

Received September 23, 2020, accepted October 21, 2020, date of publication October 27, 2020, date of current version November 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3034183

# 3D Learning and Reasoning in Link Prediction Over Knowledge Graphs

MOJTABA NAYYERI<sup>1,2</sup>, MIRZA MOHTASHIM ALAM<sup>1,2</sup>, JENS LEHMANN<sup>1,3</sup>,  
AND SAHAR VAHDATI<sup>2</sup>

<sup>1</sup>Smart Data Analytics Group (SDA), University of Bonn, 53115 Bonn, Germany

<sup>2</sup>Nature-Inspired Machine Intelligence, Institute for Applied Informatics (InfAI), 01069 Dresden, Germany

<sup>3</sup>Fraunhofer IAIS, 01069 Dresden, Germany

Corresponding author: Mojtaba Nayyeri (anayyeri@cs.uni-bonn.de)

This work was supported in part by the Speech Recognition Project (SPEAKER) under Grant BMWi FKZ 01MK20011A, in part by the Speech Recognition Project (JOSEPH) through the Fraunhofer Zukunftsstiftung, in part by the H2020-EU Learning, Applying, Multiplying Big Data Analytics (LAMBDA) Project under Grant EU 809965, in part by the Competence Center Machine Learning Rhine-Ruhr (ML2R) Project under Grant BmBF FKZ 01 15 18038 A/B/C, in part by the Maschinelles Lernen mit Wissensgraphen (MLwin) Project under Grant BmBF 01IS18050 D/F, in part by the Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Project under Grant BmBF 01/IS18026A-F, in part by the Trustworthy AI integrating Learning, Optimization and Reasoning (TAILOR) Project under Grant EU 952215, and in part by the H2020-EU Digital Platform and Analytic Tools for Energy (PLATOON) Project under Grant 872592.

**ABSTRACT** Knowledge Graph Embeddings (KGE) are used for representation learning in Knowledge Graphs (KGs) by measuring the likelihood of a relation between nodes. Rotation-based approaches, specially axis-angle representations, were shown to improve the performance of many Machine Learning (ML)-based models in different tasks including link prediction. There is a perceived disconnect between the topics of KGE models and axis-angle rotation-based approaches. This is particularly visible when considering the ability of KGEs to learn relational patterns such as symmetry, inversion, implication, equivalence, composition, and reflexivity considering axis-angle rotation-based approaches. In this article, we propose *RodE*, a new KGE model which employs an axis-angle representation for rotations based on Rodrigues' formula. *RodE* inherits the main advantages of 3-dimensional rotation from angle, orientation and distance preservation in the embedding space. Thus, the model efficiently captures the similarity between the nodes in a graph in the vector space. Our experiments show that *RodE* outperforms state-of-the-art models on standard datasets.

**INDEX TERMS** Link prediction, knowledge graph embedding, 3D rotation, learning and reasoning Rodrigues formula.

## I. INTRODUCTION

N-dimensional rotations have been used for several AI-based applications such as motion detection in computer vision and kinematics descriptions in robotics. More specifically, such rotations have been recently employed in the development of ML models resulting in powerful learners for neural networks [7], spectral clustering [15], regression [33], ensemble learning [25], and representation learning in knowledge graphs (KGs) [27]. Among N-dimensional rotations, the 3D case is particularly important due to the compatibility of its coordinate system with real world. Therefore, the study of rotations with three degrees of freedom has reached a state of maturity including quantum mechanics [3]. Different representations [1], [21] for three-dimensional rotations can be selected per application such as Euler angles [5], axis-angle

representation (Rodrigues formula) [8], exponential coordinates [11], and matrices [24]. Each of these rotation-based approaches have specific characteristics in preserving distance, angle and orientation of the rotated objects. particularly the axis-angle representation has some general advantages over the other representations [12]: 1) in comparison to the rotation matrices which contain 9 ( $= 3 \times 3$ ) parameters, the Rodrigues formula represents rotation with four parameters (three parameters for axis, one parameter for angle), 2) the Rodrigues representation avoids the problem of parallel configuration of axes known as Gimbal lock [14] (reduction in degrees of freedom) caused by using Euler angles in 3D space, 3) Rodrigues has less computational cost compared to the Euler angles which uses three matrices of rotation.

One of the recent domains in which using rotations led to achieving state-of-the-art results is learning and reasoning over Knowledge Graphs (KGs) using embeddings (KGEs). Generally, KGEs are designed for graph completion by

The associate editor coordinating the review of this manuscript and approving it for publication was Alicia Fornés.

predicting missing links on knowledge graphs. A Knowledge Graph is a multi-relational directed graph with nodes and edges represented in the form of triples (*head, relation, tail*), such that head and tail represent nodes and the relation is the edge between two nodes. For a given KG, a KGE model creates head, relation, and tail vectors as (**head, relation, tail**) for which we use the notation (**h, r, t**) in this article. Recent studies in the KGEs have shown that nodes and edges of KGs often form relational patterns [27]. The advantages of Rotations was particularly notable for learning relational patterns for which the previous KGE models encountered problems. The success of the link prediction task using a KGE model depends on its ability of learning and inferring different aspects of the KGs such as relational patterns. Formally, patterns are represented in the form of *Premise*  $\implies$  *Conclusion* where *Premise* is conjunction of several atoms (triples) and *conclusion* is an atom. For example, *affiliatedIn*(*X, Y*)  $\wedge$  *locatedIn*(*Y, Z*)  $\implies$  *livesIn*(*X, Z*) is a composition pattern with two atoms in the body and one atom in the head. A KGE model is supposed to infer that *X* lives in *Z* when it learns *affiliatedIn*(*X, Y*)  $\wedge$  *locatedIn*(*Y, Z*). Rotation based KGEs have shown to be more effective than non-rotation based KGEs due to the algebraic advantages of rotation in learning graph patterns [27]. However, the story does not end with adding a rotation capability to the model. The representation – in which the rotation is performed – affects the accuracy of the underlying model. Using axes-angle enables the aforementioned advantages and provides in addition a more flexible and representative rotation which consequently accelerates the capability of the models in learning a wide range of patterns including composition, symmetry, reflexivity, implication and equivalence. Moreover, a powerful axes-angle representation of rotation facilitates the preservation of the rotation angle, its orientation and the distance of rotated objects in an accurate way. This happens in a geometric space which subsequently provides a powerful node-clustering in the embedding space after mapping nodes of a graph to vector. Despite the positive effects of axis-angle representations and its advantages, it was so far unexplored by the already existing rotation-based KGEs (RotatE and QuatE).

In this work, the mentioned advantages of the Rodrigues rotations are exploited for learning and reasoning over Knowledge Graphs (KGs) with the objective of link prediction for graph completion. We focus on Knowledge Graph Embedding (KGE) models as one of the most used techniques for this problem. We aim at modelling relations as Rodrigues rotation to map the head to the tail vector for positive triples to determine that the triple is plausible in the vector space. Our model is called RodE, a 3D dimensional embedding model which is based on learning low-dimensional representations of the entities and relations of the underlying KG. We represent relations in a KG by axis-angle rotation to map the nodes or the underlying graph, in a vector space. The performance on link prediction tasks improves with our contributions, since RodE inherits the general characterises

of rotations. The aforementioned advantages of Rodrigues rotations can be mapped to Knowledge graphs as:

- By preserving *distance, angle* and *orientation* of vectors simultaneously, RodE enables capturing similarity of nodes in the embedding space, comprehensively.
- As the problem of parallel configuration of axes (Gimbal lock) is avoided by RodE, the vectors corresponding to head entities are mapped to the tail vectors without errors in the mappings of the vectors caused by reduced dimensions.
- Memory consumption and computational costs are optimized due to the reduced number of parameters considered for modelling rotations.

The evaluation was performed on four standard benchmarks namely WN18, FB15k, WN18RR, and FB15k-237. The experimental results demonstrate that our method outperforms the state-of-the-art models with the same embedding dimensions. The evaluation results indicate that this is still the case even if we reduce the embedding dimension.

## II. RELATED WORK

The promising results of embedding models in solving link prediction tasks has resulted in the development of several models during the last couple of years. Generally, such embedding models can be grouped into three categories [32]: Semantic-matching models, Distance-based models, and Neural Network-based (NN) models. The remaining part of this section provides a general overview of KGEs and includes a detailed description about the models that are relevant to this work.

### A. SEMANTIC-MATCHING EMBEDDING

This category includes models that measure the plausibility of a triple comparing the similarity of the latent features of the entities and relations. Therefore, the scoring functions of such KGE models are designed based on relation-specific semantic similarities between nodes. Among the semantic-matching models, QuatE [37], TuckEr [2], RESCAL [23], ComplEx [31], DisMult [35] and HolE [22] measure plausibility of edges by matching latent semantics. We look into the details of a base model and a state-of-the-art of this category later in this section. **QuatE** [37] extends embeddings from complex space into quaternion space. Similar to RotatE, QuatE represents relations as rotations. However, a rotation in quaternion space is more expressive than rotation in a complex space. A product of two quaternions  $\mathbf{q}_1 \otimes \mathbf{q}_2$  ( $\otimes$  is Hamilton product between two quaternion vectors) is equivalent to first scaling  $\mathbf{q}_1$  by magnitude  $|\mathbf{q}_2|$  and then rotating it in four dimensions. QuatE represents entities and relations as quaternion vectors  $\mathbf{h} = \mathbf{h}_1 + \mathbf{h}_2\mathbf{i} + \mathbf{h}_3\mathbf{j} + \mathbf{h}_4\mathbf{k}$ ,  $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4 \in \mathbb{R}^d$  ( $\mathbf{i}, \mathbf{j}, \mathbf{k}$  are fundamental quaternion units). The scoring function of QuatE is then

$$f(h, r, t) = \mathbf{h}' \cdot \mathbf{t} = \langle \mathbf{h}'_1, \mathbf{t}_1 \rangle + \langle \mathbf{h}'_2, \mathbf{t}_2 \rangle + \langle \mathbf{h}'_3, \mathbf{t}_3 \rangle + \langle \mathbf{h}'_4, \mathbf{t}_4 \rangle \quad (1)$$

where  $\langle \cdot, \cdot \rangle$  is an inner product of real vectors and  $\cdot$  is a quaternion vectors product.

$\mathbf{h}'$  is computed by first, normalizing relation embedding  $\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2\mathbf{i} + \mathbf{r}_3\mathbf{j} + \mathbf{r}_4\mathbf{k}$  to a unit quaternion:

$$\mathbf{r}^{(n)} = \frac{\mathbf{r}}{|\mathbf{r}|} = \frac{\mathbf{r}_1 + \mathbf{r}_2\mathbf{i} + \mathbf{r}_3\mathbf{j} + \mathbf{r}_4\mathbf{k}}{\sqrt{\mathbf{r}_1^2 + \mathbf{r}_2^2 + \mathbf{r}_3^2 + \mathbf{r}_4^2}} \quad (2)$$

and then computing Hamiltonian product between  $\mathbf{r}^{(n)}$  and  $\mathbf{h} = \mathbf{h}_1 + \mathbf{h}_2\mathbf{i} + \mathbf{h}_3\mathbf{j} + \mathbf{h}_4\mathbf{k}$ . However, the Hamiltonian product for modelling rotations is in 4D dimensional space. While the model subsumes 2D rotation, it cannot express the 3D rotation (with axis-angle representation) due to its formulation.

**Tucker** [2] encodes KG as binary tensors  $\mathcal{W} \in \mathbb{R}^{(d_e \times d_r \times d_e)}$ , where  $d_e$  is embedding dimension for entities and  $d_r$  - embedding dimension of relations. The  $ijk$ -th elements of  $\mathcal{W}$  is determined to be 1, if triple  $(h_i, r_j, t_k)$  holds, otherwise its value is -1. The scoring function is defined as:

$$f(h, r, t) = \mathcal{W} \times_1 h \times_2 r \times_3 t \quad (3)$$

where  $\times_n$  represents the mode- $n$  product [2]. The logistic sigmoid function is then applied to the score to compute probability of the triple being true. Claimed by the authors, Tucker obtains state-of-the-art performance with a relatively smaller embedding dimension for  $\mathbf{h}, \mathbf{r}, \mathbf{t}$ . However, the total number of adjustable parameters hugely increases due to using the core tensor ( $\mathcal{W}$ ) which is problematic in large-scale KGs. Among the models under this category, only QuatE uses rotations but not considering axis-angle.

### B. DISTANCE-BASED EMBEDDING MODELS

This category includes those models in which the plausibility of a triple is computed based on a distance function, which generally, is either translation-based or rotation-based. A family of translation-based models have been proposed under the category of distance-based such as TransE [6], TransD [16], TransH [34], TransR [18]. Recently, rotation-based models such as RotatE [27] are also proposed under this category. Here, we address two of the highly-performed models, one from the fundamental aspects (TransE) and one recently developed rotational model (RotatE). **TransE** [6] is one of the early embedding models which stayed with its surprisingly high performance for modelling multi-relational data despite simplicity. In this model, relations are interpreted as translations between entities to find embeddings of a triple  $(h, r, t)$  as  $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ . The scoring function is defined as:

$$f(h, r, t) = -\|\mathbf{h} + \mathbf{r} - \mathbf{t}\| \quad (4)$$

The minus sign in front of norm corresponds to high numbers showing plausibility. The simplicity of TransE and its functionality in real space makes it highly scalable. However, as can be deduced from its scoring function, TransE has difficulties representing 1-MANY, MANY-1, and MANY-MANY relations [34].

**RotatE** [27] embeds entities and relations into complex vector space and uses the following scoring function:

$$f(h, r, t) = -\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\| \quad (5)$$

where  $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^d$  ( $\mathbb{C}$  is complex space),  $\mathbf{r} = [\mathbf{r}_1, \dots, \mathbf{r}_d]$  is a unit vector, i.e.  $|\mathbf{r}_i| = 1, i = 1, \dots, d$ , and  $\circ$  is Hadamard (element-wise) product. Relation  $\mathbf{r}$  is then interpreted as a rotation from  $\mathbf{h}$  to  $\mathbf{t}$ . Unlike other KGEs, RotatE is capable of encoding relational patterns such as composition and symmetry of relations. RotatE uses complex vectors to model rotation in 2D dimensional space. Using distance in the score function, the model becomes limited to 1 to  $n$  scoring in learning (covered by Tucker [2]). In [29], a high-level analysis is represented as a non-abelian group in a generalized framework which addressed rotations in 3D. Although it uses several KGEs as examples of the addressed groups, the comparison with the recent state-of-the-art of rotation-based embedding models, namely QuatE is not covered and the reported results are sensibly lower than the performance of the QuatE model in its original report [36].

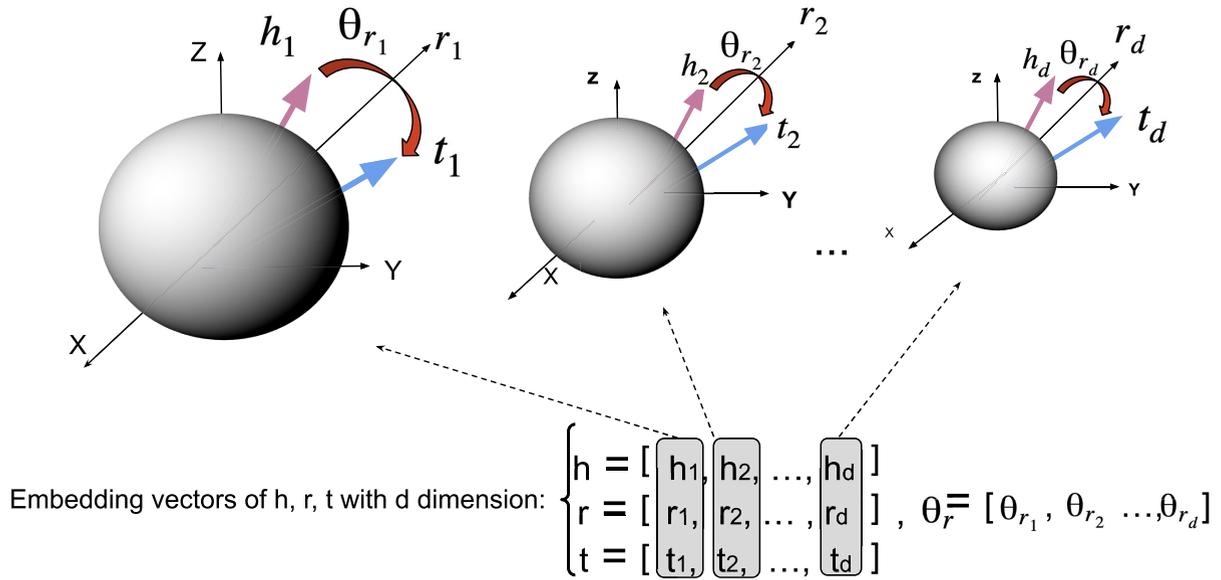
### C. NEURAL NETWORK-BASED EMBEDDING MODELS

This category considers the models in which the plausibility of a triple is computed based on a multi-layer Neural Network (NN). ConvE [9] is an example model in this category that uses a global 2D convolution operation for head and relation embedding. Then, these vectors are reshaped into matrices and concatenated. Furthermore, a linear layer and an inner product is taken with all vectors of tail entities to generate a score for each triple. Another NN-based model is NTN [26]. Since these models are not in the scope of this work and in addition, their performance are not among the state-of-the-art, we skip further details. Most of the models in this class are using ReLU activation function which results in same score for most of the triples during the evaluation. This causes an unrealistic jump in the performances of the models which is not an advantage caused by proper formulation, but it is a disadvantage caused by getting a same scoring for most of the triples – regardless of being positive or negative [28].

### III. RODRIGUES FORMULA

In mathematics, the Special Orthogonal group ( $\mathbf{SO}(q)$ ) is the group of all rotations around the origin of  $\mathbb{R}^q$ , i.e.  $q$  dimensional Euclidean space, by composition of operators for rotations [13]. Considering rotations, the following reminders are noted here from which the characteristics of RodE will be driven:

- a rotation preserves origin, angle, distance and orientation,
- a non-trivial rotation is a rotation over axis  $r$  by  $\theta_r$  degrees,
- composition of two rotations is another rotation,
- a rotation has a unique inverse rotation,
- identity map (the output is same as input) satisfies the definition of rotation.



**FIGURE 1.** Rodrigues rotation of embedding vectors. Using a relation-specific Rodrigues rotation in 3-dimensional space (x, y, z), an element-wise mapping of head (h) to tail (t) is performed in each embedding dimension [1, 2, ..., d]. Each of the 3D spheres refers to one element of (h, r, t). The relation is represented as axis and the angle is its corresponding rotation around the axis (no semantics in the size of the sphere).

We specifically consider the prominent rotation of the  $SO(3)$  group, namely Rodrigues' formula [10] as a rotation in  $\mathbb{R}^3$  (3D rotation). Rodrigues rotates a 3D vector  $\mathbf{e} \in \mathbb{R}^3$  around the axis  $\mathbf{r} \in \mathbb{R}^3$  ( $\mathbf{r}$  is a unit vector) by  $\theta$  degrees according to the right hand rule. We denote the rotation by adding the index "rot", therefore,  $\mathbf{e}_{rot}$  is the rotated vector of  $\mathbf{e}$ . The Rodrigues' formula enables the rotation as follows

$$\mathbf{e}_{rot} = \mathbf{e} \cos(\theta) + (\mathbf{r} \times \mathbf{e}) \sin(\theta) + \mathbf{r}(\mathbf{r}, \mathbf{e})(1 - \cos(\theta)), \quad (6)$$

where  $\times$  is the cross product.

Rotation can be also represented through matrix-vector multiplication as follows

$$\mathbf{e}_{rot} = \mathbf{R} \mathbf{e}, \quad (7)$$

where  $\mathbf{R}$  is an orthogonal matrix, i.e.,  $\mathbf{R} \mathbf{R}^T = \mathbf{I}$  where  $\mathbf{I}$  is the identity matrix. The Rodrigues' rotation matrix is defined as follows:

$$\begin{aligned} \mathbf{R} &= \mathbf{I} \cos(\theta) + \sin(\theta) \mathbf{K} + (1 - \cos(\theta)) \mathbf{K}^2, \\ \mathbf{K} &= \begin{bmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{bmatrix}. \end{aligned} \quad (8)$$

$\mathbf{K}$  is a skew-symmetric matrix and  $r_i$  with  $i = 1, \dots, 3$  denotes the  $i$ -th element of the axis  $\mathbf{r}$ .

#### IV. RodE EMBEDDING MODEL

The formulation of RodE uses Rodrigues' rotation which enables the mapping of nodes in a graph through edges into a vector space. This facilitates RodE to cover different categories of embedding models. In order to inherit the characteristics of different embedding categories, RodE is formulated for:

- Semantic-matching (with 1-n scoring) models: applied on KGs with a low number of entities, in particular relative to the number of triples.
- Distance-based models: best practice in large scale KGs with relatively high number of entities.

In the following sections, we first present these two formulations of RodE. Then, we show the advantages of utilizing the Rodrigues' rotation and the core formulation of the RodE score for learning the main relational patterns (e.g., symmetric, inverse) of the graph. We furthermore describe its connection to the current closest state-of-the-art model (QautE). For the implementation of the RodE model, we use the **Pytorch** framework.

#### A. RodE FOR SEMANTIC-MATCHING BASED KGES

Given a triple  $(h, r, t)$  in a KG, we aim at mapping the embedding of head  $\mathbf{h}$  to the embedding of tail  $\mathbf{t}$  via a relation-specific algebraic operation. Here, we model the relation  $r$  as an element-wise 3D rotation induced by Rodrigues' formula, shown in Figure 1. In element-wise rotation, each element of the head vector  $(\mathbf{h}_i, i = 1, \dots, d)$  is rotated by  $\theta_{r_i}$  degree around the relation axis  $\mathbf{r}_i$ . This is performed to map the head element  $\mathbf{h}_i$  to the tail element  $\mathbf{t}_i$  in  $\mathbb{R}^3$  for each fact  $(h, r, t)$  (with the purpose of approving plausibility based on the underlying model formulation). In other words, if the triple  $(h, r, t)$  is positive, the following equality approximately holds in the vector space

$$\mathbf{h}_i^T \mathbf{R}_{r_i} \approx \mathbf{t}_i^T, \quad i = 1, \dots, d. \quad (9)$$

Therefore, the scoring function is defined as

$$f(h, r, t) = \langle \mathbf{h}^T * \mathbf{R}_r, \mathbf{t} \rangle, \quad (10)$$

where  $\mathbf{h}^T$  is transpose of  $\mathbf{h}$ ,  $*$  is a block-matrix product.  $h$  is a  $d$  dimensional head embedding in which each of the elements  $h_i$  is a three dimensional vector. The same holds for the tail embedding.  $R_r$  is block-diagonal matrix

$$R_r = \text{diag}(R_{r1}, \dots, R_{rd}) = \begin{bmatrix} R_{r1} & 0 & \dots & 0 \\ 0 & R_{r2} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & R_{rd} \end{bmatrix},$$

$$h = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_d \end{bmatrix}, \quad t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_d \end{bmatrix}, \quad h_i, t_i \in \mathcal{R}^3, R_{ri} \in \mathcal{R}^{3 \times 3},$$

$$i = 1, \dots, d.$$

and  $R_{ri}$  is the Rodrigues' rotation matrix corresponding to the  $i$ th element of the embedding.

### B. RodE FOR DISTANCE-BASED KGES

Generally, distance-based models map triples of a KG into vector space by minimizing the distance of algebraically transformed head vector and the tail vector. Using RodE, in order to encode a positive sample  $(h, r, t)$  into the vector space, we first rotate each element of the head vector ( $\mathbf{h}_i$ ) by  $\theta_{ri}$  degrees. This rotation turns around the relation axis  $r_i$  to obtain a new vector  $\mathbf{h}_i^{\theta_{ri}}$ . In the next step, we translate  $\mathbf{h}_i^{\theta_{ri}}$  by another relation vector  $\mathbf{p}_r$  to meet each element of the tail vector  $\mathbf{t}_i$  as shown here:

$$\mathbf{h}_i^T \mathbf{R}_{ri} + \mathbf{p}_r^T \approx \mathbf{t}_i^T, \quad i = 1, \dots, d. \quad (11)$$

This equation represents constraints in the vector space for each positive triples  $(h, r, t)$ . Based on the constraints, we introduce the score function as:

$$f_r(h, t) = -\|\mathbf{h}^T * \mathbf{R}_r + \mathbf{p}_r^T - \mathbf{t}^T\|, \quad (12)$$

$$\text{where } \mathbf{p}_r = \begin{bmatrix} \mathbf{p}_{r1} \\ \mathbf{p}_{r2} \\ \vdots \\ \mathbf{p}_{rd} \end{bmatrix}, \quad \mathbf{p}_i \in \mathbb{R}^3, \quad i = 1, \dots, d.$$

### C. MODELLING RELATIONAL PATTERNS

Due to the main characteristics inherited from the Rodrigues rotation, RodE is capable of modelling the main patterns in graphs such as composition (e.g.  $\text{affiliatedIn}(X, Y) \wedge \text{locatedIn}(Y, Z) \implies \text{livesIn}(X, Z)$ ), and inverse ( $\text{supervisedBy}(Y, Z) \implies \text{studentOf}(Z, Y)$ ) studied in the literature. Here, we discuss pattern encoding in RodE on a set of relational patterns namely Composition, Inverse, Symmetric, Reflexive, Implication, and Equivalence. Furthermore, the encoding of these relational patterns by RodE is discussed for each pattern individually:

- **Composition:** As mentioned in Section III, the result of the composition for two rotations represents another rotation. Therefore, any rotation-based model inherits these characteristics which enables the encoding

of composition. Using axis-angle, RodE additionally employs an element-wise rotation in 3-dimension space to express the composition patterns. To show this mathematically, let  $r_1, r_2$  represent the two relations that their composition gives another relation  $r_3$ .

The following equation shows composition in RodE:

$$\begin{aligned} \mathbf{h}_i^T \mathbf{R}_{r1i} &\approx \mathbf{t}_i^T, \quad i = 1, \dots, d, \\ \mathbf{t}_i^T \mathbf{R}_{r2i} &\approx \mathbf{t}_i^T, \quad i = 1, \dots, d, \end{aligned} \quad (13)$$

from these constraints, we conclude that

$$\mathbf{h}_i^T \mathbf{R}_{r3i} \approx \mathbf{t}_i^T, \quad i = 1, \dots, d. \quad (14)$$

The above equations hold if

$$\mathbf{R}_{r1i} \mathbf{R}_{r2i} = \mathbf{R}_{r3i}, \quad i = 1, \dots, d. \quad (15)$$

Since naturally the composition of two rotations always results in another rotation, this equation holds in rotation-based approaches.

- **Inverse:** Every rotation has a unique inverse rotation. In RodE, each of these inverse rotations can be expressed as an element-wise rotation. Therefore, following this characteristic, RodE is capable of encoding inverse patterns. To show this characteristic mathematically, let us assume  $r_1, r_2$  as two relations in a KG which are in inverse relation to each other. RodE enforces

$$\begin{aligned} \mathbf{h}_i^T \mathbf{R}_{r1i} &\approx \mathbf{t}_i^T, \quad i = 1, \dots, d, \\ \mathbf{t}_i^T \mathbf{R}_{r2i} &\approx \mathbf{h}_i^T, \quad i = 1, \dots, d. \end{aligned} \quad (16)$$

By substituting the above equations and considering the inverse relation between the two relations, we can conclude that

$$\mathbf{R}_{r1i} = \mathbf{R}_{r2i}^{-1}, \quad i = 1, \dots, d. \quad (17)$$

This shows that RodE model is fully capable of encoding the inverse patterns.

- **Symmetric:** In rotation-based models, for a symmetric relation  $r$ , if the angle becomes 0 or  $\pi$ , then  $\mathbf{R}_{ri} = \mathbf{I}$  or  $-\mathbf{I}$ . The symmetry of relation  $r$  forces the following constraints in the vector space:

$$\begin{aligned} \mathbf{h}_i^T &\approx \mathbf{t}_i^T \text{ (or } -\mathbf{t}_i^T), \quad i = 1, \dots, d \\ \mathbf{t}_i^T &\approx \mathbf{h}_i^T \text{ (or } -\mathbf{h}_i^T), \quad i = 1, \dots, d. \end{aligned} \quad (18)$$

Therefore, for each positive triple  $(h, r, t)$ , RodE can express its symmetric  $(t, r, h)$  if each element of head and tail becomes either equal to the original vector or its mirror image. This enables the model to embed  $2^d$  entities in the vector space with distinct embedding vectors.

- **Reflexive:** Considering a reflexive relation  $r$ , then  $h = t = e$ . Therefore, given a reflexive relation, RodE enforces the following constraints in the vector space for every entity the underlying KG:

$$\mathbf{R}_{ri} \mathbf{e}_i \approx \mathbf{e}_i, \quad i = 1, \dots, d. \quad (19)$$

We can rewrite the approximate equality of the previous equation as an exact equality such that

$$\mathbf{R}_{ri}\mathbf{e}_i = \alpha_i\mathbf{e}_i, \alpha_i \in [0, \epsilon], \quad i = 1, \dots, d, \quad (20)$$

where  $\epsilon$  is an arbitrary small value. The result is a  $3 \times 3$  matrix  $\mathbf{R}_{ri}$  with three eigenvalues(vectors). Therefore, there are three options (equal to number of eigenvectors) for the  $i$ th element of the embedding in an entity. Since in reflexive patterns, the aim is to assign different vectors for each entity, the number of possible eigenvalues in each vector provides  $3^d$  diffident vectors. This shows that RodE has a high capacity for representation of entities when reflexivity constraint is enforced.

- **Implication:** The implication pattern holds between two relations  $r_1$  and  $r_2$  when  $r_1$  implies  $r_2$ . We show that if a model represents  $(h, r_1, t)$  as a positive triple, then it can also also be positively represented as the plausibility of  $(h, r_2, t)$  triple. The representation of this fact in RodE is the following equation:

$$\langle \mathbf{R}_{r_1i}\mathbf{h}_i, \mathbf{t}_i \rangle \leq \langle \mathbf{R}_{r_2i}\mathbf{h}_i, \mathbf{t}_i \rangle, \quad i = 1, \dots, d \quad (21)$$

The resultant output is:

$$\langle (\mathbf{R}_{r_1i} - \mathbf{R}_{r_2i})\mathbf{h}_i, \mathbf{t}_i \rangle \leq 0, \quad i = 1, \dots, d. \quad (22)$$

If we assume that  $\mathbf{R}_{12} = \mathbf{R}_{r_1i} - \mathbf{R}_{r_2i}$ , then considering only equality for the equation (i.e., = 0), we have

$$\langle \mathbf{R}_{12}\mathbf{h}_i, \mathbf{t}_i \rangle = 0, \quad i = 1, \dots, d. \quad (23)$$

For a  $(\mathbf{t}_i)$  as a three dimensional vector, there exist at least two other vectors  $\mathbf{t}_{i1}, \mathbf{t}_{i2}$  with which they are mutually orthogonal. Let us focus on  $\mathbf{t}_{i1}$ , therefore we have

$$\mathbf{R}_{12}\mathbf{h}_i = \mathbf{t}_{i1}, \quad i = 1, \dots, d. \quad (24)$$

If  $\mathbf{R}_{12}$  is invertible, then a unique solution for  $\mathbf{h}_i$  exists. Due to having two options  $\mathbf{t}_{i1}, \mathbf{t}_{i2}$ , there will be at least  $2^d$  possible options for assigning entity embedding when implication constraint is enforced. This also proves the broad ability of RodE in encoding of the implication relational patterns.

- **Equivalence:** In order to encode equivalence, same procedure as implication pattern can be followed. Assuming that  $\mathbf{R}_{12}$  is invertible, there will be exactly  $2^d$  possible options for entity embedding assignment in order to represent entities uniquely.

#### D. CONNECTION TO QuatE

QuatE considers relations as a 4D rotation. The model also covers 2D rotation due to having more degrees of freedom. However, it cannot express 3D rotation even with more degrees of freedom due to limitation of its scoring function. In order to model 3D rotation using Quaternion numbers, the following equation is proposed as an equivalent equation to (9):

$$(\cos(\theta_{ri}) + \mathbf{r}_i \sin(\theta_{ri})) \otimes \mathbf{h}_i \otimes (\cos(\theta_{ri}) - \mathbf{r}_i \sin(\theta_{ri})) \approx \mathbf{t}_i, \quad (25)$$

where  $\otimes$  is Hamiltonian product.  $(\cos(\theta_{ri}) + \mathbf{r}_i \sin(\theta_{ri}))$  is quaternion representation of relation and

$(\cos(\theta_{ri}) + \mathbf{r}_i \sin(\theta_{ri})) \otimes \mathbf{h}_i \otimes (\cos(\theta_{ri}) - \mathbf{r}_i \sin(\theta_{ri}))$  is equivalent to the Rodrigues rotation of  $\mathbf{h}_i$  with  $\theta_{ri}$  degree around the axis  $\mathbf{r}_i$ .

#### E. OPTIMAL ROTATION BY USING RodE

Let  $\mathcal{H}_r = \{\mathbf{h}_i\}_{j=1}^{n_h}$  and  $\mathcal{T}_r = \{\mathbf{t}_i\}_{j=1}^{n_t}$  be the set of all head and tail entities which are connected with the relation  $r$  in the graph. The mapping from  $\mathcal{H}_r$  to  $\mathcal{T}_r$  is performed by relation specific rotation as follows

$$\max_{\mathbf{R}_{ri}} \sum_{j=1}^n \langle \mathbf{h}_i^{jT} \mathbf{R}_{ri}, \mathbf{t}_j^j \rangle, \quad i = 1, \dots, d. \quad (26)$$

We rewrite the term  $\langle \mathbf{h}_i^{jT} \mathbf{R}_{ri}, \mathbf{t}_j^j \rangle$  in the quaternion form

$$q_{ri} \otimes \mathbf{h}_i \otimes q_{ri}^* \cdot \mathbf{t}_i. \quad (27)$$

where  $q_{ri} = \cos(\theta_{ri}) + \mathbf{r}_i \sin(\theta_{ri})$ , “.” and  $\langle \cdot, \cdot \rangle$  are quaternion and vector inner products. We have the following equality

$$q_{ri} \otimes \mathbf{h}_i \otimes q_{ri}^* \cdot \mathbf{t}_i = (q_{ri} \otimes \mathbf{h}_i) \cdot (\mathbf{t}_i \otimes q_{ri}). \quad (28)$$

The Hamilton product can be rewritten as a matrix-vector multiplication. Therefore we have

$$q_{ri} \otimes \mathbf{h}_i \otimes q_{ri}^* \cdot \mathbf{t}_i = \langle (\mathbf{H}_i \mathbf{q}_{ri}), (\mathbf{T}_i \mathbf{q}_{ri}) \rangle, \quad (29)$$

where  $\mathbf{H}_i = \begin{bmatrix} 0 & -h_{i3} & h_{i2} \\ h_{i3} & 0 & -h_{i1} \\ -h_{i2} & h_{i1} & 0 \end{bmatrix}$  and  $\mathbf{T}_i = \begin{bmatrix} 0 & -t_{i3} & t_{i2} \\ t_{i3} & 0 & -t_{i1} \\ -t_{i2} & t_{i1} & 0 \end{bmatrix}$ .

Considering Equation (26), we have

$$\begin{aligned} \max_{\mathbf{R}_{ri}} \sum_{j=1}^n \langle \mathbf{h}_i^{jT} \mathbf{R}_{ri}, \mathbf{t}_j^j \rangle &= \max_{\mathbf{q}_{ri}} \sum_{j=1}^n \langle (\mathbf{H}_i^j \mathbf{q}_{ri}), (\mathbf{T}_i^j \mathbf{q}_{ri}) \rangle \\ &= \max_{\mathbf{q}_{ri}} \sum_{j=1}^n \langle \mathbf{q}_{ri}^T \mathbf{H}_i^j \mathbf{t}_j^j \rangle \\ &= \max_{\mathbf{q}_{ri}} \mathbf{q}_{ri}^T \sum_{j=1}^n (\mathbf{H}_i^j \mathbf{T}_i^j) \mathbf{q}_{ri}, \quad i = 1, \dots, d. \end{aligned} \quad (30)$$

$\mathcal{W} = \sum_{j=1}^n \mathbf{H}_i^j \mathbf{T}_i^j$  is a  $4 \times 4$  symmetric matrix. Let  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  be the eigenvalues of the matrix and  $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4$  be the corresponding eigenvectors which are mutually orthogonal.  $\mathbf{q}_{ri}$  is linear combination (with coefficient  $\beta_1, \dots, \beta_4$ ) of the eigenvectors as

$$\mathbf{q}_{ri} = \beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \beta_3 \mathbf{u}_3 + \beta_4 \mathbf{u}_4. \quad (31)$$

Therefore,

$$\begin{aligned} \mathbf{q}_{ri}^T \mathcal{W} \mathbf{q}_{ri} &= (\beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \beta_3 \mathbf{u}_3 + \beta_4 \mathbf{u}_4)^T \mathcal{W} \\ &\quad (\beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \beta_3 \mathbf{u}_3 + \beta_4 \mathbf{u}_4) \\ &= (\beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \beta_3 \mathbf{u}_3 + \beta_4 \mathbf{u}_4) \\ &\quad (\lambda_1 \beta_1 \mathbf{u}_1 + \lambda_2 \beta_2 \mathbf{u}_2 + \lambda_3 \beta_3 \mathbf{u}_3 + \lambda_4 \beta_4 \mathbf{u}_4)^T \\ &= \lambda_1 \beta_1^2 + \lambda_2 \beta_2^2 + \lambda_3 \beta_3^2 + \lambda_4 \beta_4^2 \end{aligned} \quad (32)$$

**TABLE 1. Statistical information of the datasets. Number of entity and Number of relation present in the datasets are represented based on division between training, test triple, and validation sets.**

Dataset	Number of entity	Number of relation	Training triple	Test triple	Validation triple
FB15k	14951	1345	483142	59071	50000
FB15k-237	14541	237	272115	20466	17535
WN18	40943	18	141442	5000	5000
WN18RR	40943	11	86835	3134	3034

**TABLE 2. Entity, relation and information of rules.**

Dataset	FB15k	FB15k-237	WN18	WN18RR
# of entities	14951	14505	40943	40943
# of relations	1345	237	18	11
# of reflexive triples in training set	1803	1625	7	7
# of reflexive triples per relation in test	7780	2606	1074	1074
# of symmetric triples in training set	36331	35597	29667	29665
# of symmetric triples per relation in test	15863	3305	1397	1286
# of implication triples in training set	250406	58158	554	218
# of implication triples per relation in test	52149	14194	4959	2920
# of inverse triples in training set	391805	76132	102504	268
# of inverse triples per relation in test	57156	14012	4997	3107

The maximum of  $\mathbf{q}_{ri}^T \mathcal{W} \mathbf{q}_{ri}$  is obtained with  $\beta_1 = 1, \beta_2 = \beta_3 = \beta_4 = 0$ . Therefore, the optimal relation embedding (angle and rotation axis) will be the eigenvector corresponding to the largest eigenvalue of  $\mathcal{W}$ .

## V. EXPERIMENTS

In this section, we present the results of our experiments evaluating the performance of RodE model in comparison to other models.

### A. UNDERLYING KNOWLEDGE GRAPHS

For the training and evaluation of RodE model, the experiments are done over four standard datasets. Each of these underlying KGs for the experiments is originally divided by its provider into three sub-datasets: train, test and validation. This is a necessity for the evaluation of the learning part for KGEs. The statistics are shown in Table 1.

The list of used datasets are:

- **FB15k** is a dataset extracted from the larger **Free-Base** dataset [4] in the form of a standard KG to be used for such experiments. FB15k includes general knowledge in different domains such as movies, actors, and sports [9].
- **FB15k-237** is a refined subset of FB15k [30]. In this version of FB15k, most of the relational patterns in the form of inverse patterns are removed from the training set.
- **WN18** is a subset of WordNet dataset [20]. It contains mostly hyponym and hypernym types of relation and the data is in a strict structural hierarchy [9].
- **WN18RR** is another refined version of WN18 [9]. Similar to FB15k-237, most of the inverse patterns are removed in the WN18RR dataset.

Since the core of this research is about relational patterns, the datasets have been initially analysed in order to secure the existence of the possible studied patterns encoded by RodE. Table 2 shows the statistics about the relational patterns. For the test set, we first count the number of triples that their corresponding relations are involved in the patterns of the training set.

Additionally, we note that there might be triples in the test set whose corresponding relation have been already involved in several patterns in the training set.

Therefore, most of the triples are involved in the considered patterns for the evaluation of inference capabilities of our model. Considering fairness aspects for evaluations, here we note that the same datasets have been in-use in the original evaluation of the models considered for comparisons in our work.

### B. BASELINE MODELS FOR COMPARISONS

A set of selected state-of-the-art models are used for experiments from all the three categories of KGEs explained in Section II. This list contains **TransE** [6] and **RotatE** [27] from Distance-based model; **TuckEr** [2], **Complex** [31], and **QuatE** [37] from the Semantic-matching category, and the **ConvE** [9] model from NN-based models.

Considering completeness aspects of experiments, we employed two different loss functions for the optimization of different KGE categories. For Distance-based models, we use RotatE Loss function [27]. RotatE Loss is a variant of Margin Ranking Loss (MRL) in which the plausibility of triples are measured by putting a margin between positive and negative samples.

The multi-class logarithmic loss [17] with Nuclear 3-Norm Regularization is employed as loss function of Semantic-matching category.

**TABLE 3.** Results of empirical evaluations of RodE for FB15k and FB15k-237 in comparison to the result of other KGE models taken from TransE [37], ComplEx [37], QuatE [37], and Simple [37]. For RotatE, the original evaluation is done in [27] setting which is hardly possible to provide. The result of TuckEr is obtained from [2]. Therefore, we provided the evaluations with the same setting of our model.

Model Type	Model	FB15k					FB15k-237				
		MR	MRR	Hit@1	Hit@3	Hit@10	MR	MRR	Hit@1	Hit@3	Hit@10
Distance Based	TransE	-	0.463	0.297	0.578	0.749	357	0.294	-	-	0.465
	RotatE	32	0.699	0.585	0.788	0.872	200	0.327	0.233	0.363	0.517
	RodE	37	0.784	0.724	0.825	0.882	171	0.338	0.244	0.375	0.530
Semantic Based	TuckEr	-	0.795	0.741	0.833	0.892	-	0.358	0.266	0.394	0.544
	ComplEx	-	0.692	0.599	0.759	0.840	349	0.247	0.158	0.275	0.428
	QuatE	-	0.833	0.800	0.859	0.900	-	0.366	0.271	0.401	0.556
	Simple	-	0.727	0.660	0.773	0.838	-	-	-	-	-
	RodE	-	0.856	0.825	0.875	0.911	-	0.368	0.273	0.405	0.560
Neural Network Based	ConvE	51	0.657	0.558	0.723	0.831	244	0.325	0.237	0.356	0.501

**TABLE 4.** Results of empirical evaluations of RodE for WN18 and WN18RR in comparison to the result of other KGE models taken from TransE [37], ComplEx [37], QuatE [37], and Simple [37]. For RotatE, the original evaluation is done in [27] setting which is hardly possible to provide. The result of TuckEr is obtained from [2]. Therefore, we provided the evaluations with the same setting of our model.

Model Type	Model	WN18					WN18RR				
		MR	MRR	Hit@1	Hit@3	Hit@10	MR	MRR	Hit@1	Hit@3	Hit@10
Distance Based	TransE	-	0.495	0.113	0.888	0.943	3384	0.226	-	-	0.501
	RotatE	697	0.908	0.894	0.919	0.931	3901	0.473	0.427	0.491	0.560
	RodE	142	0.926	0.904	0.944	0.957	3100	0.479	0.433	0.494	0.575
Semantic Based	TuckEr	-	0.953	0.949	0.955	0.958	-	0.470	0.443	0.482	0.526
	ComplEx	-	0.941	0.936	0.945	0.94	5261	0.44	0.41	0.46	0.51
	QuatE	-	0.950	0.944	0.954	0.960	-	0.482	0.436	0.499	0.572
	Simple	-	0.942	0.939	0.944	0.947	-	-	-	-	-
	RodE	-	0.946	0.936	0.951	0.962	-	0.487	0.441	0.501	0.579
Neural Network Based	ConvE	374	0.943	0.935	0.946	0.956	4187	0.43	0.40	0.44	0.52

### C. HYPERPARAMETER SETTINGS

Considering validity aspects of experiments, KGE models need to be trained with multiple hyperparameter settings. The optimal hyperparameter setting used for RodE depends on the KGE category. Considering fairness aspects of the experiments, we also provide the hyperparameters used by the authors of other models, which have been provided in the section VI.

#### 1) SETTING OF RodE FOR SEMANTIC-MATCHING

The results of evaluating RodE model against a list of state-of-the-art models are shown in Table 3, and 4. For both FB15k and FB15k-237, the learning rate  $\alpha$  is 0.005 and 0.05 accordingly. The batch size  $\beta$  is 1000. The N3 regularization parameter is 0.0025 (FB15k) and 0.05 (FB15k237) [17]. In addition, FB15k has an embedding dimension  $d \in \{2000\}$ , whereas, for FB15k-237 it is 4000.

For WN18 and WN18RR, the embedding dimension  $d$ , 250 and 400 have been used. For WN18, the learning rate  $\alpha$  is kept at 0.5 and the batch size  $\beta$  is 1024. WN18RR uses a learning rate of 0.1 and a batch size of 100. In both of the cases, N3 regularizer has been used with regularization parameter  $\theta$  with value 0.01 (WN18) and 0.1 (WN18RR). For all the datasets, Adaptive Gradient Descent (Adagrad) has been used for optimization purpose.

Table 7 shows the result of RodE in Semantic-matching category with a fixed dimension  $d$  equal to 100.

This evaluation is in order to concretely show the performance of RodE in low dimensions. For FB15k, the learning rate  $\alpha$  is 0.1, batch size  $\beta$  is 1000 and regularization parameter  $\theta$  is 0.01. For FB15k-237, the batch size  $\beta$  is 100, learning rate  $\alpha$  is 0.05 and the regularization parameter  $\theta$  is 0.05. For WN18, the learning rate is 0.5, the regularization parameter is 0.1, and the batch size is 1024. For WN18RR, the batch size  $\beta$  is 100, learning rate  $\alpha$  is 0.1 and the regularization parameter  $\theta$  is 0.09.

#### 2) SETTING OF RodE FOR DISTANCE-BASED

Similar to the result of RodE in Semantic-matching, Table 3 and 4, also provide the comparison of Distance-based RodE to other models. For FB15k, a range of embedding dimension,  $d \in \{200, 250\}$  has been used. The learning rate is in the grid settings of  $\alpha \in \{0.01, 0.05, 0.1\}$ . The batch size  $\beta$  is set to 1024. Number of negative samples  $N$  is set to 50, the RotatE Loss parameter  $\gamma$  is 24.

For FB15k-237, the embedding dimension is in the range of  $d \in \{200, 250\}$ . The batch size and learning rate have been set to the same range as FB15k, and the number of negative samples is set to  $N \in \{10, 50\}$ . The optimal gamma is 10.

For WN18, the embedding dimension is kept at  $d \in \{50\}$  and  $\gamma$  is fixed to 6.0. We kept the batch size  $\beta$  to 512 and learning rate  $\alpha$  is 0.01. The number of negative sample  $N$  is 50.

**TABLE 5. Comparison of distance-based models. Evaluation on the symmetric, inverse, and implication patterns on FB15k.**

Pattern Type	Model	FB15k				
		MR	MRR	Hit@1	Hit@3	Hit@10
Symmetric	TransE	7	0.385	0.000	0.732	0.906
	RotatE	5	0.474	0.022	0.924	0.972
	RodE	3	0.743	0.527	0.952	0.978
Inverse	TransE	15	0.775	0.698	0.828	0.911
	RotatE	12	0.864	0.816	0.901	0.943
	RodE	10	0.906	0.884	0.919	0.948
Implication	TransE	5	0.908	0.859	0.952	0.979
	RotatE	5	0.920	0.879	0.959	0.978
	RodE	3	0.938	0.913	0.959	0.982

**TABLE 6. Comparison of semantic-based models. Evaluation on the symmetric, inverse, and implication patterns on FB15k-237.**

Pattern Type	Model	FB15k-237				
		MRR	Hit@1	Hit@3	Hit@10	
Symmetric	ComplEx	0.206	0.017	0.244	0.690	
	QuatE	0.205	0.010	0.256	0.698	
	RodE	0.207	0.012	0.257	0.701	
Inverse	ComplEx	0.227	0.034	0.268	0.723	
	QuatE	0.227	0.025	0.283	0.749	
	RodE	0.223	0.025	0.284	0.750	
Implication	ComplEx	0.204	0.038	0.241	0.616	
	QuatE	0.204	0.033	0.256	0.623	
	RodE	0.218	0.020	0.265	0.727	

For WN18RR, the embedding dimension,  $d \in \{300\}$ , number of negative sample  $N$  is 50, batch size  $\beta$  is 512,  $\gamma$  is 6.0, learning rate  $\alpha$  is 0.1.

### 3) SETTING FOR OTHER MODELS

In Table 3 and Table 4, we provide comparisons of other models with RodE (colored cells in all of the tables point to the best performing models). In both tables, we compare the results to report the performances obtained by **TransE** from [37], and **QuatE** from [37]. The result of RotatE is obtained by running it with the optimal hyper parameters used in RotatE [27], but with reduced embedding dimension (in this case 250). The results of **TuckEr**, shown in Table 4, have been reported from [2]. **QuatE** fixes the embedding dimension  $d$  to 1000 and uses multi-class log-loss with N-3 norm regularization. In **TuckEr** for FB15k, entity embedding is 200 and relation dimension is 30. For WN18 and WN18RR both entity and relation embeddings have an embedding dimension  $d \in \{200\}$ .

### D. SETTINGS FOR CLUSTERING TYPES

As an additional evaluation, we consider reporting the similarities of type entities, (in Section VI-B). The plots use **T-SNE** [19] for clustering of entity types, both is Semantic-matching, and distance based matching. We have selected 6 different types to be considered for the visualization of the clustering mechanism of RodE inherited from strength

of 3d rotation. The list of these types and their corresponding number of entities are:

- /film/film (1947 entities fall into this type)
- /tv/tv\_program (277 entities fall into this type).
- /music/instrument (99 entities fall into this type).
- /location/statistical\_region (1947 entities fall into this type).
- /music/group\_member (1118 entities fall into this type).
- /organization/endowed\_organization (450 entities fall into this type).

In order to select these clusters, we considered semantics of types. In our strategy, the initial step was to avoid selection of distinct types. For example, the type */business/employer* is in meaning the same or very similar to */organization/endowed-organization*. Secondly, we analysed the clusters after plotting them, in order to derive reasoning behind RodE for clustering output.

T-SNE is used to reduce high dimensional entity embedding to 2-dimension embeddings. We set the perplexity hyperparameter to 50 and the number of iteration to 1000 while initialization of T-SNE.

### E. EVALUATION METRICS

Standard link prediction measurements namely Mean Rank (MR), Mean Reciprocal Rank (MRR), and Hits at (1, 3 and 10) are the metrics used in our evaluations. The link prediction results are obtained in filtered settings as it is done in [27], where ranking has been performed for the test triples against all the other corrupted triples (by either corrupting the head or tail of the triplet). These candidates of the corrupted triples do not reside in training, test or, validation set.

## VI. RESULTS AND DISCUSSION

In this section, we present the results of the evaluation of RodE. The methods employed in our evaluation gives a different perspective about RodE's performance in link prediction tasks:

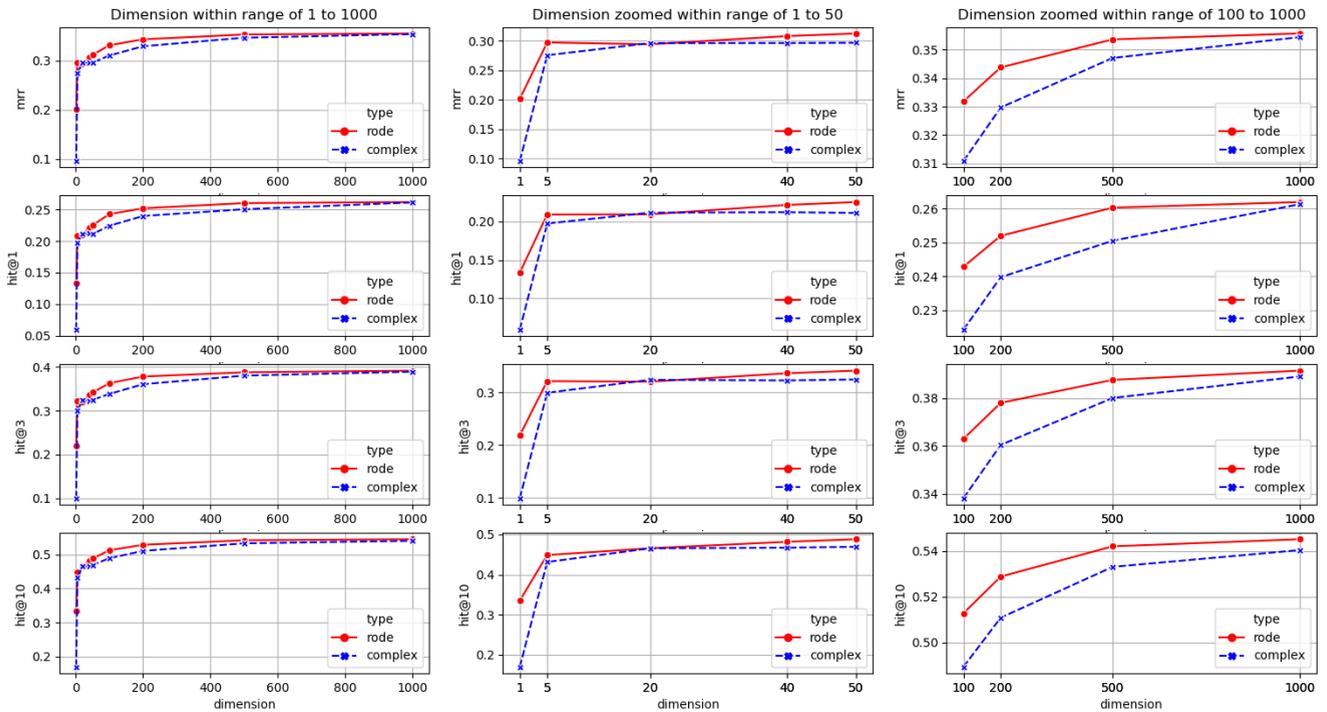
- Empirical Evaluation in High Dimension
- Empirical Evaluation in Low Dimension
- Comparison of Performance w.r.t. Changes in Dimension
- Similarity Evaluation of Types forming Clusters

The remainder of this section includes the detailed results corresponding to each of these aspects.

### A. EMPIRICAL EVALUATION

The results of running experiments on FB15k and FB15k-237 considering RodE both for Distance-based and Semantic-Matching categories are represented in Table 3. Similarly, Table 4 contains the results for WN18 and WN18RR datasets.

RodE outperforms the recent Distance-based model RotatE in all of the comparison metrics on the FB15k and FB15k-237 datasets. For this evaluation, we reproduced RotatE and used the same dimension for comparisons. For example on FB15k-237, RodE obtains 0.338 in MRR while



**FIGURE 2.** Incremental changes of dimensions and their influence on the performance of RodE and Complex models is depicted. A general visualization is provided for dimensions ranging from 1 to 1000 (left column). Zoomed version of performance changes are shown in middle column. High dimensions ranging from 100 to 1000 are in the third column visualizations (first column in the right side). Each row shows one of the metrics of MRR, Hits@1, Hits@3, and Hits@10.– stronger ones are marked darker).

**TABLE 7.** Evaluation Results for 100 Dimensions. RodE is analysed with a focus on low dimension considering MRR, Hit@1, Hit@3, and Hit@10 as comparison metrics.

Model Type	FB15k				FB15k-237				WN18				WN18RR			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
ComplEx [17]	0.83	0.79	0.85	0.89	0.35	0.26	0.39	0.54	0.95	0.94	0.95	0.95	0.47	0.43	0.49	0.56
Semantic Based RodE	0.83	0.78	0.86	0.90	0.36	0.26	0.39	0.55	0.95	0.94	0.95	0.96	0.48	0.44	0.49	0.57

RotatE obtains 0.327. On WN18 and WN18RR, RodE outperforms RotatE in all comparison metrics.

In the Semantic-matching category, RodE outperforms recent state-of-the-art models namely TuckEr and QuatE on FB15k and FB15k-237 datasets. The Hits@3 of RodE on FB15k-237 is 0.405 while QuatE achieves 0.401. Similar to these results are observed in FB15k such that the MRR for RodE is 0.856 while QuatE is 0.833. For the same category (semantic-matching models) but on WN18RR, the Hits@10 of RodE, 0.579, significantly outperforms TuckEr (0.526). However, RodE only slightly performs better than QuatE in this setting. On WN18, TuckEr performs slightly better than RodE in all the metrics except Hits@10. In addition, we provide an evaluation for the performance of RodE also in low dimensional embeddings in Table 7. For this purpose, we compare RodE with ComplEx model which is chosen because it outperforms other models that could be considered for this evaluation. ComplEx is trained with N-3 norm regularization. The results show that RodE outperforms ComplEx in most of the metrics. In Hits@10 on FB15k, we obtain 90% while ComplEx obtains 89%. In WN18RR,

RodE shows a performance of 48%, 44%, 57% respectively in MRR, Hits@1 and Hits@10; whereas, in ComplEx, these results are 47%, 43%, 56% accordingly.

In order to further investigate on how our model performs in a wide range of dimensions, we provide results for comparison of RodE’s performance w.r.t. dimension changes shown in Figure 2. This evaluation includes a set of dimensions ranging from 1 to 1000. The sub-figures in the first column are gathered based on dimension withing range of 1 to 1000. In both models of RodE and ComplEx, the performance improves incrementally by increasing the dimension. The changes are more visible when the dimension is still under 500, and slowly converge after this threshold. Sub-figures collected in second column, are a zoomed version of the evaluation below dimension 50. This additionally confirms that RodE outperforms ComplEx in low dimensions. Third column is a collective visualization of performances in dimensions between 100 to 1000. These results show that in high dimensions, the performance of other models get closer to RodE while their differences are more significant in low dimensions. In mapping and measuring of embeddings for

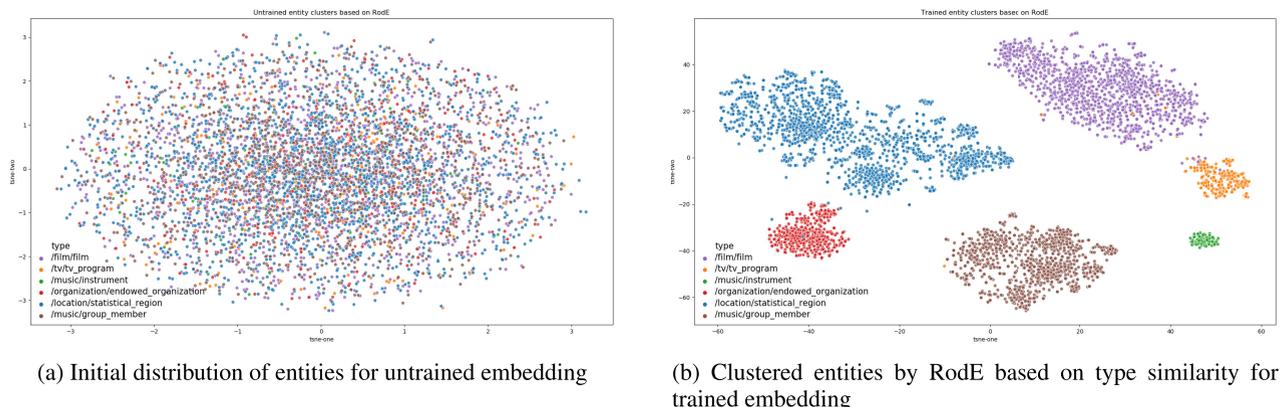


FIGURE 3. Initial distribution of entities against clustered types by Distance-based RodE.

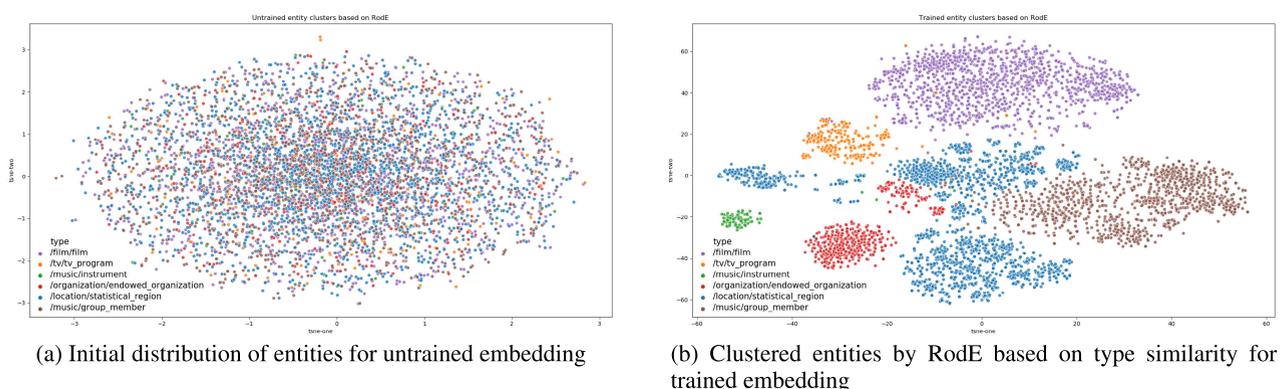


FIGURE 4. Initial distribution of entities against clustered types by Semantic-matching RodE.

large-scale KGs, high performance in low dimensions has advantages for memory usage for scalability issues.

### B. NODES DISTRIBUTIONS BASED ON TYPES

With a focus on nodes in a KG, we provide an evaluation and visualization for distribution of types. Our motivation is to assess the strength of the main advantages of 3D rotation from angle, orientation and distance preservation in the embedding space. With this advantage, the model is expected to efficiently capture the similarity between the nodes of a graph in the vector space. In order to do so, we selected a list of types (Section V-D) which appear in FB15k and filtered the nodes of these types. Furthermore, we retrieved the embedding vectors of these nodes, before and after training with RodE. The visualizations in the left hand side of the Figure 3 and Figure 4 are the raw distribution of untrained vectors, for both KGE categories of RodE. This shows that the corresponding vectors of such types are mixed in the initial learning phase. This is particularly important since prior knowledge (about type of entities) is not used neither in the initialization phase nor in the training phase. However, our model is able to properly distinguish the types. This state is a real world problems because not every KG has the information about types available for a learning algorithm.

The figures in the right side of 3 and 4 show the harmonious clustering of the embedding vectors based on their inner-type similarities after they are trained by RodE. For the 6 chosen types, RodE is capable of clearly splitting nodes belonging to each of these types into exactly 6 clusters. Our insights are that Distance-based version of RodE has a slightly better clustering function than the semantic based version because of inclusive design for translation and rotation at the same time.

### C. PATTERN-BASED PERFORMANCE ANALYSIS

In Table 5, we focus on evaluating distance-based models on symmetric, inverse, and implication patterns. For this evaluation, the test set is designed to include the top frequent relations with highest number of triples in the considered pattern per dataset (statistics are in Table 2). For example, in symmetric patterns of FB15k, the relation “/award/award\_nominee.../award\_nominee” with 12958 triples has the majority of involvement in this pattern, the second relation is “/award/award...\_honor/award\_winner” with 6866 triples, and the third relation is “/music/performance\_role/.../track\_contribution/role” with 3068 number of triples. The distance-based RodE significantly outperforms other models of the same category.

For example, considering symmetric pattern in FB15K dataset, it achieves a performance of 0.527 in Hits@1 while TransE shows a 0.000 performance and RotatE reaches on 0.022 percentage.

In Table 6, we report the results of semantic-based RodE in comparison to ComplEx and QuatE on FB15k-237 for the same patterns. RodE and QuatE show a close performance in learning of symmetric patterns. However, for the Hits@10 of RodE in learning the implication patterns significantly outperforms the other two models. RodE in also shows a better performance in comparison to ComplEx both in implication and inverse.

## VII. CONCLUSION

In this work, we showed how to use the advantages of 3-dimensional rotations with axis-angle representation. The methodology of this work is based on Rodrigues Formula used in Knowledge Graph Embedding models for link prediction tasks. Overall, our results show that using Rodrigues rotation both in semantic-matching and distance-based embeddings improved the results. In terms of accuracy in link prediction, the semantic-matching version of RodE outperforms the distance-based version on RodE. Experiments on the entity clustering based on entity type over the FB15K dataset show that RodE inherits the advantages of Rodrigues Rotation namely distance, angle and orientation preservation. Both versions of RodE perform accurate entity type clustering even without any prior knowledge about the entity type that could be injected in the model. However, the distance-based version on RodE provided a more accurate results in clustering types in comparison to its semantic-based version. In future work, we will further investigate the NN-based category of KGEs with a focus on Rodrigues rotation. Furthermore, we target applications of RodE for large-scale KGs. Given that RodE performs well even with low embedding dimensions, it may be particularly suitable for large-scale applications. Therefore, we plan to use it for large-scale KGs.

## REFERENCES

- [1] S. L. Altmann, *Rotations, Quaternions, and Double Groups*. New York, NY, USA: Courier Corporation, 2005.
- [2] I. Balažević, C. Allen, and T. M. Hospedales, "Tucker: Tensor factorization for knowledge graph completion," 2019, *arXiv:1901.09590*. [Online]. Available: <http://arxiv.org/abs/1901.09590>
- [3] G. Baym, *Lectures on Quantum Mechanics*. Boca Raton, FL, USA: CRC Press, 2018.
- [4] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A collaboratively created graph database for structuring human knowledge," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, 2008, pp. 1247–1250.
- [5] I. Bonev, D. Zlatanov, and C. Gosselin, "Advantages of the modified euler angles in the design and control of pkms," in *Proc. Parallel Kinematic Mach. Int. Conf.*, 2002, pp. 171–188.
- [6] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2787–2795.
- [7] K. Choromanski, A. Pacchiano, J. Pennington, and Y. Tang, "KAMA-NNs: Low-dimensional rotation based neural networks," in *Proc. 22nd Int. Conf. Artif. Intell. Statist.*, 2019, pp. 236–245.
- [8] J. S. Dai, "Euler–Rodrigues formula variations, quaternion conjugation and intrinsic connections," *Mechanism Mach. Theory*, vol. 92, pp. 144–152, Oct. 2015.
- [9] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, "Convolutional 2D knowledge graph embeddings," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1–10.
- [10] M. Franciscus, *Rodrigues' Rotation Formula*. Pennsylvania, PA, USA: Strupress, 2011.
- [11] G. Gallego and A. Yezzi, "A compact formula for the derivative of a 3-D rotation in exponential coordinates," *J. Math. Imag. Vis.*, vol. 51, no. 3, pp. 378–384, Mar. 2015.
- [12] J. W. Gibbs, *Vector Analysis: A Text-Book for the Use of Students of Mathematics and Physics, Founded Upon the Lectures of J. Willard Gibbs*. New Haven, CT, USA: Yale Univ. Press, 1922.
- [13] P. R. Girard, *Quaternions, Clifford Algebras and Relativistic Physics*. Birkhäuser, Basel: Springer, 2007.
- [14] E. G. Hemingway and O. M. O'Reilly, "Perspectives on euler angle singularities, gimbal lock, and the orthogonality of applied forces and applied moments," *Multibody Syst. Dyn.*, vol. 44, no. 1, pp. 31–56, Sep. 2018.
- [15] J. Huang, F. Nie, and H. Huang, "Spectral rotation versus K-means in spectral clustering," in *Proc. 27th AAAI Conf. Artif. Intell.*, 2013, pp. 431–437.
- [16] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, "Knowledge graph embedding via dynamic mapping matrix," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2015, pp. 687–696.
- [17] T. Lacroix, N. Usunier, and G. Obozinski, "Canonical tensor decomposition for knowledge base completion," in *Proc. ICML*, 2018, pp. 2863–2872.
- [18] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2181–2187.
- [19] L. V. D. Maaten and G. Hinton, "Visualizing data using t-sne," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
- [20] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [21] R. M. Murray, Z. Li, S. S. Sastry, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*. Boca Raton, FL, USA: CRC Press, 1994.
- [22] M. Nickel, L. Rosasco, and T. Poggio, "Holographic embeddings of knowledge graphs," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 1955–1961.
- [23] M. Nickel, V. Tresp, and H.-P. Kriegel, "A three-way model for collective learning on multi-relational data," in *Proc. ICML*, vol. 11, 2011, pp. 809–816.
- [24] L.-P. Rivest and T. Chang, "Regression and correlation for  $3 \times 3$  rotation matrices," *Can. J. Statist.*, vol. 34, no. 2, pp. 187–202, 2006.
- [25] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.
- [26] R. Socher, D. Chen, C. D. Manning, and A. Ng, "Reasoning with neural tensor networks for knowledge base completion," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 926–934.
- [27] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang, "RotatE: Knowledge graph embedding by relational rotation in complex space," 2019, *arXiv:1902.10197*. [Online]. Available: <http://arxiv.org/abs/1902.10197>
- [28] Z. Sun, S. Vashishth, S. Sanyal, P. Talukdar, and Y. Yang, "A re-evaluation of knowledge graph completion methods," 2019, *arXiv:1911.03903*. [Online]. Available: <http://arxiv.org/abs/1911.03903>
- [29] T. Yang, L. Sha, and P. Hong, "NagE: Non-abelian group embedding for knowledge graphs," 2020, *arXiv:2005.10956*. [Online]. Available: <http://arxiv.org/abs/2005.10956>
- [30] K. Toutanova and D. Chen, "Observed versus latent features for knowledge base and text inference," in *Proc. 3rd Workshop Continuous Vector Space Models Compositionality*, 2015, pp. 57–66.
- [31] T. Trouillon, J. Welbl, S. Riedel, E. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 2071–2080.
- [32] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2724–2743, Dec. 2017.
- [33] R. Y. Wang, K. Pulli, and J. Popović, "Real-time enveloping with rotational regression," in *Proc. ACM SIGGRAPH Papers*, 2007, p. 73.

[34] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014., pp. 1112–1119.

[35] B. Yang, W.-T. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," 2014, *arXiv:1412.6575*. [Online]. Available: <http://arxiv.org/abs/1412.6575>

[36] S. Zhang, Y. Tay, L. Yao, and Q. Liu, "Quaternion knowledge graph embeddings," 2019, *arXiv:1904.10281*. [Online]. Available: <http://arxiv.org/abs/1904.10281>

[37] S. Zhang, Y. Tay, L. Yao, and Q. Liu, "Quaternion knowledge graph embeddings," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 2731–2741.



**MOJTABA NAYERI** received the B.S. and M.S. degrees in computer engineering from the Ferdowsi University of Mashhad, Mashhad, Iran, in 2011 and 2014, respectively. He is currently pursuing the Ph.D. degree with the Smart Data Analysis Group, University of Bonn, Germany. Since 2020, he has been a Research Assistant with the Nature-Inspired Machine Intelligence Research Group, Institute of Applied Informatics (InfAI). His current research interests include machine learning, knowledge graphs, pattern recognition, and the semantic web.



**MIRZA MOHTASHIM ALAM** received the bachelor's degree from the Department of Computer Science, BRAC University, Bangladesh. He is currently pursuing the master's degree with the Department of Computer Science, University of Bonn. He is currently a student Assistant with the Nature-Inspired Machine Intelligence Research Group, Institute of Applied Informatics (InfAI). His research interests include knowledge graph analysis, machine learning, and pattern recognition. He was awarded the Vice-Chancellors Gold Medal for his outstanding performance.



**JENS LEHMANN** received the master's degree in computer science from the Technical University of Dresden and the University of Bristol and the Ph.D. degree (*summa cum laude*) from the University of Leipzig. He is currently the Head of the Smart Data Analysis Research Group, a Full Professor with the University of Bonn, and the Lead Scientist of Fraunhofer IAIS. He contributed to various open-source projects such as DL-Learner, SANSa, LinkedGeoData, and DBpedia. He has authored more than 100 publications, which were cited more than 18 000 times and won 12 international awards. His research interests include semantic web technologies, question answering, machine learning, and knowledge graph analysis.



**SAHAR VAHDATI** received the M.Sc. and Ph.D. degrees in computer science from the University of Bonn. She has been a Senior Researcher and a Postdoctoral Researcher with Oxford University, U.K. She is currently leading the Nature-Inspired Machine Intelligence Research Group, Institute of Applied Informatics (InfAI), University of Leipzig. Her research interests include using knowledge representation, analyze knowledge graphs, and artificial intelligence (AI).

...