

# Embedding Knowledge Graphs in Function Spaces

Louis Mozart Kamdem Teyou

louis888@mail.uni-paderborn.de

Paderborn University

Electronic, Informatic and Mathematic

Paderborn, North Rhine-Westphalia, Germany

Caglar Demir

caglar.demir@uni-paderborn.de

Paderborn University

Electronic, Informatic and Mathematic

Paderborn, North Rhine-Westphalia, Germany

Axel-Cyrille Ngonga Ngomo

axel.ngonga@upb.de

Paderborn University

Electronic, Informatic and Mathematic

Paderborn, North Rhine-Westphalia, Germany

## ABSTRACT

We introduce a novel embedding method diverging from conventional approaches by operating within function spaces of finite dimension rather than finite vector space, thus departing significantly from standard knowledge graph embedding techniques. Initially employing polynomial functions to compute embeddings, we progress to more intricate representations using neural networks with varying layer complexities. We argue that employing functions for embedding computation enhances expressiveness and allows for more degrees of freedom, enabling operations such as composition, derivatives and primitive of entities representation. Additionally, we meticulously outline the step-by-step construction of our approach and provide code for reproducibility, thereby facilitating further exploration and application in the field.

## CCS CONCEPTS

• **Computing methodologies** → **Learning latent representations.**

## KEYWORDS

Function space, Knowledge graph, Knowledge graph embedding, Polynomial function, Neural Networks

### ACM Reference Format:

Louis Mozart Kamdem Teyou, Caglar Demir, and Axel-Cyrille Ngonga Ngomo. 2024. Embedding Knowledge Graphs in Function Spaces. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)*, October 21–25, 2024, Boise, ID, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3627673.3679819>

## 1 INTRODUCTION

A knowledge graph (KG) serves as a structured representation of data, aiming to encapsulate and convey real-world knowledge [19]. Comprising triples that articulate facts about the world, KG data are pivotal, in organizing information. For example, the triple

$\langle \textit{Obama}, \textit{president\_of}, \textit{America} \rangle$  represents the fact that Barack Obama was the president of the United States. Consequently, KGs find applications in various domains such as web search [35], recommender systems [28, 40], and natural language processing [31].

Given that KGs are essentially composed of strings or characters, knowledge graph embedding (KGE) involves mapping entities and relations from a KG into a vector space [39]. This transformation facilitates computational operations, enabling the application of machine learning and deep learning techniques to extract insights from the KG. Hence, an effective KGE model should strive to maintain the properties and semantics inherent in the original KG.

Knowledge graphs are typically represented as  $K \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ , where  $\mathcal{E}$  and  $\mathcal{R}$  represent sets of entities and relations respectively. They are commonly embedded in  $d$ -dimensional vector spaces  $\mathbb{V}$  such as  $\mathbb{R}^d$  (real numbers),  $\mathbb{C}^d$  (complex numbers), or even  $\mathbb{H}^d$  (quaternions) [3, 7, 38]. While such embeddings offer a low-dimensional representation, they treat entities and relations as static vectors, which may limit their ability to capture some dynamics in the knowledge graph. For instance, in real-world applications, relationships between entities can change over time exemplifying temporal dynamics (e.g. transition from “*is\_friends\_with*” to “*was\_friends\_with*.”) [20, 27]. They can also depend on context, such as in movie recommendations, where the relevance of a movie may be context-dependent (e.g., “*is\_interested\_in\_watching*” may vary based on user preferences [5]). Static embeddings may fail to represent such dynamic behaviours accurately, leading to suboptimal performance in tasks like link prediction or knowledge graph completion [25].

To mitigate the aforementioned drawback of the existing embedding approaches, in this work, we propose the use of functions to represent entities and relations of a KG, i.e., embedding them in function space. Functional representations of entities and relations provide a compelling alternative to static embeddings. For instance, functions can be time-dependent, they can support compositionality, enabling the combination of simpler functions to represent more complex relationships and entities. Moreover, functions offer a richer and more expressive representation compared to vectors, allowing the modelling of complex interactions and dependencies within KGs. Also, functions offer greater interpretability, as they can be analyzed and understood based on their mathematical properties and behaviour. By examining the functional forms, it would be



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

CIKM '24, October 21–25, 2024, Boise, ID, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0436-9/24/10

<https://doi.org/10.1145/3627673.3679819>

possible to gain insights into the underlying structure of the knowledge representation, thereby facilitating better understanding and interpretation.

Building upon these advantages, we hypothesize that embedding KGs in a function space offers increased degrees of freedom in the embeddings. Therefore, in this paper, we leverage the advantages of **function multiplication** and composition to compute the embeddings in function space. Our methods, dubbed  $\text{FMULT}_n$ ,  $\text{FMULT}_n^i$ , and  $\text{FMULT}$ , compute embeddings by representing entities and relations as functions.  $\text{FMULT}_n$  employs polynomials to model the interactions between entities and relations, capturing non-linear relationships that static vectors might miss.  $\text{FMULT}_n^i$  utilizes trigonometric functions to incorporate periodic patterns and cyclic behaviours inherent in the KG. Finally,  $\text{FMULT}$  leverages neural networks to learn complex, high-dimensional embeddings that can adapt dynamically to various contexts.

This paradigm shift allows for a more flexible and expressive representation of entities and relations. Notably, to the best of our knowledge, our model is the first link prediction model to use functions for representing entities and relations in KGs. Our results have shown better performance compared to traditional static embedding methods, highlighting the effectiveness of functional embeddings in capturing the complex dynamics of KGs.

In summary, we make the following contributions in this work:

- We propose three different functional embedding approaches  $\text{FMULT}_n$ ,  $\text{FMULT}_n^i$ , and  $\text{FMULT}$  to embed the entities and relations of KGs using polynomials, trigonometric functions, and neural networks, respectively.
- We perform a comprehensive evaluation of our proposed methods on multiple benchmark datasets, demonstrating their superior performance in link prediction tasks compared to state-of-the-art models.
- We make our implementations and experimental results publicly available to facilitate further research and reproducibility in the field.<sup>1</sup>

The remainder of the paper is organized as follows: In the next section, we present related work in the KGE field. In Sections 3 and 4, we provide preliminaries on function space and show how we can use functions to compute the embeddings. In Sections 5 and 6, we present the results of our approaches and compare them with baseline models.

## 2 RELATED WORK

As a reminder, several categories of KGE models can be found in the literature. This includes:

**Translational models.** The foundational work in translational models was established by TransE [3]. In TransE, given a triple  $\langle h, r, t \rangle \in K$ , the optimization objective involves minimizing the score  $\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$ , where  $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$ . A higher score is assigned if  $\langle h, r, t \rangle$  holds in  $K$ , and a lower score otherwise. To resolve TransE shortcomings e.g. TransE cannot model reflexive relationship [8, 21, 41]. Some variants have been proposed with similar ideas but different projection strategies. TransH [41] embeds knowledge graphs by projecting entities and relations onto hyperplanes

specific to each relation. TransR [30] introduces separate spaces for entities and relations, connected by a shared projection matrix. Meanwhile, TransD [21] employs independent projection vectors for each entity and relation, reducing computational complexity compared to TransR. RotatE [36] embeds entities and relations into complex space and replaces the addition operation in TransE with complex multiplication. TransG [42] enhances this idea by incorporating probabilistic principles, integrating Bayesian nonparametric Gaussian mixture models and Gaussian distribution covariance. Meanwhile, TransSparse [22] introduces adaptive sparsity to transfer matrices, aiming to address heterogeneity issues within KGs.

**Rotational models.** In contrast to translational models, rotational models [5], use the power of bilinear transformation for capturing complex interactions and correlations between entities and relations in knowledge graphs. This category of model was inaugurated by RESCAL [31] which models entities and relations using matrices, representing relationships as bilinear interactions between entity vectors. DistMult [43] simplifies RESCAL by Representing entities and relations as low-dimensional vectors instead of matrices, employing the dot product to compute scores for triples. Although DistMult is very accurate in handling symmetric relations, it performs poorly for anti-symmetric relations. ComplEx [38] tackles this drawback and uses complex embeddings to model asymmetric relations, capturing both the interactions and correlations between entities and relations. Similarly to ComplEx, RotatE [36] embeds entities and relations into complex space, representing relationships as rotations from head to tail entities. A variation of RotatE is RotE [7] which focuses on learning rotation operations to capture relational patterns. QuatE [45], extends RotatE, DistMult and ComplEx by using quaternion embeddings, enabling more expressive representations of relationships in hypercomplex space. OctE Further extends QuatE by employing octonion embeddings, offering even richer and more complex representations of relationships. However, the scaling effect in the octonion and quaternion space can be a bottleneck for QuatE and OctE hence, QMult and OMult [10] solve this issue thanks to the batch normalization technique. Dual quaternion methods like DualE [5] use dual quaternions [24] to embed entities and relations, offering a representation that combines the advantages of translation and rotation.

**Hyperbolic models.** The literature also features hyperbolic embedding methods which leverage the properties of hyperbolic geometry to capture hierarchical structures in KGs. Among them, RotH [7] extends RotatE by representing entities and relations in hyperbolic space. MuRP [1], which maps entities and relations from a knowledge graph onto a hyperbolic space. MuRE [1], which is a variant of MuRP, operates in Euclidean space, offering a simpler alternative for certain types of knowledge graphs.

**Deep learning models.** This group of models uses the power of convolutional neural networks to encode both entity and relation information simultaneously. Among them we have ConvE [12] which applies 2D convolutional filters to capture local patterns in the entity-relation space, followed by max-pooling to extract global features. Similar to ConvE, ConvO and ConvQ [10] also employ convolutional neural networks, but they are built upon OMult and QMult respectively.

<sup>1</sup><https://github.com/dice-group/dice-embeddings>

Each of these models offers a unique approach to learning representations of entities and relations in knowledge graphs. For a more detailed review of KGE models, we recommend recent surveys [23, 39]. However, existing methods typically represent entities and relations as static vectors, focusing mainly on operations such as multiplication, addition, scaling, and batch normalization with these vectors.

*Notations:* In the following parts of the paper, unless stated elsewhere,

- We denote entities and relations with minuscule letters, e.g.:  $h$ . Embeddings are denoted using a bold font, e.g.: the embedding of  $h$  is denoted  $\mathbf{h}$ .
- $\mathcal{F}(X, Y)$  the set of all function defined from  $X$  to  $Y$ .
- $\Omega$  is a bounded and closed domain of  $\mathbb{R}$  and  $\|\Omega\|$  represent its length. e.g.:  $\|[a, b]\| = b - a$ .
- $\mathcal{F}(X)$  is the set of all function defined from  $X$  to  $X$ .
- $e^{ix} = \cos x + i \sin x$  with  $i^2 = -1$ .
- $\mathcal{Re}$  is the real part of a complex number i.e.  $\mathcal{Re}(a + ib) = a$ .
- For any vector  $u$  and  $v \in \mathbb{R}^n$ , the product  $uv$  represent the element wise multiplication between  $u$  and  $v$ .

### 3 FUNCTION SPACE

A function [16, 18, 26] can be defined as a mathematical relationship or correspondence between two sets of elements. In other words, a function maps each input value to a unique output value. Formally, a function  $f$  from a set  $X$  (the domain) to a set  $Y$  (the range) is denoted as  $f : X \rightarrow Y$ . It associates each element  $x$  in the domain  $X$  with one element  $y$  in the range  $Y$ , written as  $y = f(x)$ .

A function space [26] denoted  $\mathcal{F}$  is a space that consists of functions as its elements. In other words, it is a set of functions with certain properties defined on a given domain. That is,  $\mathcal{F}(X, Y) = \{f, f : X \rightarrow Y\}$ . They can vary widely depending on the specific properties and structures imposed on the functions within them. In this work we focus on function space that consists of  $p$ -integrable functions denoted  $\mathcal{L}^p$  [6].

#### 3.1 Integrable Functions

A function  $f$  is said to be  $p$ -integrable ( $p \in \mathbb{N}$ ) over a domain  $\Omega$  and we note  $f \in \mathcal{L}^p(\Omega)$  iff the function  $f$  to the power  $p$  is Lebesgue integrable [26] that is,

$$f \in \mathcal{L}^p(\Omega) \iff \int_{\Omega} \|f(x)\|^p dx < +\infty. \quad (1)$$

For any  $1 \leq p < \infty$ ,  $\mathcal{L}^p$  spaces are also Banach space. That is, they are all vector space with an equipped norm  $\|\cdot\|_p$  defined as:

$$\forall f \in \mathcal{L}^p(\Omega), \|f\|_p = \left( \int_{\Omega} \|f(x)\|^p dx \right)^{\frac{1}{p}}. \quad (2)$$

However, it's noteworthy that they do not universally qualify as Hilbert spaces, except in the special case where  $p = 2$ , as highlighted by Young [44]. In such case they are equipped with a scalar product  $\langle \cdot, \cdot \rangle_{\mathcal{L}^2(\Omega)}$  defined as:

$$\forall f, g \in \mathcal{L}^2(\Omega), \langle f, g \rangle_{\mathcal{L}^2(\Omega)} = \int_{\Omega} f(x)g(x)dx. \quad (3)$$

A straight example of integrable functions is polynomial functions.

**Table 1: Points and weights for adaptive Gaussian quadrature**

$k$	1	2	3	4	5
Points ( $x_k$ )	$-\frac{1}{3}\sqrt{5-2\sqrt{\frac{10}{7}}}$	$-\frac{1}{3}\sqrt{5+2\sqrt{\frac{10}{7}}}$	0	$\frac{1}{3}\sqrt{5-2\sqrt{\frac{10}{7}}}$	$\frac{1}{3}\sqrt{5+2\sqrt{\frac{10}{7}}}$
Weights ( $w_k$ )	$\frac{322+13\sqrt{70}}{900}$	$\frac{322-13\sqrt{70}}{900}$	$\frac{128}{225}$	$\frac{322+13\sqrt{70}}{900}$	$\frac{322-13\sqrt{70}}{900}$

#### 3.2 Polynomial Functions

A polynomial function represents a distinctive class of mathematical functions expressed as the sum of individual terms. We denote the space  $\mathbb{R}_{deg}[x]$  the space of all polynomial functions of degree at most  $deg \in \mathbb{N}$  with coefficients in  $\mathbb{R}$ . That is,

$$\forall P \in \mathcal{F}(\mathbb{R}), P \in \mathbb{R}_{deg}[x] \iff \exists (a_i)_{i=0}^{deg} \in \mathbb{R}, P(x) = \sum_{i=0}^{deg} a_i x^i.$$

Clearly, for any bounded domain  $\Omega \subseteq \mathbb{R}$ ,  $\mathbb{R}_{deg}[x] \subseteq \mathcal{L}^2(\Omega)$ ,  $\forall deg \in \mathbb{N}$ . Therefore, given  $P_1, P_2 \in \mathbb{R}_{deg}[x]$  such that  $P_1(x) = \sum_{i=0}^{deg} a_i x^i$ , and  $P_2(x) = \sum_{i=0}^{deg} b_i x^i$  their scalar product can be computed as:

$$\langle P_1(x), P_2(x) \rangle_{\mathcal{L}^2(\Omega)} = \int_{\Omega} P_1(x)P_2(x)dx \quad (4)$$

$$= \sum_{i,j=0}^{deg} \frac{a_i b_j (M^{i+j} - m^{i+j})}{1+i+j}. \quad (5)$$

Here,  $M$  and  $m$  are respectively the upper and lower bounds of  $\Omega$ . See the appendix in Section 7 to see how this is derived. We can then derive the norm of all polynomial functions as:

$$\|P_1(x)\|_{\mathcal{L}^2[0,1]}^2 = \langle P_1(x), P_1(x) \rangle_{\mathcal{L}^2[0,1]} \quad (6)$$

$$= \sum_{i=0}^{deg} \frac{a_i^2}{1+2i} + \sum_{\substack{i,j=0 \\ i \neq j}}^{deg} \frac{a_i a_j}{1+i+j}. \quad (7)$$

These properties of polynomial functions are interesting, as their scalar product and norm can be computed without approximations, which is not always possible in  $\mathcal{L}^2$ . Therefore, for other types of functions, we suggest using integral approximations.

#### 3.3 Integral Approximation

Integral approximation is a numerical technique used to approximate the value of a definite integral by partitioning the interval of integration into subintervals and approximating the area under the curve using rectangles (for rectangle methods) and trapezoid (for trapezoid methods) [4].

However, these approaches have the main disadvantage of not guaranteeing a maximal error. To address this limitation, we follow the methodology of works on numerical integration, by employing adaptive Gaussian quadrature [17] with 5 nodes for the approximation of integrals. In adaptive Gaussian quadrature, an integral over the domain  $[-1, 1]$  can be approximated as follows:

$$\int_{-1}^1 f(x)dx \approx \sum_{k=1}^5 w_k f(x_k), \quad (8)$$

**Table 2: Embedding spaces and scoring function of some state-of-the-art KGE models**

Models	Embedding Space	Scoring Function
<b>FMULT<sub>n</sub></b>	$\mathbb{R}_n[x]$	$\langle \mathbf{h}(x) \otimes \mathbf{r}(x), \mathbf{t}(x) \rangle_{\mathcal{L}^2(\Omega)}$
<b>FMULT<sub>n</sub><sup>i</sup></b>	$\mathbb{C}_n[x]$	$\langle \mathbf{h}(x) \otimes \mathbf{r}(x), \mathbf{t}(x) \rangle_{\mathcal{L}^2(\Omega)}$
<b>FMULT</b>	$\mathcal{L}^2(\Omega)$	$\langle (\mathbf{h} \circ \mathbf{r})(x), \mathbf{t}(x) \rangle_{\mathcal{L}^2(\Omega)}$
TransE	$\mathbb{R}^d$	$\ \mathbf{h} + \mathbf{r} - \mathbf{t}\ $
DistMult	$\mathbb{R}^d$	$\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle$
RotatE	$\mathbb{R}^d$	$\ \mathbf{h} \circ \mathbf{r} - \mathbf{t}\ $
CompEx	$\mathbb{C}^{d/2}$	$\mathcal{R}e\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle$
QMult	$\mathbb{H}^{d/4}$	$\mathbf{h} \otimes \mathbf{r}^p \cdot \mathbf{t}$
OMult	$\mathbb{O}^{d/8}$	$\mathbf{h} \otimes \mathbf{r} \cdot \mathbf{t}$
DualE	$\mathbb{H}^{d/4}$	$\langle \mathbf{h} \otimes \mathbf{r}, \mathbf{t} \rangle$
HolE	$\mathbb{R}^d$	$\langle \mathbf{r}, \mathbf{h} \star \mathbf{t} \rangle$

where  $w_k \in \mathbb{R}$  are weights and  $x_k \in [-1, 1]$  are the nodes. The nodes and weights for the interval  $[-1, 1]$  are provided in Table 1. Using a change of variable, we can derive the Gaussian quadrature formula for an integral over any finite bounds,  $a$  to  $b$ , as:

$$\int_a^b f(x)dx \approx \frac{b-a}{2} \sum_{k=1}^5 w_k f\left(\frac{b-a}{2}x_k + \frac{b+a}{2}\right). \quad (9)$$

## 4 METHODOLOGY

Let us consider the problem of embedding a KG into a  $d$ -dimensional vector space, ensuring compatibility with existing methods. In Table 2, we show the embedding space where each of the state-of-the-art models operates. We initiate the exploration of functional embeddings by employing polynomial functions. This choice is grounded in the fundamental theorem of real analysis, known as the Weierstrass approximation theorem, which asserts that any function, no matter how complex, can be accurately approximated by polynomial functions [32, 33]. Given their simplicity and universal approximating power, polynomial functions serve as an ideal starting point for our investigation into functional embeddings. We then extend this idea into complex spaces, leveraging the expressive power of trigonometric functions. This serves as a natural progression before delving into more complex function representations, notably employing Neural Networks which is a kind of generalization of polynomial embeddings.

### 4.1 Embedding with Polynomial Functions

We first show how to embed KGs using polynomial functions. Specifically, when embedding in polynomial space  $\mathbb{R}_n[x]$ , our approach, FMULT, seamlessly transforms into FMULT<sub>n</sub>, where  $n$  denotes the degree of the polynomial utilized for representing the embeddings.

- Let  $\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle \in \mathcal{K}$
- We compute the embedding of  $\mathbf{h}$ ,  $\mathbf{r}$  and  $\mathbf{t}$  in  $\mathbb{R}_{deg}^m[x]$  where  $m = \frac{d}{deg+1}$  and  $deg \in \mathbb{N}$  is a hyper-parameter s.t.  $0 \leq deg \leq$

$d - 1$  as follows:

$$\mathbf{h}(x) = \sum_{i=0}^{deg} a_i x^i, \mathbf{r}(x) = \sum_{i=0}^{deg} b_i x^i, \mathbf{t}(x) = \sum_{i=0}^{deg} c_i x^i$$

where  $a_{(\cdot)}$ ,  $b_{(\cdot)}$  and  $c_{(\cdot)} \in \mathbb{R}^m$ . Hence,

$$\text{FMULT}_{deg}(\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle) = \langle \mathbf{h}(x) \otimes \mathbf{r}(x), \mathbf{t}(x) \rangle_{\mathcal{L}^2(\Omega)} \quad (10)$$

$$= \int_{\Omega} \mathbf{h}(x)\mathbf{r}(x)\mathbf{t}(x)dx \quad (11)$$

$$= \int_m^M \sum_{i,j,k=0}^{deg} a_i b_j c_k x^{i+j+k} dx \quad (12)$$

$$= \sum_{i,j,k=0}^{deg} \frac{a_i b_j c_k (M^{1+i+j+k} - m^{1+i+j+k})}{1+i+j+k}. \quad (13)$$

$M$  and  $m$  represent the upper and lower bounds of  $\Omega$ .

### 4.2 Embedding with Trigonometric Functions

As shown with the ComplEx model [38], the complex space is a prominent space to compute the embeddings as the imaginary part can be used to model anti-symmetric relations. We extend this idea by using a trigonometric function to represent the embedding. Let be  $\mathbb{C}[x] = \{f_r + if_c, \text{ where } i^2 = -1 \text{ and } f_r, f_c \in \mathcal{L}^2(\Omega)\}$ . For any  $f \in \mathbb{C}[x]$ , We have

$$\|f\|^2 = \int_0^1 f(x)f^*(x)dx = \|f_r\|_{\mathcal{L}^2(\Omega)}^2 + \|f_c\|_{\mathcal{L}^2(\Omega)}^2, \quad (14)$$

where  $f^* = f_r - if_c$ .

For  $m \in 2\mathbb{N}$ , we define the space

$$\mathbb{C}_n^m[x] = \left\{ \sum_{k=0}^n \alpha_k e^{ikx}, \text{ where } \alpha_k \in \mathbb{C}^{m/2} \right\} \subseteq \mathbb{C}[x]$$

When computing the embedding in complex space, FMULT reduces to FMULT<sub>n</sub><sup>i</sup>. For simplicity, here we choose  $\Omega = [0, 1]$ .

- Let  $\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle \in K$
- We compute the embedding of  $\mathbf{h}$ ,  $\mathbf{r}$  and  $\mathbf{t}$  in  $\mathbb{C}_{deg}^m[x]$  where  $m = \frac{d}{deg+1}$  and  $deg \in \mathbb{N}$  is a hyper-parameter s.t.  $0 \leq deg \leq d - 1$ .

$$\mathbf{h}(x) = \sum_{k=0}^{deg} a_k e^{kix}, \mathbf{r}(x) = \sum_{k=0}^{deg} b_k e^{kix}, \mathbf{t}(x) = \sum_{k=0}^{deg} c_k e^{kix}$$

where,  $a_{(\cdot)}$ ,  $b_{(\cdot)}$  and  $c_{(\cdot)} \in \mathbb{C}^{m/2}$ . Hence,

$$\begin{aligned} \text{FMULT}_n^i(\mathbf{h}, \mathbf{r}, \mathbf{t}) &= \mathcal{R}e\left(\langle \mathbf{h}(x) \otimes \mathbf{r}(x), \mathbf{t}(x) \rangle_{\mathcal{L}^2(\Omega)}\right) \\ &= \mathcal{R}e\left(a_0 b_0 c_0 + \sum_{u,v,w=1}^{deg} \frac{a_u b_v c_w}{u+v+w} (1 - e^{i(u+v+w)})\right) \end{aligned}$$

Step-by-step derivation of FMULT<sub>n</sub><sup>i</sup> is provided in the Appendix. Some direct consequences of these scoring formulations are the following theorems:

THEOREM 4.1.  $\forall \langle h, r, t \rangle \in \mathcal{K}$ ,

$$\text{If } \|\Omega\| = 1, FMULT_0(\langle h, r, t \rangle) = DistMult(\langle h, r, t \rangle)$$

$$\text{If } \|\Omega\| = 1, FMULT_0^i(\langle h, r, t \rangle) = ComplEx(\langle h, r, t \rangle)$$

THEOREM 4.2.  $\forall \langle h, r, t \rangle \in \mathcal{K}$  s.t.  $\langle t, r, h \rangle \in \mathcal{K}$ ,

$$FMULT_n(\langle h, r, t \rangle) = FMULT_n(\langle t, r, h \rangle)$$

$$FMULT_n^i(\langle h, r, t \rangle) = FMULT_n(\langle t, r, h \rangle)$$

Apart from being able to generalize DistMult and ComplEx (see Theorem 4.1),  $FMULT_n$  and  $FMULT_n^i$  also have the advantage to model symmetric relations as highlighted in Theorem 4.2. However, this last advantage is also their main disadvantage as it will be assigned the same score to any permutation of  $\langle h, r, t \rangle$  even if this does not belong to  $\mathcal{K}$  for instance,  $FMULT_n(\langle h, r, t \rangle) = FMULT_n(\langle r, t, h \rangle) \forall n \in \mathbb{N}$ . To solve this problem, we consider using two approaches:

- Leveraging function composition in lieu of conventional function multiplication.
- Incorporating an activation function to amplify non-linearity within the embeddings.

### 4.3 Embedding with Neural Networks

Building upon the foundation laid by polynomial functions, we now delve into a more sophisticated approach by harnessing the power of Neural Networks for embedding. While polynomial functions offer simplicity and certain approximation capabilities, Neural Networks present a paradigm shift with their ability to capture highly complex and nonlinear relationships inherent in the data as mentioned by Dongare et. al. in [13]. Here, we explore how entities and relations in KGs can be represented as Neural Networks. Figure 1 shows a summarization of FMULT with a single layer.

**4.3.1 Entities representation.** Considering the problem of embedding into a  $d$ -dimensional vector space, for  $\mathbf{u} = \mathbf{h}, \mathbf{r}$  or  $\mathbf{t}$ , we initialize the weights and bias of the neural network as follows:

$$W_{\mathbf{u}} = \begin{pmatrix} W_{\mathbf{u}_1} \\ W_{\mathbf{u}_2} \\ \vdots \\ W_{\mathbf{u}_{d/2}} \end{pmatrix} \text{ and } b_{\mathbf{u}} = \begin{pmatrix} b_{\mathbf{u}_1} \\ b_{\mathbf{u}_2} \\ \vdots \\ b_{\mathbf{u}_{d/2}} \end{pmatrix} \quad (15)$$

i.e.  $W_{\mathbf{u}}, b_{\mathbf{u}} \in \mathbb{R}^{d/2}$ .

We represent  $\mathbf{u}$  as a Neural Network with  $n$  layers as:

$$\mathbf{u}(x) = \begin{pmatrix} \underbrace{\sigma(W_{\mathbf{u}_n} x + b_{\mathbf{u}_n}) \circ \dots \circ \sigma(W_{\mathbf{u}_1} x + b_{\mathbf{u}_1})}_{(n-1)} \\ \underbrace{\sigma(W_{\mathbf{u}_{2n}} x + b_{\mathbf{u}_{2n}}) \circ \dots \circ \sigma(W_{\mathbf{u}_{n+1}} x + b_{\mathbf{u}_{n+1}})}_{(n-1)} \\ \vdots \\ \underbrace{\sigma(W_{\mathbf{u}_{kn}} x + b_{\mathbf{u}_{kn}}) \circ \dots \circ \sigma(W_{\mathbf{u}_{(k-1)n+1}} x + b_{\mathbf{u}_{(k-1)n+1}})}_{(n-1)} \end{pmatrix} \quad (16)$$

where  $k \in \mathbb{N}$  and such that  $k \times n = \frac{d}{2}$ . That is, if  $n = d/2$  (Maximal number of layers)

$$\mathbf{u}(x) = \sigma(W_{\mathbf{u}_n} x + b_{\mathbf{u}_n}) \circ \dots \circ \sigma(W_{\mathbf{u}_1} x + b_{\mathbf{u}_1}) \quad (17)$$

**4.3.2 Scoring Function Derivation.** Given a triple  $\langle h, r, t \rangle \in \mathcal{K}$ , we define the scoring function of FMULT as follow:

$$FMULT(\langle h, r, t \rangle) = \langle \mathbf{h} \circ \mathbf{r}(x), \mathbf{t}(x) \rangle_{\mathcal{L}^2(\Omega)} \quad (18)$$

$$= \sum_{i=1}^{(k-1)n+1} \langle \mathbf{h}^{(i)} \circ \mathbf{r}^{(i)}(x), \mathbf{t}^{(i)}(x) \rangle_{\mathcal{L}^2(\Omega)} \quad (19)$$

here,  $\mathbf{h}^{(i)}(\cdot)$ ,  $\mathbf{r}^{(i)}(\cdot)$  and  $\mathbf{t}^{(i)}(\cdot)$  represent the  $i$ -th components of  $\mathbf{h}(\cdot)$ ,  $\mathbf{r}(\cdot)$  and  $\mathbf{t}(\cdot)$  respectively i.e.

$$\mathbf{h}^{(i)}(x) = \sigma(W_{\mathbf{h}_{in}} x + b_{\mathbf{h}_{in}}) \underbrace{\circ \dots \circ \sigma(W_{\mathbf{h}_{(i-1)n+1}} x + b_{\mathbf{h}_{(i-1)n+1}})}_{(n-1)} \quad (20)$$

$$\mathbf{r}^{(i)}(x) = \sigma(W_{\mathbf{r}_{in}} x + b_{\mathbf{r}_{in}}) \underbrace{\circ \dots \circ \sigma(W_{\mathbf{r}_{(i-1)n+1}} x + b_{\mathbf{r}_{(i-1)n+1}})}_{(n-1)} \quad (21)$$

$$\mathbf{t}^{(i)}(x) = \sigma(W_{\mathbf{t}_{in}} x + b_{\mathbf{t}_{in}}) \underbrace{\circ \dots \circ \sigma(W_{\mathbf{t}_{(i-1)n+1}} x + b_{\mathbf{t}_{(i-1)n+1}})}_{(n-1)} \quad (22)$$

thus,

$$\begin{aligned} FMULT(\langle h, r, t \rangle) &= \sum_{i=1}^{(k-1)n+1} \int_{\Omega} \mathbf{h}^{(i)}(\mathbf{r}^{(i)}(x)) \times \mathbf{t}^{(i)}(x) dx \\ &\approx \sum_{j=1}^5 \sum_{i=1}^{(k-1)n+1} w_j \mathbf{h}^{(i)}(\mathbf{r}^{(i)}(x_j)) \times \mathbf{t}^{(i)}(x_j) dx, \end{aligned}$$

where  $w_j$  and  $x_j$  are the weights and points in Table 1.

## 5 EXPERIMENTS

### 5.1 Datasets

In our experiments, we consider our approaches for link prediction tasks, thereby providing a compelling comparison with state-of-the-art models. Prior to performing this task, we first performed an exhaustive grid search to fine-tune FMULT's and  $FMULT_n$ 's parameters, seeking the optimal configuration. This phase was the most challenging aspect of our research, and the results presented on large datasets do not reflect the optimal parameters due to the substantial time investment required for this process.

We perform the evaluations considering benchmark datasets KINSHIP, COUNTRIES, UMLS, NELL-995-h100, NELL-995-h50 and NELL-995-h75. The KINSHIP dataset is a KG data that focuses on familial relationships [14]. It typically contains information about familial connections between individuals, such as parent-child relationships, sibling relationships, grandparent-grandchild relationships, and so on. The UMLS dataset is a KG data which focuses on bio-medicine [37]. Each entity is a medical concept, and the edges represent semantic relationships between these concepts, such as "is-a," "part-of," "treats," etc. The COUNTRIES dataset typically refers to a collection of data related to countries around the

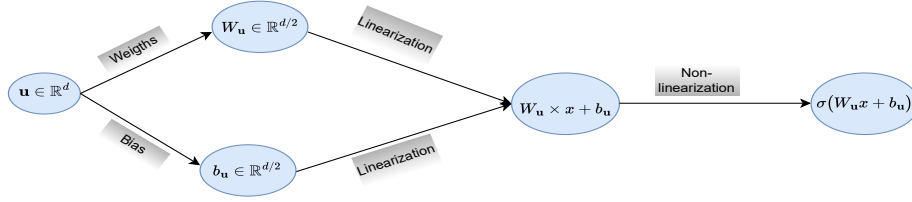


Figure 1: FMULT architecture with a single layer

world [15]. Finally, the NELL-995-h100, NELL-995-h50 and NELL-995-h75 datasets are subsets of the Never-Ending Language Learning (NELL) dataset designed for multi-hop reasoning [29]. The statistics related to each dataset can be found in Table 3.

Table 3: Overview of benchmark datasets

Dataset	$ \mathcal{E} $	$ \mathcal{R} $	$ \mathcal{G}_{Train} $	$ \mathcal{G}_{Validation} $	$ \mathcal{G}_{Test} $
WN18-RR	40,943	22	86,835	3,034	3,134
FB15k-237	14,541	237	272,115	17,535	20,466
NELL-995-h50	34,667	86	72,767	5,440	5,393
NELL-995-h75	28,085	114	59,135	4,441	4,389
NELL-995-h100	22,411	86	50,314	3,763	3,746
UMLS	135	46	5,216	652	661
KINSHIP	104	25	8,544	1,068	1,074

## 5.2 Training Strategy and Experimental Setup

Throughout the experiments, all models are trained on each KG data to minimize the binary cross-entropy loss function. The link prediction performance of each KGE model is evaluated using the filtered MRR (Mean Reciprocal rank), and Hits at n (H@n) with n = 1, 3 10. To ensure a fair comparison, each model is trained using a negative sampling scoring technique, maintaining a consistent ratio of 50%. For benchmarking purposes, we adopt embedding dimensions  $d$  chosen from the set {16, 32, 64, 100, 128}, as commonly used in related literature. Optimization is performed using the Adam optimizer with a learning rate of 0.02, and each model is trained for 500 epochs with a batch size of 1024.

For our approaches, the uniformity in model complexity is maintained by employing the  $L_2$  regularization technique [9]. The degree  $n$  of  $FMULT_n$  is chosen between {0, 1, 3, 7} and for FMULT, we found that using two layers helps the model to generalize better on some datasets (KINSHIP, NELL-995-h75) however, due to limited time complexity, we considered using only a single layer with a tanh activation function to represent all entities and relations see Figure 1.

## 6 RESULTS AND DISCUSSION

### 6.1 Embedding Dynamics

In Figure 2, we show the embeddings of the COUNTRIES dataset in two dimensions, showcasing the representation of each entity across different variants of the FMULT method and the DistMult model. As expected, DistMult displays the embeddings as 2-dimensional vectors. When applying a sigmoid activation function to FMULT,

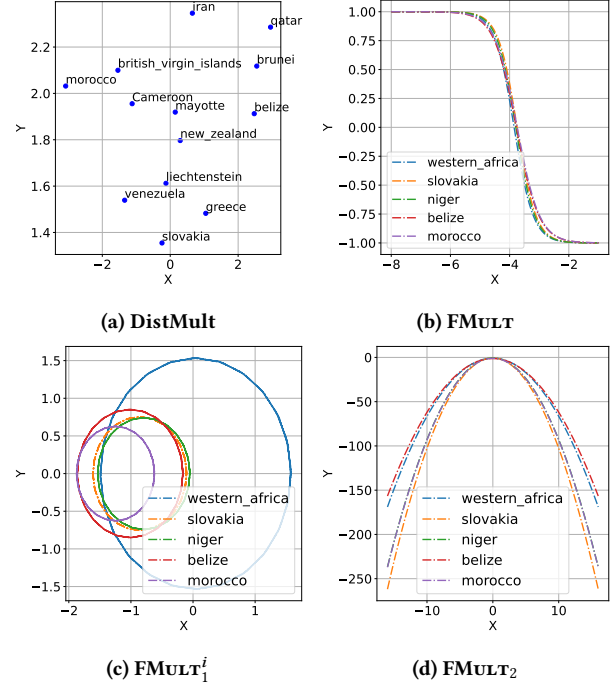


Figure 2: Entities representation for the COUNTRIES dataset in 2-dimension

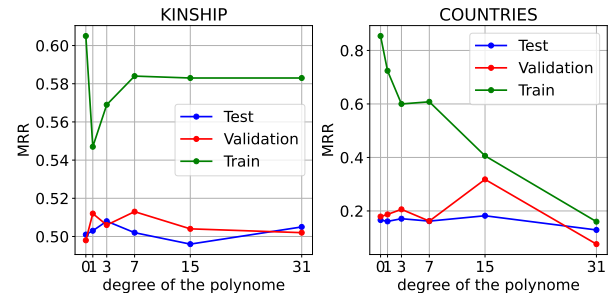


Figure 3: Impact of degrees on polynomial embeddings

it effectively uses this function to represent the computed embeddings. Conversely,  $FMULT_1^i$  uses trigonometric functions, resulting in circular embeddings with varying centres and diameters determined by the trained embeddings. Further,  $FMULT_2$  employs squared functions to represent the embeddings, resulting in shapes

**Table 4: Link prediction results on UMLS, KINSHIP and NELL-995-h100. Results are taken from the corresponding paper. Bold and underlined results indicate the best and second-best results respectively. The dash (-) denotes values missing in the paper**

Models	UMLS				KINSHIP				NELL-995-h100			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TransE	.72	.55	.87	.96	.14	.05	.13	.30	.51	.37	.59	.77
TransR	.89	.79	.98	<u>.99</u>	.67	.53	.77	.96	.42	.31	.48	.62
DistMult	.81	.74	.86	.95	.67	.54	.75	.93	.77	.66	.86	.96
ComplEx	.96	.92	.99	<b>1.0</b>	<u>.89</u>	<u>.82</u>	<u>.96</u>	<b>.99</b>	<u>.83</u>	.74	.91	<b>.98</b>
QMult [10]	<u>.96</u>	<u>.93</u>	<u>.98</u>	<b>1.0</b>	.88	.81	.94	<b>.99</b>	<u>.83</u>	.73	.91	<b>.98</b>
OMult [10]	.95	.91	.98	1.0	.87	.80	.94	<b>.99</b>	.75	.64	.82	.93
MuRE	<b>.97</b>	<b>.95</b>	<b>.99</b>	<b>1.0</b>	.82	.72	.89	<u>.98</u>	<u>.83</u>	.74	.90	<b>.98</b>
MuRP	.89	.80	.97	<u>.99</u>	.75	.62	.84	.96	<b>.85</b>	<u>.75</u>	<u>.90</u>	<b>.98</b>
HolE	.95	.91	.99	<b>1.0</b>	.89	.82	.96	<b>.99</b>	.71	.59	.79	.92
RotatE	.87	.77	.95	.98	.76	.65	.84	.96	.46	.38	.49	.63
TuckER	.87	.78	.96	.98	.58	.42	.68	.92	.33	.23	.37	.54
DualE	.94	.89	.99	<b>1.0</b>	.82	.71	.91	<b>.99</b>	.77	.67	.85	.96
ConvO [10]	.90	.82	.98	<b>1.0</b>	.86	.77	.93	<u>.98</u>	-	-	-	-
ConvQ [10]	.92	.86	.98	<b>1.0</b>	.86	.77	.93	<u>.98</u>	-	-	-	-
<b>FMULT<sub>n</sub></b>	.75	.65	.82	.93	.59	.44	.68	.91	<u>.83</u>	<b>.76</b>	.88	.96
<b>FMULT</b>	<b>.97</b>	<b>.95</b>	<b>.99</b>	<b>1.0</b>	<b>.90</b>	<b>.83</b>	<b>.98</b>	<b>.99</b>	<b>.85</b>	<b>.76</b>	<b>.93</b>	<u>.97</u>

characterized by a parabola. Notably, all embeddings share a common focus and vertex, highlighting a consistent feature across the diverse visualization techniques.

We also investigate in Figure 3 how the degree of  $FMULT_n$  impacts the embedding on the KINSHIP and COUNTRIES datasets. More precisely, we plot the MRR achieved by  $FMULT_n$  on those datasets for  $n = \{1, 3, 7, 15, 31\}$  at the training, test and validation phase. On the KINSHIP data, it is clear that a 3-degree polynomial function generalizes better than other degrees. An explanation for this is due to the nature of the KINSHIP data itself which mainly consists of symmetric relationships. Meanwhile, on the COUNTRIES dataset, a 15-degree polynomial offers a better generalization compared to others.

## 6.2 Link Prediction Results: Comparison with the state-of-the-arts

Table 4 presents the results on the UMLS, KINSHIP, and NELL-995-h100 datasets. On the UMLS dataset,  $FMULT$  excels, achieving the highest scores across all metrics, tying with MuRE and outperforming other state-of-the-art models.  $FMULT_n$  does not have high performance, however, remains competitive. Since the UMLS data contains hierarchical relationships that are often complex, we can conclude that representing entities as neural networks is highly effective in capturing complex and non-linear relationships, which is likely why  $FMULT$  achieved outstanding results across all metrics. However, polynomial functions may not be as flexible as neural networks in capturing the nuances of hierarchical and medical data. This might explain why  $FMULT_n$  performed lower than  $FMULT$  and other state-of-the-art models. On the KINSHIP data  $FMULT$  leads in MRR, H@1, and H@3, outperforming all other models, while  $FMULT_n$  lags behind most state-of-the-art models. Note that the KINSHIP dataset is known for its numerous symmetric relationships, such as parent-child and sibling relationships. The result suggests that neural networks can seamlessly handle symmetry,

**Table 5: Link prediction results for WN18-RR and FB15k-237 datasets. Results are taken from corresponding papers. Bold and underlined results indicate the best and second-best results respectively.**

Models	WN18-RR				FB15k-237			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TransE [34]	.23	.05	.37	.52	.31	.22	.35	.50
DistMult [12]	.43	.39	.44	.49	.24	.16	.26	.42
ComplEx [12]	.44	.41	.46	.51	.25	.16	.28	.43
QMult [10]	.44	.39	.45	.54	.35	.25	.38	<u>.54</u>
OMult [10]	.45	.41	.48	.54	<u>.38</u>	.25	.38	.53
MuRE[7]	.46	.42	.47	.53	.31	.23	.34	.49
MuRP [5]	<u>.48</u>	<u>.55</u>	<u>.49</u>	.44	.34	<u>.52</u>	.37	.25
RotatE [5]	<u>.48</u>	.43	<u>.49</u>	<u>.57</u>	.34	.24	.38	.53
TuckER [2]	.47	.44	.48	.53	.36	.27	.39	.54
DualE [5]	<b>.49</b>	<b>.58</b>	.51	.44	.37	<b>.56</b>	<u>.40</u>	.27
ConvO [10]	.46	.43	.47	.52	.37	.27	<u>.40</u>	<u>.54</u>
ConvQ [10]	.46	.42	.47	.53	.34	.25	.38	.53
<b>FMULT<sub>n</sub></b>	.43	.39	.44	.50	<b>.47</b>	.31	<b>.56</b>	<b>.59</b>
<b>FMULT</b>	.43	.27	<b>.52</b>	<b>.60</b>	.03	.01	.02	.08

which might explain why  $FMULT$  leads in most metrics. However, polynomial functions may not be particularly adept at capturing symmetric relationships, which could explain the middling performance of  $FMULT_n$  on this dataset leading to lower scores compared to neural networks. A similar observation can be found on the NELL-995-h100 where  $FMULT$  demonstrates outstanding performance, leading in MRR, H@1, and H@3, and achieving near-perfect scores in H@10.  $FMULT_n$  also performs well on this dataset, closely following the top models.

In Table 5, we describe the results of the WN18RR and FB15k-237 datasets. On the WN18RR dataset,  $FMULT_n$  is trained with  $n = 0$  and performed as expected similarly to DistMult but does not outperform the state-of-the-art models. However,  $FMULT$  shows a stronger

**Table 6: Link prediction results on NELL-995-h75, NELL-995-h50 and Countries. Results are taken from the corresponding paper Bold and underlined results indicate the best and second-best results respectively.**

Models	NELL-995-h75				NELL-995-h50				COUNTRIES			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TransE	.55	.41	.63	.78	.56	.43	.63	.80	.70	.41	.99	1.0
TransR	.44	.32	.50	.63	.44	.34	.49	.62	.68	.42	.93	.99
DistMult	.77	.67	.84	.95	.75	.65	.82	.95	<u>.98</u>	.97	.99	.99
ComplEx	.83	.75	.89	<u>.97</u>	.79	.70	.87	<b>.97</b>	<u>.98</u>	<u>.97</u>	.99	1.0
QMult [11]	<b>.94</b>	<b>.91</b>	<b>.97</b>	<b>.98</b>	.67	.58	.73	.82	<u>.98</u>	.96	.99	.99
OMult[11]	.66	.57	.74	.83	.67	.58	.73	.82	<u>.98</u>	<u>.97</u>	.99	1.0
MuRE	.79	.69	.86	.96	.78	.68	.85	.95	<b>.99</b>	<b>.98</b>	1.0	1.0
HolE	.71	.60	.78	.91	<u>.81</u>	<u>.72</u>	<b>.89</b>	<b>.97</b>	<u>.98</u>	.96	.99	1.0
RotatE	.46	.39	.49	.63	.46	.37	.49	.63	.77	.61	.93	.99
TuckER	.34	.23	.37	.54	.33	.22	.36	.53	.09	.03	.09	.21
DualE	.77	.68	.85	.94	.75	.63	.83	.95	.95	.91	.98	.99
<b>FMULT<sub>n</sub></b>	.77	.68	.83	.95	<b>.82</b>	<b>.74</b>	<u>.88</u>	<u>.96</u>	<u>.98</u>	<b>.98</b>	.99	.99
<b>FMULT</b>	<u>.85</u>	<u>.77</u>	<u>.93</u>	<b>.98</b>	.77	.64	.87	<b>.97</b>	.93	.86	.99	1.0

performance in H@3 and H@10, indicating it is particularly effective in retrieving the correct tail entity within the top 3 and top 10 predictions of the dataset. This implies that while it may struggle to predict the exact head entity (lower H@1), it captures useful information that positions the correct entity closer to the top in ranked lists. DualE and RotatE outperform both FMULT and FMULT<sub>n</sub> in terms of MRR and H@1, which highlights the efficiency of models that capture rotational and dual embeddings for this dataset. On the FB15k-237, FMULT<sub>n</sub> is trained using one-degree polynomials ( $n = 1$ ) and this significantly outperforms other models. Meanwhile, we observe a very poor performance of FMULT. We believe that this extreme suboptimal performance is due to inappropriate hyperparameter settings for this dataset. Given the large size of the data, experimenting with various settings is very time-consuming.

Table 6 presents the link prediction results on the NELL-995-h75, NELL-995-h50, and Countries datasets. On the NELL-995-h75, FMULT excels with second-best performance across all metrics, closely following QMult. FMULT<sub>n</sub> also performs competitively, especially in H@10. On the NELL-995-h50 dataset, FMULT and FMULT<sub>n</sub> both show strong performance, but in slightly different ways: FMULT is most effective at predicting H@10 showing best performance with HolE and ComplEx due to the flexibility of neural networks. While FMULT<sub>n</sub> excels in accurately identifying the most relevant connections like H@1 and H@3 and maintaining a high overall performance (MRR), indicating the efficacy of polynomial functions in capturing a mix of relationship complexities contained in the NELL-995-h50 dataset. This analysis highlights the complementary strengths of the two approaches, with neural networks providing broader predictive power and polynomial functions offering precision in ranking relevance on this particular dataset. On the Countries data, apart from TuckER, all other approaches perform incredibly well which is because the data itself contains only two relations. However, both FMULT and FMULT<sub>n</sub> demonstrate strong performance, with FMULT<sub>n</sub> achieving near-perfect scores across all metrics. Overall, both FMULT and FMULT<sub>n</sub> demonstrate competitive performance against state-of-the-art models, with FMULT

often achieving second-best results and FMULT<sub>n</sub> excelling in several metrics, particularly in the NELL-995-h50 and Countries datasets.

FMULT excels in datasets with complex, hierarchical (e.g. UMLS), or symmetric relationships (KINSHIP, WN18-RR), outperforming state-of-the-art models, while FMULT<sub>n</sub> remains competitive but less effective due to its reliance on polynomial functions. Conversely, FMULT<sub>n</sub> significantly outperforms in datasets suited to polynomial representations (e.g. NELL-995-h50) but may underperform if hyperparameters are not well-tuned. Both approaches demonstrate robust performance in datasets with fewer relationships (e.g. COUNTRIES), with FMULT<sub>n</sub> particularly excelling in precision.

## 7 CONCLUSION

In this work, we have developed three novel embedding methods that operate in function space: FMULT<sub>n</sub>, FMULT<sub>n</sub><sup>i</sup>, and FMULT. This marks the first time that functional representations have been employed for embedding entities and relations in knowledge graphs. Our experimental results on eight benchmark datasets demonstrate that either polynomial-based embeddings (FMULT<sub>n</sub>) or neural network-based embeddings (FMULT) can significantly improve state-of-the-art results in knowledge graph completion.

The promising outcomes suggest that functional representations provide a more flexible and expressive framework for capturing the complex dynamics within knowledge graphs. While this study has focused on FMULT<sub>n</sub> and FMULT, the potential of the trigonometric function-based approach (FMULT<sub>n</sub><sup>i</sup>) remains unexplored and will be the subject of future work.

## ACKNOWLEDGMENTS

This project received funding from the European Union’s Horizon Europe research and innovation programme through two grants (Marie Skłodowska-Curie grant No. 101073307 and grant No. 101070305). It was also supported by the "WHALE" project (LFN 1-04), funded by the Lamarr Fellow Network programme and the Ministry of Culture and Science of North Rhine-Westphalia (MKW NRW).



## REFERENCES

- [1] Ivana Balazevic, Carl Allen, and Timothy Hospedales. Multi-relational poincaré graph embeddings. *Advances in Neural Information Processing Systems*, 32, 2019.
- [2] Ivana Balazević, Carl Allen, and Timothy M Hospedales. Tucker: Tensor factorization for knowledge graph completion. *arXiv preprint arXiv:1901.09590*, 2019.
- [3] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.
- [4] Richard L Burden and J Douglas Faires. *Numerical analysis*. Brooks Cole, 1997.
- [5] Zongsheng Cao, Qianqian Xu, Zhiyong Yang, Xiaochun Cao, and Qingming Huang. Dual quaternion knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 6894–6902, 2021.
- [6] Claudio Carmeli, Ernesto De Vito, and Alessandro Toigo. Vector valued reproducing kernel hilbert spaces of integrable functions and mercer theorem. *Analysis and Applications*, 4(04):377–408, 2006.
- [7] Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. Low-dimensional hyperbolic knowledge graph embeddings. *arXiv preprint arXiv:2005.00545*, 2020.
- [8] Zhe Chen, Yuehan Wang, Bin Zhao, Jing Cheng, Xin Zhao, and Zongtao Duan. Knowledge graph completion: A review. *Ieee Access*, 8:192435–192456, 2020.
- [9] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. L2 regularization for learning kernels. *arXiv preprint arXiv:1205.2653*, 2012.
- [10] Caglar Demir, Diego Moussallem, Stefan Heindorf, and Axel-Cyrille Ngonga Ngomo. Convolutional hypercomplex embeddings for link prediction. In *Asian Conference on Machine Learning*, pages 656–671. PMLR, 2021.
- [11] Caglar Demir and Axel-Cyrille Ngonga Ngomo. Clifford embeddings—a generalized approach for embedding in normed algebras. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 567–582. Springer, 2023.
- [12] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.
- [13] AD Dongare, RR Kharde, Amit D Kachare, et al. Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology (IJEIT)*, 2(1):189–194, 2012.
- [14] Ruogu Fang, Andrew C Gallagher, Tsuhan Chen, and Alexander Loui. Kinship classification by modeling facial feature heredity. In *2013 IEEE international conference on image processing*, pages 2983–2987. IEEE, 2013.
- [15] Andrés Fernández, Michael W Klein, Alessandro Rebutti, Martin Schindler, and Martin Uribe. Capital control measures: A new dataset. Technical report, National Bureau of Economic Research, 2015.
- [16] Izrail Moiseevich Gelfand, Elena Georgievna Glagoleva, and Emmanuil Elevich Shnol. *Functions and graphs*, volume 1. Springer Science & Business Media, 1990.
- [17] Gene H Golub and John H Welsch. Calculation of gauss quadrature rules. *Mathematics of computation*, 23(106):221–230, 1969.
- [18] JW Hesselgreaves. Functions and graphs. by im gelfand, eg glagoleva, and ee shnol. pp. v, 105. 56s. 1969.(massachusetts institute of technology; great russell street, london, wci). *The Mathematical Gazette*, 54(390):418–418, 1970.
- [19] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d’Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan F. Sequeda, Steffen Staab, and Antoine Zimmermann. Knowledge graphs. *ACM Comput. Surv.*, 54(4):71:1–71:37, 2022.
- [20] Nitisha Jain. *Representation and curation of knowledge graphs with embeddings*. PhD thesis, Universität Potsdam, 2022.
- [21] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 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)*, pages 687–696, 2015.
- [22] Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. Knowledge graph completion with adaptive sparse transfer matrix. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.
- [23] Shaoyong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Networks Learn. Syst.*, 33(2):494–514, 2022.
- [24] Yan-Bin Jia. Dual quaternions. *Iowa State University: Ames, IA, USA*, 2013.
- [25] Seyed Mehran Kazemi and David Poole. Simple embedding for link prediction in knowledge graphs. *Advances in neural information processing systems*, 31, 2018.
- [26] Alois Kufner, Oldrich John, and Svatopluk Fucik. *Function spaces*, volume 3. Springer Science & Business Media, 1977.
- [27] Julien Leblay, Melisachew Wudage Chekol, and Xin Liu. Towards temporal knowledge graph embeddings with arbitrary time precision. In *proceedings of the 29th acm international conference on information & knowledge management*, pages 685–694, 2020.
- [28] Dongze Li, Hanbing Qu, and Jiaqiang Wang. A survey on knowledge graph-based recommender systems. In *2023 China Automation Congress (CAC)*, pages 2925–2930. IEEE, 2023.
- [29] Xi Victoria Lin, Richard Socher, and Caiming Xiong. Multi-hop knowledge graph reasoning with reward shaping. *arXiv preprint arXiv:1808.10568*, 2018.
- [30] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29, 2015.
- [31] Maximilian Nickel, Volker Tresp, Hans-Peter Kriegel, et al. A three-way model for collective learning on multi-relational data. In *icml*, volume 11, pages 3104482–3104584, 2011.
- [32] Kenneth A Ross. *Elementary analysis*. Springer, 2013.
- [33] Walter Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1964.
- [34] Daniel Ruffinelli, Samuel Broscheit, and Rainer Gemulla. You can teach an old dog new tricks! on training knowledge graph embeddings. 2020.
- [35] A. Singhal. Introducing the knowledge graph: Things, not strings. Google Blog, 2012. Retrieved from <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/>.
- [36] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.
- [37] Théo Trouillon, Christopher R Dance, Éric Gaussier, Johannes Welbl, Sebastian Riedel, and Guillaume Bouchard. Knowledge graph completion via complex tensor factorization. *Journal of Machine Learning Research*, 18(130):1–38, 2017.
- [38] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR, 2016.
- [39] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.*, 29(12):2724–2743, 2017.
- [40] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 950–958, 2019.
- [41] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28, 2014.
- [42] Han Xiao, Minlie Huang, and Xiaoyan Zhu. Knowledge semantic representation: a generative model for interpretable knowledge graph embedding. *arXiv preprint arXiv:1608.07685*, 2016.
- [43] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*, 2014.
- [44] Nicholas Young. *An introduction to Hilbert space*. Cambridge university press, 1988.
- [45] Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. Quaternion knowledge graph embeddings. *Advances in neural information processing systems*, 32, 2019.

## A SCORES DERIVATION

### A.1 Scalar Product and Norm Derivation of Polynomial Functions

Given  $P_1, P_2 \in \mathbb{R}_{deg}[x]$  such that  $P_1(x) = \sum_{i=0}^{deg} a_i x^i$ , and  $P_2(x) = \sum_{i=0}^{deg} b_i x^i$  we have:

$$\langle P_1(x), P_2(x) \rangle_{\mathcal{L}^2(\Omega)} = \int_{\Omega} P_1(x) P_2(x) dx \quad (23)$$

$$= \int_{\Omega} \sum_{i=0}^{deg} a_i x^i \times \sum_{j=0}^{deg} b_j x^j dx \quad (24)$$

$$= \int_{\Omega} \sum_{i,j=0}^{deg} a_i b_j x^{i+j} dx \quad (25)$$

$$= \sum_{i,j=0}^{deg} a_i b_j \int_{\Omega} x^{i+j} dx \quad (26)$$

$$= \sum_{i,j=0}^{deg} \frac{a_i b_j (\sup \Omega^{i+j} - \inf \Omega^{i+j})}{1 + i + j}. \quad (27)$$

Here,  $\sup \Omega$  and  $\inf \Omega$  are respectively the upper and lower bounds of  $\Omega$ .

### A.2 Full Scoring Function Derivation of $\text{FMULT}_n^i(h, r, t)$

- Let  $\langle h, r, t \rangle \in K$
- We compute the embedding of  $h, r$  and  $t$  in  $\mathbb{C}_{deg}^m[x]$  where  $m = \frac{d}{deg+1}$  and  $deg \in \mathbb{N}$  is a hyper-parameter s.t.  $0 \leq deg \leq d - 1$ .

$$\mathbf{h}(x) = \sum_{k=0}^{deg} a_k e^{kix}, \mathbf{r}(x) = \sum_{k=0}^{deg} b_k e^{kix}, \mathbf{t}(x) = \sum_{k=0}^{deg} c_k e^{kix}$$

where,  $a_{(\cdot)}, b_{(\cdot)}$  and  $c_{(\cdot)} \in \mathbb{C}^{m/2}$ . Hence,

$$\text{FMULT}_n^i(h, r, t) = A$$

$$\begin{aligned} A &= \text{Re} \left( \langle \mathbf{h}(x) \otimes \mathbf{r}(x), \mathbf{t}(x) \rangle_{\mathcal{L}_2(\Omega)} \right) \\ &= \text{Re} \left( \int_{\Omega} \sum_{k=0}^{deg} a_k e^{kix} \times \sum_{k=0}^{deg} b_k e^{kix} \times \sum_{k=0}^{deg} b_k e^{kix} dx \right) \\ &= \text{Re} \left( \int_{\Omega} \sum_{u,v,w=0}^{deg} a_u b_v c_w e^{i(u+v+w)x} dx \right) \\ &= \text{Re} \left( \int_{\Omega} a_0 b_0 c_0 dx + \sum_{u,v,w=1}^{deg} \int_{\Omega} a_u b_v c_w e^{i(u+v+w)x} dx \right) \\ &= \text{Re}(a_0 b_0 c_0 \|\Omega\|) + \text{Re} \left( \sum_{u,v,w=1}^{deg} \frac{a_u b_v c_w}{u+v+w} (e^{i(u+v+w) \times \inf \Omega} - e^{i(u+v+w) \times \sup \Omega}) \right) \end{aligned}$$

## B PROOF OF THEOREMS

- If  $\|\Omega\| = 1$ ,  $\text{FMULT}_0(\langle h, r, t \rangle) = \text{DistMult}(\langle h, r, t \rangle)$

PROOF. We assume  $\|\Omega\| = 1$ . If  $deg = 0$ , the polynomials reduce to  $d$ -dimensional constant vectors (as  $m = \frac{d}{0+1} = d$ ) i.e.,

$$\mathbf{h}(x) = a_0, \mathbf{r}(x) = b_0, \mathbf{t}(x) = c_0.$$

Therefore,

$$\text{FMULT}_0(\langle h, r, t \rangle) = \int_{\Omega} a_0 b_0 c_0 dx \quad (28)$$

$$= a_0 b_0 c_0 \|\Omega\| \quad (29)$$

$$= a_0 b_0 c_0 \quad (30)$$

$$= \text{DistMult}(\langle h, r, t \rangle) \quad (31)$$

□

- If  $\|\Omega\| = 1$ ,  $\text{FMULT}_0^i(\langle h, r, t \rangle) \equiv \text{ComplEx}(\langle h, r, t \rangle)$

PROOF.

From Subsection A.2, we can directly derive that if  $deg = 0$ , the second term of the summation vanish. We therefore remain with  $\text{FMULT}_0^i(h, r, t) = \text{Re}(a_0 b_0 c_0 \|\Omega\|)$  and since by hypothesis,  $\|\Omega\| = 1$  we get  $\text{FMULT}_0^i(h, r, t) = \text{Re}(a_0 b_0 c_0)$  with  $a_0, b_0$  and  $c_0 \in \mathbb{C}^{d/2}$  which is the definition of  $\text{ComplEx}$ . □

- $\text{FMULT}_n(\langle h, r, t \rangle) = \text{FMULT}_n(\langle t, r, h \rangle)$

PROOF. Using the definition of  $\text{FMULT}_n$ :

$$\text{FMULT}_n(\langle h, r, t \rangle) = \int_{\Omega} \mathbf{h}(x) \mathbf{r}(x) \mathbf{t}(x) dx \quad (32)$$

By swapping  $h$  and  $t$ :

$$\text{FMULT}_n(\langle t, r, h \rangle) = \int_{\Omega} \mathbf{t}(x) \mathbf{r}(x) \mathbf{h}(x) dx \quad (33)$$

$$= \int_{\Omega} \mathbf{h}(x) \mathbf{r}(x) \mathbf{t}(x) dx \quad (34)$$

$$= \text{FMULT}_n(\langle h, r, t \rangle) \quad (35)$$

□

- $\text{FMULT}_n^i(\langle h, r, t \rangle) = \text{FMULT}_n^i(\langle t, r, h \rangle)$

PROOF. Using the definition of  $\text{FMULT}_n^i$ :

$$\text{FMULT}_n^i(\langle h, r, t \rangle) = \text{Re} \left( \int_{\Omega} \mathbf{h}(x) \mathbf{r}(x) \mathbf{t}(x) dx \right)$$

By swapping  $h$  and  $t$ :

$$\text{FMULT}_n^i(\langle t, r, h \rangle) = \text{Re} \left( \int_{\Omega} \mathbf{t}(x) \mathbf{r}(x) \mathbf{h}(x) dx \right) \quad (36)$$

$$= \text{Re} \left( \int_{\Omega} \mathbf{h}(x) \mathbf{r}(x) \mathbf{t}(x) dx \right) \quad (37)$$

$$= \text{FMULT}_n^i(\langle h, r, t \rangle). \quad (38)$$

□