NestE: Modeling Nested Relational Structures for Knowledge Graph Reasoning

Bo Xiong, Mojtaba Nayyeri, Linhao Luo, Zihao Wang, Shirui Pan, Steffen Staab

1University of Stuttgart, Stuttgart, Germany
2Monash University, Melbourne, Australia
3Griffith University, Queensland, Australia
4University 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., ((President 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 $1 \times 3$ matrix, and each nested relation is modeled as a $3 \times 3$ matrix that rotates the $1 \times 3$ 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 (Joe-Biden, 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).

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 ($T_1$, works_for_when, $T_2$). Such a fact about facts is referred to as a nested fact 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.

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).

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
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 Donald Trump?" because the query about the primary fact (President, holds position, President) depends on another fact (Donald Trump, 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 $\phi \rightarrow \psi$. As illustrated in Fig. 1, (Location A, uses Language B) implies (Language B, Location 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. 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 $1 \times 3$ hypercomplex matrix, with each element signifying a component of the atomic fact. Furthermore, each nested relation is modeled through a $3 \times 3$ hypercomplex matrix that rotates the $1 \times 3$ 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 variability of the triples. Hyper-relational facts (Guan et al. 2020; Rosso, Yang, and Cudré-Mauroux 2020; Galkin et al. 2020; Xiong et al. 2023) attach a set of key-value qualifiers to the primal triple, where each qualifier specifies certain semantics of the primary 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; Melielie et al. 2017; Demeester, Rocktaschel, 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., $\forall e_1, e_2, e_3 : (e_1, r_1, e_2) \land (e_2, r_2, e_3) \Rightarrow (e_1, r_3, e_3)$, where there 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 $(x, r_1, y) \Rightarrow (p, r_2, q)$. The proposed embeddings (i.e., BiVE-Q and BiVE-B) concatenate the embeddings of the head, relation, and tail, subsequently embedding them via an MLP. However, such modeling does 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 $G = (V, R, T)$, where $V$ represents the set of entities, $R$ stands for the set of relation names, and $T = \{(h, r, t) : h, t \in V, r \in R\}$ represents the set of triples. We refer to $G$ as an atomic factual KG, and each $(h, r, t) \in T$ 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 $G = (V, R, T)$, a set of nested triples is defined by $\tilde{T} = \{(T_i, \tilde{r}, T_j) : T_i, T_j \in T, \tilde{r} \in \tilde{R}\}$, where $\tilde{T}$ is the set of atomic triples and $\tilde{R}$ is the set of nested relation names.

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

---

2In their work, the KG is referred to as a bi-level KG, and the term "high-level facts" is synonymous with nested facts.
set of nested triples $\hat{T}$ defined on $G$ and $\hat{R}$, a nested factual KG is defined as $\hat{G} = (V, R, T, \hat{R}, \hat{T})$.

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 $\hat{G} = (V, R, T, \hat{R}, \hat{T})$, the triple prediction problem involves answering a query $(T_i, \hat{r}, ?t)$ or $([h, \hat{r}, t], i)$ with $T_i, \hat{T}_j \in T$ and $\hat{r} \in \hat{R}$, where the variable $?h$ or $?t$ needs to be found at an atomic triple within $\hat{G}$.

**Definition 4** (Conditional Link Prediction). Given a nested factual KG $\hat{G} = (V, R, T, \hat{R}, \hat{T})$, 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 $(T_i, \hat{r}, (h_j, r_j, ?))$, $(T_i, ?, (r_j, t_j))$, $((h_i, r_i, ?), \hat{r}, T_j)$, or $((?, r_i, t_i), \hat{r}, T_j)$, where the variables need to be bound to entities within $\hat{G}$.

**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 hypercomplex number systems lies in their multiplication rules of the imaginary components.

**(Spherical) quaternions** $Q$ follow the multiplication rules:

\[
i^2 = j^2 = k^2 = 1,
ij = k = -ji,\quad kj = i = -jk,\quad ki = j = -ik.
\]  

**(Hyperbolic) quaternions** $H$ follow the multiplication rules:

\[
i^2 = -1, j^2 = k^2 = 1,
ij = k = -ji,\quad kj = i = -jk,\quad ki = j = -ik.
\]  

**(Split) quaternions** $S$ follow the multiplication rules:

\[
i^2 = -1, j^2 = k^2 = 1,
i j = k = i = j = -k,\quad kj = i = -jk.
\]

**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 hyperspheres, 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:

\[
|Q| = s^2 + x^2 + y^2 + z^2 = c > 0 \quad \text{(hypersphere)}
\]
\[
|H| = s^2 - x^2 - y^2 - z^2 = c > 0 \quad \text{(Lorentz hyperbolic space)}
\]
\[
|S| = s^2 + x^2 - y^2 - z^2 = c > 0 \quad \text{(pseudo-hyperboloid)}.
\]

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 4D hypercomplex numbers $s + xi + yj + zk$, where $s, x, y, z \in \mathbb{R}$. 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 $r_h$ and a translation hypercomplex vector $r_t$. For a given triple $(h, r, t)$, we apply the following operation:

\[
h' = (h \oplus r_h) \otimes r_t,
\]

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 $r_\theta = s^{\theta} + x^{\theta}i + y^{\theta}j + z^{\theta}k$ by $r_\theta = \frac{s^{\theta} + x^{\theta}i + y^{\theta}j + z^{\theta}k}{\sqrt{s^{\theta2} + x^{\theta2} + y^{\theta2} + z^{\theta2}}}$. Hamilton product is defined by combining the components of the hypercomplex numbers.

\[
h' = h \otimes r_\theta = \begin{cases} 
   (s_h \circ s_{\theta} \circ 1 + x_h \circ x_{\theta} \circ i^2 + y_h \circ y_{\theta} \circ j^2 + z_h \circ z_{\theta} \circ k^2) \\
   + (s_h \circ x_{\theta} \circ i + x_h \circ s_{\theta} \circ i + y_h \circ z_{\theta} \circ j + z_h \circ y_{\theta} \circ k) \\
   + (s_h \circ y_{\theta} \circ j + x_h \circ z_{\theta} \circ ik + y_h \circ s_{\theta} \circ j + z_h \circ x_{\theta} \circ ik) \\
   + (s_h \circ z_{\theta} \circ k + x_h \circ y_{\theta} \circ ij + y_h \circ x_{\theta} \circ ij + z_h \circ s_{\theta} \circ k) 
\end{cases}
= s_{h_\theta} + x_{h_\theta}i + y_{h_\theta}j + z_{h_\theta}k,
\]

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:

\[
\phi(h, r, t) = \langle h', t \rangle = \langle s_{h_\theta}, s_t \rangle + \langle x_{h_\theta}, x_t \rangle + \langle y_{h_\theta}, y_t \rangle + \langle z_{h_\theta}, z_t \rangle,
\]

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

$$T' = T_i \otimes_{3 \times 3} \hat{r}_\theta = \begin{bmatrix} h_i & r_i & t_i \end{bmatrix}^T \times \begin{bmatrix} \hat{r}_{11} & \hat{r}_{12} & \hat{r}_{13} \\ \hat{r}_{21} & \hat{r}_{22} & \hat{r}_{23} \\ \hat{r}_{31} & \hat{r}_{32} & \hat{r}_{33} \end{bmatrix} = \begin{bmatrix} h'_i \\ r'_i \\ t'_i \end{bmatrix}^T,$$

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 $3 \times 3$ 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 $(T_i, \hat{r}, T_j)$, we calculate the inner product between the transformed head $T'_i$ fact and the tail fact $T_j$ as:

$$\rho(T_i, \hat{r}, T_j) = \langle T'_i, T_j \rangle,$$

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

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

$$L = L_{\text{atomic}} + \lambda_1 L_{\text{nested}} + \lambda_2 L_{\text{aug}}$$

where $\lambda_1$ and $\lambda_2$ 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:

$$L_{\text{atomic}} = \sum_{(h, r, t) \in \mathcal{T}} g(-\phi(h, r, t) + \sum_{(h', r', t') \in \mathcal{T}} g(\phi(h', r', t')))$$

$$L_{\text{aug}} = \sum_{(h, r, t) \in \mathcal{T}} g(-\phi(h, r, t) + \sum_{(h', r', t') \in \mathcal{T}} g(\phi(h', r', t')))$$

$$L_{\text{nested}} = \sum_{(T_i, \hat{r}, T_j) \in \mathcal{T}} g(-\rho(T_i, \hat{r}, T_j)) + \sum_{(T'_i, \hat{r}, T'_j) \in \mathcal{T}} g(\rho(T'_i, \hat{r}, T'_j)),$$

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

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 $\psi$ and $\phi$ 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 $\forall h, t: (h, r, t) \rightarrow (t, r, h)$, anti-symmetry $\forall h, t: (h, r, t) \rightarrow \neg(h, r, t)$, inversion $\forall h, t: (h, r_1, t) \rightarrow (t, r_2, h)$ and composition $\forall e_1, e_2, e_3: (e_1, r_1, e_2) \land (e_2, r_2, e_3) \rightarrow (e_1, r_3, e_3)$.

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 $\psi \rightarrow^* \phi$.

- Relational symmetry (R-symmetry): an atomic relation $r$ is symmetric w.r.t a nested relation $\hat{r}$ if $\forall x, y \in \mathcal{E}, \langle x, r, y \rangle \hat{r} \langle y, r, x \rangle$.

- Relational inverse (R-inverse): two atomic relations $r_1$ and $r_2$ are inverse w.r.t a nested relation $\hat{r}$ if $\forall x, y \in \mathcal{E}, \langle x, r_1, y \rangle \hat{r} \langle y, r_2, x \rangle$.

- Relational implication (R-implication): an atomic relation $r_1$ implies a atomic relation $r_2$ w.r.t a nested relation $\hat{r}$ if $\forall x, y \in \mathcal{E}, \langle x, r_1, y \rangle \hat{r} \rightarrow \langle y, r_2, x \rangle$.

- Relational inverse implication (R-Inv-implication): an atomic relation $r_1$ inversely implies an atomic relation $r_2$ w.r.t a nested relation $\hat{r}$ if $\forall x, y \in \mathcal{E}, \langle x, r_1, y \rangle \hat{r} \rightarrow \langle y, r_2, x \rangle$.

- 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 \( \tilde{r} \) if \( \forall y \in \mathcal{E}, ((x, r, y) \xrightarrow{\tilde{r}} (x', r, y')) \) (resp. \( \forall x \in \mathcal{E}, ((x, r, y_1) \xrightarrow{\tilde{r}} (x', r, y_2)) \)).

- **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 \( \tilde{r} \) if \( \forall y \in \mathcal{E}, ((x_1, r_1, y) \xrightarrow{\tilde{r}} (x_2, r_2, y)) \) (resp. \( \forall x \in \mathcal{E}, ((x, r, y_1) \xrightarrow{\tilde{r}} (x_2, r_2, y_2)) \)).

- **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 \( \tilde{r} \) if \( \forall y \in \mathcal{E}, ((x_1, r_1, y) \xrightarrow{\tilde{r}} (y_2, r_2, x_2)) \) (resp. \( \forall x \in \mathcal{E}, ((x, r, y_1) \xrightarrow{\tilde{r}} (y_2, r_2, x)) \)).

- **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 alma Mater and transfers.fo. 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 \( T \) and \( \tilde{T} \) into training, validation, and test sets in an 8:1:1 ratio.

**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.

### Table 1: Statistics of \( \tilde{G} = (V, R, T, \tilde{R}, \tilde{T}) \).

|  | \( |V| \) | \( |R| \) | \( |T| \) | \( |\tilde{R}| \) | \( |\tilde{T}| \) | \( |T'| \) |
|---|---|---|---|---|---|---|
| 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 |

\(^4\)https://github.com/thunlp/OpenKE
\(^5\)https://github.com/bdi-lab/BiVE/
Table 3 shows the outcomes 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 taken from (Chung and Whang 2023).

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 taken from (Chung and Whang 2023).

Table 4 illustrates the results of base 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 taken 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
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).

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 agreement No. 01MK20008F.

References
for Knowledge Graph Completion. In WWW, 1761–1771. ACM / IW3C2.


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


