NestE: Modeling Nested Relational Structures for Knowledge Graph Reasoning

Bo Xiong[1], Mojtaba Nayyeri[1], Linhao Luo[2], Zihao Wang[1], Shirui Pan[3], Steffen Staab[1,4]

[1]University of Stuttgart, Stuttgart, Germany
[2]Monash University, Melbourne, Australia
[3]Griffith University, Queensland, Australia
[4]University of Southampton, Southampton, United Kingdom
bo.xiong@ki.uni-stuttgart.de

Abstract

Reasoning with knowledge graphs (KGs) has primarily focused on triple-shaped facts. Recent advancements have been explored to enhance the semantics of these facts by incorporating more potent representations, such as hyper-relational facts. However, these approaches are limited to *atomic facts*, which describe a single piece of information. This paper extends beyond *atomic facts* and delves into *nested facts*, represented by quoted triples where subjects and objects are triples themselves (e.g., ((*BarackObama*, *holds_position*, *President*), *succeed_by*, (*DonaldTrump*, *holds_position*, *President*))). These nested facts enable the expression of complex semantics like *situations* over time and *logical patterns* over entities and relations. In response, we introduce NestE, a novel KG embedding approach that captures the semantics of both atomic and nested factual knowledge. NestE represents each atomic fact as a 1×3 matrix, and each nested relation is modeled as a 3×3 matrix that rotates the 1×3 atomic fact matrix through matrix multiplication. Each element of the matrix is represented as a complex number in the generalized 4D hypercomplex space, including (spherical) quaternions, hyperbolic quaternions, and split-quaternions. Through thorough analysis, we demonstrate the embedding's efficacy in capturing diverse logical patterns over nested facts, surpassing the confines of first-order logic-like expressions. Our experimental results showcase NestE's significant performance gains over current baselines in triple prediction and conditional link prediction. The code and pre-trained models are open available at https://github.com/xiongbo010/NestE.

Introduction

Knowledge graphs (KGs) depict relationships between entities, commonly through triple-shaped facts such as (*JoeBiden*, *holds_position*, *VicePresident*). KG embeddings map entities and relations into a lower-dimensional vector space while retaining their relational semantics. This empowers the effective inference of missing relationships between entities directly from their embeddings. Prior research (Bordes et al. 2013; Trouillon et al. 2016) has primarily centered on embedding triple-shaped facts and predicting the missing elements of these triples. Yet, to augment the triple-shaped representations, recent endeavors explore knowledge that extends beyond these triples. For instance, n-ary facts (Liu, Yao, and Li 2020; Fatemi et al. 2020) describe relationships between multiple entities, and hyper-relational facts (Galkin et al. 2020; Xiong et al. 2023a) augment primal triples with key-value qualifiers that provide contextual information. These approaches allow for expressing complex semantics and enable answering more sophisticated queries with additional knowledge (Alivanistos et al. 2022).

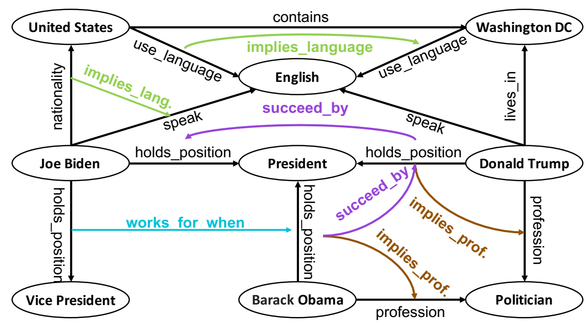

Figure 1: An example of a nested factual KG consisting of 1) a set of atomic facts describing the relationship between entities and 2) a set of nested facts describing the relationship between atomic facts. Nested factual relations are colored and they either describe situations in/over time (e.g., *succeed_by* and *works_for_when*) or logical patterns (e.g., *implies_profession* and *implies_language*).

However, these beyond-triple representations typically focus only on relationships between entities that jointly define an *atomic fact*, overlooking the significance of relationships that describe multiple facts together. Indeed, within a KG, each atomic fact may have a relationship with another atomic fact. Consider the following two atomic facts: T_1=(*JoeBiden*, *holds_position*, *VicePresident*) and T_2=(*BarackObama*, *holds_position*, *President*). We can depict the scenario where *JoeBiden* held the position of *VicePresident* under the *President BarackObama* using a triple $(T_1, \textit{works\_for\_when}, T_2)$. Such a fact about facts is referred to as a *nested fact* [1] and the relation connecting these two facts is termed a *nested relation*. Fig. 1 provides an illustration of a KG containing both atomic and nested facts.

[1]This is also called a *quoted triple* in RDF star (Champin 2022).

These nested relations play a crucial role in expressing complex semantics and queries in two ways: 1) **Expressing situations involving facts in or over time**. This facilitates answering complex queries that involve multiple facts. For example, KG embeddings face challenges when addressing queries like "*Who was the president of the USA after* DonaldTrump*?*" because the query about the primary fact (?, *holds_position*, *President*) depends on another fact (*DonaldTrump*, *holds_position*, *President*). As depicted in Fig. 1, *succeed_by* conveys such temporal situation between these two facts, allowing the direct response to the query through conditional link prediction. 2) **Expressing logical patterns (implications) using a non-first-order logical form** $\psi \xrightarrow{\hat{r}} \phi$. As illustrated in Fig. 1, (*Location A*, *uses_language*, *Language B*) $\xrightarrow{\textit{implies\_language}}$ (*Location in A*, *uses_language*, *Language B*) represents a logical pattern, as it holds true for all pairs of (*Location A*, *Location in A*). Modeling such logical patterns is crucial as it facilitates generalization. Once these patterns are learned, new facts adhering to these patterns can be inferred. A recent study (Chung and Whang 2023) explored link prediction over nested facts.[2] However, their method embeds facts using a multilayer perceptron (MLP), which fails to capture essential logical patterns and thus has limited generalization capabilities.

In this paper, we introduce NestE, an innovative approach designed to embed the semantics of both atomic facts and nested facts that enable representing temporal situations and logical patterns over facts. NestE represents each atomic fact as a 1×3 hypercomplex matrix, with each element signifying a component of the atomic fact. Furthermore, each nested relation is modeled through a 3×3 hypercomplex matrix that rotates the 1×3 atomic fact matrix via a matrix-multiplicative Hamilton product. Our matrix-like modeling for facts and nested relations demonstrates the capacity to encode diverse logical patterns over nested facts. The modeling of these logical patterns further enables efficient modeling of logical rules that extend beyond the first-order-logic-like expressions (e.g., Horn rules). Moreover, we propose a more general hypercomplex embedding framework that extends the quaternion embedding (Zhang et al. 2019) to include hyperbolic quaternions and split-quaternions. This generalization of hypercomplex space allows for expressing rotations over hyperboloid, providing more powerful and distinct inductive biases for embedding complex structural patterns (e.g., hierarchies). Our experimental findings on triple prediction and conditional link prediction showcase the remarkable performance gain of NestE.

Related Work

Beyond-Triple KGs To enrich the semantics of triple-base KGs, several lines of work have explored more powerful representations (Xiong et al. 2023b). Temporal KGs (Dasgupta, Ray, and Talukdar 2018; Leblay and Chekol 2018; Wang et al. 2023) introduce an additional timestamp to each triple to specify the temporal validity of the fact. Hyper-relational facts (Guan et al. 2020; Rosso, Yang, and Cudré-Mauroux 2020; Galkin et al. 2020; Xiong et al. 2023a) attach a set of key-value qualifiers to the primal triple, where each qualifier specify certain semantics of the primal fact. N-ary facts (Liu, Yao, and Li 2020, 2021; Fatemi et al. 2020) represent a fact as an abstracted relationship between n entities. Bilinear models are generalized to n-ary facts by replacing the bilinear product with multilinear products (Liu, Yao, and Li 2020, 2021). These representations capture relationships between entities or between entities and facts, but they do not capture the relationships between multiple facts.

Describing relationships between facts Rule-based approaches (Niu et al. 2020; Meilicke et al. 2019; Demeester, Rocktäschel, and Riedel 2016; Guo et al. 2016; Yang, Yang, and Cohen 2017; Sadeghian et al. 2019) consider relationships between facts, but they are confined to first-order-logic-like expressions (i.e., Horn rules), i.e., $\forall e_1, e_2, e_3$: $(e_1, r_1, e_2) \wedge (e_2, r_2, e_3) \Rightarrow (e_1, r_3, e_3)$, where there must exist a path connecting e_1, e_2, and e_3 in the KG. Notably, (Chung and Whang 2023) marked an advancement by examining KG embeddings with relationships between facts as nested facts, denoted as $(x, r_1, y) \stackrel{\hat{r}}{\Rightarrow} (p, r_2, q)$. The proposed embeddings (i.e., BiVE-Q and BiVE-B[3]) concatenate the embeddings of the head, relation, and tail, subsequently embedding them via an MLP. However, such modeling does do not explicitly capture crucial logical patterns over nested facts, which bear significant importance in KG embeddings.

Algebraic and geometric embeddings Algebraic embeddings like QuatE (Zhang et al. 2019) and BiQUE (Guo and Kok 2021) represent relations as algebraic operations and score triples using inner products. They can be viewed as a unification of many earlier functional (Bordes et al. 2013) and multiplication-based (Trouillon et al. 2016) models. Geometric embeddings like hyperbolic embeddings (Chami et al. 2020; Balazevic, Allen, and Hospedales 2019) further extend the functional models to non-Euclidean hyperbolic space, enabling the representation of hierarchical relations.

Preliminaries

A KG is denoted as a graph $G = (\mathcal{V}, \mathcal{R}, \mathcal{T})$, where $\mathcal{V}$ represents the set of entities, $\mathcal{R}$ stands for the set of relation names, and $\mathcal{T} = \{(h, r, t) : h, t \in \mathcal{V}, r \in \mathcal{R}\}$ represents the set of triples. We refer to G as an atomic factual KG, and each $(h, r, t) \in \mathcal{T}$ is referred to as an atomic triple. The nested triple and nested factual KG are defined as follows.

Definition 1 (Nested Triple)**.** *Given an atomic factual KG $G = (\mathcal{V}, \mathcal{R}, \mathcal{T})$, a set of nested triples is defined by $\widehat{\mathcal{T}} = \{\langle T_i, \hat{r}, T_j\rangle : T_i, T_j \in \mathcal{T}, \hat{r} \in \widehat{\mathcal{R}}\}$, where $\mathcal{T}$ is the set of atomic triples and $\widehat{\mathcal{R}}$ is the set of nested relation names.*

Definition 2 (Nested Factual Knowledge Graph)**.** *Given a KG $G = (\mathcal{V}, \mathcal{R}, \mathcal{T})$, a set of nested relation names $\widehat{\mathcal{R}}$, and a*

[2]In their work, the KG is referred to as a bi-level KG, and the term "high-level facts" is synonymous with nested facts.

[3]Note that BiVE-B, despite being described as based on the biquaternian–BiQUE, employs quaternion space with an additional translation component based on our analysis of the code.

set of nested triples $\widehat{\mathcal{T}}$ defined on G and $\widehat{\mathcal{R}}$, a nested factual KG is defined as $\widehat{G} = (\mathcal{V}, \mathcal{R}, \mathcal{T}, \widehat{\mathcal{R}}, \widehat{\mathcal{T}})$.

We can now define triple prediction and conditional link prediction (Chung and Whang 2023) as follows.

Definition 3 (Triple Prediction)**.** *Given a nested factual KG $\widehat{G} = (\mathcal{V}, \mathcal{R}, \mathcal{T}, \widehat{\mathcal{R}}, \widehat{\mathcal{T}})$, the triple prediction problem involves answering a query $\langle T_i, \widehat{r}, ?t\rangle$ or $\langle h?, \widehat{r}, T_j\rangle$ with $T_i, T_j \in \mathcal{T}$ and $\widehat{r} \in \widehat{\mathcal{R}}$, where the variable $?h$ or $?t$ needs to be bounded to an atomic triple within $\widehat{G}$.*

Definition 4 (Conditional Link Prediction)**.** *Given a nested factual KG $\widehat{G} = (\mathcal{V}, \mathcal{R}, \mathcal{T}, \widehat{\mathcal{R}}, \widehat{\mathcal{T}})$, let $T_i = (h_i, r_i, t_i)$ and $T_j = (h_j, r_j, t_j)$. The conditional link prediction problem involves queries $\langle T_i, \widehat{r}, (h_j, r_j, ?)\rangle$, $\langle T_i, \widehat{r}, (?, r_j, t_j)\rangle$, $\langle (h_i, r_i, ?), \widehat{r}, T_j\rangle$, or $\langle (?, r_i, t_i), \widehat{r}, T_j\rangle$, where the variables need to be bound to entities within $\widehat{G}$.*

NestE: Embedding Atomic and Nested Facts

Unified Hypercomplex Embeddings

We first extend QuatE (Zhang et al. 2019), a KG embedding in 4D hypercomplex quaternion space, into a more general 4D hypercomplex number system including three variations: (spherical) quaternions, hyperbolic quaternions, and split quaternions. Each of these 4D hypercomplex numbers is composed of one real component and three imaginary components denoted by $s + xi + yj + zk$ with $s, x, y, z \in \mathbb{R}$ and i, j, k being the three imaginary parts. The distinctive feature among these hyper-complex number systems lies in their multiplication rules of the imaginary components.

(Spherical) quaternions $\mathcal{Q}$ follow the multiplication rules:

$$\begin{aligned} &i^2 = j^2 = k^2 = 1, \\ &ij = k = -ji, jk = i = -kj, ki = j = -ik. \end{aligned} \tag{1}$$

Hyperbolic quaternions $\mathcal{H}$ follow the multiplication rules:

$$\begin{aligned} &i^2 = -1, j^2 = k^2 = 1, \\ &ij = k, jk = -i, ki = j,\ ji = -k, kj = i, ik = -j. \end{aligned} \tag{2}$$

Split quaternions $\mathcal{S}$ follow the multiplication rules:

$$\begin{aligned} &i^2 = -1, j^2 = k^2 = 1, \\ &ij = k, jk = -i, ki = j,\ ji = -k, kj = i, ik = -j. \end{aligned} \tag{3}$$

Geometric intuitions The distinctions in the multiplication rules of various hypercomplex numbers give rise to different geometric spaces that provide suitable inductive biases for representing different types of relations. Specifically, spherical quaternions, hyperbolic quaternions, and split quaternions with the same norm c correspond to 4D hypersphere, Lorentz model of hyperbolic space (i.e., the upper part of the double-sheet hyperboloid), and pseudo-hyperboloid (i.e., one-sheet hyperboloid, with curvature $\sqrt{c}$, respectively. These are denoted as follows:

$$\begin{aligned} &|\mathcal{Q}| = s^2 + x^2 + y^2 + z^2 = c > 0 \text{ (hypersphere)} \\ &|\mathcal{H}| = s^2 - x^2 - y^2 - z^2 = c > 0 \text{ (Lorentz hyperbolic space)} \\ &|\mathcal{S}| = s^2 + x^2 - y^2 - z^2 = c > 0 \text{ (pseudo-hyperboloid).} \end{aligned} \tag{4}$$

These spaces have well-known characteristics: spherical spaces are adept at modeling cyclic relations (Wang et al. 2021), hyperbolic spaces provide geometric inductive biases for hierarchical relations (Chami et al. 2020), and the pseudo-hyperboloid (Xiong et al. 2022) offers a balance between spherical and hyperbolic spaces, making it suitable for embedding both cyclic and hierarchical relations. Moreover, by representing relations as geometric rotations over these spaces (i.e., Hamilton product), fundamental logical patterns such as symmetry, inversion, and compositions can be effectively inferred (Zhang et al. 2019; Chami et al. 2020; Xiong et al. 2022). Our proposed embeddings can be viewed as a unification of previous approaches that leverages these geometric inductive biases in these geometric spaces within a single geometric algebraic framework.

For convenience, we parameterize each entity and relation as a Cartesian product of d 4D hypercomplex numbers $\mathbf{s} + \mathbf{x}i + \mathbf{y}j + \mathbf{z}k$, where $\mathbf{s}, \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^d$. This enables us to define all algebraic operations involving these hypercomplex vectors in an element-wise manner.

Atomic Fact Embeddings

Each atomic relation is represented by a rotation hypercomplex vector $\mathbf{r}_\theta$ and a translation hypercomplex vector $\mathbf{r}_b$. For a given triple (h, r, t), we apply the following operation:

$$\mathbf{h}' = (\mathbf{h} \oplus \mathbf{r}_b) \otimes \mathbf{r}_\theta, \tag{5}$$

where $\oplus$ and $\otimes$ stand for addition and Hamilton product between hypercomplex numbers, respectively. The addition involves an element-wise sum of each hypercomplex component. The Hamilton product rotates the head entity. To ensure proper rotation on the unit sphere, we normalize the rotation hypercomplex number $\mathbf{r}_\theta = \mathbf{s}_r^\theta + \mathbf{x}_r^\theta i + \mathbf{y}_r^\theta j + \mathbf{z}_r^\theta k$ by $\mathbf{r}_\theta = \frac{\mathbf{s}_r^\theta + \mathbf{x}_r^\theta \boldsymbol{i} + \mathbf{y}_r^\theta \boldsymbol{j} + \mathbf{z}_r^\theta \boldsymbol{k}}{\sqrt{\mathbf{s}_r^{\theta 2} + \mathbf{x}_r^{\theta 2} + \mathbf{y}_r^{\theta 2} + \mathbf{z}_r^{\theta 2}}}$. Hamilton product is defined by combining the components of the hypercomplex numbers.

$$\begin{aligned} \mathbf{h}' &= \mathbf{h} \otimes \mathbf{r}_\theta \\ &= \left(\mathbf{s}_h \circ \mathbf{s}_r^\theta \circ 1 + \mathbf{x}_h \circ \mathbf{x}_r^\theta \circ i^2 + \mathbf{y}_h \circ \mathbf{y}_r^\theta \circ j^2 + \mathbf{z}_h \circ \mathbf{z}_r^\theta \circ k^2\right) \\ &+ \left(\mathbf{s}_h \circ \mathbf{x}_r^\theta \circ i + \mathbf{x}_h \circ \mathbf{s}_r^\theta \circ i + \mathbf{y}_h \circ \mathbf{z}_r^\theta \circ jk + \mathbf{z}_h \circ \mathbf{y}_r^\theta \circ kj\right) \\ &+ \left(\mathbf{s}_h \circ \mathbf{y}_r^\theta \circ j + \mathbf{x}_h \circ \mathbf{z}_r^\theta \circ ik + \mathbf{y}_h \circ \mathbf{s}_r^\theta \circ j + \mathbf{z}_h \circ \mathbf{x}_r^\theta \circ ik\right) \\ &+ \left(\mathbf{s}_h \circ \mathbf{z}_r^\theta \circ k + \mathbf{x}_h \circ \mathbf{y}_r^\theta \circ ij + \mathbf{y}_h \circ \mathbf{x}_r^\theta \circ ij + \mathbf{z}_h \circ \mathbf{s}_r^\theta \circ k\right) \\ &= \mathbf{s}_{h'} + \mathbf{x}_{h'} i + \mathbf{y}_{h'} j + \mathbf{z}_{h'} k, \end{aligned} \tag{6}$$

where the multiplication of imaginary components follows the rules (Eq.1-3) of the chosen hypercomplex systems.

The scoring function $\phi(h, r, t)$ is defined as:

$$\phi(h, r, t) = \langle \mathbf{h}', \mathbf{t}\rangle = \langle \mathbf{s}_{h'}, \mathbf{s}_t\rangle + \langle \mathbf{x}_{h'}, \mathbf{x}_t\rangle + \langle \mathbf{y}_{h'}, \mathbf{y}_t\rangle + \langle \mathbf{z}_{h'}, \mathbf{z}_t\rangle, \tag{7}$$

where $\langle \cdot, \cdot \rangle$ represents the inner product.

Nested Fact Embeddings

To represent an atomic fact (h, r, t) without losing information, we embed each atomic triple as a 1×3 matrix, where each column corresponds to the embedding of the respective element. Consequently, we have $\mathbf{T}_i = [\mathbf{h}_i, \mathbf{r}_i, \mathbf{t}_i]$.

To embed various shapes of nested relations between T_i and T_j with relation $\widehat{r}$, we model each nested relation using a 1×3 translation matrix and a 3×3 rotation matrix, where each element is a 4D hypercomplex number. Specifically, we first translate the head triple $\mathbf{T}_i$ with $\widehat{\mathbf{r}}_b$, followed by applying a matrix-like rotation of $\widehat{\mathbf{r}}_\theta$, as defined by

$$\mathbf{T}i' = (\mathbf{T}i \oplus_{1\times3} \widehat{\mathbf{r}}_b) \otimes_{3\times3} \widehat{\mathbf{r}}_\theta, \tag{8}$$

where the matrix addition $\oplus_{1\times3}$ is performed through an element-wise summation of the hypercomplex components within the matrices. The matrix-like Hamilton product $\otimes_{3\times3}$ is defined as a product akin to matrix multiplication:

$$\mathbf{T}'_i = \mathbf{T}_i \otimes_{3\times3} \widehat{\mathbf{r}}_\theta = \begin{bmatrix} \mathbf{h}_i \\ \mathbf{r}_i \\ \mathbf{t}_i \end{bmatrix}^\top \times \begin{bmatrix} \widehat{\mathbf{r}}^\theta_{11} & \widehat{\mathbf{r}}^\theta_{12} & \widehat{\mathbf{r}}^\theta_{13} \\ \widehat{\mathbf{r}}^\theta_{21} & \widehat{\mathbf{r}}^\theta_{22} & \widehat{\mathbf{r}}^\theta_{23} \\ \widehat{\mathbf{r}}^\theta_{31} & \widehat{\mathbf{r}}^\theta_{32} & \widehat{\mathbf{r}}^\theta_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_i \otimes \widehat{\mathbf{r}}^\theta_{11} + \mathbf{r}_i \otimes \widehat{\mathbf{r}}^\theta_{21} + \mathbf{t}_i \otimes \widehat{\mathbf{r}}^\theta_{31} \\ \mathbf{h}_i \otimes \widehat{\mathbf{r}}^\theta_{12} + \mathbf{r}_i \otimes \widehat{\mathbf{r}}^\theta_{22} + \mathbf{t}_i \otimes \widehat{\mathbf{r}}^\theta_{32} \\ \mathbf{h}_i \otimes \widehat{\mathbf{r}}^\theta_{13} + \mathbf{r}_i \otimes \widehat{\mathbf{r}}^\theta_{23} + \mathbf{t}_i \otimes \widehat{\mathbf{r}}^\theta_{33} \end{bmatrix}^\top = \begin{bmatrix} \mathbf{h}'_i \\ \mathbf{r}'_i \\ \mathbf{t}'_i \end{bmatrix}^\top, \tag{9}$$

where $\otimes$ is the Hamilton product.

Remarks This matrix-like modeling of nested facts provides flexibility to capture diverse shapes of logical patterns inherent in nested relations. In essence, different shapes of situations or patterns can be effectively modeled by manipulating the 3×3 rotation matrix. For instance, relational implications can be represented using a diagonal matrix, while inversion can be captured using an anti-diagonal matrix. See **theoretical justification** for further analysis.

To assess the plausibility of the nested fact $(T_i, \widehat{r}, T_j)$, we calculate the inner product between the transformed head $\mathbf{T}'_i$ fact and the tail fact $\mathbf{T}_j$ as:

$$\rho(T_i, \widehat{r}, T_j) = \langle \mathbf{T}'_i, \mathbf{T}_j \rangle, \tag{10}$$

where $\langle \cdot, \cdot \rangle$ denotes the matrix inner product.

Learning objective We sum up the loss of atomic fact embedding $\mathcal{L}_{\text{atomic}}$, the loss of nested fact embedding $\mathcal{L}_{\text{meta}}$, and additionally the loss term $\mathcal{L}_{\text{aug}}$ for augmented triples generated by random walking as used in (Chung and Whang 2023). The overall loss is defined as

$$\mathcal{L} = \mathcal{L}_{\text{atomic}} + \lambda_1 \mathcal{L}_{\text{nested}} + \lambda_2 \mathcal{L}_{\text{aug}} \tag{11}$$

where λ_1 and λ_2 are the weight hyperparameters indicating the importance of each loss. Negative sampling is applied by randomly replacing one of the head or tail entity/triple. These losses are defined as follows:

$$\begin{aligned} \mathcal{L}_{\text{atomic}} &= \sum_{(h,r,t)\in\mathcal{T}} g\left(-\phi\left(h,r,t\right)\right) + \sum_{g((h',r',t'))\notin\mathcal{T}} g\left(\phi\left(h',r',t'\right)\right) \\ \mathcal{L}_{\text{aug}} &= \sum_{(h,r,t)\in\mathcal{T}'} g\left(-\phi\left(h,r,t\right)\right) + \sum_{(h',r',t')\notin\mathcal{T}'} g\left(\phi(h',r',t')\right) \\ \mathcal{L}_{\text{nested}} &= \sum_{(T_i,\widehat{r},T_j)\notin\widehat{\mathcal{T}}} g\left(-\rho(T_i,\widehat{r},T_j)\right) + \sum_{(T'_i,\widehat{r},T'_j)\notin\widehat{\mathcal{T}}} g\left(\rho(T'_i,\widehat{r},T'_j)\right), \end{aligned} \tag{12}$$

where $g = \log(1 + \exp(x))$ and $\mathcal{T}'$ is the set of augmented triples.

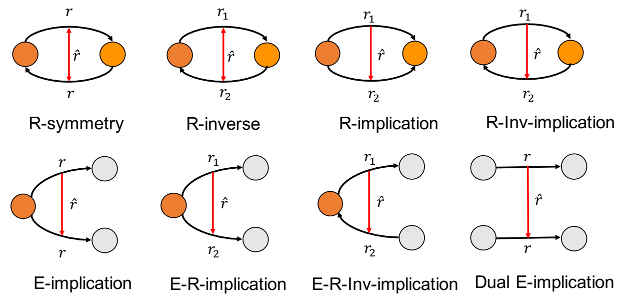

Figure 2: A structural illustration of different shapes of logical patterns, where the colored circles are free variables.

Theoretical Justification

Modeling logical patterns is of great importance for KG embeddings because it enables generalization, i.e., once the patterns are learned, new facts that respect the patterns can be inferred. A logical pattern is a logical form $\psi \rightarrow \phi$ with ψ and ϕ being the body and head, implying that if the body is satisfied then the head must also be satisfied.

First-order-logic-like logical patterns Existing KG embeddings studied logical patterns expressed in the first-order-logic-like form. Prominent examples include symmetry $\forall h, t\colon (h, r, t) \rightarrow (t, r, h)$, anti-symmetry $\forall h, t\colon (h, r, t) \rightarrow \neg(h, r, t)$, inversion $\forall h, t\colon (h, r_1, t) \rightarrow (t, r_2, h)$ and composition $\forall e_1, e_2, e_3\colon (e_1, r_1, e_2) \wedge (e_2, r_2, e_3) \rightarrow (e_1, r_3, e_3)$.

Proposition 1. *NestE can infer symmetry, anti-symmetry, inversion, and composition, regardless of the specific choices of hypercomplex number systems.*

This proposition holds because NestE subsumes ComplEx (Trouillon et al. 2016) (i.e., 4D complex numbers generalize 2D complex numbers).

Logical patterns over nested facts We extend the vanilla logical patterns in KGs to include nested facts. This can be expressed in a non-first-order-logic-like form $\psi \xrightarrow{\widehat{r}} \phi$.

- **Relational symmetry (R-symmetry):** an atomic relation r is symmetric w.r.t a nested relation $\widehat{r}$ if $\forall x, y \in \mathcal{E}, \langle x, r, y\rangle \overset{\widehat{r}}{\leftrightarrow} \langle y, r, x\rangle$.
- **Relational inverse (R-inverse):** two atomic relations r_1 and r_2 are inverse w.r.t a nested relation $\widehat{r}$ if $\forall x, y \in \mathcal{E}$, $(\langle x, r_1, y\rangle \overset{\widehat{r}}{\leftrightarrow} \langle y, r_2, x\rangle)$.
- **Relational implication (R-implication):** an atomic relation r_1 implies a atomic relation r_2 w.r.t a nested relation $\widehat{r}$ if $\forall x, y \in \mathcal{E}, (\langle x, r_1, y\rangle \xrightarrow{\widehat{r}} \langle x, r_2, y\rangle)$.
- **Relational inverse implication (R-Inv-implication):** an atomic relation r_1 inversely implies an atomic relation r_2 w.r.t a nested relation $\widehat{r}$ if $\forall x, y \in \mathcal{E}, (\langle x, r_1, y\rangle \xrightarrow{\widehat{r}} \langle y, r_2, x\rangle)$.
- **Entity implication (E-implication):** an entity x_1 (resp. y_1) implies entity x_2 (resp. y_2) w.r.t an atomic rela-

tion r and a nested relation $\widehat{r}$ if $\forall y \in \mathcal{E}, (\langle x_1, r, y\rangle \xrightarrow{\widehat{r}} \langle x_2, r, y\rangle)$ (resp. $\forall x \in \mathcal{E}, (\langle x, r, y_1\rangle \xrightarrow{\widehat{r}} \langle x, r, y_2\rangle)$).

- **Entity relational implication (E-R-implication):** an entity x_1 and relation r_1 (resp. y_1 and relation r_1) implies entity x_2 and relation r_2 (resp. y_2 and relation r_2) w.r.t a nested relation $\widehat{r}$ if $\forall y \in \mathcal{E}, (\langle x_1, r_1, y\rangle \xrightarrow{\widehat{r}} \langle x_2, r_2, y\rangle)$ (resp. $\forall x \in \mathcal{E}, (\langle x, r_1, y_1\rangle \xrightarrow{\widehat{r}} \langle x, r_2, y_2\rangle)$).
- **Entity relational inverse implication (E-R-Inv-implication):** an entity x_1 and relation r_1 (resp. y_1 and relation r_1) inversely implies entity x_2 and relation r_2 (resp. y_2 and relation r_2) w.r.t a nested relation $\widehat{r}$ if $\forall y \in \mathcal{E}, (\langle x_1, r_1, y\rangle \xrightarrow{\widehat{r}} \langle y, r_2, x_2\rangle)$ (resp. $\forall x \in \mathcal{E}, (\langle x, r_1, y_1\rangle \xrightarrow{\widehat{r}} \langle y_2, r_2, x\rangle)$).
- **Dual Entity implication (Dual E-implication):** an entity pair (x_1, x_2) implies another entity pair (y_1, y_2) iff both (x_1, y_1) and (x_2, y_2) satisfy E-implication.

Fig. 2 illustrates the structure of the introduced patterns and Table 8 in the Appendix presents exemplary patterns of nested facts.

Proposition 2. *NestE can infer R-symmetry, R-inverse, R-implication, R-Inv-implication, E-implication, E-R-implication, E-R-Inv-implication, and Dual E-implication.*

Proof Sketch. To infer different logical patterns via different free variables, we can set some elements of the relation matrix to be zero-valued or one-valued complex numbers. For example, the implication and inverse implication relations can be inferred by setting the matrix to be diagonal or anti-diagonal. See Appendix for details. □

Experimental Results

Experiment Setup

Datasets We utilize three benchmark KGs: FBH, FBHE, and DBHE, that contain nested facts and are constructed by (Chung and Whang 2023). FBH and FBHE are based on FB15K237 from Freebase (Bollacker et al. 2008) while DBHE is based on DB15K from DBpedia (Auer et al. 2007). FBH contains only nested facts that can be inferred from the triple facts, e.g., *prerequisite_for* and *implies_position*, while FBHE and DBHE further contain externally-sourced knowledge crawled from Wikipedia articles, e.g., *next_almaMater* and *transfers_to*. The authors of (Chung and Whang 2023) spent six weeks manually defining these nested facts and adding them to the KGs. The dataset details are presented in Table 1. We split $\mathcal{T}$ and $\widehat{\mathcal{T}}$ into training, validation, and test sets in an 8:1:1 ratio.

	$\|V\|$	$\|R\|$	$\|\mathcal{T}\|$	$\|\widehat{\mathcal{R}}\|$	$\|\widehat{\mathcal{T}}\|$	$\|\mathcal{T}\|'$
FBH	14,541	237	310,117	6	27,062	33,157
FBHE	14,541	237	310,117	10	34,941	33,719
DBHE	12,440	87	68,296	8	6,717	8,206

Table 1: Statistics of $\widehat{G} = (V, R, \mathcal{T}, \widehat{\mathcal{R}}, \widehat{\mathcal{T}})$. $|\mathcal{T}|'$ denotes the number of atomic triples involved in the nested triples.

Baselines We consider BiVE-Q and BiVE-B (Chung and Whang 2023) as our major baselines as they are specifically designed for KGs with nested facts and have demonstrated significant improvements over triple-based methods. We also compare some rule-based approaches as they indirectly consider relations between facts in first-order-logic-like expression, including Neural-LP (Yang, Yang, and Cohen 2017), DRUM (Sadeghian et al. 2019), and AnyBURL (Meilicke et al. 2019). We further include QuatE (Zhang et al. 2019) and BiQUE (Guo and Kok 2021) as they are the SoTA triple-based methods and they are also based on 4D hypercomplex numbers. However, these triple-based methods do not directly apply to the nested facts. Following (Chung and Whang 2023), we create a new triple-based KG G_T where the atomic facts are converted into entities and nested facts are converted into triples (see Appendix for details). For our approach, we implement three variants of NestE: NestE-Q (using quaternions), NestE-H (using hyperbolic quaternions), NestE-S (split quaternions), as well as their counterparts with translations: NestE-QB, NestE-HB, and NestE-SB. In the Appendix, we also extend BiVE-Q and BiVE-B to other hypercomplex numbers: BiVE-H, BiVE-HB BiVE-S, and BiVE-SB for further comparison. We employ three standard metrics: Filtered MR (Mean Rank), MRR (Mean Reciprocal Rank), and Hit@10. We report the mean performance over 10 random seeds for each method, and the relatively small standard deviations are omitted.

Implementation details We implement the framework based on OpenKE [4] and the code [5]. We train our methods on triple prediction and evaluate them on other tasks. The detailed hyperparameter settings can be found in the Appendix.

Main Results

Triple prediction Table 2 presents the results of triple prediction. First, it shows that all triple-based approaches yield relatively modest results compared to BiVE-Q and BiVE-B, designed specifically for KGs with nested facts. Our approach, NestE-Q, the quaternionic version, already outperforms the baselines across most metrics. Particularly notable are the pronounced enhancements in FBHE and DBHE, with MRR improvements of 14.1% and 17.7% respectively, underscoring the efficacy of the proposed NestE model. Furthermore, NestE-H and NestE-S demonstrate heightened performance over NestE-Q across various evaluation metrics, particularly in terms of MR. This highlights the advantages that hyperbolic quaternions and split quaternions offer over standard quaternions. Impressively, the split quaternionic version attains the highest performance, followed closely by the hyperbolic quaternionic variant. Moreover, through the incorporation of a hypercomplex translation component, NestE-QB, Fact-HB, and NestE-SB consistently outperform their non-translation counterparts, showing the advantages of combining multiple transformations (rotation and translation) within the hypercomplex space.

[4]https://github.com/thunlp/OpenKE
[5]https://github.com/bdi-lab/BiVE/

	FBH			*FBHE*			*DBHE*		
	MR (↓)	MRR (↑)	Hit@10 (↑)	MR (↓)	MRR (↑)	Hit@10 (↑)	MR (↓)	MRR (↑)	Hit@10 (↑)
QuatE*	145603.8	0.103	0.114	94684.4	0.101	0.209	26485.0	0.157	0.179
BiQUE*	81687.5	0.104	0.115	61015.2	0.135	0.205	19079.4	0.163	0.185
Neural-LP*	115016.6	0.070	0.073	90000.4	0.238	0.274	21130.5	0.170	0.209
DRUM*	115016.6	0.069	0.073	90000.3	0.261	0.274	21130.5	0.166	0.209
AnyBURL*	108079.6	0.096	0.108	83136.8	0.191	0.252	20530.8	0.177	0.214
BiVE-Q	6.20	0.855	0.941	8.35	0.711	0.866	3.63	0.687	0.958
BiVE-B	8.63	0.833	0.924	9.53	0.705	0.860	4.66	0.718	0.945
NestE-Q (Ours)	6.56	0.863	0.953	5.77	0.811	0.943	3.51	0.809	0.960
NestE-H (Ours)	4.69	0.858	0.964	3.99	0.781	0.943	2.65	0.806	0.969
NestE-S (Ours)	3.87	0.867	0.977	3.60	0.795	0.947	2.55	0.809	0.966
NestE-QB (Ours)	6.04	0.898	0.958	5.55	0.845	0.947	2.54	0.847	0.973
NestE-HB (Ours)	3.82	0.899	0.971	3.53	0.828	0.955	2.62	0.842	0.972
NestE-SB (Ours)	**3.34**	**0.922**	**0.982**	**3.05**	**0.851**	**0.962**	**2.07**	**0.862**	**0.984**

Table 2: Results of triple prediction. Shaded numbers are better results than the best baseline. The best scores are boldfaced and the second best scores are underlined. * denotes results taking from (Chung and Whang 2023).

	FBH			*FBHE*			*DBHE*		
	MR (↓)	MRR (↑)	Hit@10 (↑)	MR (↓)	MRR (↑)	Hit@10 (↑)	MR (↓)	MRR (↑)	Hit@10 (↑)
QuatE*	163.7	0.346	0.494	1546.4	0.124	0.189	551.6	0.208	0.309
BiQUE*	111.0	0.423	0.641	90.1	0.387	0.617	29.5	0.378	0.677
Neural-LP*	185.9	0.433	0.648	146.2	0.466	0.716	32.2	0.517	0.756
DRUM*	262.7	0.394	0.555	207.6	0.413	0.620	49.0	0.470	0.732
AnyBURL*	228.5	0.380	0.563	166.0	0.418	0.607	81.7	0.403	0.594
BiVE-Q	4.33	0.826	0.948	6.56	0.761	0.886	2.69	0.852	0.971
BiVE-B	5.34	0.836	0.940	7.49	0.761	0.872	2.91	0.858	0.967
NestE-Q (Ours)	1.70	0.930	0.986	2.89	0.863	0.948	**1.68**	0.930	0.987
NestE-H (Ours)	1.68	0.909	0.987	2.87	0.843	0.945	1.82	0.912	0.986
NestE-S (Ours)	1.54	0.925	**0.991**	3.04	0.850	0.941	1.76	0.910	0.988
NestE-QB (Ours)	1.71	**0.935**	0.987	3.00	0.865	0.949	1.70	**0.931**	0.986
Fact-HB (Ours)	1.60	0.924	0.989	2.76	0.855	0.950	1.92	0.918	0.981
NestE-SB (Ours)	**1.52**	0.934	**0.991**	**2.61**	**0.867**	**0.951**	1.72	0.919	**0.990**

Table 3: Results of conditional link prediction. Shaded numbers are better results than the best baseline. The best scores are boldfaced and the second best scores are underlined. * denotes results taking from (Chung and Whang 2023).

Conditional link prediction Table 3 shows the outcomes of conditional link prediction. It is evident that all three NestE variants substantially outperform the two SoTA baselines, BiVE-Q and BiVE-B, across all datasets. Notably, the best NestE variant surpasses the baselines by 11.8%, 13.9%, and 8.5% in terms of MRR for FBH, FBHE, and DBHE, respectively. This remarkable performance gain underscores the effectiveness of the proposed method. Similar to the trends observed in triple prediction, the incorporation of translation components in NestE-QB, Fact-HB, and Fact-SB leads to further improvements over their counterparts without translation components. This reaffirms the advantages gained from the integration of multiple hypercomplex transformations. Intriguingly, we noticed that varying hypercomplex number systems yield the best performance on different datasets, contrasting the observations from triple prediction. We conjecture that this stems from the inherent variance in inductive biases offered by different hypercomplex number systems, making them more suitable for certain datasets over others. We believe the choices of spaces can be linked to a hyperparameter that offers flexibility in adapting to diverse dataset characteristics.

Base link prediction Table 4 illustrates the results of base link prediction. Among our approaches, namely NestE-Q, NestE-H, and NestE-S, we observe competitive or improved results in comparison to SoTA embedding-based and rule-based methods on the FBHE and DBHE datasets. The best performance is achieved by NestE-QB, which outperforms the baselines across a majority of metrics. This outcome substantiates the fact that the incorporation of nested facts into triple-based KGs indeed enhances the inference capabilities for base link prediction.

Ablation Analysis

Embedding analysis of logical patterns To verify whether the learned embeddings capture the inference of logical patterns over nested facts, we visualized the real part of the embeddings of the 8 relations in DBHE. The analysis of the embeddings yields insightful observations. As shown

	FBHE		*DBHE*	
	MRR (↑)	Hit@10 (↑)	MRR (↑)	Hit@10 (↑)
QuatE*	0.354	0.581	0.264	0.440
BiQUE*	0.356	0.583	0.274	**0.446**
Neural-LP*	0.315	0.486	0.233	0.357
DRUM*	0.317	0.490	0.237	0.359
AnyBURL*	0.310	0.526	0.220	0.364
BiVE-Q	0.369	0.603	0.271	0.428
BiVE-B	0.370	0.607	0.274	0.422
NestE-Q (Ours)	0.365	0.605	0.284	**0.446**
Fact-H (Ours)	0.349	0.593	0.266	0.423
NestE-S (Ours)	0.350	0.592	0.272	0.432
NestE-QB (Ours)	**0.371**	**0.608**	**0.289**	0.443
Fact-HB (Ours)	0.353	0.594	0.271	0.423
NestE-SB (Ours)	0.355	0.594	0.273	0.431

Table 4: Results of base link prediction. The best scores are boldfaced and the second best scores are underlined. * denotes results taking from (Chung and Whang 2023).

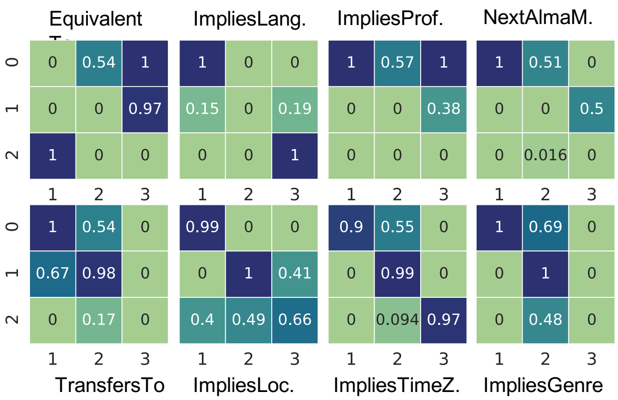

Figure 3: The visualization of the average of the real component embeddings of the 8 nested relations in DBHE.

in Fig. 3, the lower left element and upper right element of the embedding *EquivalentTo* are 1, showcasing that *EquivalentTo* predominantly adheres to R-symmetry or R-inverse. On the other hand, the upper left element and lower right element of the *ImpliesLang.* are 1, affirming its alignment with the R-implication rule. Similarly, the embeddings of *NextAlmaM.*, *TransfersTo*, and *ImpliesGenre* indicate high adherence to E-implicationsas as only one of the corners is 1. We find that the embedding of relation *ImpliesProf.* does not have a significant pattern. We conjecture that this is because *ImpliesProf.* follows many rule patterns and there exists no global solution that satisfies all rules. See the Appendix for the statistics of the logical patterns in the datasets.

Influence of nested fact embeddings To evaluate the influence of nested fact embeddings, we perform a comparison by excluding the loss associated with nested fact embeddings (i.e., setting $\lambda_1 = 0$). The outcomes presented in Table 5 underscore the significant enhancements achieved by incorporating nested fact embeddings, particularly evident in the improvements in MR and H@10 for DBHE.

	FBHE		*DBHE*	
	MRR (↑)	Hit@10 (↑)	MRR (↑)	Hit@10 (↑)
NestE-Q ($\lambda_1 = 0$)	**0.368**	0.604	0.281	0.431
NestE-Q ($\lambda_1 = 0.5$)	0.365	**0.605**	**0.284**	**0.446**
NestE-H ($\lambda_1 = 0$)	0.347	0.589	0.267	0.420
NestE-H ($\lambda_1 = 0.5$)	**0.349**	**0.593**	0.266	**0.423**
NestE-S ($\lambda_1 = 0$)	0.347	0.589	**0.272**	0.427
NestE-S ($\lambda_1 = 0.5$)	**0.350**	**0.592**	**0.272**	**0.432**

Table 5: Ablation study on the nested fact embeddings for base link prediction. Best results are boldfaced.

$\hat{r}$	NestE-Q	-QB	-H	-HB	-S	-SB
Equiv.To	0.994	**0.997**	0.992	**0.997**	**0.997**	0.995
ImpliesLang.	0.671	0.602	**0.680**	0.662	0.614	0.622
ImpliesProf.	0.807	0.916	0.830	0.935	0.832	**0.936**
ImpliesLocat.	0.929	0.869	0.893	0.810	0.929	**0.958**
ImpliesTime.	0.305	0.297	0.307	**0.329**	0.290	0.293
ImpliesGenre	0.726	0.762	0.719	0.741	0.742	**0.796**
NextAlmaM.	0.770	**0.812**	0.689	0.688	0.751	0.795
Transf.To	**0.977**	0.952	0.964	0.949	0.921	0.953

Table 6: Performance per relation on triple prediction.

Relation-specific performance In Table 6, we present the performance results for each relation within the DBHE dataset. Notably, the diverse hypercomplex number systems lead to optimal performance for different relations. This reiterates our conjecture that distinct benefits are offered by varying hypercomplex number systems, catering to the specific characteristics of different relation types. Remarkably, our findings reveal that the incorporation of a hypercomplex translation component (as seen in NestE-QB, NestE-HB, and NestE-SB) notably enhances the embeddings of relations such as *ImpliesProf.* and *ImpliesGenre* across all variants of hypercomplex number systems. However, this does not extend to relations like *ImpliesLocat.* and *ImpliesLang.*, suggesting a more complex relationship between these specific relations and the hypercomplex translation.

Conclusion

This paper considers a novel perspective by extending traditional atomic factual knowledge representation to include nested factual knowledge. This enables the representation of both temporal situations and logical patterns that go beyond conventional first-order logic expressions (Horn rules). Our proposed approach, NestE, presents a family of hypercomplex embeddings capable of embedding both atomic and nested factual knowledge. This framework effectively captures essential logical patterns that emerge from nested facts. Empirical evaluation demonstrates the substantial performance enhancements achieved by NestE compared to existing baseline methods. Additionally, our generalized hypercomplex embedding framework unifies previous algebraic and geometric embedding methods, offering versatility in embedding diverse relation types.

Acknowledgments

The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Bo Xiong and Zihao Wang. This research has been partially funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 - 390740016, the Stuttgart Center for Simulation Science (SimTech), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No: 860801, and the the Bundesministerium für Wirtschaft und Energie (BMWi), grant aggrement No. 01MK20008F.

References

Alivanistos, D.; Berrendorf, M.; Cochez, M.; and Galkin, M. 2022. Query Embedding on Hyper-Relational Knowledge Graphs. In *ICLR*. OpenReview.net.

Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; and Ives, Z. G. 2007. DBpedia: A Nucleus for a Web of Open Data. In *ISWC/ASWC*, volume 4825 of *Lecture Notes in Computer Science*, 722–735. Springer.

Balazevic, I.; Allen, C.; and Hospedales, T. M. 2019. Multi-relational Poincaré Graph Embeddings. In *NeurIPS*, 4465–4475.

Bollacker, K. D.; Evans, C.; Paritosh, P. K.; Sturge, T.; and Taylor, J. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *SIGMOD Conference*, 1247–1250. ACM.

Bordes, A.; Usunier, N.; García-Durán, A.; Weston, J.; and Yakhnenko, O. 2013. Translating Embeddings for Modeling Multi-relational Data. In *NIPS*, 2787–2795.

Chami, I.; Wolf, A.; Juan, D.; Sala, F.; Ravi, S.; and Ré, C. 2020. Low-Dimensional Hyperbolic Knowledge Graph Embeddings. In *ACL*, 6901–6914. Association for Computational Linguistics.

Champin, P. 2022. RDF-star: Paving the Way to the Next generation of Linked Data. *ERCIM News*, 2022(128).

Chung, C.; and Whang, J. J. 2023. Learning Representations of Bi-Level Knowledge Graphs for Reasoning beyond Link Prediction. In *AAAI*.

Dasgupta, S. S.; Ray, S. N.; and Talukdar, P. P. 2018. HyTE: Hyperplane-based Temporally aware Knowledge Graph Embedding. In *EMNLP*, 2001–2011. Association for Computational Linguistics.

Demeester, T.; Rocktäschel, T.; and Riedel, S. 2016. Lifted Rule Injection for Relation Embeddings. In *EMNLP*, 1389–1399. The Association for Computational Linguistics.

Fatemi, B.; Taslakian, P.; Vázquez, D.; and Poole, D. 2020. Knowledge Hypergraphs: Prediction Beyond Binary Relations. In *IJCAI*, 2191–2197. ijcai.org.

Galkin, M.; Trivedi, P.; Maheshwari, G.; Usbeck, R.; and Lehmann, J. 2020. Message Passing for Hyper-Relational Knowledge Graphs. In *EMNLP (1)*, 7346–7359. Association for Computational Linguistics.

Guan, S.; Jin, X.; Guo, J.; Wang, Y.; and Cheng, X. 2020. NeuInfer: Knowledge Inference on N-ary Facts. In *ACL*, 6141–6151. Association for Computational Linguistics.

Guo, J.; and Kok, S. 2021. BiQUE: Biquaternionic Embeddings of Knowledge Graphs. In *EMNLP (1)*, 8338–8351. Association for Computational Linguistics.

Guo, S.; Wang, Q.; Wang, L.; Wang, B.; and Guo, L. 2016. Jointly Embedding Knowledge Graphs and Logical Rules. In *EMNLP*, 192–202. The Association for Computational Linguistics.

Leblay, J.; and Chekol, M. W. 2018. Deriving Validity Time in Knowledge Graph. In *WWW (Companion Volume)*, 1771–1776. ACM.

Liu, Y.; Yao, Q.; and Li, Y. 2020. Generalizing Tensor Decomposition for N-ary Relational Knowledge Bases. In *WWW*, 1104–1114. ACM / IW3C2.

Liu, Y.; Yao, Q.; and Li, Y. 2021. Role-Aware Modeling for N-ary Relational Knowledge Bases. In *WWW*, 2660–2671. ACM / IW3C2.

Meilicke, C.; Chekol, M. W.; Ruffinelli, D.; and Stuckenschmidt, H. 2019. Anytime Bottom-Up Rule Learning for Knowledge Graph Completion. In *IJCAI*, 3137–3143. ijcai.org.

Nayyeri, M.; Xiong, B.; Mohammadi, M.; Akter, M. M.; Alam, M. M.; Lehmann, J.; and Staab, S. 2023. Knowledge Graph Embeddings using Neural Ito Process: From Multiple Walks to Stochastic Trajectories. In *ACL (Findings)*, 7165–7179. Association for Computational Linguistics.

Nickel, M.; Tresp, V.; and Kriegel, H. 2011. A Three-Way Model for Collective Learning on Multi-Relational Data. In *ICML*, 809–816. Omnipress.

Niu, G.; Zhang, Y.; Li, B.; Cui, P.; Liu, S.; Li, J.; and Zhang, X. 2020. Rule-Guided Compositional Representation Learning on Knowledge Graphs. In *AAAI*, 2950–2958. AAAI Press.

Rosso, P.; Yang, D.; and Cudré-Mauroux, P. 2020. Beyond Triplets: Hyper-Relational Knowledge Graph Embedding for Link Prediction. In *WWW*, 1885–1896. ACM / IW3C2.

Sadeghian, A.; Armandpour, M.; Ding, P.; and Wang, D. Z. 2019. DRUM: End-To-End Differentiable Rule Mining On Knowledge Graphs. In *NeurIPS*, 15321–15331.

Sun, Z.; Deng, Z.; Nie, J.; and Tang, J. 2019. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In *ICLR (Poster)*. OpenReview.net.

Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; and Bouchard, G. 2016. Complex Embeddings for Simple Link Prediction. In *ICML*, volume 48 of *JMLR Workshop and Conference Proceedings*, 2071–2080. JMLR.org.

Wang, J.; Wang, B.; Qiu, M.; Pan, S.; Xiong, B.; Liu, H.; Luo, L.; Liu, T.; Hu, Y.; Yin, B.; and Gao, W. 2023. A Survey on Temporal Knowledge Graph Completion: Taxonomy, Progress, and Prospects. *CoRR*, abs/2308.02457.

Wang, S.; Wei, X.; dos Santos, C. N.; Wang, Z.; Nallapati, R.; Arnold, A. O.; Xiang, B.; Yu, P. S.; and Cruz, I. F. 2021. Mixed-Curvature Multi-Relational Graph Neural Network for Knowledge Graph Completion. In *WWW*, 1761–1771. ACM / IW3C2.

Wang, Z.; Zhang, J.; Feng, J.; and Chen, Z. 2014. Knowledge Graph Embedding by Translating on Hyperplanes. In *AAAI*, 1112–1119. AAAI Press.

Xiong, B.; Nayyer, M.; Pan, S.; and Staab, S. 2023a. Shrinking Embeddings for Hyper-Relational Knowledge Graphs. In *ACL*. International Committee on Computational Linguistics.

Xiong, B.; Nayyeri, M.; Daza, D.; and Cochez, M. 2023b. Reasoning beyond Triples: Recent Advances in Knowledge Graph Embeddings. In *CIKM*, 5228–5231. ACM.

Xiong, B.; Zhu, S.; Nayyeri, M.; Xu, C.; Pan, S.; Zhou, C.; and Staab, S. 2022. Ultrahyperbolic Knowledge Graph Embeddings. In *KDD*, 2130–2139. ACM.

Yang, B.; Yih, W.; He, X.; Gao, J.; and Deng, L. 2015. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *ICLR (Poster)*.

Yang, F.; Yang, Z.; and Cohen, W. W. 2017. Differentiable Learning of Logical Rules for Knowledge Base Reasoning. In *NIPS*, 2319–2328.

Zhang, S.; Tay, Y.; Yao, L.; and Liu, Q. 2019. Quaternion Knowledge Graph Embeddings. In *NeurIPS*, 2731–2741.