Low-Dimensional Hyperbolic Knowledge Graph Embedding for Better Extrapolation to Under-Represented Data

Zhuoxun Zheng[1,2](✉), Baifan Zhou[2,3], Hui Yang[4], Zhipeng Tan[1], Arild Waaler[2], Evgeny Kharlamov[1,2], and Ahmet Soylu[2,3]

[1] Bosch Center for AI, Renningen, Germany
zhengzx712@gmail.com
[2] University of Oslo, Oslo, Norway
[3] Oslo Metropolitan University, Oslo, Norway
[4] Laboratoire Interdisciplinaire des Sciences du Numérique, Paris, France

Abstract. Past works have shown knowledge graph embedding (KGE) methods learn from facts in the form of triples and extrapolate to unseen triples. KGE in hyperbolic space can achieve impressive performance even in low-dimensional embedding space. However, existing work limitedly studied extrapolation to under-represented data, including under-represented entities and relations. To this end, we propose HolmE, a general form of KGE method on hyperbolic manifolds. HolmE addresses extrapolation to under-represented entities through a special treatment of the bias term, and extrapolation to under-represented relations by supporting strong composition. We provide empirical evidence that HolmE achieves promising performance in modelling unseen triples, under-represented entities, and under-represented relations. We prove that mainstream KGE methods either: (1) are special cases of HolmE and thus support strong composition; (2) do not support strong composition. The code and data are open-sourced at https://github.com/nsai-uio/HolmE-KGE.

Keywords: Knowledge graph embedding · Knowledge graph

1 Introduction

Knowledge graphs (KGs) refer to multi-relational graphs that represent facts in the form of subject-relation-object triples. Studies in KG embedding (KGE) strive to represent KGs numerically via link prediction, enabling the possibility to leverage powerful machine learning (ML) in many practical applications, such as predicting interactions between drugs and targets [12] or diseases [16].

The extrapolation of KGE models to unseen triples (i.e., predicting unseen links) is the focus of current studies. Many past works have achieved promising results. Translational models have good performance as well as transparent geometrical explanations [3]. KGE models in hyperbolic space achieve impressive

A. Meroño Peñuela et al. (Eds.): ESWC 2024, LNCS 14664, pp. 100–120, 2024.
https://doi.org/10.1007/978-3-031-60626-7_6

performance even in low-dimensional space, due to their expressiveness for hierarchical structures [6]. However, *extrapolation* is multi-faceted; past work limitedly discussed the extrapolation to *under-represented data*, i.e., long-tail data, which includes *under-represented entities*, and *relations* [33,35].

We argue that for good extrapolation to under represented entities, a special treatment to the bias terms in KGE models is needed. Many past models adopt scoring functions with entity-specific bias terms to evaluate the probability of predicted entities (higher score, higher chance to be predicted) [1,5,6]. We observe that the bias terms learned from the training set are highly correlated with the entity distribution in the training set (Fig. 1b). In benchmark datasets such as WN18RR and FB15k-237, the latter is also highly correlated with the entity distribution in the test set (Fig. 1a). This makes exploitation possible: to simply assign higher scores to entities that are more frequent via giving higher biases to these entities. In this regard, the bias terms essentially serves as prior probabilities of the entities in the training set. Adopting the bias terms assumes that the entity distribution of the training set is similar to that in the test set, which is true for benchmark datasets , but not necessarily true for many real-world cases, where most entities in test data are under-represented. In these cases, the use of bias terms can lead to performance deterioration.

On extrapolation to *under-represented relations*, we propose the notion of *strong composition*, The entire embedding space for relations of *strong composition* KGE model supports composition (Fig. 2), contrasting to *weak composition* KGE, whose embedding space only has a sub-space that supports composition. The key difference is that, *weak composition* KGE would assume relations to be not compositional and tend to put them into the sub-space not supporting composition, when the compositional relations are under-represented in the training set. Practically, *strong composition* KGEs can better model under-represented relations, as these relations are not embedded in sub-spaces where composition patterns are not supported.

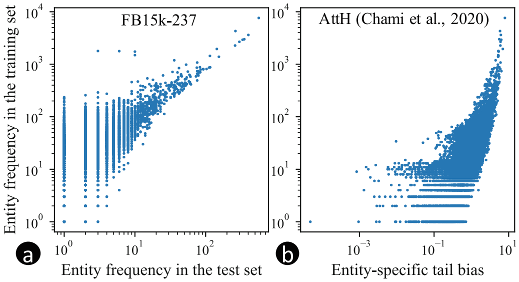

Fig. 1. (a) High correlations of the entity frequency (occurrence of triples) in the training set with that in the test set, and (b) with the scoring bias terms b_t of a KGE model [6].

To these ends, we propose Holme that addresses multi-fold extrapolation: to unseen triples, to under-represented entities, and to under-represented relations. Our contributions are summarised as below:

- We give a detailed analysis of *extrapolation* in multiple aspects (Sect. 3), including extrapolation to: (1) unseen triples; (2) under-represeted entities; (3) under-represented relations;

- We propose a general form of KGE method HolmE, consisting of rotation and translation in product space of manifolds (Sect. 4). We provide extensive empirical evidence (Sect. 5) that HolmE outperforms SotA in low-dimensional space in multiple aspects of extrapolation.
- We give in-depth analysis of the influence of biases of KGE models (Sect. 5.3) and argue that the bias term should be treated with special care, depending on the assumption whether the test data has an entity distribution similar to that of the training data.
- We provide theoretical proof that HolmE supports strong composition, and that main stream KGE methods are either special cases of HolmE and thus support strong composition or they do not satisfy strong composition (Sect. 4).

2 Related Work

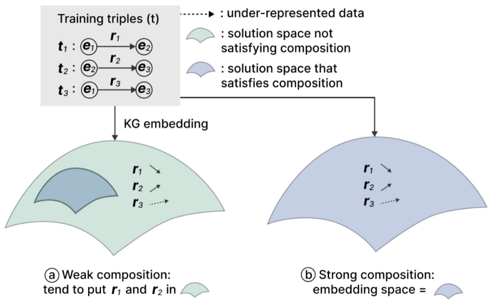
Training triples (t)
t_1 : e_1 $\xrightarrow{r_1}$ e_2
t_2 : e_2 $\xrightarrow{r_2}$ e_3
t_3 : e_1 $\xrightarrow{r_3}$ e_3
: under-represented data
: solution space not satisfying composition
: solution space that satisfies composition
KG embedding
r_1
r_2
r_3
r_1
r_2
r_3
(a) Weak composition: tend to put r_1 and r_2 in
(b) Strong composition: embedding space =

Fig. 2. Difference between weak composition and strong composition: models satisfying weak composition will tend to embed r_1 and r_2 in a subspace that does not satisfy composition if r_3 (or the composition pattern) is under-represented in the training set; models satisfying strong composition do not have this problem, because their whole embedding space $\mathcal{G}$ satisfies composition.

In general, KGE studies have explored in two directions: exploration of expressive spaces for entity embeddings, and appropriate mathematical transformation that model relations as mappings.

Geometric Space. The first group of work has focused on Euclidean spaces, including TransE [3], TransR [21]. Another group of work has explored 2D complex space, e.g., ComplEx [27], RotatE [25]. Hyperbolic space refers Riemannian distance spaces with negative curvature. Hyperbolic embedding achieves good performance even in low-dimensional embedding due to its high expressivity and suitability for hierarchical structure [6], which is common in KGs. Past work explored various curvature spaces, such as MurP [1], AttH [6], TorusE [14], and GIE [5].

Mapping Modelling. Translational or rotational models have clear geometrical explanations, and support many relation patterns, e.g., TransE [3], RotatE [25], and some variants such as TransH [30], TransD [17]. Bilinear models [2,34] as another group of KGE models treat both entities and relations as embedding vectors/matrices and perform multiplication scoring.

Neural Networks. These works, including ConvE [11], ConvKB [10] and CompGCN [28] use graph neural networks to encode KGs [9]. Recently transformer, which exhibits strong modelling power in natural language models has already been used for learning embeddings for various KG-related link prediction tasks, such as RETRA [31]. These methods can also have good performance, although they usually lack explicitly geometrical explanation on relation patterns and do not explicitly support relation patterns [5,20,24].

Extrapolation to Under-represented Data. The fairness by the view of the representation bias is becoming a hot topic in research [4,8,22]. In KGE-related fields, this issue refers to the modelling of entities or relations that are long-tail, i.e. under-represented in KGs. The existence of those long-tail data deteriorate not only the modelling capability for those specific items, but also the general performance of the mothods [9,13]. KGEs for long-tail entities are studied in [19, 33]. The work [33] uses neural networks to learn KG structure and attribute information, and tested on datasets with attribute information, falling in a rather different framework than KGE. Another work [19] tested their models on long-tail entities of FB15k-237, but insufficiently explained what factors are important. KGE for under-represented relations is partially covered in few-shot relation KG completion. A series of work [29,35] trains neural networks or bilinear models on triples with few-shot relations tested on datasets such as NELL, Wiki. This also falls in a rather different framework. In summary, exploring fairness from the view of representation bias, which focus on extrapolation to under-represented data in the KGE framework still remains to be explored. Our work also falls within this category.

3 Problem Setup

3.1 Knowledge Graph Embedding (KGE)

A knowledge graph (KG) is a multi-relational graph, denoted as $G = (\mathcal{E}, \mathcal{R}, \mathcal{F})$, where $\mathcal{E}$ is a set of entities (nodes), $\mathcal{R}$ is a set of binary relations (edge types) between entities, and $\mathcal{F}$ is a set of facts (edges) given in the triple form of $(h, r, t) \in \mathcal{F} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$, with h, t and r denoting the head entity, tail entity, and the relation in between respectively.

A common setting of KGE problem seeks to solve the problem of link prediction: $(h, r, ?)$ (or $(?, r, t)$), namely given the *query* of the head entity (or tail entity) and the relation, to find the most probable tail entity (or head entity). For simplicity, we denote the query in both two directions as $(h, r) \rightarrow t$.

Let $(\mathcal{P}, \mathcal{G})$ be an embedding space, where $\mathcal{P}$ is a distance space, $\mathcal{G}$ is a set of mappings that have domain and range defined on $\mathcal{P}$, and s is a scoring function: $s : \mathcal{P} \times \mathcal{G} \times \mathcal{P} \rightarrow \mathbb{R}$. The KGE problem aims to find an embedding from G to $(\mathcal{P}, \mathcal{G})$ that (i) maps entities $h, t \in \mathcal{E}$ to points (vectors) $\mathbf{e}_h, \mathbf{e}_t \in \mathcal{P}$; (ii) maps each relation $r \in \mathcal{R}$ to a map $g_r \in \mathcal{G}$ such that $s(\mathbf{e}_h, g_r, \mathbf{e}_t)$ ranks how probable that $(h, r, t) \in \mathcal{F}$.

3.2 Riemannian Geometry

We briefly introduce Riemannian geometry and refer the readers to textbooks [32]. A Riemannian manifold is a distance space where the distance between two points are characterised by a Riemannian metric. A hyperbolic space is a Riemannian manifold with constant negative curvature $-\kappa$ and dimension d as $\mathbb{H}^{d,\kappa} := \{\mathbf{x} \in \mathbb{R}^d \mid ||\mathbf{x}||^2 < \frac{1}{\kappa}\}$, where $||\cdot||$ denotes the L2 norm.

For each point on a non-Euclidean manifold $\mathbf{x} \in \mathbb{H}^{d,\kappa}$, the tangent space $\mathcal{T}_{\mathbf{x}}^c$ is a d-dimensional vector space containing all possible directions of paths in $\mathbb{H}^{d,\kappa}$ leaving from $\mathbf{x}$. The transformation of $\mathbf{x}$ on $\mathbb{H}^{d,\kappa}$ to the $\mathcal{T}_{\mathbf{x}}^c$ is referred to as the *logarithmic map*, and from $\mathcal{T}_{\mathbf{x}}^c$ to $\mathbb{H}^{d,\kappa}$ the *exponential map*. Their closed-form expressions at the origin:

$$exp_0^{\kappa}(\mathbf{e}) = tanh(\sqrt{\kappa}||\mathbf{e}||)\frac{\mathbf{e}}{\sqrt{\kappa}||\mathbf{e}||} \tag{1}$$

$$log_0^{\kappa}(\mathbf{e}) = arctanh(\sqrt{\kappa}||\mathbf{e}||)\frac{\mathbf{e}}{\sqrt{\kappa}||\mathbf{e}||} \tag{2}$$

Following previous work, the translation in hyperbolic space is defined as the Möbius addition [15], denoted as $\oplus^{\kappa}$, which provides an analogue to Euclidean addition for hyperbolic space:

$$\mathbf{x} \oplus^{\kappa} \mathbf{y} = \frac{(1 + 2\kappa \langle \mathbf{x}, \mathbf{y} \rangle + \kappa||\mathbf{y}||^2)\mathbf{x} + (1 - \kappa||\mathbf{x}||^2)\mathbf{y}}{1 + 2\kappa \langle \mathbf{x}, \mathbf{y} \rangle + \kappa^2||\mathbf{x}||^2||\mathbf{y}||^2} \tag{3}$$

The distance on $\mathbb{H}^{d,\kappa}$ is defined as:

$$d^{\kappa}(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{\kappa}} arctanh(\sqrt{\kappa}||-\mathbf{x} \oplus^{\kappa} \mathbf{y}||) \tag{4}$$

Product Space of Hyperbolic Space: $\mathbb{P}_{m,n,\kappa}$ of dimension $d = m \cdot n$ consists of m component spaces $\mathbb{H}^{n,\kappa}$ of dimension n: $\mathbb{P}_{m,n,\kappa} = \overbrace{\mathbb{H}^{n,\kappa} \times \ldots \times \mathbb{H}^{n,\kappa}}^{m \text{ times}}$.

A vector on $\mathbb{P}_{m,n,\kappa}$ can be decomposed into m sub-vectors in $\mathbb{H}^{n,\kappa}$ of dimension n. For any $\mathbf{x}, \mathbf{y} \in \mathbb{P}_{m,n,\kappa}$, where $\mathbf{x} = (\mathbf{x}^1, \ldots, \mathbf{x}^m)$, $\mathbf{y} = (\mathbf{y}^1, \ldots, \mathbf{y}^m)$, the Möbius addition $\oplus^{\kappa}$ can be extended to $\oplus_{\mathcal{P}}^{\kappa}$ on $\mathbb{P}_{m,n,\kappa}$:

$$\mathbf{x} \oplus_{\mathcal{P}}^{\kappa} \mathbf{y} = (\mathbf{x}^1 \oplus^{\kappa} \mathbf{y}^1, \ldots, \mathbf{x}^m \oplus^{\kappa} \mathbf{y}^m). \tag{5}$$

The distance between$\mathbf{x}, \mathbf{y}$ is calculated as the sum of all Riemannian distances between the sub-vectors $\mathbf{x}^i, \mathbf{y}^i$ (Eq. 6).

$$d_{\mathcal{P}}^{\kappa}(\mathbf{x}, \mathbf{y})^2 = \sum_{i=1}^{m} \left(\frac{2}{\sqrt{\kappa}} arctanh(\sqrt{\kappa}||-\mathbf{x}^i \oplus^{\kappa} \mathbf{y}^i||) \right)^2. \tag{6}$$

3.3 Extrapolation

Extrapolation to Unseen Triples. This is the most studied part in the past work, which refers to the *ability of models that can extrapolate from seen facts to unseen facts.* Commonly, a KG is split into the training set $\mathcal{F}_{tr}$, validation set $\mathcal{F}_{val}$, and test set $\mathcal{F}_{tst}$. The $\mathcal{F}_{tr}$ is used to train the KGE model, the $\mathcal{F}_{val}$ is to select the hyper-parameters of the KGE model, and $\mathcal{F}_{tst}$ is to test the extrapolation, where the three sets do not share any triples.

Extrapolation to Under-represented Entities. Under-represented entities are those entities that have a limited number of occurrence in the training set $\mathcal{F}_{tr}$. We formulate it as entities with frequency less than a threshold ϵ_e: $e \in \{e|f_e < \epsilon_e\}$, where ϵ_e should be a reasonable number depending on actual situations. A good KGE model should still perform well on triples that contain under-represented entities. Past work rely on learning biases for different entities, which essentially service as a term that adjusts the scores to match the prior probability of the entities. This approach has a strong drawback: it assumes the prior probabilities of the entities are similar across the $\mathcal{F}_{tr}$ and $\mathcal{F}_{tst}$ as in Fig. 1. This is not guaranteed to be true in real-world situations. For example in industry the KGE models can trained on large general dataset but applied on specific test sets. This significantly limits the extrapolation of KGE models to datasets whose prior probabilities are different from the training set. A better solution would be to decide the usage of the prior probabilities of entities depending on the actual situations.

Extrapolation to Under-represented Relations. Under-represented relations are relations that have a limited presence in the training data $\mathcal{F}_{tr}$. We formulate it as relations with a frequency less than a threshold ϵ_r: $r \in \{r|f_r < \epsilon_r\}$.

3.4 Composition Patterns

We differentiate between *weak composition patterns* and *strong composition patterns*. An intuitive understanding of both is provided in Fig. 2, with formal definitions presented below.

Weak Composition. A KGE model $(\mathcal{P}, \mathcal{G}, s)$ satisfies *weak composition* if there exist $g_1, g_2, g_3 \in \mathcal{G}$ such that $\forall \mathbf{e}, \mathbf{e}', \mathbf{e}'' \in \mathcal{M}$, we have $g_1(\mathbf{e}) = \mathbf{e}' \wedge g_2(\mathbf{e}') = \mathbf{e}'' \Rightarrow g_3(\mathbf{e}) = \mathbf{e}''$.

Strong Composition. A KGE model $(\mathcal{P}, \mathcal{G}, s)$ satisfies *strong composition* if for any $g_1, g_2 \in \mathcal{G}$, there exists $g \in \mathcal{G}$ such that $\forall \mathbf{e}, \mathbf{e}', \mathbf{e}'' \in \mathcal{M}_{ent}$, we have $g_1(\mathbf{e}) = \mathbf{e}' \wedge g_2(\mathbf{e}') = \mathbf{e}'' \Rightarrow g(\mathbf{e}) = \mathbf{e}''$.

4 Method: HolmE

We aim to design a KGE model that can (1) encode strong relation patterns such as symmetry, inversion and strong composition; (2) learns KG embeddings in Riemannian space that allows better expressivity in low-dimensional space; (3) possesses good extrapolation to unseen triples, under-represented entities, and relations. We elaborate on the formulas of HolmE and the rationale behind them; then prove HolmE supports strong relation patterns; and then compare HolmE to representative KGE methods.

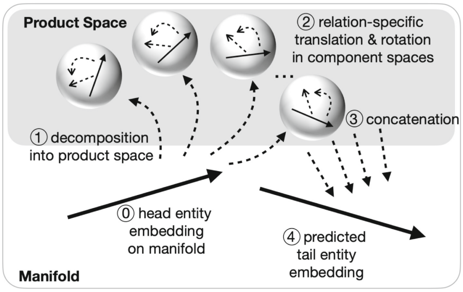
Product Space
② relation-specific translation & rotation in component spaces
...
③ concatenation
① decomposition into product space
⓪ head entity embedding on manifold
④ predicted tail entity embedding
Manifold

Fig. 3. HolmE performs relation-specific translation and rotation in component spaces of the product space (illustrated with 3D balls, with the shape $\mathbb{P}_{m,n,\kappa}, m, n, \kappa$ indicate the number of component spaces, the component space ($\mathbb{H}^{n,\kappa}$) dimension and the curvature of the hyperbolic space respectively.)

4.1 HolmE: Holomorphic KG Embedding

Intuition. HolmE decomposes the embeddings of the head entity in the manifold $\mathbb{P}_{m,n,\kappa}$ (Fig. 3.0) to sub-vectors in the component spaces of the product space $\mathbb{H}^{n,\kappa}$ (Fig. 3.1), and performs relation-specific translation and rotation to the sub-vectors (Fig. 3.2); then, the sub-vectors are concatenated (Fig. 3.3) to the predicted embedding of the tail entity (Fig. 3.4). We give the general form of HolmE as follows:

Definition 1. *Let $(\mathcal{P}, \mathcal{G})$ be the embedding space, where $\mathcal{P} = \mathbb{P}_{m,n,\kappa}$, and $\mathcal{G}$ is a group of mappings:*

$$g_{(\boldsymbol{a},\boldsymbol{A})}: \mathbb{P}_{m,n,\kappa} \rightarrow \mathbb{P}_{m,n,\kappa} \tag{7}$$

$$\boldsymbol{e} \mapsto \boldsymbol{a} \oplus_{\mathcal{P}}^{\kappa} (\boldsymbol{A} \cdot \boldsymbol{e}). \tag{8}$$

and the scoring function of a tuple (h, r, t) is defined by

$$s(\boldsymbol{e}_h, g_r, \boldsymbol{e}_t) = -d_{\mathcal{P}}^{\kappa}(g_r(\boldsymbol{e}_h), \boldsymbol{e}_t)\,[+b_h + b_t], \tag{9}$$

$$g_r(\boldsymbol{e}_h) = \boldsymbol{a}_r \oplus_{\mathcal{P}}^{\kappa} (\boldsymbol{A}_r \cdot \boldsymbol{e}_h), \tag{10}$$

where $\boldsymbol{e}_h, \boldsymbol{e}_t \in \mathcal{P}$ are embeddings of the head entity and tail entity on the manifold; $g_r = g_{(\boldsymbol{a}_r,\boldsymbol{A}_r)} \in \mathcal{G}$ is the relation-specific mapping, with parameters of translation $\boldsymbol{a}_r \in \mathcal{P}_{m,n,\kappa}$ and rotation $\boldsymbol{A}_r \in \boldsymbol{SO}(d)$ (special orthogonal matrix); $\oplus_{\mathcal{P}}^{\kappa}$ is the extended Möbius addition on product space $\mathcal{P}$ with curvature κ; b_h and b_t are biases of the head entity and the tail entity (they are optional and thus in brackets, details see the bias paragraph and Sect. 5.3); the distance $d_{\mathcal{P}}^{\kappa}$ is calculated on the product space $\mathcal{P}$.

We now elaborate on each part of HolmE and the rationales behind them in the following paragraphs.

Entity Embedding with Learned Curvature. The entities are embedded as vectors in a hyperbolic manifold $\mathcal{M}$ with a learned constant curvature κ (Fig. 3.0). They are decomposed into sub-vectors in a product space $\mathcal{P}$ with a series of n dimensional component spaces (Fig. 3.1). These component spaces also share the same κ. The curvature is learned because it adjusts the embedding space to better distribute points throughout the space [6]. We adopt a constant curvature because it is required to support strong composition[1].

Relation Embedding. The relation mapping g_r consists of two components, introduced as below:

Rotation in Product Space. A tempting choice is to model rotation simply as a high dimensional rotation matrix, which contains a high number of parameters and is very difficult to learn. HolmE decomposes a high dimensional rotation into the product space with a series of n dimensional component spaces of curvature κ (Fig. 3.2). The rotation matrix $\mathbf{A}_r$ is a special orthogonal matrix (namely $|\mathbf{A}_r| = 1$), in the form of a diagonal rotation matrix consisting of a series of rotation matrices of n dimensions: $\mathbf{A}_r = diag[\mathbf{R}^n(\theta_{r,1}), \ldots, \mathbf{R}^n(\theta_{r,d/n})]$, where $\mathbf{R}^n(\theta_r)$ is a n dimensional rotation matrix, $(n = 2, 3, \ldots)$. When $n = 2$, $\mathbf{R}^n(\theta_r)$ is the special case of rotation in complex space (Eq. 11), and HolmE becomes a holomorphic function (this is where the name HolmE comes from).

$$\mathbf{R}^n(\theta_r) := \begin{bmatrix} cos(\theta_r) & -sin(\theta_r) \\ sin(\theta_r) & cos(\theta_r) \end{bmatrix}. \tag{11}$$

Translation in Product Space. The translation of HolmE is performed with extended *Möbius addition* (Eq. 5) in the component spaces with the curvature κ (Fig. 3.2). This means the Möbius addition is performed between the sub-vectors of $\mathbf{a}_r$ and the resulting vectors of the rotation $(\mathbf{A}_r \cdot \mathbf{e}_h)$, where each sub-vector has the shape $\mathcal{P}_{d,n}$. We adopt the translation in the product space instead of high dimensional translation, different from past work [6], because translation in the product space matches the rotation in the product space and thus geometrically makes more sense, and this matching is required by strong composition. Important to note is that the Möbius addition here is the *left addition*, namely the translation vector $\mathbf{a}_r$ must be on the left hand side, for ensuring strong composition (See footnote 1).

Scoring Function and Biases. HolmE has two forms of scoring functions: (1) one is simply the distance $s = d^{\kappa}_{\mathcal{P}}(g_r(\mathbf{e}_h), \mathbf{e}_t)$, which gives HolmE; (2) the other one is the distance adding the biases of head entity and tail entity $s = d^{\kappa}_{\mathcal{P}}(g_r(\mathbf{e}_h), \mathbf{e}_t) + b_h + b_t$, which results in HolmE-b. The two forms of HolmE should be used in different scenarios. HolmE-b should be used when the test set is assumed to have an entity distribution similar to the training set; HolmE should be used in other cases. Details see Sect. 5.3.

[1] Proof see https://github.com/nsai-uio/HolmE-KGE/blob/main/Proof.pdf.

Loss and Training of HolmE. We adopt cross-entropy loss with uniform negative sampling as past works:

$$\mathcal{L} = \sum_{(h,r,t)\in\mathcal{F}\cup\mathcal{F}'} log(1 + exp(-ys(\mathbf{e}_h, g_r, \mathbf{e}_t))),$$

where $\mathcal{F}$ and $\mathcal{F}'$ denote training examples and negative examples respectively. The negative examples are sampled uniformly from all possible triplets obtained by perturbing the tail entities in triplets (h, r, t). The label y is 1 if the triplet is training example, else is −1. The training setting follows the initialisation in tangent space as [7]. In particular, all parameters are initialised and optimised in the tangent space of the manifold using standard Euclidean techniques. These parameters are mapped to the manifold with exponential map, which ensures the embeddings are in the desired space [7].

4.2 Model Analysis

Benefits. HolmE has two major benefits and several supporting benefits: (1) The design of HolmE gives clear geometric meaning and relatively good transparency (illustrated in Fig. 3). (1.1) HolmE decomposes high dimensional computation in product space and saves parameters. (1.2) HolmE has a matching rotation, translation and distance function on the product spaces. (2) HolmE has good extrapolation to unseen triples, under-represented entities and relations (Sect. 5). (2.1) HolmE provides optional biases applied according to actual scenarios. (2.2) HolmE supports strong composition (Theorem 1).

Theorem 1. *The KGE defined by Definition 1 satisfies strong composition*[2]*.*

Complexity Analysis. The time complexity of HolmE mainly comes from scoring calculation and cross entropy loss calculation. While the operation for calculating loss is consistent across various KGE models, the complexity of the score function in d dimentional HolmE is linear, i.e., $O(d)$ for each triple. This complexity is comparable to that of traditional KGE models such as TransE and lower than that of neural network-based models such as ConvE, as no additional graph aggregation operations are incorporated, making it more efficient in scenarios where computational simplicity and speed are critical. In terms of space complexity, for a KG with $|\mathcal{E}|$ entities and $|\mathcal{R}|$ relations, HolmE need $d\,|\mathcal{E}|+2d\,|\mathcal{R}|$ parameters, while HolmE-b need more d parameters for entity-specific bias term. Compared with other Riemannian KGE, such as AttH and GIE, (both need at least $(d+1)\,|\mathcal{E}| + 3d\,|\mathcal{R}|$ parameters), HolmE need fewer parameters. Compared with other Euclidean KGE methods, HolmE achieves similar representation performance in a lower embedding dimension [1,6].

[2] Proof see https://github.com/nsai-uio/HolmE-KGE/blob/main/Proof.pdf..

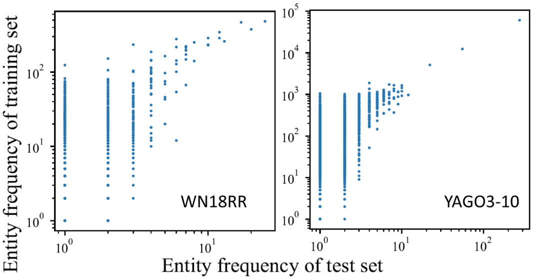

Fig. 4. Entity frequency in training and test set.

Limitations. HolmE is limited in providing theoretically strict mappings satisfying the *transitivity* pattern and mapping patterns of *one-to-many* (1-to-N), *many-to-one* (N-to-1), and *many-to-many* (N-to-N). However, empirical studies show that HolmE still has good performance on these relation patterns or mapping patterns, due to the mechanism in KGE that selects the best scored predicted tail embedding instead of the exact tail embedding (see Appendix).

Comparison to TransE/RotatE. HolmE adopted similar components of translation and rotation as in other translational KGEs. HolmE is a general form that unites KGEs supporting strong composition, including TransE and RotatE. KGEs like neural networks, bilinear models, mixture models cannot guarantee to satisfy strong composition because they lack explicit geometric explanation and the notion of *relation embedding as mapping* is not clearly defined.

Comparison to Hyperbolic KGE. MurP [1] adopts translation on high dimensional hyperbolic space and does not adopt rotation. AttH [6] adopts rotation in product space, but translation and distance function on high dimensional manifolds, with the result that AttH does not support strong composition. GIE [5] is a mixture model of different curvature spaces and not pure hyperbolic, and also does not support strong composition. In contrast, HolmE has rotation, translation, distance on product space and supports strong composition; HolmE also has a special treatment to the bias terms.

5 Experiments

We evaluates these hypotheses: (H1) HolmE outperforms SotA on extrapolation to unseen triples in low-dimensional space; (H2) HolmE outperforms SotA on extrapolation to under-represented data; (H2.1) in scenarios where entity's frequency distribution between the training and testing sets are sufficiently different, HolmE outperforms others; (H2.2) in scenarios where relation's distribution between the training and testing sets are sufficiently different, HolmE outperforms others.

Table 1. Dataset statistics. $\bar{f}_e$, $\bar{f}_r$: average of entity and relation frequency.

Dataset	#ent	#rel	#tri	$\bar{f}_e$	$\bar{f}_r$
WN18RR	41k	11	93k	4.3	7.9k
FB15k-237	15k	237	310k	37.5	1.1k
YAGO3-10	123k	37	1M	17.5	29.1k

5.1 Overall Experiment Design

To test H1, we conduct experiments (Sect. 5.2) to train HolmE and baseline KGE models on $\mathcal{F}_{tr}$ and test on unseen triples $\mathcal{F}_{tst}$. This follows the standard setting of past work. Moreover, we design experiments evaluating the influence of strong composition. To test H2 (H2.1 in Sect. 5.3, H2.2 in Sect. 5.4), we plot the MRR along entity (Sect. 5.3) or relation frequency to evaluate performance of various models. In addition, we split the benchmark data to create scenarios where the frequency distribution of entities and relations are sufficiently different, we test the model performance on these scenarios. In all of our experiments, we conducted a hyperparameter search for the batch size b (chosen from {500, 1000, 2000}); optimizer op (selected from Adam or Adagrad); the learning rate lr (for Adam ranging from [0.0001, 0.005], for Adagrad ranging from [0.01, 0.1]); negative sampling size neg (chosen from {50, 150, 200}). For the high-dimensional embedding, the embedding size is searched in the range of $\{200, 400, 500\}$. All the experiments are done using GPU Nvidia A100-80GB (training and testing models) and CPU Core(TM) i7-11850H (other scripts).

5.2 Extrapolation to Unseen Triples

Benchmark Datasets. We use 3 standard datasets: WN18RR [11], FB15k-237 [26], YAGO3-10 [23]. The statistics of these datasets are reported in Table 1. Figure 1a and Fig. 4 indicate in all three benchmark datasets, there is high correlation between entity frequency in the training and test set.

Datasets of under Represented Entities. The datasets WN18RR-under-e and FB15k-237-under-e are generated by selecting all triples with under-represented target entities, where under-represented entities are entities with frequency less than a threshold and stand for about 80% of total entities. Table 2 reports the statistics of these new test data. Appendix shows that the different values of threshold do not influence the validity of claims in the paper.

The series of test sets with different percentage of under-represented entities shown in Fig. 7 is generated by random sampling. We denote the set of triples of under-represented entities in the WN18RR or FB15k-237 as $\mathcal{F}_{ue}$ (with cardinality f_{ue}) and the set of other triples as $\mathcal{F}_{we}$ (with cardinality f_{we}). We want to generate a dataset with the percentage of triples with under-represented entities of p_{tst}. then the percentage of other triples is $1 - p_{tst}$. For each p_{tst}, we would like to generate the test set repeatedly in five different scales $sc \in \{20\%, 40\%, 60\%, 80\%, 100\%\}$. Then the random sampling is performed as: For a given p_{tst}, for a scale $sc \in \{20\%, 40\%, 60\%, 80\%, 100\%\}$, we uniformly random sample $f_{ue} \times po$ triples from $\mathcal{F}_{ue}$ and uniformly random sample $(f_{ue} \times po)/p_{tst} * (1 - p_{tst})$ triples from $\mathcal{F}_{we}$.

Datasets of Under-represented Relations. Similarly, the datasets WN18RR-under-r and FB15k-237-under-r are generated by selecting all triples with under-represented relations, where under-represented relation are relations

Table 2. Dataset statistics for test sets with under-represented entities and relations. f_u/f_{total} indicates the number of test queries in the dataset defined by the number of test queries in their original test sets of WN18RR/FB15k-237. Threshold is the frequency to determine under-represented entities/relations such that under-represented entities/relations always take about 80% of the total entities/relations.

Dataset	#entities	entity percentage	#test queries	f_u/f_{total}	threshold
WN18RR-under-e	33k	81.0%	3487	55.6%	5
FB15k-237-under-e	11k	80.4%	18k	44.9%	47
Dataset	#relations	relation percentage	#test queries	f_u/f_{total}	threshold
WN18RR-under-r	9	81.8%	1618	25.8%	7402.0
FB15k-237-under-r	189	79.7%	10k	26.5%	1282.4

Table 3. HolmE outperforms SotA in link prediction for extrapolation to unseen triples in low-dimensional space (d = 32). Best score in bold and second best underlined. Sources are indicated by citations, or generated/reproduced by us with open source code.

$\mathcal{M}$	Model	WN18RR				FB15k-237				YAGO3-10			
		MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
$\mathbb{R}$	TransE	.177	.031	.269	.447	.110	.066	.114	.196	–	–	–	–
$\mathbb{C}$	RotatE [6]	.387	.330	.417	.491	.290	.208	.316	.458	–	–	–	–
$\mathbb{C}$	ComplEx-N3 [6]	.420	.390	.420	.460	.294	.211	.322	.463	.336	.259	.367	.484
$\mathbb{R}$	ConvE	.438	.410	.440	.520	.325	.237	.356	.501	.430	.335	.479	.580
$\mathbb{R}$	CompGCN	.467	.433	.481	.532	.323	.237	.351	.500	.438	.329	.502	.586
$\mathbb{H}$	MurP [6]	.465	.420	.484	.544	.323	.235	.353	.501	.230	.150	.247	.392
$\mathbb{H}$	AttH [6]	.466	.419	.484	.551	.324	.236	.354	.501	.397	.310	.437	.566
$\mathbb{M}$	GIE	.472	.424	.492	.558	.320	.230	.350	.503	–	–	–	
$\mathbb{P}$	HolmE-b	**.486**	**.440**	**.502**	**.575**	.329	**.238**	.358	.511	**.454**	**.349**	**.518**	**.655**
$\mathbb{P}$	HolmE	.466	.415	.489	.561	**.331**	.237	**.366**	**.517**	.441	.333	.507	.641

with frequency less than a threshold and stand for about 80% of total relations. Table 2 reports the statistics of these new test data.

Baselines. We compare HolmE and the variant with bias (HolmE-b) with three types of baselines: (1) representative works that provide insights in this field, including TransE [3], RotatE [25], and ComplEx-N3 [18]; (2) graph neural network based models, including ConvE [11] and CompGCN [28] (2) Riemannian KGE, including MurP [1], AttH [6], and a mixture model of Euclidean, sphere, and hyperbolic space, GIE [5].

Evaluation Metrics. We adopt *mean reciprocal rank* (MRR), calculated as the mean of reciprocal ranks of the predicted entity; hits at k (H@K, $k \in \{1, 3, 10\}$), calculated as the percentage of the correct triples among the top k predictions.

Results Analysis. The results (Table 3) show that the HolmE models can outperform baselines in low-dimensional embedding space. On WN18RR, FB15k-237, YAGO3-10, the best HolmE model outperform the best published baseline

Table 4. Verifying the influence of strong composition with pretraining and fine-tuning

Training set:	Pretrained on train-remain			Fine tuned on train-com		
Test set	test-remain	test-com	total-test	test-remain	test-com	total-test
AttH	0.324	0.064	0.294	0.227	0.272	0.228
AttH-ub	0.318	**0.105**	0.294	0.228	0.248	0.229
HolmE-b	0.337	0.090	**0.342**	**0.246**	**0.285**	**0.247**
HolmE	**0.331**	0.050	0.299	0.236	0.276	0.237

by 3%, 2.2%, and 14.4%, respectively. HolmE is sometimes worse than models with bias (GIE, HolmE-b), because the three datasets have similar entity distribution in the training set and test set (Fig. 1a and Fig. 4). Though, HolmE can still perform impressively well, especially on FB15k-237. We postulate in this case, the gain of supporting strong composition sometimes outmatches the loss of dropping the bias terms.

Link Prediction Performance along Embedding Dimension. Following [6], we test HolmE with different embedding sizes and compare with representative baselines. The results (Fig. 5) reveal that HolmE achieves impressive performance in low dimensions; and it consistently outperform these baselines in a wide range of dimensions.

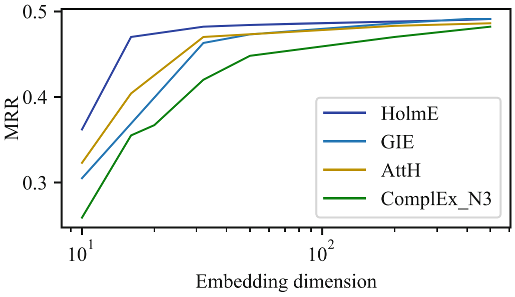

Fig. 5. MRR for KGE models with $d \in \{10, 16, 20, 32, 50, 200, 500\}$ on WN18RR.

Influence of Strong Composition. To verify that strong composition can really improve model performance for under-represented relations, we design experiments of fine tuning. The rationale behind is that strong composition has entire relation embedding space supporting composition, thus under-represented relations will also be embedded to support composition. To verify this, we remove a subset of triples from the training set of FB15k-237, such that some relations (denoted as r^{com}) that support composition pattern become under-represented in the remaining training set (train-remain). The removed subset is denoted as train-com. Correspondingly, we also remove all triples that contain r^{com} from the test set of FB15k-237 (test-total), the remaining test set is denoted as test-remain, and the removed subset is denoted as test-com. KGE models will be first pre-trained on FB15k-237, and tested on test-remain and test-com. Then these pre-trained models will be fine-tuned with train-com and again tested on test-remain and test-com. The results (Table 4) shows after fine-tuning, HolmE always outperform their counterpart in the case of with bias or

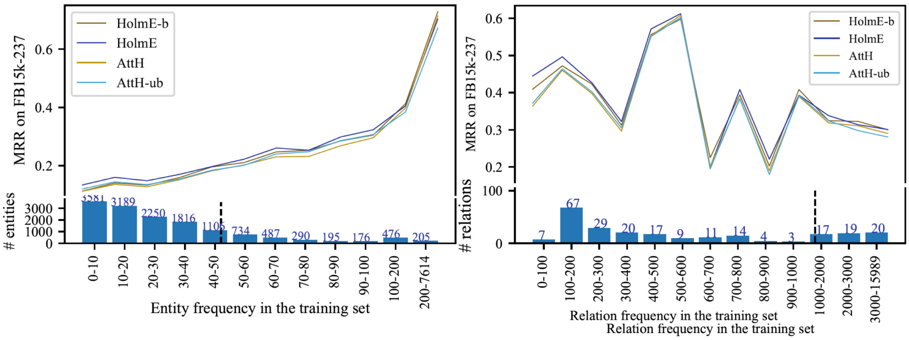

Fig. 6. MRR of four models (Table 3) along entity and relation frequency for FB15k-237. Left: for triples with under-represented entities, unbiased model (HolmE, AttH-ub) outperform biased model (HolmE-b, AttH), and HolmE outperform other models. Right: HolmE achieves better performance for triples with under-represented relations.

without bias: HolmE-b > AttH, HolmE > AttH-ub. The performance of all model deteriorate on test-remain and increases significantly on test-com. We postulate that HolmE models can easily adapt the embeddings learned on *train-remain* for under-represented relations to new embeddings when the composition patterns are revealed by *train-com*. While AttH models are worse in this regard, we postulate the reason is the relation embeddings learned by AttH are in the sub-space that does not support composition after pretraining on *train-remain*, and need more change to be adapted to the sub-space that supports composition when the composition patterns are revealed by *test-com*. HolmE models are easier to adapt the relation embeddings because the its entire relation embeddings space supports strong composition.

5.3 Extrapolation to Under-Represented Entities

Correlations between Entity Distribution and Bias Terms. To verify H2.1, we test the correlations between entity frequency in the training set and test set ($F_{tr} - F_{tst}$), between the training set and the bias terms ($F_{tr} - b_t$) for the WN18RR and FB15k-237. The results show that these distributions are highly correlated (Fig. 1 and Appendix). Especially FB15k-237 exhibits a Pearson correlation coefficient of 0.91 for $F_{tr} - F_{tst}$. The other correlations are also evident, and sometimes of higher orders than simple linear correlation.

Model Performance Influenced by Entity Frequency. To verify H2.1, we plot the number of entities along the entity frequency (occurrence of triples) in the training set of FB15k-237 (Fig. 6.left), which has a reasonably rich set of relations and entities. We also plot performance of four KGE models in Table 3 along the entity frequency in Fig. 6.left. We choose to compare AttH with HolmE because it is a representative hyperbolic KGE model and it is the best SotA model in low-dimensional space for FB15k-237. In addition to the original AttH

Table 5. Testing extrapolation to under-represented entities and relations. Models without bias (HolmE, AttH-ub) outperform models with bias (HolmE-b, AttH) on test set of triples only with under-represented entities. HolmE outperforms SotA models.

Model	WN18RR-under-e				FB15k-237-under-e				WN18RR-under-r				FB15k-237-under-r			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
AttH	.371	.335	.383	.437	.142	.078	.148	.269	.312	.242	.338	.434	.390	.301	.426	.566
AttH-ub	.391	**.355**	.402	.463	.148	.086	.154	.271	.276	.209	.300	.408	.392	.304	.425	.568
HolmE-b	.391	.350	.403	.472	.148	.081	.157	.280	**.341**	**.265**	**.374**	**.494**	.403	.311	.443	.579
HolmE	**.401**	**.355**	**.419**	**.483**	**.162**	**.093**	**.175**	**.297**	.308	.227	.342	.461	**.413**	**.322**	**.452**	**.590**

version with bias, we add an extra version of AttH without bias (denoted as AttH-ub). Figure 6.left reveals several observations: (1) Most entities are under-represented and have very low frequency compared to the mean frequency of 37.5 and the maximum frequency of 7614; over 10k entities have samples below 50. (2) All model performance is relatively poor for triples with under-represented entities, and increases as the entity frequency increases; (3) Models without bias (HolmE, AttH-ub) outperform models with bias (HolmE-b, AttH) for triples with under-represented entities; and the other way around for other triples. This is further studied in the next paragraph.

Scenarios Where Entity Frequency Differs between the Training and the Test Set. To better understand the influence of bias and under-represented entities on model performance and further verify H2.1, we create subsets of test data of the two benchmark data sets, which include only triples with under-represented entities (WN18RR-test-under-e and FB15k-237-test-under-e). To quantify under-represented data, we consider the Pareto principle (80%-20% principle) and choose a frequency threshold (Fig. 6.left) such that 80% entities are regarded as under-presented entities (the choice of threshold does not influence the validity of the observations and our claims, see influence of threshold in Appendix). We test the same models in Fig. 6.left on these two test sets. The results (Table 5) shows that models without bias (HolmE, AttH-ub) are persistantly better than models with bias (HolmE-b, AttH), and HolmE outperforms other models.

To ensure the created test sets reveal systematic effect, we additionally create a series of test sets with different percentage of triples with under-represented entities via random sampling from the test set of FB15k-237 (Table 2). The same four models are tested on these test sets. Figure 7 reveals the tendency of model performance along the percentage of triples with under-represented entities in the test set. including these observations: (1) As the percentage of triples with under-represented entities (p_{tst}) increases in the test sets, all model performance deteriorates; (2) As p_{tst} increases, AttH-ub (without bias) gradually outmatches AttH (with bias), and the degree that HolmE (no bias) outmatches HolmE-b increases; (3) HolmE always outperforms other models.

5.4 Extrapolation to Under-Represented Relations

Model Performance Influenced by Relation Frequency. To verify H2.2, we plot the performance of KGE model trained on FB15k-237-train along the relation frequency (Fig. 6.right). We can see for under-represented relations (rare relations) HolmE models have better performance than AttH. For well-represented relations (the very right bar), the performance of models is very close. The model performance does not change monotonously along the relation frequency. This is likely because the results are aggregated from both under-represented entities and well-represented entities, leading to a complex trend.

Scenarios Where Relation Frequency Differs between the Training and the Test Set. To verify H2.2, we create test sets of triples only with under represented relations (WN18RR-test-under-r and FB15k-237-test-under-r). The under-represented relations are relations with frequency lower than a threshold. Similar to Sect. 5.3 we adopt a threshold such that 80% relations are under-represented (the choice of threshold does not influence the validity of the observations and our claims, see influence of threshold in Appendix). We test the same four models as in Fig. 6.left on these two datasets. The results (Table 5) shows that sometimes models with bias are better than without bias (for WN18RR). We postulate it is because the results are aggregated from both under-represented entities and well-represented entities, and sometimes well-represented entities takes a major effect. HolmE models consistently outperform AttH models for under-represented relations for both the cases of with bias or without bias. This confirms H2.2.

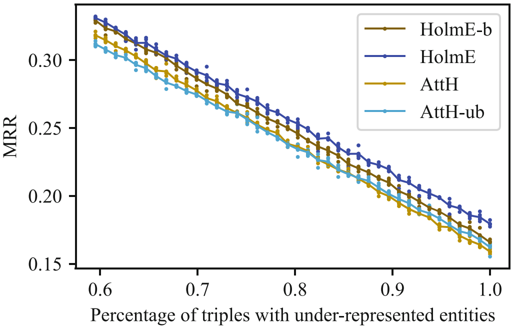

Fig. 7. Each dot is an experiment on test sets created based on FB15k-237. The lines are average performance of five repeated experiments. Performance of unbiased model (HolmE) increases as percentage of triples with under-represented entities increases. The results on the very left hand side are namely in Table 3. The results on the very right hand side are namely in Table 5.

6 Conclusion

This work proposes HolmE, a general form of hyperbolic KG embedding method that addresses multi-fold extrapolation: to unseen triples, under-represented entities, and relations (the latter two limitedly discussed in past work). We give in-depth analysis of the influence of bias terms on model performance on under-represented entities. We also show that HolmE supports strong composition, with the entire relation embedding space that support composition. We prove that main stream KGE methods are either special cases of HolmE and thus support strong composition, or they do not support strong composition.

A Result Details

Correlation between Entity Frequency and Bias Term. Fig. 8 shows that in different data sets and models, there is strong correlation between bias term and entity frequency.

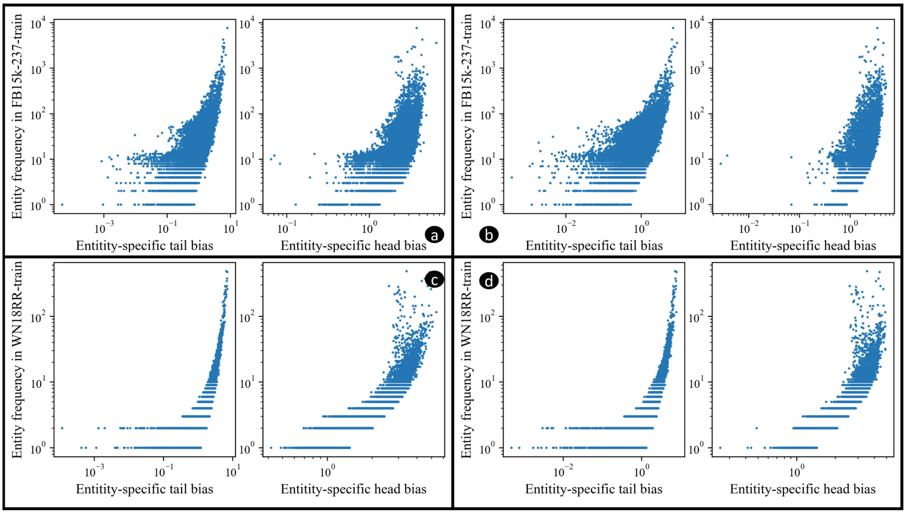

Fig. 8. Correlation between bias term and entity frequency. a,b: Atth and HolmE in FB15k-237 respectively; c,d: Atth and HolmE in WN18RR respectively.

Influence of Entity and Relation Frequency Threshold. Fig. 9 shows that results for different choice of entity and relation frequency threshold does not influence the validity of the observations and claims: (Fig. 9a) models without bias (HolmE, AttH-ub) outperform their counterparts with bias (HolmE-b, AttH), and HolmE outperform other models for triples with under-represented entities; (Fig. 9b) the majority of entities are under-represented entities as long as the threshold is not chosen to be extremely small, considering that the mean of entity frequency is 37.5 and the maximum is 7614. The results are obtained on FB15k-237. HolmE outperforms other models for triples with under-represented relations (Fig. 9c); the majority of relations are under-represented relations as long as the threshold is not chosen to be extremely small (Fig. 9d), considering that the mean of relation frequency is 1148.2 and the maximum is 15989. The results are obtained on FB15k-237.

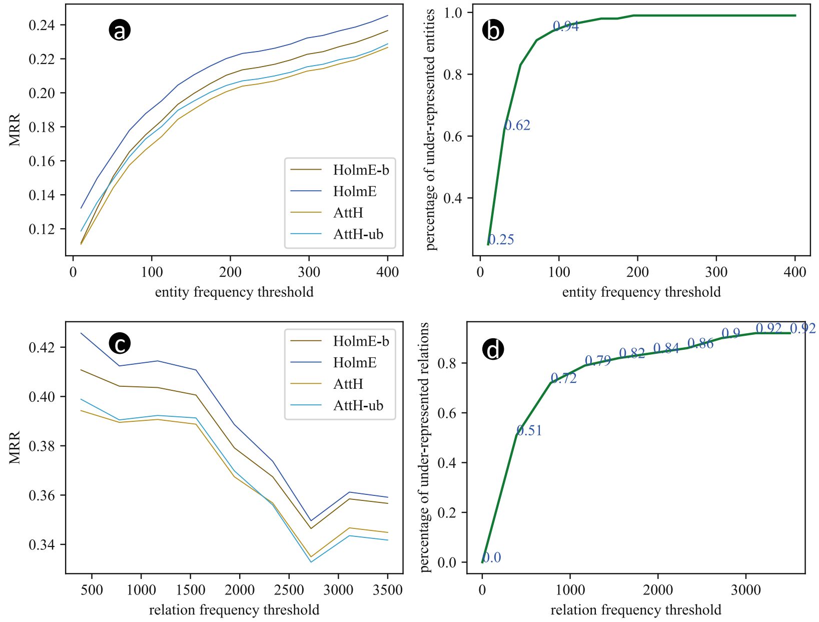

Fig. 9. The choice of threshold does not influence the observations or claims

Link Prediction Results in High Dimension. As expected, embeddings in different spaces achieve similar results in high dimensional space (Table 6), because both Euclidean and hyperbolic spaces become expressive enough to represent complex hierarchies in KGs [6]. Similar to the results in low-dimensional space, HolmE (without bias) is slightly worse than the models with bias (GIE, HolmE-b), although HolmE does not use the advantage of prior probabilities provided by the bias term (Tables 7 and 8).

Table 6. HolmE achieves comparable results as SotA hyperbolic KGE in high dimensional embedding space. Best score in bold and second best underlined. Sources are indicated by citations, or reproduced by us with open source code.

$\mathcal{M}$	Model	WN18RR				FB15k-237				YAGO3-10			
		MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
$\mathbb{R}$	TransE [3]	.226	–	–	.501	.294	–	–	.465	–	–	–	–
$\mathbb{C}$	RotatE [25]	.476	.428	.492	.571	.338	.241	.375	.533	.495	.402	.550	.670
$\mathbb{C}$	ComplEx-N3 [18]	.480	.435	.495	.572	<u>.357</u>	<u>.264</u>	<u>.392</u>	<u>.547</u>	.569	<u>.498</u>	.609	.701
$\mathbb{R}$	ConvE	.43	.40	.44	.52	.325	.237	.356	.501	–	–	–	–
$\mathbb{R}$	CompGCN	.479	.443	.494	.546	.355	.264	.390	.535	–	–	–	–
$\mathbb{H}$	MurP [1]	.481	.440	.495	.566	.335	.243	.367	.518	.354	.249	.400	.567
$\mathbb{H}$	AttH [6]	.486	.443	.499	.573	.348	.252	.384	.540	.568	.493	<u>.612</u>	<u>.702</u>
$\mathbb{M}$	GIE [5]	**.491**	**.452**	<u>.505</u>	<u>.575</u>	**.362**	**.271**	**.401**	**.552**	**.579**	**.505**	**.618**	**.709**
$\mathbb{H}$	HolmE-b	**.491**	<u>.445</u>	**.509**	**.584**	.352	.260	.383	.542	<u>.570</u>	.497	<u>.612</u>	.701
$\mathbb{H}$	HolmE	.479	.433	.497	.567	.352	.259	.389	.545	.546	.469	.597	.694

Table 7. Detailed analysis for relation patterns on FB15k-237. Sym. : Symmetry, Asym.: ASymmetry, Inv.: Inversion, Tran.: Transitivity, Comp.: Composition.

	AttH	AttH-ub	HolmE	HolmE-b
Sym	0.322	0.312	0.330	**0.336**
Asym	0.326	0.319	**0.338**	0.336
Inv	0.319	0.312	**0.331**	0.329
Tran	0.314	0.296	0.317	**0.320**
Comp	0.279	0.273	**0.292**	**0.292**

Table 8. Detailed analysis for relation mapping properties in low-dimensional space on FB15k-237.

Task	RMPs	AttH	AttH-ub	HolmE	HolmE-b
Predicting Head (MRR)	1-to-1	.460	.458	**.485**	.478
	1-to-N	.430	.411	.456	**.459**
	N-to-1	.079	.077	**.099**	.085
	N-to-N	.243	.235	**.253**	**.253**
Predicting Tail (MRR)	1-to-1	.452	.467	.475	**.477**
	1-to-N	.060	.059	**.077**	.072
	N-to-1	.728	.725	**.743**	.742
	N-to-N	.350	.343	**.360**	**.360**

References

1. Balažević, I., Allen, C., Hospedales, T.: Multi-relational poincaré graph embeddings. Adv. Neural Inf. Process. Syst. **32** (2019)
2. Balažević, I., Allen, C., Hospedales, T.: TuckER: tensor factorization for knowledge graph completion. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 5185–5194 (2019)
3. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data, vol. 26 (2013)
4. Cao, E., Wang, D., Huang, J., Hu, W.: Open knowledge enrichment for long-tail entities. In: Proceedings of The Web Conference 2020, pp. 384–394 (2020)
5. Cao, Z., Xu, Q., Yang, Z., Cao, X., Huang, Q.: Geometry interaction knowledge graph embeddings. In: AAAI Conference on Artificial Intelligence (2022)
6. Chami, I., Wolf, A., Juan, D.C., Sala, F., Ravi, S., Ré, C.: Low-dimensional hyperbolic knowledge graph embeddings. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 6901–6914 (2020)
7. Chami, I., Ying, Z., Ré, C., Leskovec, J.: Hyperbolic graph convolutional neural networks. Adv. Neural Inf. Process. Syst. **32** (2019)
8. Chu, P., Bian, X., Liu, S., Ling, H.: Feature space augmentation for long-tailed data. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12374, pp. 694–710. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58526-6_41
9. Dai, E., et al.: A comprehensive survey on trustworthy graph neural networks: privacy, robustness, fairness, and explainability. arXiv preprint arXiv:2204.08570 (2022)
10. Dai Quoc Nguyen, T.D.N., Nguyen, D.Q., Phung, D.: A novel embedding model for knowledge base completion based on convolutional neural network. In: Proceedings of NAACL-HLT, pp. 327–333 (2018)
11. Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S.: Convolutional 2D knowledge graph embeddings. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
12. Djeddi, W.E., Hermi, K., Ben Yahia, S., Diallo, G.: Advancing drug-target interaction prediction: a comprehensive graph-based approach integrating knowledge graph embedding and protbert pretraining. BMC Bioinformatics **24**(1), 488 (2023)
13. Dong, Y., Ma, J., Wang, S., Chen, C., Li, J.: Fairness in graph mining: a survey. IEEE Transactions on Knowledge and Data Engineering (2023)

14. Ebisu, T., Ichise, R.: Toruse: knowledge graph embedding on a lie group. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
15. Ganea, O., Bécigneul, G., Hofmann, T.: Hyperbolic neural networks. Adv. Neural Inf. Process. Syst. **31** (2018)
16. Islam, M.K., Amaya-Ramirez, D., Maigret, B., Devignes, M.D., Aridhi, S., Smaïl-Tabbone, M.: Molecular-evaluated and explainable drug repurposing for COVID-19 using ensemble knowledge graph embedding. Sci. Rep. **13**(1), 3643 (2023)
17. Ji, G., He, S., Xu, L., Liu, K., Zhao, J.: Knowledge graph embedding via dynamic mapping matrix. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (volume 1: Long papers), pp. 687–696 (2015)
18. Lacroix, T., Usunier, N., Obozinski, G.: Canonical tensor decomposition for knowledge base completion. In: International Conference on Machine Learning, pp. 2863–2872. PMLR (2018)
19. Li, M., Sun, Z., Zhang, S., Zhang, W.: Enhancing knowledge graph embedding with relational constraints. Neurocomputing **429**, 77–88 (2021)
20. Li, R., et al.: House: knowledge graph embedding with householder parameterization. arXiv preprint arXiv:2202.07919 (2022)
21. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
22. Liu, X., Zhao, F., Gui, X., Jin, H.: LeKAN: extracting long-tail relations via layer-enhanced knowledge-aggregation networks. In: Bhattacharya, A., et al. (eds.) Database Systems for Advanced Applications. DASFAA 2022. LNCS, vol. 13245, pp. 122–136. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-00123-9_9
23. Mahdisoltani, F., Biega, J., Suchanek, F.: YAGO3: a knowledge base from multilingual wikipedias. In: 7th Biennial Conference on Innovative Data Systems Research. CIDR Conference (2014)
24. Pavlović, A., Sallinger, E.: Expressive: a spatio-functional embedding for knowledge graph completion. arXiv preprint arXiv:2206.04192 (2022)
25. Sun, Z., Deng, Z.H., Nie, J.Y., Tang, J.: RotatE: knowledge graph embedding by relational rotation in complex space. In: International Conference on Learning Representations (2018)
26. Toutanova, K., Chen, D.: Observed versus latent features for knowledge base and text inference. In: Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality, pp. 57–66 (2015)
27. Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., Bouchard, G.: Complex embeddings for simple link prediction. In: International Conference on Machine Learning, pp. 2071–2080. PMLR (2016)
28. Vashishth, S., Sanyal, S., Nitin, V., Talukdar, P.: Composition-based multi-relational graph convolutional networks. arXiv preprint arXiv:1911.03082 (2019)
29. Wang, S., Huang, X., Chen, C., Wu, L., Li, J.: Reform: error-aware few-shot knowledge graph completion. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 1979–1988 (2021)
30. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 28 (2014)
31. Werner, S., Rettinger, A., Halilaj, L., Lüttin, J.: RETRA: recurrent transformers for learning temporally contextualized knowledge graph embeddings. In: Verborgh, R., et al. (eds.) ESWC 2021. LNCS, vol. 12731, pp. 425–440. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77385-4_25

32. Willmore, T.J.: An Introduction to Differential Geometry. Courier Corporation, Honolulu (2013)
33. Xu, Y.W., Zhang, H.J., Cheng, K., Liao, X.L., Zhang, Z.X., Li, Y.B.: Knowledge graph embedding with entity attributes using hypergraph neural networks. Intell. Data Anal. **26**(4), 959–975 (2022)
34. Yang, B., Yih, S.W.t., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: Proceedings of the International Conference on Learning Representations (ICLR) 2015 (2015)
35. Zhang, C., Yao, H., Huang, C., Jiang, M., Li, Z., Chawla, N.V.: Few-shot knowledge graph completion. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 3041–3048 (2020)