

# Non-translational Alignment for Multi-relational Networks

Shengnan Li<sup>1</sup>, Xin Li<sup>\*1</sup>, Rui Ye<sup>1</sup>, Mingzhong Wang<sup>2</sup>, Haiping Su<sup>1</sup>, Yingzi Ou<sup>1</sup>

<sup>1</sup> School of Computer Science, Beijing Institute of Technology, China

<sup>2</sup> School of Business, University of the Sunshine Coast, Australia

{shengnanli,xinli,yerui}@bit.edu.cn, mwan@usc.edu.au, {haipingsu,ouyingzi}@bit.edu.cn

## Abstract

Most existing solutions for the alignment of multi-relational networks, such as multi-lingual knowledge bases, are “translation”-based which facilitate the network embedding via the trans-family, such as TransE. However, they cannot address triangular or other structural properties effectively. Thus, we propose a non-translational approach, which aims to utilize a probabilistic model to offer more robust solutions to the alignment task, by exploring the structural properties as well as leveraging on anchors to project each network onto the same vector space during the process of learning the representation of individual networks. The extensive experiments on four multi-lingual knowledge graphs demonstrate the effectiveness and robustness of the proposed method over a set of state-of-the-art alignment methods.

## 1 Introduction

Network structure provides an efficient abstract for organizing interconnected data, such as social networks, knowledge bases [Bollacker *et al.*, 2008], and graph structure data [Pan *et al.*, 2016] [Pan *et al.*, 2017]. Most information networks are domain-specific or serve specific purposes. For example, Pinterest enables image sharing among users, Foursquare helps users to share their location interests, and DBLP mainly manages co-author relationship on publications. Moreover, there are large number of vertical knowledge bases and monolingual knowledge bases in various types.

Given these individual information networks, the benefits are significant if they can be connected or aligned to construct more complete/compact networks, enabling knowledge transferring across networks. For example, aligning different social networks can help to increase the accuracy and robustness for applications like link prediction and cross-domain recommendation. Moreover, constructing multi-lingual knowledge bases from monolingual knowledge graphs can greatly enhance the coherence and completeness of the knowledge.

<sup>\*</sup>Xin Li (xinli@bit.edu.cn) is the corresponding author. This work has been partially supported by National Key R&D Program of China under Grant No. 2017YFB0803300, NSFC under Grant No. 61772074.

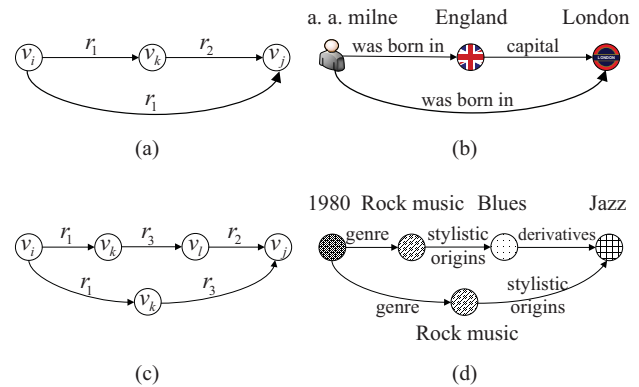


Figure 1: Two types of structures v.s. their instances in the dataset

To accomplish such tasks, the key is to detect and locate the correspondence among nodes and relations from different networks, such as identifying the same user across different social networks, and finding the same entity across knowledge bases in different languages.

Other than utilizing specific features to detect the correspondence between nodes, e.g., comparing the similarity between user-names in different social networks, the more general and practical solution is to apply the representation learning approaches to construct the low-dimensional representations (vectors) of nodes/edges for alignment tasks.

For the alignment tasks over single-relational networks, in which the relations/edges between any two nodes are considered to be of the same type, [Liu *et al.*, 2016] proposed to render a common subspace for the embeddings of multiple networks while preserving the proximity of users with “similar” sets of followers and followees in the embedded space. A relevance measurement was then proposed for selecting the candidates.

[Man *et al.*, 2016] proposed to embed each network into a low-dimensional space to obtain the representations of nodes. Then the authors employed a Multi-Layer Perceptron (MLP) to learn the mappings between nodes across networks with learned representations as features, and the MLP was supervised by the observed anchor links.

For multi-relational networks, in which the edges between nodes may be of different relation types, most so-

lutions apply the trans-family, such as TransE [Bordes *et al.*, 2013], TransH [Wang *et al.*, 2014b] and pTransE [Lin *et al.*, 2015a], to generate vector representations for nodes and relations in multi-relational network consisting of triplets ( $head(h), relation(r), tail(t)$ ). [Chen *et al.*, 2017] adapted TransE with various hand-crafted alignment loss functions to encode entities and relations of each language in a separated embedding space. [Zhu *et al.*, 2017] utilized pTransE to jointly encode both entities/nodes and relations into a low-dimensional subspace, and then aligned entities according to their distance in the subspace.

However, the trans-family are constrained by the relation  $h + r = t$ . That is, if a head node  $h$  moves to a tail node  $t$  with a relation  $r$ , then other possible relations leading  $h$  to  $t$  will be ignored. [Liu *et al.*, 2017] proved that the trans-family are unable to capture the triangular structures with a single type of relation, which frequently appear in multi-relational networks. Consequently, the performance of alignment is inevitable compromised if the network representation is learned with the trans-family. We also found that triangular structures with multiple relations and other local structures with more than one path from  $h$  to  $t$  have negative impact on the performance of alignment when the trans-family are adopted.

Fig.1(a) and Fig.1(c) illustrate two example structures and their corresponding instances (Fig.1(b) and Fig.1(d)) occurred in the real-world data set WK31-15k (detailed in Sec.4). Fig.1(b) shows three self-explanatory facts. Fig.1(d) shows two kinds of knowledge hierarchy with regard to the origin of several styles of music, which are co-existing knowledge with no inconsistency and acknowledged by the public. According to Wikipedia, Rock music drew on African-American genres of blues, while Jazz roots in blues (the upper path of Fig.1(d)). Rock music also incorporated influence from Jazz (the lower path of Fig.1(d)). We can observe that both local structures (Fig.1(a) and Fig.1(c)) have two paths from  $v_i$  to  $v_j$ , but  $r_2$  will be ignored with the hard constraint  $h + r = t$  in the trans-family ( $r_2 = 0$ ).

Therefore, we argue that the trans-family may not be effective enough for the tasks of the alignment across multi-relational networks where the number of triangular and other local structures they fail to capture is non-trivial. Thus, in this paper, we propose a non-translational approach to perform the alignment across multi-relational networks, e.g., multi-lingual knowledge graphs.

The meaning of “non-translation” is two-fold: 1. the proposed approach does not rely on the trans-family to learn robust representations of networks; 2. the alignment for multi-lingual knowledge bases in our experiments is addressed by exploring the network structural information but not by machine translation.

Specifically, the objective function is optimized to preserve network structures as well as to locate the correspondence across networks. Thus, the vector representation of networks and the alignment can be achieved simultaneously with the unified optimization framework. Extensive experiments were conducted on four real world datasets to demonstrate the effectiveness and robustness of the proposed method over several state-of-the-art methods.

## 2 Related Work

There are two lines of research related to our work, namely *network embedding* and *network alignment*.

### 2.1 Network Embedding

In general, network embedding aims to learn vector representations for nodes/relations of a network. [Perozzi *et al.*, 2014] propose a skip-gram based vertex embedding approach - DeepWalk, which takes vertexes as words and random walks as sentences. LINE [Tang *et al.*, 2015] deals with both directed and undirected graphs by explicitly preserving so called first-order and second-order proximity.

For multi-relational networks, TransE [Bordes *et al.*, 2013] is the pioneer work on utilizing translation-based approach to represent both the nodes and their relations onto a low-dimensional vector space. Given the triplet  $(h, r, t)$ , the representation vector of the tail node  $t$  is required to be as close as possible to the representation vector of the head node  $h$  plus the representation of the relation  $r$ . That is, the goal is to minimize  $\|h+r-t\|$ . Its extensions, including TransH [Wang *et al.*, 2014b], TransR [Lin *et al.*, 2015b], and pTransE [Wang *et al.*, 2014a] (referred to as “trans-family”), try to project the nodes and relations onto different subspaces to deal with various mapping properties of relations, such as reflexive, one-to-many, many-to-one, and many-to-many.

However, the trans-family are inefficient in dealing with triangular and other complex link structures with more than one path from  $h$  to  $t$ , which frequently appear in multi-relational networks. [Liu *et al.*, 2017] proposed a structural representation learning approach for multi-relational networks to address the weakness of trans-family that render them incapable of preserving the triangular structures.

### 2.2 Network Alignment

Network alignment with structural matching is a key research topic in bioinformatics for aligning protein-protein interaction networks [Klau, 2009]. Recent advance of network alignment methodologies has mainly been driven by social networks and knowledge graphs with the aim of fusing networks for more complete/compact information.

[Zhang and Yu, 2015b] proposed an unsupervised approach to align multiple anonymous social networks by utilizing user profiles (username, locations) and structural features (mutual friends). The alignment is computed by minimizing the *friendship inconsistency* and the anchor links are selected according to a “confidence” score. The fusion of anchor link prediction and social link prediction via random walk was found effective for network alignment tasks [Zhang and Yu, 2015a]. [Tan *et al.*, 2014] proposed to model each social network with a hypergraph, and then to embed the hypergraphs into the same space in a way the nodes on each hyper-edge are drawn together as close as possible in the embedded space. Eigenvalue decomposition based on the hyper-edges is required, which however is computational expensive. [Man *et al.*, 2016] proposed to learn the embeddings of two networks by preserving the first-order proximity within them respectively. Then the alignment is performed in the embedded space by transforming alignment task into a bi-classification task with the known anchor pairs as positive instances.

For knowledge graph alignment, [Zhu *et al.*, 2017] proposed to utilize pTransE to jointly encode both entities and relations into a low-dimensional subspace, then aligned entities according to their distance in the subspace. Predicted anchors during the execution are iteratively included to supervise the alignment. [Chen *et al.*, 2017] adapted TransE to address a specific task of multilingual knowledge graph alignment. They simultaneously encoded entities and relations of each language into separate low-dimensional spaces with various loss functions, and then constructed transitions between vector spaces. A seed set of anchors across knowledge bases were utilized to supervise the transition for alignment.

Our work applies the similar technique of projecting multiple networks onto a common embedded space. However, our approach utilizes a probabilistic model to explore the structural properties for network representation learning, instead of applying the trans-family. Therefore, we alleviate the constraint in the conventional translation-based representation learning. Consequently, we achieve more accurate network representations, and more precise and robust alignments.

### 3 Model Framework

Let  $G = (V, E, R)$  denote a multi-relational network where  $V := \{v_i\}$  is the set of nodes,  $R := \{r_i\}$  is the set of relation labels, and  $E := \{(v_s, r_m, v_t)\}$  is the set of directed edges. A directed edge  $(v_s, r_m, v_t)$  denotes there is a relation  $r_m$  between the source node  $v_s$  and the target node  $v_t$  in the network. In the task of alignment for multi-relational networks, if the alignment among two nodes or relations between network  $X$  and  $Y$  is known in advance, it is denoted as aligned (anchor) node-pair  $(v_i^X, v_j^Y) \in V_a$  or aligned (anchor) relation-pair  $(r_i^X, r_j^Y) \in R_a$  respectively.

To accomplish the alignment task, two main stages are deployed in our model: the knowledge embedding, which learns the representations of nodes and relations for each individual network while preserving local connectivities of these networks, and the alignment, which joins the embeddings of different networks onto the same subspace by leveraging on the anchor nodes and anchor relations.

#### 3.1 Individual Embedding

Inspired by [Liu *et al.*, 2017], in this paper we adapt a probabilistic method which avoids using hard constraints in the trans-family to encode the multi-relational networks, such as knowledge graph. For each network, embeddings of nodes and relations are learned in a  $d$ -dimensional space  $\mathcal{R}^d$ .

Given a node  $v_i$ , the probability that it has relation  $r_s$  with  $v_j$ , considering the relation between  $v_i$  and other nodes, is defined as:

$$p(v_j|v_i, r_s) = \frac{\exp(\vec{u}_j^T(\vec{u}_i + \vec{u}_{r_s}))}{\sum_{(v_i, r_s, v_x) \in E'} \exp(\vec{u}_x^T(\vec{u}_i + \vec{u}_{r_s}))} \quad (1)$$

where  $E' = \{(v_i, r_s, v_x) | v_x \in V\}$ ,  $\vec{u}_i, \vec{u}_j$  and  $\vec{u}_{r_s}$  denote the corresponding vectors for the node  $v_i, v_j$  and the relation  $r_s$  respectively. In Eq.(1), the source vector  $\vec{u}_i$  of  $v_i$ , the target vector  $\vec{u}_j$  of  $v_j$ , and the relation vector  $\vec{u}_{r_s}$  of  $r_s$  of the directed edge  $(v_i, r_s, v_j)$  are related by merely adding  $\vec{u}_{r_s}$  to

$\vec{u}_i$ , instead of enforcing the constraints as in trans-family. Besides, the direction of the edge is implicitly defined in the usage of addition. Therefore, only one vector is required to represent a node. This is an improvement than the work in [Liu *et al.*, 2017], [Liu *et al.*, 2016] and [Tang *et al.*, 2015], which require two vector representations  $\vec{u}_i \in \mathcal{R}^d$  and  $\vec{u}_i' \in \mathcal{R}^d$  to distinguish that if a node  $v_i$  is a source node or a target node in a triplet or in a directed graph. As a result, the number of parameters are reduced which helps to avoid overfitting.

Based on Eq.(1), in addition to preserving the connectivity between nodes and their directed relations in triplets, three distinct non-isomorphic motifs found in [Liu *et al.*, 2017] are also preserved in the low-dimensional vector spaces by defining their corresponding probabilities. For example,  $p_1$  is defined to capture the motif in which the out-degree of  $v_i$  is 2, given as:

$$p_1(v_j^{r_s}, v_k^{r_t} | v_i) = \frac{\exp(\vec{u}_j^T(\vec{u}_i + \vec{u}_{r_s}) + \vec{u}_k^T(\vec{u}_i + \vec{u}_{r_t}))}{\sum_{\substack{(v_i, r_p, v_x) \in E' \\ \wedge (v_i, r_q, v_y) \in E'}} \exp(\vec{u}_x^T(\vec{u}_i + \vec{u}_{r_p}) + \vec{u}_y^T(\vec{u}_i + \vec{u}_{r_q}))} \quad (2)$$

where  $v_j^{r_s}$  is a concise expression of the pair  $(v_j, r_s)$ , which indicates the case that the node  $v_j$  is connected with others through the relation  $r_s$ .  $p_2$  and  $p_3$  are defined in the same manner for other two motifs, which are not presented here due to the page limit.

The objective function for knowledge embedding needs to minimize the KL-divergence between  $p, p_1, p_2, p_3$  and their corresponding empirical counterparts  $\hat{p} = \frac{\omega_{ij}}{d_{i r_s}^{out}}$ ,  $\hat{p}_1 = \frac{\omega_{ij} * \omega_{ik}}{d_i^{out} * d_i^{out}}$ ,  $\hat{p}_2 = \frac{\omega_{ji} * \omega_{ik}}{d_i^{in} * d_i^{out}}$  and  $\hat{p}_3 = \frac{\omega_{ji} * \omega_{ki}}{d_i^{in} * d_i^{in}}$ , where  $d_i^{out} = \sum_{j \in N_{out}(v_i)} \omega_{ij}$ ,  $d_i^{in} = \sum_{j \in N_{in}(v_i)} \omega_{ji}$ .  $N_{out}(v_i)$  and  $N_{in}(v_i)$  are the sets of one-hop neighbors of  $v_i$  as the source and the target nodes respectively.  $d_{i r_s}^{out} = \sum_{j \in N_{out}(v_i, r_s)}$ , and  $N_{out}(v_i, r_s)$  denotes the set of one-hop neighbors of  $v_i$  as the source node which are also connected via relation  $r_s$ .  $\omega_{ij}$  is the weight of the edge  $(v_i, v_j)$ . The weight is set as the number of the connections between two nodes  $v_i$  and  $v_j$ . When there is no relation between them,  $\omega_{ij}$  equals to 0. Thereafter, the objective function is defined as:

$$\begin{aligned} O_1 = & - \sum_{L \in \{X, Y\}} \sum_{(v_i, r_s, v_j) \in E_L} \omega_{ij}^L * \log p(v_j^L | v_i^L, r_s^L) \\ & - \sum_{L \in \{X, Y\}} \sum_{\substack{(v_i, r_s, v_j) \in E_L \\ \wedge (v_i, r_t, v_k) \in E_L}} \omega_{ij}^L * \omega_{ik}^L * \log p_1(v_j^{r_s^L}, v_k^{r_t^L} | v_i^L) \\ & - \sum_{L \in \{X, Y\}} \sum_{\substack{(v_j, r_s, v_i) \in E_L \\ \wedge (v_i, r_t, v_k) \in E_L}} \omega_{ji}^L * \omega_{ik}^L * \log p_2(v_j^{r_s^L}, v_k^{r_t^L} | v_i^L) \\ & - \sum_{L \in \{X, Y\}} \sum_{\substack{(v_j, r_s, v_i) \in E_L \\ \wedge (v_k, r_t, v_i) \in E_L}} \omega_{ji}^L * \omega_{ki}^L * \log p_3(v_j^{r_s^L}, v_k^{r_t^L} | v_i^L) \end{aligned} \quad (3)$$

Minimizing the above objective function will lead the structural information in individual networks to be preserved in the low-dimensional vector space, during the process of model inference.

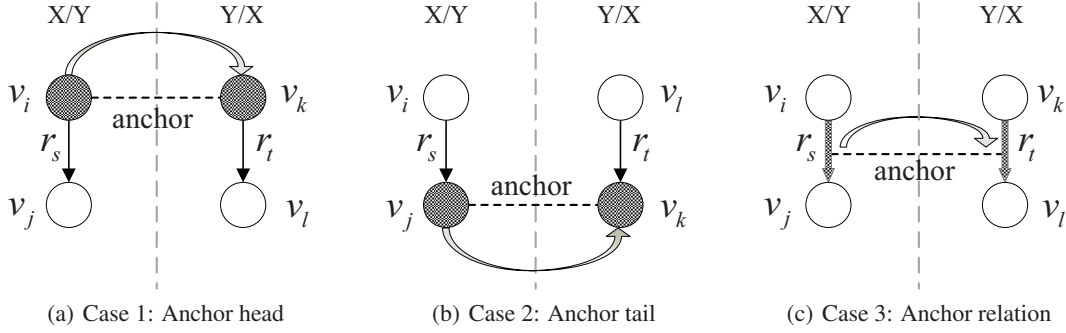


Figure 2: Three cases of leveraging on anchors in triplets

### 3.2 Joint Embedding

With the embedding of each individual multi-relational network obtained, we locate the anchor nodes/relations in the embedded space and join the embeddings of different networks into the same subspace by drawing the anchor nodes/relations and their counterparts coincide in the embedded space. For example, if the task is to align multi-lingual knowledge graphs, then the embeddings of different languages need to be joined into the same semantic space.

Given two networks  $X$  and  $Y$ , an anchor is defined as an entity or a relation in  $X$  which has a unique counterpart in  $Y$ . In multi-lingual knowledge graphs, the counterpart means the two corresponding anchors have the same meaning in different languages.

Anchors are used to supervise the alignment task. Triplets  $(v_i^{X/Y}, r_s^{X/Y}, v_j^{X/Y})$  that include anchor nodes or anchor relations are selected because they help to preserve the connectivity information in the low-dimensional space. With regard to anchors' different roles, such as a source node, a target node, or a relation as depicted in Fig.2, we deliberately replace the current learned representation of a node/relation with the presentation of its counterpart learned in another network. Therefore, the objective function is defined as:

$$\begin{aligned}
 O_2 = & \\
 & - \sum_{v_k \in Y/X} \sum_{T \in E^{X/Y}} \omega_{ij}^{X/Y} \mathbb{1}_{V_a}(v_i^{X/Y}, v_k^{Y/X}) \log p(v_j^{X/Y} | v_k^{Y/X}, r_s^{X/Y}) \\
 & - \sum_{v_k \in Y/X} \sum_{T \in E^{X/Y}} \omega_{ij}^{X/Y} \mathbb{1}_{V_a}(v_j^{X/Y}, v_k^{Y/X}) \log p(v_k^{Y/X} | v_i^{X/Y}, r_s^{X/Y}) \\
 & - \sum_{r_t \in Y/X} \sum_{T \in E^{X/Y}} \omega_{ij}^{X/Y} \mathbb{1}_{R_a}(r_s^{X/Y}, r_t^{Y/X}) \log p(v_j^{X/Y} | v_i^{X/Y}, r_t^{Y/X})
 \end{aligned} \quad (4)$$

where  $\mathbb{1}_{V_a}$  and  $\mathbb{1}_{R_a}$  are the indicator functions to test which triplet can be used in the alignment framework, and  $T$  stands for the triplet  $(v_i^{X/Y}, r_s^{X/Y}, v_j^{X/Y})$ .

### 3.3 Model Inference

To infer the vector representations of networks, stochastic gradient descent is applied for optimization. Taking node  $v_i$  in network  $X$  as an example, the gradient is defined as:

$$\frac{\partial O}{\partial \vec{u}_i^X} = \frac{\partial O_1}{\partial \vec{u}_i^X} + \frac{\partial O_2}{\partial \vec{u}_i^X}. \quad (5)$$

To efficiently estimate the probabilities in Eq.(3-4), Negative sampling [Mikolov *et al.*, 2013] is applied to avoid summation over the entire set of nodes(refer to Eq.(1)), thus reducing the computational complexity. The equivalent counterparts in the objective function can then be computed as:

$$\begin{aligned}
 \log p(v_j^X | v_i^X, r_s^X) & \propto \log \sigma(\vec{u}_j^{X^T} (\vec{u}_i^X + \vec{u}_{r_s}^X)) \\
 & + \sum_{m=1}^K E_{v_n \sim P_n(v)} \log \sigma(-\vec{u}_n^{X^T} (\vec{u}_i^X + \vec{u}_{r_s}^X))
 \end{aligned} \quad (6)$$

$$\begin{aligned}
 \log p(v_j^Y | v_i^X, r_s^Y) & \propto \log \sigma(\vec{u}_j^{Y^T} (\vec{u}_i^X + \vec{u}_{r_s}^Y)) \\
 & + \sum_{m=1}^K E_{v_n \sim P_n(v)} \log \sigma(-\vec{u}_n^{Y^T} (\vec{u}_i^X + \vec{u}_{r_s}^Y))
 \end{aligned} \quad (7)$$

$$\begin{aligned}
 \log p_1(v_j^{r_s^X}, v_k^{r_t^X} | v_i^X) & \propto \log \sigma(\vec{u}_j^{X^T} (\vec{u}_i^X + \vec{u}_{r_s}^X) + \vec{u}_k^{X^T} (\vec{u}_i^X + \vec{u}_{r_t}^X)) \\
 & + \sum_{m=1}^K E_{v_n \sim P_n(v)} \log \sigma(-\vec{u}_j^{X^T} (\vec{u}_i^X + \vec{u}_{r_s}^X) - \vec{u}_n^{X^T} (\vec{u}_i^X + \vec{u}_{r_t}^X))
 \end{aligned} \quad (8)$$

where  $\sigma(x) = 1/(1 + \exp(-x))$  is the sigmoid function,  $v_n$  denotes a negative sample node from the ‘‘noisy distribution’’ of  $p_n(v) = d_v^{3/4}$ , and  $d_v$  is the output degree.  $r_l$  denotes a negative sample edge from a uniform distribution. In Eq.(6-8), each of the first component on the right side models the real structures (positive samples) and the second component models the negative samples drawn from the noise distribution. Note that  $v_i$ ,  $r_l$  and  $v_n$  should not form a fact triplet, and  $K$  is the number of negative samples.

Thereafter, the partial derivative of the objective function w.r.t.  $\vec{u}_i$  can be inferred by substituting corresponding components in Eq.(3-4) with Eq.(6-8). Detailed deductions are not included due to the page limit.

### 3.4 Alignment Determination

To determine the alignments between entities/relations across networks, cosine similarity between vectors of one node/relation from network  $X$  and another from  $Y$  is adopted.

$$\text{correl}(v_i^X, v_j^Y) = \frac{\sum_{p=1}^d u_{ip}^X \times u_{jp}^Y}{\sqrt{\sum_{p=1}^d u_{ip}^{X^2}} \times \sqrt{\sum_{p=1}^d u_{jp}^{Y^2}}} \quad (9)$$

$$correl(r_i^X, r_j^Y) = \frac{\sum_{p=1}^d u_{r_{ip}}^X \times u_{r_{jp}}^Y}{\sqrt{\sum_{p=1}^d u_{r_{ip}}^X{}^2} \times \sqrt{\sum_{p=1}^d u_{r_{jp}}^Y{}^2}} \quad (10)$$

After evaluating all possible pair combinations of nodes/relations from  $X$  and  $Y$ , the most relevant node/relation pairs between two networks can be considered as pairs of anchors.

## 4 Experiments

Multi-lingual knowledge graphs are employed to evaluate the proposed non-translational alignment model (NTAM). Specifically, two types of cross-lingual tasks, including cross-lingual entity matching and cross-lingual relation matching, are evaluated with the trilingual datasets WK31-15k and WK31-120k [Chen *et al.*, 2017], including English(En), German(De) and French(Fr) knowledge graphs which are extracted from DBpedia with known aligned entities as ground truth. The datasets are further preprocessed by removing incomplete triplets, manually aligning more relations as anchors for training and test sets. The statistics of the datasets are listed in Table 1 and Table 2.

We compare our proposed NTAM with several state-of-the-art approaches for network alignment tasks, including:

- **MTransE**: A TransE-based approach [Chen *et al.*, 2017] which performs network alignment by inferring a linear transformation in the embedding spaces for different knowledge graphs.
- **ITransE**: A pTransE-based approach for network alignment [Zhu *et al.*, 2017] which iteratively utilizes the predicted anchors to supervise future alignments.
- **IONE**: A state-of-the-art alignment approach for single-relation networks [Liu *et al.*, 2016] in which the directions of the edges are deliberately considered, and all relations are considered to be the same type.
- **NTAM-2**: A variant of proposed NTAM which requires only second-order connectivities (refer to Eq.(1)) to be preserved in knowledge embedding. Its joint embedding component is the same as NTAM.
- **NTAM-1**: Differing from NTAM-2, NTAM-1 requires only the first-order connectivity to be preserved in knowledge embedding.

### 4.1 Evaluation Metrics

In the experiments,  $Hits@N$  is utilized as the evaluation metric, which is defined as:

$$Hits^{X/Y \rightarrow Y/X}@N = \frac{|Corr@N|^{X/Y}}{|Test Set|} \quad (11)$$

where  $|Corr@N|^{X/Y}$  is the number of nodes/relations in network  $X$  with their anchors among the top- $N$  neighbors in the low-dimensional space of network  $Y$ .  $|Test Set|$  is the size of the test set. Note that  $Hits^{X \rightarrow Y}@N$  is guaranteed to be the same as  $Hits^{Y \rightarrow X}@N$  only when two networks are completely overlapped, which means that each triplet in one network has its unique counterpart in another network.

dataset	#Entity	#Relation	#Triplet
WK31-15k-En-De	15,617(En)	1,935(En)	217,860(En)
	15,167(De)	1,173(De)	156,452(De)
WK31-15k-En-Fr	16,517(En)	2,488(En)	215,301(En)
	16,676(Fr)	2,735(Fr)	189,416(Fr)
WK31-120k-En-De	68,840(En)	2,588(En)	641,046(En)
	63,392(De)	1,369(De)	431,168(De)
WK31-120k-En-Fr	125,550(En)	3,587(En)	1,502,414(En)
	124,502(Fr)	2,955(Fr)	988,763(Fr)

Table 1: Statistics of the datasets used for evaluation

dataset	#Aligned Entity	#Aligned Relation
WK31-15k-En-De	2,070	445
WK31-15k-En-Fr	3,116	598
WK31-120k-En-De	9,680	772
WK31-120k-En-Fr	42,378	1,127

Table 2: The number of anchors among datasets

### 4.2 Entity Alignment

In the experiments, entity alignment aims to find the matching entities for different language knowledge graphs. The embedding dimension is set as 100. We tabulate  $Hits@1$  and  $Hits@10$  for four datasets (with training ratios as 80% for entity anchors, 100% for relation anchors) for each approach in Table 3.

Without considering the types of relations, IONE is significantly outperformed by other approaches which target at multi-relational network alignment. NTAM performs the best in terms of higher precision<sup>1</sup>. NTAM-1 and NTAM-2, two variants of NTAM, also outperform ITransE and MTransE, proving the benefit of non-translational embedding.

ITransE is the most competitive approach in all baselines. However, the experiments show that ITransE cannot provide consistent or robust results on entity alignment bidirectionally for two networks. That is, the alignment from network  $X$  to  $Y$  may differ significantly from the alignment from  $Y$  to  $X$ , w.r.t.  $Hits^{X \rightarrow Y}@N$  and  $Hits^{Y \rightarrow X}@N$ . The reason is that ITransE deliberately preprocesses the training data to construct new triplets. These new triplets are generated by replacing head and/or tail in the triplets of one network with its counterpart in another network. Such replacements produce triplets which in fact do not exist in a network. This operation favors one network by adding more triplets in the network to compute the individual representation. On the contrary, our approach provides a more accurate and robust solution for the alignment task by utilizing non-translational embedding and performing the alignment in a probabilistic basis.

NTAM-2 performs worse than NTAM-1 for most cases, as in multi-relational networks, the neighbors defined in the second order proximity do not require their connections to be the same relations. Thus, two nodes sharing a bunch of anchors connected by different relations are also considered as close in the embedded space, while they are not.

<sup>1</sup>MTransE has been tested based on the authors' github repository with the same datasets. Since we got poor performance with lower than 10% for  $Hits@10$  for all datasets, its results are not included in the comparison.

Dataset	WK31-15k						WK31-120k					
	En-De		De-En		En-Fr		Fr-En		En-De	De-En	En-Fr	Fr-En
Aligned Language	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10	Hits@10	Hits@10
Metric(%)												
ITransE	29.4	49.6	50.4	76.1	55.0	80.0	54.6	80.8	36.4	51.9	37.9	57.8
IONE	7.2	30.8	12.5	33.5	40.4	77.2	54.0	80.9	66.0	70.2	51.7	52.7
NTAM-1	66.5	83.4	54.9	82.7	62.0	86.9	38.8	85.3	68.8	68.6	62.0	56.1
NTAM-2	58.3	76.7	42.4	73.3	60.6	86.2	39.3	84.0	67.9	70.5	62.6	56.5
NTAM	<b>66.7</b>	<b>85.8</b>	<b>73.3</b>	<b>86.7</b>	<b>73.6</b>	<b>92.9</b>	<b>84.0</b>	<b>93.9</b>	<b>74.1</b>	<b>77.8</b>	<b>68.9</b>	<b>69.6</b>

Table 3: Performance comparison on entity alignment

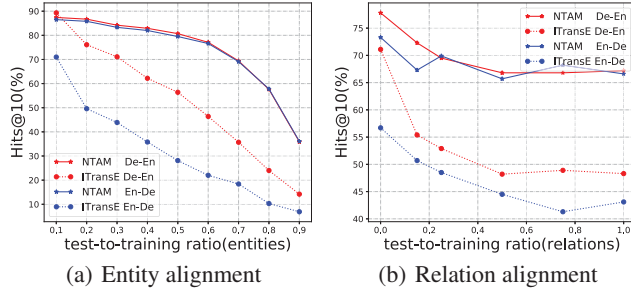


Figure 3: Effects of test-to-training ratio to alignment tasks

Dataset	WK31-15k			WK31-120k		
	En	Fr	De	En	Fr	De
Meth./#strc.						
TransE	62.4%	67.3%	74.5%	66.5%	65.6%	72.3%
NTAM	68.0%	70.8%	80.2%	66.9%	70.0%	73.6%
#strc.	14,439	76,647	946	43,950	694,481	2,606

Table 4: Link prediction within individual network

Besides, all approaches perform best over the WK31-15k-En-Fr dataset. This attributes to the highest ratio of the aligned node anchors to the entire node set in this dataset. Besides it is a rather dense network among all networks in terms of the ratio of the number of triplets to the number of nodes.

Fig.3(a) illustrates the alignment performance for  $Hits@10$  with different test-to-training ratios. It shows that NTAM consistently outperforms ITransE. With the increase of ratio value, the performance enhancement becomes more significant. ITransE iteratively applies newly “predicted anchor” to aid the alignment. However, when a smaller set of anchor seeds gets involved, the chance of improper usage of non-anchor nodes as anchor nodes increases. Thus, with more “false predicted anchors” included, the alignment performance of ITransE gets worse.

### 4.3 Relation Alignment

Relation alignment aims to find the matching relations for different language knowledge graphs. In these experiments, all entity anchors are used in the training, but set different test-to-training ratios for relation anchors. Fig.3(b) shows that the relation alignment can still be learned from the embedding and entity anchors even though training ratio is 0. NTAM consistently performs better than ITransE on different test-to-training ratios. We argue the advantage again comes from the robust representation learning method in NTAM.

### 4.4 Link Prediction within Individual Network

In addition to entity and relation alignment, link prediction tries to predict the missing  $h$  or  $t$  for a triplet fact  $(h, r, t)$  in a given knowledge base. Effective link prediction can help to improve the completeness of a network, which in turn helps to improve the alignment performance.

To testify the effectiveness of the non-translational approach, we also evaluate it on the link prediction task in a single network with the representations obtained by only using individual embedding objective (Sec. 3.1). The process is abstracted as a binary classification problem which is based on the low-dimensional vector representation of the graph. While the triplets in a knowledge base are treated as positive samples, negative samples can be generated by corrupting each fact triplet  $(h, r, t)$  with the head ( $h$ ) or tail ( $t$ ) replaced. The classification accuracy is used as the evaluation criterion.

Table 4 shows the results of link prediction by TransE and NTAM over individual knowledge graphs. The last line shows the number of structures, including the triangular structures in [Liu *et al.*, 2017] and the local structures in Fig.1. For all datasets, the non-translational approach has consistent improvements since it effectively alleviates the constraint in translation-based approaches. It is also evident TranE has similar performance to NTAM if the number of such structures is relatively small (e.g. De in WK31-15k and WK31-120k). With the increase of these structures in the network, the improvement by NTAM becomes more significant.

## 5 Conclusion

In this paper, we proposed a non-translational alignment model for multi-relational networks. It overcomes the intrinsic limitation of hard constraints in solutions based on the trans-family. A probabilistic model is utilized to explore the structural properties for networks. It is also capable of using anchors to project each network onto the same vector space while learning the representation of individual networks. Stochastic gradient descent and negative sampling are used to boost the performance of the learning process. With presentations well trained, the alignment is then performed based on the vector representation of networks with cosine similarity as the measurement. The extensive experiments on four real world datasets demonstrate the effectiveness and robustness of the proposed method over a set of state-of-the-art methods. For the future work, we plan to extend it to incorporate semantic information into the alignment.

## References

- [Bollacker *et al.*, 2008] Kurt D. Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008*, pages 1247–1250, 2008.
- [Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems*, pages 2787–2795, 2013.
- [Chen *et al.*, 2017] Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI, Melbourne, Australia, August 19-25, 2017*, pages 1511–1517, 2017.
- [Klau, 2009] Gunnar W Klau. A new graph-based method for pairwise global network alignment. *BMC bioinformatics*, 10(Suppl 1):S59, 2009.
- [Lin *et al.*, 2015a] Yankai Lin, Zhiyuan Liu, Huan-Bo Luan, Maosong Sun, Siwei Rao, and Song Liu. Modeling relation paths for representation learning of knowledge bases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP, Lisbon, Portugal, September 17-21*, pages 705–714, 2015.
- [Lin *et al.*, 2015b] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence, January 25-30, Austin, Texas, USA.*, pages 2181–2187, 2015.
- [Liu *et al.*, 2016] Li Liu, William K. Cheung, Xin Li, and Lejian Liao. Aligning users across social networks using network embedding. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence, IJCAI, New York, NY, USA, 9-15 July*, pages 1774–1780, 2016.
- [Liu *et al.*, 2017] Lin Liu, Xin Li, William K. Cheung, and Chengcheng Xu. A structural representation learning for multi-relational networks. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25*, pages 4047–4053, 2017.
- [Man *et al.*, 2016] Tong Man, Huawei Shen, Shenghua Liu, Xiaolong Jin, and Xueqi Cheng. Predict anchor links across social networks via an embedding approach. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence, IJCAI, New York, NY, USA, 9-15 July*, pages 1823–1829, 2016.
- [Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of the 27th Annual Conference on Neural Information Processing Systems 2013.*, pages 3111–3119, Lake Tahoe, Nevada, USA, December 2013.
- [Pan *et al.*, 2016] Shirui Pan, Jia Wu, Xingquan Zhu, Chengqi Zhang, and Philip S. Yu. Joint structure feature exploration and regularization for multi-task graph classification. *IEEE Transactions on Knowledge & Data Engineering*, 28(3):715–728, 2016.
- [Pan *et al.*, 2017] S. Pan, J. Wu, X. Zhu, G. Long, and C. Zhang. Task sensitive feature exploration and learning for multitask graph classification. *IEEE Transactions on Cybernetics*, 47(3):744–758, 2017.
- [Perozzi *et al.*, 2014] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: online learning of social representations. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA - August 24 - 27*, pages 701–710, 2014.
- [Tan *et al.*, 2014] Shulong Tan, Ziyu Guan, Deng Cai, Xuzhen Qin, Jiajun Bu, and Chun Chen. Mapping users across networks by manifold alignment on hypergraph. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence, July 27 -31, Québec City, Québec, Canada.*, pages 159–165, 2014.
- [Tang *et al.*, 2015] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. LINE: large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web, WWW 2015, Florence, Italy, May 18-22*, pages 1067–1077, 2015.
- [Wang *et al.*, 2014a] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph and text jointly embedding. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, October 25-29, Doha, Qatar*, pages 1591–1601, 2014.
- [Wang *et al.*, 2014b] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence, July 27 -31, Québec City, Québec, Canada.*, pages 1112–1119, 2014.
- [Zhang and Yu, 2015a] Jiawei Zhang and Philip S. Yu. Integrated anchor and social link predictions across social networks. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI, Buenos Aires, Argentina, July 25-31*, pages 2125–2132, 2015.
- [Zhang and Yu, 2015b] Jiawei Zhang and Philip S. Yu. Multiple anonymized social networks alignment. In *2015 IEEE International Conference on Data Mining, ICDM 2015, Atlantic City, NJ, USA, November 14-17*, pages 599–608, 2015.
- [Zhu *et al.*, 2017] Hao Zhu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Iterative entity alignment via joint knowledge embeddings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, August 19-25*, pages 4258–4264, 2017.