text{stru}} + C_{\text{rel}} + C_{\text{name}} + C_{\text{attr}}$. We derive π_1^* and update the anchor set $\mathcal{M}_a^1$ with the same process in Section 3.1. With $\mathcal{M}_a^1$, we can adjust C_{stru} and C_{rel} accordingly, resulting in a new OT problem and a new coupling matrix π_2^*. We repeat this process for a fixed number of epochs in order to gradually improve the completeness of the anchor set. The final coupling matrix in the second stage is denoted as π_{OT}^*.

3.3 Gromov-Wasserstein Refinement

Although the approximated GWD has the advantages mentioned above, the reliance on the anchor set may lead to accumulated error. Therefore, in the final matching stage, we consider the following

FGW objective:

$$f_{\text{FGW}} = \alpha f_{\text{WD}}(C_{\text{sum}}, \pi) + (1-\alpha) f_{\text{GWD}}(A, A', \pi). \tag{10}$$

Due to the difficulty in optimizing f_{FGW} discussed in 2.3, we only consider optimizing the second term f_{GWD} to improve stability. We employ the Bregman Proximal Gradient algorithm, introduced by Xu et al. (2019a) and shown to have a local linear convergence guarantee by Li et al. (2022). For the k-th iteration, BPG takes the form

$$\pi^{k+1} = \arg\min_{\pi \in \Pi}\{\nabla_\pi f_{\text{GWD}}(\pi^k)^T \pi + \frac{1}{\beta} \mathbf{KL}(\pi||\pi^k)\}, \tag{11}$$

where β is the step size and $\mathbf{KL}(\cdot||\cdot)$ is the Kullback-Leibler divergence. As such, the π-update is identical to the entropic OT problem, and we can invoke the Sinkhorn algorithm to tackle it.

Our GW refinement process incorporates two improvements to BPG. First, we use π_{OT}^* as the initial point rather than the uniform distribution, significantly facilitating the optimization process. Second, we employ the relative change of f_{FGW} instead of f_{GWD} as the optimization stopping criterion, which more accurately reflects the discrepancy between KGs. In the following section, we will test the effectiveness of our proposed FGWEA with the progressive optimization algorithm.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate the proposed FGWEA on four frequently used EA datasets, including two

Model	Configurations				DBP15K$_{ZH\_EN}$			DBP15K$_{JA\_EN}$			DBP15K$_{FR\_EN}$		
	Name	Attr.	Trans.	Sup.	Hit1	Hit10	MRR	Hit1	Hit10	MRR	Hit1	Hit10	MRR
JAPE	✗	✓	✗	30%	0.412	0.745	0.490	0.363	0.685	0.476	0.324	0.667	0.430
GCNAlign	✗	✓	✗	30%	0.413	0.744	0.549	0.399	0.745	0.546	0.373	0.745	0.532
FGWEA	✗	✓	✗	0%	**0.929**	**0.978**	**0.948**	**0.922**	**0.974**	**0.942**	**0.967**	**0.994**	**0.978**
GMatch	✓	✗	✗	30%	0.679	0.785	-	0.740	0.872	-	0.894	0.952	-
SelfKG	✓	✗	✗	0%	0.745	0.866	-	0.816	0.913	-	0.957	0.992	-
FGWEA	✓	✗	✗	0%	**0.926**	**0.967**	**0.942**	**0.954**	**0.981**	**0.964**	**0.996**	**0.999**	**0.997**
RDGCN	✓	✗	✓	30%	0.708	0.846	0.746	0.767	0.895	0.812	0.886	0.957	0.911
DATTI	✓	✗	✓	0%	0.890	0.958	-	0.921	0.971	-	0.979	0.990	-
SEU	✓	✗	✓	0%	0.900	0.965	0.924	0.956	0.991	0.969	0.988	0.999	0.992
EASY	✓	✗	✓	0%	0.898	0.979	0.930	0.943	0.990	0.960	0.980	0.998	0.990
CPL-OT	✓	✗	✓	0%	0.927	0.964	0.940	0.956	0.983	0.970	0.990	0.994	0.990
UED	✓	✗	✓	0%	0.915	-	-	0.941	-	-	0.984	-	-
LightEA	✓	✗	✓	0%	0.952	**0.984**	0.964	0.981	**0.997**	**0.987**	**0.995**	0.998	**0.996**
FGWEA	✓	✗	✓	0%	**0.959**	0.983	**0.969**	**0.982**	0.995	**0.987**	0.994	**0.999**	**0.996**
AttrGNN	✓	✓	✗	30%	0.796	0.929	0.845	0.783	0.921	0.834	0.919	0.978	0.910
BERT-INT*	✓	✓	✗	30%	0.968	0.990	0.977	0.964	0.991	0.975	0.992	0.998	0.995
ICLEA	✓	✓	✗	0%	0.884	0.972	-	0.924	0.978	-	0.991	0.999	-
FGWEA	✓	✓	✗	0%	**0.976**	**0.994**	**0.983**	**0.978**	**0.992**	**0.988**	**0.997**	**0.999**	**0.998**
MCLEA*	✓	✓	✓	30%	0.972	0.996	0.981	0.986	**0.999**	0.991	0.997	**1.000**	0.998
FGWEA	✓	✓	✓	0%	**0.987**	**0.997**	**0.991**	**0.991**	0.998	**0.994**	**0.998**	**1.000**	**0.999**

Table 2: Evaluation Results of all compared EA methods on DBP15K under different configurations. Name, Attr., and Trans. represent the usage of entity name, attributes, and translation information, respectively. Sup. indicates the ratio of entity links for supervision. Methods marked with * use additional information not in DBP15K.

multilingual datasets DBP15K (Sun et al., 2017) and SRPRS (Guo et al., 2019), and two monolingual multi-source datasets D-W-15K-V2 (Sun et al., 2020) and Med-BBK-9K (Qi et al., 2021). Statistics of these datasets are in Table 1. For a detailed description, please refer to Appendix A.

Baselines. A total of 21 EA methods are selected as baselines for performance comparison, spanning from supervised to unsupervised, and conventional to state-of-the-art. Detailed descriptions of most baselines can be found in Section 5. For multilingual EA, we compare with the following methods: JAPE (Sun et al., 2017), GCN-Align (Wang et al., 2018), GMatch (Xu et al., 2019b), SelfKG (Liu et al., 2022), RDGCN (Wu et al., 2019), DATTI (Mao et al., 2022a), SEU (Mao et al., 2021), EASY (Ge et al., 2021), CPL-OT (Ding et al., 2022), UED (Luo and Yu, 2022), LightEA (Mao et al., 2022b), AttrGNN (Liu et al., 2020), BERT-INT (Tang et al., 2020), ICLEA (Zeng et al., 2022a), and MCLEA (Lin et al., 2022). For monolingual multi-source EA, we compare FGWEA with MultiKE (Zhang et al., 2019), BootEA (Sun et al., 2018), RSNs (Guo et al., 2019), LogMap (Jiménez-Ruiz and Grau, 2011), PARIS (Suchanek et al., 2011), PARSE

(Qi et al., 2021), and StrMatch, a simple matching method using the string edit distance.

Evaluation Metrics. On DBP15K and SPARS, we use **HitK** and **MRR** to evaluate the performance of all EA methods. **HitK** calculates the percentage of entities in $\mathcal{G}$ whose counterparts in $\mathcal{G}'$ is in the top-K candidates of model output. **MRR** is the mean reciprocal rank. On D-W-15K-V2 and Med-BBK-9K, we adopt another evaluation protocol for a comprehensive evaluation suggested by Leone et al. (2022). We use the standard classification-based metrics, i.e., precision (**P**), recall (**R**), and F_1 scores between the set of all predicted entity pairs and that of ground truth entity pairs.

Implementation Details. Unlike most neural-based EA methods, the proposed FGWEA requires no hyper-parameter tuning and we use the same hyper-parameters across all datasets. We update 6 epochs for multi-view OT alignment and set the threshold ϵ to 1e-5. In all places where the Sinkhorn algorithm is used, we set the entropic regularization weight η to 0.1 and the number of iterations to 10. We set α in the FGW objective (10) to be the average graph density of A and A' to maintain a balance between the magnitude of

Model	SRPRS$_{EN\_FR}$		SRPRS$_{EN\_DE}$	
	Hit1	Hit10	Hit1	Hit10
BERT-INT	0.971	0.975	0.986	0.988
CPL-OT	0.974	0.988	0.974	0.989
EASY*	0.965	0.989	0.974	0.992
SEU*	0.982	0.995	0.983	0.996
LightEA*	0.986	0.994	0.988	0.995
FGWEA	**0.996**	**0.999**	**0.997**	**1.000**

Table 3: Evaluation Results on the SPARS dataset. Methods marked with * used the translated entity name.

Model	D-W-15K-V2			MED-BBK-9K		
	P	R	F_1	P	R	F_1
MultiKE	49.5	49.5	49.5	41.0	41.0	41.0
BootEA	82.1	82.1	82.1	30.7	30.7	30.7
RSNs	72.3	72.3	72.3	19.5	19.5	19.5
StrMatch	60.6	41.9	49.5	54.5	49.5	51.9
LogMap	-	-	-	86.4	44.1	58.4
PARIS	95.0	85.0	89.7	77.9	36.7	49.9
PRASE	94.8	90.0	92.3	83.7	61.9	71.1
FGWEA	**95.2**	**90.3**	**92.7**	**93.9**	**73.2**	**82.3**

Table 4: Results on cross-source EA datasets.

the WD and GWD terms. We set the step size β in BPG to 100 and the maximum iteration number to 2000. The only exception is that we encounter numerical errors on the Med-BBK-9K dataset, and thus decrease β to 50. Our model is implemented on PyTorch. All experiments are performed on a Linux server with an AMD Ryzen9 5950X CPU and an NVIDIA GeForce RTX 3090 GPU.

4.2 Results on Cross-lingual EA Datasets

DBP15K is the most widely-adopted EA dataset. Unfortunately, the experimental configurations of different baselines on this dataset are highly inconsistent, leading to unfair comparison. After a careful study of existing work, we figure out four factors that significantly effect the results: (1) the inclusion of entity names, (2) the utilization of attribute triples, (3) the use of Google translation for non-English entities, and (4) the ratio of entity links for supervision.

Based on factors (1-3), we categorize baselines into five groups and run FGWEA using the configurations for each group. The experimental settings and results of all compared baselines and FGWEA is in Table 2. As observed, FGWEA achieves the best performance in terms of Hit1 and MRR in all five groups. Specifically, the *unsupervised* FGWEA outperforms two state-of-the-art *supervised* EA approaches BERT-INT and MCLEA. SelfKG and ICLEA are two graph neural network-based methods that use the same pre-trained language model named LaBSE to encode semantic information. However, our approach outperforms them by a significant margin, demonstrating its ability of utilizing KG structures. UED and CPL-OT, which are also based on OT for alignment, do not perform as well as FGWEA, suggesting that the FGW distance we introduced is more suitable for this task.

Table 3 reports the results on the SPARS dataset. BERT-INT uses 30% entity links for training and other baselines are unsupervised. While most baselines rely on translated entity names to overcome the language barrier, FGWEA achieves the best performance with untranslated entity names. It surpasses LightEA, the current leading method on this dataset, by reducing the error rate from 1.2% to only 0.3% on SPRPS$_{EN\_DE}$.

4.3 Results on Cross-source EA Datasets

Cross-source EA poses more challenges than EA within the same knowledge source due to the larger discrepancies in schema and topology of KGs from different sources. For example, in D-W-15K-V2, we find the KG from WikiData uses OIDs as entity names. To facilitate semantic comparison in FGWEA, we replace these OIDs with entity attributes that possess linguistic information.

In Table 4, we compare FGWEA with 7 EA methods that were not included in cross-lingual EA evaluation. The results show that FGWEA consistently outperforms all the baselines on two datasets in terms of precision, recall, and F_1 scores. Remarkably, FGWEA outperforms PARSE by 11.9% in terms of F_1, which is the previous best performed method on this dataset. FGWEA also surpasses PARIS, a conventional approach that has shown superior performance to all neural-based EA in a recent study (Leone et al., 2022).

4.4 Ablation Study and Model Efficiency

To validate the effectiveness and efficiency of each component in FGWEA, we compare it with several ablations. First, we remove Gromov-Wasserstein refinement, the third matching stage in FGWEA, and refer to this new version as FGWEA w/o (without) GW. Then, we continue to remove the relational comparison and structural comparison in the

Model	DBP15K$_{ZH\_EN}$				SRPRS$_{EN\_FR}$				D-W-15K-V2				MED-BBK-9K			
	Hit1	Hit10	MRR	Time	Hit1	Hit10	MRR	Time	P	R	F_1	Time	P	R	F_1	Time
FGWEA	**0.987**	**0.997**	**0.991**	254	**0.996**	**0.999**	**0.997**	388	95.2	**90.3**	**92.7**	494	**93.9**	**73.2**	**82.3**	151
- w/o GW	0.975	0.992	0.981	57	0.979	0.989	0.983	62	95.8	84.7	89.9	49	92.4	58.5	71.7	19
- w/o S_{rel}	0.970	0.990	0.977	52	0.976	0.987	0.980	49	**98.7**	82.2	90.0	44	92.6	57.8	71.1	18
- w/o S_{stru}	0.951	0.979	0.962	40	0.966	0.982	0.972	23	97.5	56.6	71.6	25	89.4	45.2	60.1	10
GW-only	0.011	0.026	0.017	555	0.004	0.023	0.011	504	0.3	0.1	0.1	227	0.3	0.1	0.1	470
Emb-Match	0.763	0.861	0.799	1	0.915	0.959	0.931	1	57.3	57.3	57.3	1	51.1	51.1	51.1	1

Table 5: Ablation study of FGWEA. The wall-clock time is measured in seconds.

second matching stage, and obtain FGWEA w/o C_{rel} and C_{stru}, respectively. GWD-only is a baseline that directly optimizes GWD for alignment without using the progressive optimization algorithm in FGWEA. Emb-Match directly matches entities based on entity semantic embeddings.

As shown in Table 5, FGWEA performs the best compared with these variants, which validates the effectiveness of the proposed progressive optimization algorithm. Removing GW refinement in FG-WEA results in a decrease in performance on all datasets and a significant reduction in computational time. Removing either the relational comparison or the structural comparison also leads to a decline in performance, while the time consumption does not change significantly. Besides, directly optimizing GWD between KG structures is ineffective for the EA task, and aligning entity semantic embeddings alone also has poor performance. This highlights the importance of considering structural and semantic information jointly.

Note that Table 5 only calculates the time spent on the matching module. The embedding module takes approximately 5 minutes to run on DBP15K and 3 minutes on other datasets. On average, it takes approximately 10 minutes to run FGWEA on these datasets, which is relatively efficient compared to most embedding-based EA methods.

4.5 Visualization of the FGW Objective

In figure 4, we visualize the objective function in (10) and the corresponding Hit1 score for 400 epochs in GW refinement on DBP15K$_{ZH\_EN}$ without translation and attributes. We find a strong correlation between two curves—the iteration corresponding to the minimum FGW objective value is approximately that to the maximum Hit1 score. This suggests that the FGW objective can be utilized as an unsupervised metric to estimate the alignment performance and to help determine when to stop optimization in GW refinement. In this case,
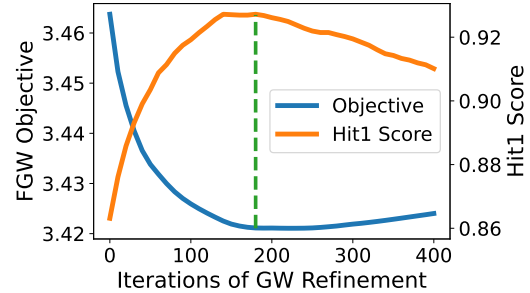

Figure 4: Visualization of the relationship between the objective function and alignment performance (Hit1) of FGWEA in the GW refinement process.

WD increases monotonically, and GWD decreases monotonically in all steps, neither of which are able to indicate model performance. More examples can be found in Appendix B.

5 Related Work

5.1 Unsupervised Entity Alignment

We categorize the existing unsupervised entity alignment methods into three groups:

(1) **Traditional heuristic EA systems**. LogMap (Jiménez-Ruiz and Grau, 2011) and PARIS (Suchanek et al., 2011) are two well-known traditional EA systems that iteratively discover entity links by logical inference, lexical matching, and probabilistic reasoning. PARSE (Qi et al., 2021) is an enhanced version of PARIS which combines probabilistic reasoning and semantic embedding.

(2) **Self-supervised neural EA methods**. SelfKG (Liu et al., 2022) uses the graph neural network to aggregate entity embeddings of one-hot neighbors, and proposes a similarity metric between the entities of two KGs for contrastive learning. ICLEA (Zeng et al., 2022a) conducts bidirectional contrastive learning via building pseudo-aligned entity pairs as pivots for cross-KG interaction.

(3) **Optimization-based non-neural EA methods**. SEU (Mao et al., 2021) transforms the EA problem into assignment problem. LightEA (Mao et al.,

2022b) is a non-neural framework which reinvents the label propagation algorithm to effectively run on KGs. Our proposed FGWEA also belongs to this group.

5.2 Optimal Transport for Entity Alignment

There have been a few approaches that use OT to improve the EA performance. OTEA (Pei et al., 2019) is a supervised method that adopts the basic TransE (Bordes et al., 2013) for KG embedding and proposes the group-level loss for embedding training based on OT theory. (Luo and Yu, 2022) develops a modified OT problem for global EA and dangling entity detection. CPL-OT (Ding et al., 2022) employs a graph convolutional network to learn entity embeddings, which are then utilized to determine the transportation cost in OT and resolve alignment conflicts. SLOTAlign (Tang et al., 2023) is an unsupervised graph alignment framework that jointly performs structure learning and optimal transport alignment. Compared with these methods, FGWEA takes the first step towards introducing the Fused Gromov-Wasserstein distance to EA, which enables better utilization of the structural information in KGs.

6 Conclusion

In this paper, we propose an unsupervised entity alignment framework named FGWEA. Instead of following the "embedding-learning-and-matching" paradigm, we invoke the Fused Gromov-Wasserstein distance to realize a more explicit and comprehensive comparison of structural and semantic information between knowledge graphs. To realize the benefits of FGWEA, we present a three-stage progressive optimization algorithm to address the challenge of optimizing the FGW objective. Experimental results show that FGWEA outperforms both supervised and unsupervised state-of-the-art entity alignment methods.

Acknowledgement

Limitations

Although the proposed FGWEA framework demonstrated the superior performance on multiple public EA datasets, there are still some limitations that require further research.

Scalability. In this paper, we have successfully extended FGW to KGs with tens of thousands of entities, which is the common size of domain-specific KGs. However, real-world general-domain KGs can be much larger and contain millions of entities. The most time-consuming step in FGWEA, the Gromov-Wasserstein refinement, has quadratic time complexity $O(|\mathcal{E}||\mathcal{T}_r'| + |\mathcal{E}'||\mathcal{T}_r|)$ and thus cannot be directly applied to million-scale KGs. There are three ways to further scale up FGWEA. First, we can remove the most time consuming step, GW refinement, while FGWEA still has competitive performance in Table 5. Second, we can use recent divide-and-conquer methods (Xin et al., 2022; Zeng et al., 2022b; Li et al., 2021) to divide large scale KGs into smaller subgraph pairs, and then apply alignment methods for each subgraph pair. Third, the coupling matrix π can be restricted to a sparse matrix which only considers top-k candidates for each entity, and the computation can be accelerated by mask OT (Gasteiger et al., 2021) or sparse Sinkhorn iteration (Mao et al., 2022b).

Dealing with dangling cases. FGWEA supposes all entities have equal probabilities to be matched in the beginning by using the uniform distribution. Therefore, it has limited ability to handle dangling entities whose counterparts are unavailable in the other KG (Sun et al., 2021). To avoid this limitation, we can invoke unbalanced OT (Chizat et al., 2018) or unbalanced GWD (Sejourne et al., 2021), which relax the assumption of equal probabilities for all entities.

References

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735. Springer.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Heng Chang, Jie Cai, and Jia Li. 2023. Knowledge graph completion with counterfactual augmentation. In *Proceedings of the ACM Web Conference 2023*, pages 2611–2620.

Jiaqi Chen, Jianheng Tang, Jinghui Qin, Xiaodan Liang, Lingbo Liu, Eric Xing, and Liang Lin. 2021. GeoQA:

A geometric question answering benchmark towards multimodal numerical reasoning. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 513–523, Online. Association for Computational Linguistics.

Penghe Chen, Yu Lu, Vincent W Zheng, Xiyang Chen, and Boda Yang. 2018. Knowedu: A system to construct knowledge graph for education. *Ieee Access*, 6:31553–31563.

Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. 2018. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 87(314):2563–2609.

Marco Cuturi. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26.

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, Minneapolis, Minnesota. Association for Computational Linguistics.

Qijie Ding, Daokun Zhang, and Jie Yin. 2022. Conflict-aware pseudo labeling via optimal transport for entity alignment. *arXiv preprint arXiv:2209.01847*.

Xin Luna Dong. 2018. Challenges and innovations in building a product knowledge graph. In *Proceedings of the 24th ACM SIGKDD International conference on knowledge discovery & data mining*, pages 2869–2869.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Johannes Gasteiger, Marten Lienen, and Stephan Günnemann. 2021. Scalable optimal transport in high dimensions for graph distances, embedding alignment, and more. In *International Conference on Machine Learning*, pages 5616–5627. PMLR.

Congcong Ge, Xiaoze Liu, Lu Chen, Baihua Zheng, and Yunjun Gao. 2021. Make it easy: An effective end-to-end entity alignment framework. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 777–786. ACM.

Lingbing Guo, Zequn Sun, and Wei Hu. 2019. Learning to exploit long-term relational dependencies in knowledge graphs. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2505–2514. PMLR.

Ernesto Jiménez-Ruiz and Bernardo Cuenca Grau. 2011. Logmap: Logic-based and scalable ontology matching. In *The Semantic Web - ISWC 2011 - 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I*, volume 7031 of *Lecture Notes in Computer Science*, pages 273–288. Springer.

Manuel Leone, Stefano Huber, Akhil Arora, Alberto García-Durán, and Robert West. 2022. A critical re-evaluation of neural methods for entity alignment. *Proceedings of the VLDB Endowment*.

Jia Li, Mengzhou Liu, Honglei Zhang, Pengyun Wang, Yong Wen, Lujia Pan, and Hong Cheng. 2021. Maskgvae: Blind denoising graphs via partition. In *Proceedings of the Web Conference 2021*, pages 3688–3698.

Jiajin Li, Jianheng Tang, Lemin Kong, Huikang Liu, Jia Li, Anthony Man-Cho So, and Jose Blanchet. 2022. Fast and provably convergent algorithms for gromov-wasserstein in graph learning. *arXiv preprint arXiv:2205.08115*.

Zhenxi Lin, Ziheng Zhang, Meng Wang, Yinghui Shi, Xian Wu, and Yefeng Zheng. 2022. Multi-modal contrastive representation learning for entity alignment. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2572–2584, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Wenge Liu, Jianheng Tang, Xiaodan Liang, and Qingling Cai. 2021. Heterogeneous graph reasoning for knowledge-grounded medical dialogue system. *Neurocomputing*, 442:260–268.

Xiao Liu, Haoyun Hong, Xinghao Wang, Zeyi Chen, Evgeny Kharlamov, Yuxiao Dong, and Jie Tang. 2022. Selfkg: Self-supervised entity alignment in knowledge graphs. In *Proceedings of the ACM Web Conference 2022*, pages 860–870.

Zhiyuan Liu, Yixin Cao, Liangming Pan, Juanzi Li, Zhiyuan Liu, and Tat-Seng Chua. 2020. Exploring and evaluating attributes, values, and structures for entity alignment. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6355–6364, Online. Association for Computational Linguistics.

Shengxuan Luo and Sheng Yu. 2022. An accurate unsupervised method for joint entity alignment and dangling entity detection. In *Findings of the Association*

for Computational Linguistics: ACL 2022, pages 2330–2339, Dublin, Ireland. Association for Computational Linguistics.

Xin Mao, Meirong Ma, Hao Yuan, Jianchao Zhu, Zongyu Wang, Rui Xie, Wei Wu, and Man Lan. 2022a. An effective and efficient entity alignment decoding algorithm via third-order tensor isomorphism. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 5888–5898.

Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. 2021. From alignment to assignment: Frustratingly simple unsupervised entity alignment. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 2843–2853.

Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. 2022b. Lightea: A scalable, robust, and interpretable entity alignment framework via three-view label propagation. In EMNLP.

Shichao Pei, Lu Yu, and Xiangliang Zhang. 2019. Improving cross-lingual entity alignment via optimal transport. In IJCAI.

Gabriel Peyré, Marco Cuturi, and Justin Solomon. 2016. Gromov-wasserstein averaging of kernel and distance matrices. In International Conference on Machine Learning, pages 2664–2672. PMLR.

Zhiyuan Qi, Ziheng Zhang, Jiaoyan Chen, Xi Chen, Yuejia Xiang, Ningyu Zhang, and Yefeng Zheng. 2021. Unsupervised knowledge graph alignment by probabilistic reasoning and semantic embedding. arXiv preprint arXiv:2105.05596.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics.

Maya Rotmensch, Yoni Halpern, Abdulhakim Tlimat, Steven Horng, and David Sontag. 2017. Learning a health knowledge graph from electronic medical records. Scientific reports, 7(1):1–11.

Apoorv Saxena, Aditay Tripathi, and Partha Talukdar. 2020. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 4498–4507, Online. Association for Computational Linguistics.

Thibault Sejourne, Francois-Xavier Vialard, and Gabriel Peyré. 2021. The unbalanced gromov wasserstein distance: Conic formulation and relaxation. In Advances in Neural Information Processing Systems, volume 34, pages 8766–8779. Curran Associates, Inc.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In AAAI.

F. M. Suchanek, S. Abiteboul, and P. Senellart. 2011. Paris: Probabilistic alignment of relations, instances, and schema. Proceedings of the Vldb Endowment, 5(3):157–168.

Zequn Sun, Muhao Chen, and Wei Hu. 2021. Knowing the no-match: Entity alignment with dangling cases. 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 3582–3593, Online. Association for Computational Linguistics.

Zequn Sun, Wei Hu, and Chengkai Li. 2017. Cross-lingual entity alignment via joint attribute-preserving embedding. In ISWC, pages 628–644.

Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. 2018. Bootstrapping entity alignment with knowledge graph embedding. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden, pages 4396–4402. ijcai.org.

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. Proceedings of the VLDB Endowment, 13(11):2326–2340.

Jianheng Tang, Jiajin Li, Ziqi Gao, and Jia Li. 2022. Rethinking graph neural networks for anomaly detection. In International Conference on Machine Learning.

Jianheng Tang, Weiqi Zhang, Jiajin Li, Kangfei Zhao, Fugee Tsung, and Jia Li. 2023. Robust attributed graph alignment via joint structure learning and optimal transport. In 39th IEEE International Conference on Data Engineering (ICDE).

Xiaobin Tang, Jing Zhang, Bo Chen, Yang Yang, Hong Chen, and Cuiping Li. 2020. BERT-INT: A bert-based interaction model for knowledge graph alignment. In IJCAI.

Vayer Titouan, Nicolas Courty, Romain Tavenard, and Rémi Flamary. 2019. Optimal transport for structured data with application on graphs. In International Conference on Machine Learning, pages 6275–6284. PMLR.

Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. 2018. Cross-lingual knowledge graph alignment via graph convolutional networks. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 349–357, Brussels, Belgium. Association for Computational Linguistics.

Yuting Wu, Xiao Liu, Yansong Feng, Zheng Wang, Rui Yan, and Dongyan Zhao. 2019. Relation-aware entity alignment for heterogeneous knowledge graphs. In IJCAI.

Kexuan Xin, Zequn Sun, Wen Hua, Wei Hu, Jianfeng Qu, and Xiaofang Zhou. 2022. Large-scale entity alignment via knowledge graph merging, partitioning and embedding. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 2240–2249.

Hongteng Xu, Dixin Luo, Hongyuan Zha, and Lawrence Carin Duke. 2019a. Gromov-wasserstein learning for graph matching and node embedding. In *International conference on machine learning*, pages 6932–6941. PMLR.

Kun Xu, Liwei Wang, Mo Yu, Yansong Feng, Yan Song, Zhiguo Wang, and Dong Yu. 2019b. Cross-lingual knowledge graph alignment via graph matching neural network. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3156–3161, Florence, Italy. Association for Computational Linguistics.

Lin Xu, Qixian Zhou, Ke Gong, Xiaodan Liang, Jianheng Tang, and Liang Lin. 2019c. End-to-end knowledge-routed relational dialogue system for automatic diagnosis. In *Proceedings of the AAAI conference on artificial intelligence*, pages 7346–7353.

Kaisheng Zeng, Zhenhao Dong, Lei Hou, Yixin Cao, Minghao Hu, Jifan Yu, Xin Lv, Lei Cao, Xin Wang, Haozhuang Liu, Yi Huang, Junlan Feng, Jing Wan, Juanzi Li, and Ling Feng. 2022a. Interactive contrastive learning for self-supervised entity alignment. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*.

Weixin Zeng, Xiang Zhao, Xinyi Li, Jiuyang Tang, and Wei Wang. 2022b. On entity alignment at scale. *The VLDB Journal*, pages 1–25.

Qingheng Zhang, Zequn Sun, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. 2019. Multi-view knowledge graph embedding for entity alignment. In *IJCAI*, pages 5429–5435.
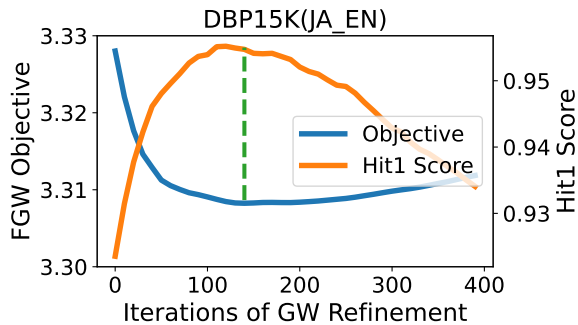
Figure 5: Visualization of the relationship between the objective function and alignment performance on DBP15K_EN_JA.

Figure 6: Visualization of the relationship between the objective function and alignment performance on SRPRS_EN_FR.

A Dataset Description

All the datasets used in our evaluation are publicly available on the Internet.

DBP15K[1] consists of three subsets of cross-lingual KG pairs extracted from DBpedia: $DBP15K_{ZH\_EN}$ (Chinese to English), $DBP15K_{JA\_EN}$ (Japanese to English), and $DBP15K_{FR\_EN}$ (French to English). Each KG pair contains 15,000 pre-aligned entity links.

SRPRS[2] is a sparse dataset that includes two cross-lingual KG pairs extracted from DBpedia: $SRPRS_{EN\_FR}$ (English to French), and $SRPRS_{EN\_DE}$ (English to German). Each subset of SRPRS also contains 15,000 entity links, but with fewer relation triples and no attribute triples.

D-W-15K-V2[3] consists of two English KGs extracted from DBpedia and WikiData, respectively, and there are 15,000 pre-aligned entity links.

MED-BBK-9K[4] is an industry dataset containing two Chinese medical KGs with 9,162 entity links, one is an authoritative human annotated KG and the other is extracted from a Chinese online encyclopedia called Baidu Baike. D-W-15K-V2 is licensed under the GNU General Public License v3.0, while other datasets are licensed under the MIT License.

B More Examples of the FGW Objective

Same as Section 4.5, in Figures 5 and 6, we visualize the objective function in (10) and the corresponding Hit1 score in GW refinement on $DBP15K_{JA\_EN}$ without translation and attributes
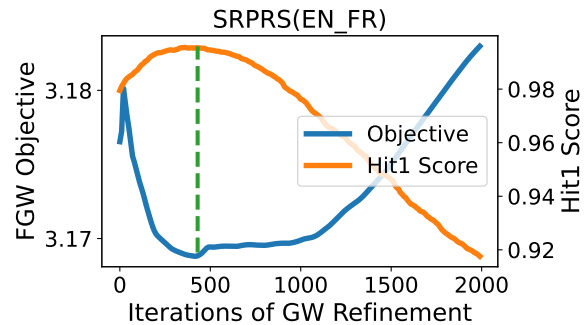
---

[1] https://github.com/nju-websoft/JAPE
[2] https://github.com/nju-websoft/RSN
[3] https://github.com/nju-websoft/OpenEA
[4] https://github.com/ZihengZZH/industry-eval-EA

and $SRPRS_{EN\_FR}$. The observation is consistent with Section 4.5. Two curves are highly correlated and the iteration corresponding to the minimum FGW objective value is approximately that to the maximum Hit1 score.

C Additional Results

Several studies have pointed out that many entities in DBP15K can be directly matched by strings to obtain aligned entities (Liu et al., 2020). In light of this, we perform additional experiments on a hard test set split of DBP15K, as introduced in (Liu et al., 2020), to minimize the influence of name bias. Furthermore, to demonstrate that FGWEA's exceptional performance cannot be solely credited to the powerful LaBSE encoder, we use the mean pooling of `bert-base-multilingual-cased`[5] as FGWEA's new semantic encoder. The embedding matching accuracy for this encoder is only 16.1% on the hard setting of $DBP15K_{ZH\_EN}$. Nonetheless, FGWEA continues to achieve competitive results as shown in Table 6, surpassing AttrGNN, the current top-performing method for this setting (Liu et al., 2020).

Model	Dataset	Hit1	Hit10	MRR
FGWEA	$DBP15K_{ZH\_EN}$	**0.756**	**0.868**	**0.796**
	$DBP15K_{JA\_EN}$	**0.788**	0.897	**0.828**
	$DBP15K_{FR\_EN}$	**0.983**	**0.997**	**0.988**
AttrGNN	$DBP15K_{ZH\_EN}$	0.662	0.818	0.719
	$DBP15K_{JA\_EN}$	0.774	**0.903**	0.821
	$DBP15K_{FR\_EN}$	0.886	0.956	0.912

Table 6: Results comparison bewteen FGWEA and AttrGNN on a hard setting of DBP15K.

---

[5] https://huggingface.co/bert-base-multilingual-cased

3332

A For every submission:

☑ A1. Did you describe the limitations of your work?
The limitation section on page 9.

☑ A2. Did you discuss any potential risks of your work?
The limitation section on page 9.

☑ A3. Do the abstract and introduction summarize the paper's main claims?
Section 1 on page 1.

☒ A4. Have you used AI writing assistants when working on this paper?
Left blank.

B ☑ Did you use or create scientific artifacts?

Section 4.1 on page 5 and Appendix A on page 13.

☑ B1. Did you cite the creators of artifacts you used?
Section 4.1 on page 5.

☑ B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
Section 4.1 on page 5 and Appendix A on page 13.

☑ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
Section 4.1 on page 5 and Appendix A on page 13.

☑ B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
Section 4.1 on page 5 and Appendix A on page 13.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Section 4.1 on page 5 and Appendix A on page 13.

☑ B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
Section 4.1 on page 5 and Appendix A on page 13.

C ☑ Did you run computational experiments?

Section 4 on pages 5-8.

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
Section 4.1 on page 5.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section 4.1 on page 5.

☑ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
Section 4.1 on page 5.

☑ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
Section 4.1

D ☒ Did you use human annotators (e.g., crowdworkers) or research with human participants?

Left blank.

☒ D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
Left blank.

☒ D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
Left blank.

☒ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
Left blank.

☒ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
Left blank.

☒ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
Left blank.