Knowledge GeoGebra: Leveraging Geometry of Relation Embeddings in Knowledge Graph Completion

Kossi Amouzouvi[◁,▷], **Bowen Song**[*], **Sahar Vahdati**[◁], **Jens Lehmann**[◁,◇]
[◁]ScaDS.AI Center, TU Dresden, Germany, {kossi.amouzouvi, sahar.vahdati}@tu-dresden.de
[▷]Department of Mathematics, KNUST, Kumasi, Ghana
[*]School of Computer Science China University of Geosciences (Wuhan), songbowen@cug.edu.cn
[◇]Amazon (work done outside of Amazon), jlehmnn@amazon.com

Abstract

Knowledge graph embedding (KGE) models provide a low-dimensional representation of knowledge graphs in continuous vector spaces. This representation learning enables different downstream AI tasks such as link prediction for graph completion. However, most embedding models are only designed considering the algebra and geometry of the entity embedding space, the algebra of the relation embedding space, and the interaction between relation and entity embeddings. Neglecting the geometry of relation embedding limits the optimization of entity and relation distribution leading to suboptimal performance of knowledge graph completion. To address this issue, we propose a new perspective in the design of KGEs by looking into the geometry of relation embedding space. The proposed method and its variants are developed on top of an existing framework, RotatE, from which we leverage the geometry of the relation embeddings by mutating the unit circle to an ellipse, and further generalize it with the concept of a butterfly curve, consecutively. Besides the theoretical abilities of the model in preserving topological and relational patterns, the experiments on the WN18RR, FB15K-237 and YouTube benchmarks showed that this new family of KGEs can challenge or outperform state-of-the-art models.

Keywords: Knowledge graph, Knowledge graph embeddings, Knowledge graph representation learning, Knowledge geogebra, Link prediction, Many-to-many problem.

1. Introduction

Knowledge Graphs (KGs) are one of the leading knowledge representation methods where every fact about entities and the relations between them is represented in the form $(head, relation, tail)$ of triples (Weikum et al., 2021). Following this structure, very many KGs have been published in different domains, such as WordNet (Miller, 1995), Freebase (Bollacker et al., 2008), WikiData (Vrandečić and Krötzsch, 2014), DBpedia (Lehmann et al., 2015), and PrimeKG (Chandak et al., 2023). KGs play an important role in AI, specifically NLP downstream tasks, including question answering, language modelling, entity linking (Bollacker et al., 2008). Despite the large quantities of triples, knowledge graphs remain incomplete (i.e., they do not contain all facts about the world or the particular domain). This is due to the evolving nature of knowledge and limitations in capturing existing ones. One of the most prominent approaches to deal with KG incompleteness is *knowledge graph embeddings* (KGEs). These models use a scoring function that defines the degree to which a relation between two entities is plausible. In order to do so, the KG is transferred into a low dimensional continuous vector space by assigning latent feature vectors to each entity and relation of the underlying KG. The choice of the vector space is important from the algebraic and geometric point of views as different spaces provide diverse interactions between relation and entity embeddings through the algebraic products they are mathematically equipped with. Therefore, different representation spaces such as real, complex ($\mathbb{C}$), and hypercomplex ($\mathbb{H}$) are used in the design of state-of-the art KGE models.

However, the algebraic and geometric aspects of the underlying spaces are not used to their full potential when dealing with relation embeddings. Firstly, in most of the KGE models the relation embeddings are parameterized by real vectors regardless of the main vector space of the model. For instance, since the complex space $\mathbb{C}^n$ is isomorphic to $\mathbb{R}^{2n}$, KGEs in this space use implicitly the Cartesian product of 2-dimensional Euclidean spaces. Despite this, the geometry of relation embeddings has never aroused any interest in the existing works. Secondly, while both the algebra and geometry aspects of the space are considered in entity embeddings (e.g., TransE, RotatE (focused on the algebra), ATTH (focused on the geometry), QuatE (focused on both)), only the algebra (e.g., Lie Group Embeddings) is utilized for relation embeddings. These two algebraic and geometrical limitations of the existing KGE models lead to suboptimal performance of the downstream tasks depending on knowledge graph completion. We propose a paradigm shift in design of KGEs and introduce a new perspective of considering geometry of relation embedding space. For this, we introduce the concept of *Knowledge GeoGebra*,

LREC-COLING 2024, pages 9832–9842

and investigate the theoretical aspects of TransE (a pioneer KGE model) and RotatE (a state-of-the-art KGE model) from the aspects of probing the geometry and algebra. We further extend the RotatE framework by considering Ellipse for relation embeddings, and generalize it based on the concept of butterfly curve. A family of KGE models with 4 versions is proposed as EllipsE and Butterfly KGE models. The evaluation show the impact of considering the geometry of relation embedding beside algebra. The main contributions of this work can be listed:
(1) introducing a new paradigm of *Knowledge GeoGebra* in KGE models
(2) proposing a family of KGE models with extending geometry of relation embeddings,
(3) analysing theoretical aspects of the EllipsE and Butterfly,
(4) empirical performance analysis of the proposed models in per,
(5) empirical analysis of the interaction between the baseline and our model characteristics and dataset properties.

2. Related Work

In general, KGE models are divided into three categories according to relation-entity interactions, mainly: tensor decomposition, deep learning, and geometric models. Geometric models can be further classified upon the geometry of the entity embedding space, mainly: Euclidean, and non-Euclidean geometry based models, or upon the algebra of the embedding spaces, mainly: real-, complex-, and hypercomplex-valued models. In this section, we briefly present geometric models.
Geometric models see relations as geometric transformations between head and the tail entities. TransE (Bordes et al., 2013), a pioneering work of art, encodes relations by straight line translation transformations in Euclidean space. Despite its success in Knowledge graph completion, it is limited in preserving some sub-graph structures. These limitations give rise to new translational models such as TransH (Wang et al., 2014), TransA (Xiao et al., 2015), TransR (Lin et al., 2015), and RotatE (Sun et al., 2019a). In contrary to TransE, RotatE uses complex product to rotate head entities. It therefore populates the Euclidean space by concentric circles and achieves translation of head entities on circles, which enables RotatE to preserve symmetry and antisymmetric relational patterns. Subsequently, QuatE (Zhang et al., 2019) extends the algebraic structure of RotatE to hypercomplex structure; relation-entity interaction is represented by quaternion product. DualE (Zhang et al., 2019) uses the dual-quaternion product to simultaneously achieve rotation and translation of head entities by relations. DualE and QuatE subsume RotatE, and moreover, they introduced non-commutative relation representation learning. Recently, it has been shown that hyperbolic geometry has the potential to facilitate hierarchical relational pattern learning (Nickel and Kiela, 2017; Sala et al., 2018; Le et al., 2019). These models define and project relation and entity embeddings onto relation-specific manifolds in the Poincaré ball. MuRP (Balazevic et al., 2019) models relations by Möbius addition and Möbius matrix-vector multiplication of the tail and head entities, respectively. They assessed the importance of using hyperbolic space by comparing the model against its reduced version, MuRE, which uses real addition and real matrix multiplication in Euclidean. Chami et al. (2020) introduced, in their model ATTH, hyperbolic rotation and reflection to conjointly learn relational and hierarchical patterns. Cao et al. (2022) extend ATTH to GIE by considering its Euclidean and spherical space versions in addition to its original version in the hyperbolic space. By interactively learning the spatial structures in the three spaces, GIE is able to simultaneously learn multiple types of geometric structures. However, they only learn relation-specific geometry in the entity embedding vector space.

3. Knowledge GeoGebra

Let us consider a knowledge graph as a multi-relational directed graph $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}_+)$ where $\mathcal{E}, \mathcal{R}$, and $\mathcal{T}_+$ are the set of nodes (entities), edges (relations between entities), and triples formed as (head entity, relation, tail entity) respectively. We use h to refer to the head entity and t to the tail entity. Borrowed from GeoGebra (a portmanteau of geometry and algebra), we introduce *Knowledge GeoGebra*, a concept that consists of probing the geometry and algebra induced by relation embeddings of KGE models. By geometry, we mean geometrical shapes (manifold) defined by relational position vectors when varying relation embeddings along a number of dimensions; and by algebra, we mean the algebraic structure exhibited by the set of relational position vectors when equipped with a composition product. This is called *relational set* throughout this work. If relation embeddings are in $\mathbb{R}^{k\times n}$ or a vector space isomorphic to the Euclidean space $\mathbb{R}^{k\times n}$, then in *Knowledge GeoGebra (KGG)* we consider the manifold of the relational set, defined in $\mathbb{X} = \mathbb{R}^k$. However, for the sake of simplicity, we only focus on studying one- or two-dimensional manifolds; that is $k \leq 2$. In the following, we first look into the Knowledge GeoGebra of KGEs using the TransE, and RotatE models as examples.

3.1. Knowledge GeoGebra of Line

In most of the translation-based models (such as TransE, TransH, TransR), relations are embedded

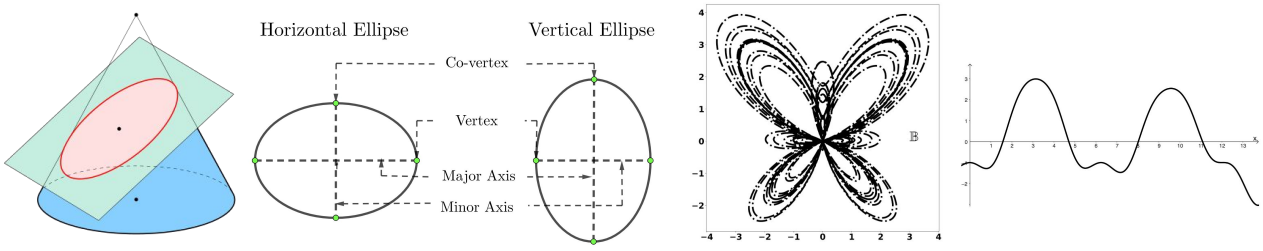

Figure 1: Ellipse defined by a conic section and type of elliptical curves on a plane (2 figures on the left side) and Butterfly curve with graph of the function $\rho(-\theta)\rho(\theta) - 1$ (2 figures on the right side).

with $\mathbf{r} \in \mathbb{R}^n$. Thus, $\mathbb{X} = \mathbb{R}$ and the relational sets $\mathcal{L}_k = \{\mathbf{r}_k \in \mathbb{R}\}$ where $\mathbf{r}_k$ represents the projection of $\mathbf{r}$ onto the k'th real space. Therefore, the manifold of the relational position vectors in the TransE model is a line. TransE uses relation-specific translation to model the relation-entity interactions in the KG. Two consecutive triples (h, r_1, m) and (m, r_2, t) induce composition of the involved relation embeddings: $\mathbf{r}_1, \mathbf{r}_2$. Since in TransE, the composition of $\mathbf{r}_1$ and $\mathbf{r}_2$ is $\mathbf{r}_2 + \mathbf{r}_1$, the relational set $\mathcal{L}_k$ is therefore equipped with the addition operation and is closed under this operation. The existence of an inverse of a translation, and the commutativity property of translations, induces an Abelian group structure on the relational sets. The implicit geometry and algebra of the relational sets empower TransE with a remarkable performance on the KG completion task, regardless of the scale of the underlying KG. Nevertheless, TransE fails to learn relational and structural patterns such as symmetry, and loop.

3.2. Knowledge GeoGebra of Circle

To mitigate the limitations of TransE in modelling more complex relational and topological patterns, the RotatE model represents relations by unit complex numbers. By doing so, RotatE changes the underlying manifold from a line to a circle, consequently the algebraic product becomes a complex product instead of addition.
RotatE embeds a relation r with a low dimension unit vector $\mathbf{r} = \mathbf{e}^{i\theta^r}$ where $\theta^r \in (-\pi, \pi]^n$. We denote by $\mathbf{r}_k = e^{i\theta^r_k} = \cos\theta^r_k + i\sin\theta^r_k$ the projection of $\mathbf{r}$ onto the k'th complex space. It follows that θ^r_k is the rotation angle in the corresponding complex plane. Relational sets are defined and denoted by,

$$\mathcal{C}_k := \{\cos\theta^r_k + i\sin\theta^r_k \mid r \in \mathcal{G}\}. \tag{1}$$

Composite transformation r of r_2 and r_1 is given by the complex number $\mathbf{r} = \mathbf{r}_2 \odot \mathbf{r}_1$ where $\odot$ is the element-wise complex product. We obtain $\mathbf{r}_k = [\mathbf{r}_2 \odot \mathbf{r}_1]_k = \cos(\theta^{r_2}_k + \theta^{r_1}_k) + i\sin(\theta^{r_2}_k + \theta^{r_1}_k)$. In other words, the sets $\mathcal{C}_k$ are endowed with complex products. Elements in $\mathcal{C}_k$ and their mutual composition lay on the unit circle centred at the origin, since $\mathcal{C}_k$ is closed under complex product. In this way, the manifold is a circle which remains invariant under arbitrary variation of relation embeddings. RotatE preserves triple plausibility by evaluating the distance between the transformed head

$$\mathbf{rh} = \mathbf{e}^{i\theta^r} \odot \mathbf{h} \tag{2}$$

and the embedded tail $\mathbf{t}$. Equation 2 could be expressed in terms of a matrix product. In fact, the k'th projection of Equation 2 is

$$\begin{pmatrix} \mathrm{Re}(\mathbf{rh})_k \\ \mathrm{Im}(\mathbf{rh})_k \end{pmatrix} = \begin{pmatrix} \cos\theta^r_k & -\sin\theta^r_k \\ \sin\theta^r_k & \cos\theta^r_k \end{pmatrix} \begin{pmatrix} \mathrm{Re}(\mathbf{h})_k \\ \mathrm{Im}(\mathbf{h})_k \end{pmatrix} \tag{3}$$

where Re and Im are the element-wise real and imaginary part operators. Complex embedding of entities allows RotatE to explore both magnitude and direction during representation learning compare to TransE which explores only the magnitude. Despite this, RotatE inherits the 1-to-1 property of the rotation transformation, which prevents the model to properly learn the underlying structure of datasets rich in topological patterns.

4. Methodolog

In this section, we extend the study of lines and circles done in the state-of-the-art models, to ellipse and butterfly curves in terms of Knowledge GeoGebra concept. The proposed models [1] consider the geometry of relation embeddings.

4.1. Knowledge GeoGebra of Ellipse

Geometrically, an ellipse (Fig. 1) is a generalization of a circle. It is a collection of complex point vectors defined as follows:

$$\mathbb{E}_{ab} = \{a\cos\theta + ib\sin\theta \mid \theta \in (-\pi, \pi]\} \tag{4}$$

where $a > 0$ and $b > 0$ are the vertex and co-vertex respectively. The real number $e = a^2 - b^2$ is called linear eccentricity. A negative or positive value of e means the ellipse major axis is the y-axis or the x-axis respectively; otherwise the ellipse is a circle of radius $a = b$. We denote by $e_{ab\theta}$ elements of the ellipse $\mathbb{E}_{ab}$. For a fixed value of a and b, the complex product $e_{xy\sigma}$ of $e_{ab\theta_1}$ and $e_{ab\theta_2}$ is given by,

$$e_{xy\sigma} = x\cos\sigma + yi\sin\sigma \tag{5}$$

where $\sigma = \theta_1 + \theta_2$, $x\cos\sigma = \frac{1}{2}\Big((a^2 + b^2)\cos(\theta_1 + \theta_2) + (a^2 - b^2)\cos(\theta_1 - \theta_2)\Big), y = ab$. This implies

[1] https://github.com/amouzard/KGeoGebra

that the sets $\mathbb{E}_{ab}$ is not invariant under complex product as $\mathcal{C}$ does. On the one hand side, one could attempt to enforce the invariance by devising an algebraic product that would preserve the elliptical shape. On the other hand, one could consider $\mathbb{E}_{**} = \cup_{a,b}\mathbb{E}_{ab}$ which is close under complex product. We adopt the latter approach which seems simpler to us and design two KGE models, namely EllipsE and EllipsEs.

4.1.1. EllipsE Model

We propose EllipsE as our first *Knowledge GeoGebra* model. EllipsE embeds entities in a d-dimensional complex space just like RotatE, but relations are embedded in $\mathbb{E}^d_{ab}$. That is, EllipsE defines two d-dimensional learnable positive real vectors $\mathbf{a}$ and $\mathbf{b}$ whose k'th components represent the vertex and co-vertex of a planar ellipse. Following Equation 3, we define the transformed head as

$$\mathbf{rh}_k := \begin{pmatrix} \mathbf{a}_k \cos\theta^r_k & -\mathbf{b}_k \sin\theta^r_k \\ \mathbf{b}_k \sin\theta^r_k & \mathbf{a}_k \cos\theta^r_k \end{pmatrix} \begin{pmatrix} \mathrm{Re}(\mathbf{h})_k \\ \mathrm{Im}(\mathbf{h})_k \end{pmatrix}. \quad (6)$$

4.1.2. EllipsEs Model

In contrary to EllipsE, EllipsEs embeds relations in $\mathbb{E}^d_{**}$ by defining three relation embeddings; the vertex and co-vertex representations $\mathbf{a}^r, \mathbf{b}^r \in \mathbb{R}^d_+$, and $e^{i\theta^r} \in \mathbb{C}^d$ to represent high dimension rotation. It is noteworthy that in EllipsEs, the vertex and co-vertex embeddings are relation-specific, whereas they are hyperparameters in EllipsE. EllipsEs transforms head entities following Equation 6. EllipsE and EllipsEs used RotatE scoring function to evaluate triple plausibility, and Adam optimizer to optimize the lost function based on adversarial negative sampling.

4.2. Knowledge GeoGebra of Butterfly Curve

The butterfly curve (Fig. 1) is among the planar curves which could not be described algebraically. One of the many equations to define points lying on the butterfly curve is the polar equation defined as follows,

$$\rho(\theta) = e^{\sin\theta} - 2\cos(4\theta) + \sin^5\left(\frac{2\theta - \pi}{24}\right) \quad (7)$$

where ρ is the distance between the origin and a point on the butterfly curve, and θ is the angle between the vector representing the point and the x-axis. Let

$$\mathbb{B} = \{e_{\theta_1} = \rho_\theta(\cos\theta + i\sin\theta) \mid \theta \in (0, 24\pi]\} \quad (8)$$

be the butterfly curve. $\mathbb{B}$ is not invariant under complex product since

$$\begin{aligned} e_{\theta_1} \cdot e_{\theta_2} &= \rho(\theta_2)\rho(\theta_1)e^{i(\theta_2+\theta_1)} \\ &\neq \rho(\theta_2+\theta_1)e^{i(\theta_2+\theta_1)} = e_{\theta_2+\theta_1} \end{aligned} \quad (9)$$

holds for some $(\theta_1, \theta_2) \in (0, 24\pi]^2$.

The butterfly model, as it can be deduced from its name, is a model which uses $\mathbb{B}^d$ for relation embeddings. Thus, the d-dimensional complex head entity embeddings are transformed as follows

$$\mathbf{rh} = \left(\rho(\theta^r) \otimes e^{i\theta^r}\right) \odot \mathbf{h} \quad (10)$$

where $\otimes$ is element-wise scalar multiplication. (For simplicity, we may omit an explicit use of the scalar product.) We deemed this model ButtErfly. ButtErfly is also a distance based model which evaluates $-\|\mathbf{rh} - \mathbf{t}\|$ to distinguish the true and false triples. Furthermore, we develop ButtErflies a variant of ButtErfly to address the problem of non-invariance of $\mathbb{B}$. In addition to θ^r, ButtErflies has relation-specific bias $\mathbf{b}^r \in \mathbb{R}^d_+$. These bias adjust the modulus ρ in such away that

$$(\mathbf{b}^{r_2} \otimes \rho(\theta^{r_2})) \otimes (\mathbf{b}^{r_1} \otimes \rho(\theta^{r_1})) = \mathbf{b}^r \otimes \rho(\theta^{r_2} \oplus \theta^{r_1}). \quad (11)$$

In Equation 11, $\oplus$ is element-wise complex addition, and r is the composition of r_2 and r_1. With this formalism, $\mathbb{B}$ becomes invariant under complex product, and the transformed head becomes

$$\mathbf{rh} = (\mathbf{b}^r \otimes \rho(\theta^r))\, e^{i\theta^r} \odot \mathbf{h}. \quad (12)$$

5. Theoretical Analysis

In this section, we analysed properties of our models which enable them to preserve structural and relational patterns.

Theoretically, EllipsE and EllipsEs are reduced to RotatE when the vertices and co-vertices are set to 1. Thus they are able to preserve (anti-)symmetry, commutativity and inversion of relations. The dependence of the modulus ρ in terms of the rotation angle, prevent our Butterfly related models to subsume RotatE. However, the transcendental function used to express the modulus enriched the analysis of these two models.

Instead of simply ending our analysis with the subsumption argument, which will not profit from the change of manifolds, we conduct additional study on the effective use of elliptical and butterfly manifolds. The commutativity of the complex product and the use of trigonometric functions to define relation embeddings entail the ability of our four models to infer commutative, symmetric, and antisymmetric relational patterns. In the following subsections, we only focus on proving the ability of our models to preserve inversion and composition relational patterns, and many-to-many structural and topological patterns.

5.1. Relational Patterns

Inversion: A relation r' is the inverse of the relation r if (t, r', h) holds whenever (h, r, t) holds. Let $\mathbf{M}^r$ be the block diagonal matrix made of the blocks

$M_k^r = \begin{pmatrix} \mathbf{a}_k^r \cos\theta_k^r & -\mathbf{b}_k^r \sin\theta_k^r \\ \mathbf{b}_k^r \sin\theta_k^r & \mathbf{a}_k^r \cos\theta_k^r \end{pmatrix}$. We denote by Δ^r its block-wise determinant, i.e. $\Delta_k^r = \det M_k^r$. EllipsEs infers inverse relations since $\det M_k^r$ is non-zero for all k. The inverse embeddings of $\mathbf{a}^r, \mathbf{b}^r$, and θ^r are $\mathbf{a}^r \oslash \Delta^r, \mathbf{b}^r \oslash \Delta^r$, and $-\theta^r$ where $\oslash$ is the element-wise division. The expression of the inverse embeddings, shows relations and their inverses do lay on the same elliptical manifold if only $\Delta^r = 1$. Therefore, EllipsE enforces $a^2 + b^2 \geq 1$ to enable inversion inference on $\mathbb{E}_{ab}$.

On the one hand, the complex inverse representation of $\rho(\theta^r)e^{i\theta^r}$ is $e^{-i\theta^r} \oslash \rho(\theta^r)$ which does not necessarily lay on the butterfly curves for all θ_k^r. That is, we do not always have $\rho(-\theta^r) = 1 \oslash \rho(\theta^r)$. On the other hand, Fig. 1 shows that the equation $\rho(-\theta)\rho(\theta) - 1 = 0$ admits many non-trivial roots. In summary, the variants of butterfly models proposed in this work infer inversion relational patterns.

Composition: A relation r is the compose relation of r_2 and r_1 if the triple (h, r, t) exists whenever (h, r_1, m) and (m, r_2, t) exist. The closure properties of $\mathbb{E}_{ab}, \mathbb{E}_{**}$, and $\mathbb{B}$ under complex product carried on in Sections 4.1 and 4.2 are the roots to our discussion about relational composition learning by our models. Stating that the relation embedding spaces are closed under complex product means composition of two relation embeddings exists and belongs to the same space. Therefore, EllipsEs and ButtErflies preserve composition patterns, whereas Ellipse and ButtErfly fail to.

5.2. Topological Patterns: Many-to-Many

In the following, we discuss how the butterfly models theoretically address the many-to-many problem. Relations are correspondences between head and tail entities. A relation that can assign only one (1) or many (M) tails to a single head is called a 1-to-1 or 1-to-M relation, respectively. Conversely, the relation is called 1-to-1 or N-to-1. In case, many heads (N) can be assigned to many tails (M), the relation is called N-to-M. The class of a relation is known by solving the equation $\|\mathbf{rh} - \mathbf{t}\| = 0$. Straightforwardly, it can be seen that this equation implies $\rho(\theta^r)r^h = r^t$ and $\theta^t - \theta^h = \theta^r$ modulo 2π; where r^x and θ^x are the modulus and the angle of the complex vector x for $x = h, t, r$. This is equivalent to solving $\rho(\theta^t - \theta^h + 2k\pi)r^h = r^t$, with k an integer. Given the fact that ρ has a period of 24π, for a fixed tail (respectively head) entity, the relation r could assign at most 12 head (respectively tail) entity candidates. Thus, the butterfly models are able to preserve many-to-many structural and topological patterns.

6. Experiments

We evaluated our models on the downstream task of KG completion (KGC) through link prediction. The goal of this task is to predict the missing entity (?) in the queries $(h, r, ?)$ or $(?, r, t)$. The missing entity is replaced by all entities is the KG to obtain corrupted triples. We filter the set of corrupted triples following TransE model (Bordes et al., 2013). The models score these triples and rank them in a decreasing order. The missing entity is finally chosen from the lowest ranked corrupted triples, i.e. the triple whose score is the closest to 0. To compare our models against the baselines, we used the mean reciprocal rank (MRR) given by $\sum_{j=1}^{n_t} \frac{1}{r_j}$, where r_j is the rank of the j-th test triple and n_t - the number of triples in the test dataset; and Hits@N which is the percentage of triples whose rank is equal or smaller than N ($N = 1, 3, 10$). We conduct an extensive hyperparameters grid search of the four models on the selected datasets and the optimal hyperparameter sets were fixed based on the best MRR. The batch size β is tuned in the range of $\{256, 512, 1024\}$, the number of negative sample η in $\{128, 256\}$, the margin γ in $\{1, 6, 9, 12, 24\}$, the self-adversarial temperature α in $\{0, 0.5, 1\}$, and the learning rate λ in $\{0.00005, 0.0001, 0.0005, 0.001\}$. We selected the embedding dimensions $d \in \{500, 800\}$ on FB15k-237 and WN18RR, and we fixed $d = 300$. The maximum step size σ is set to 120000. The Adam optimizer was used as the optimization function. Table 4 contains the optimal hyperparameter set for the datasets of interest.

6.1. Datasets, Pattern Specific Datasets, and Baselines

In our experiments, we used three benchmark datasets: WN18RR (Dettmers et al., 2018), FB15k-237 (Toutanova and Chen, 2015), YouTube (Cen et al., 2019). Tables 2 summarizes the statistical information of the size of the benchmark datasets. WN18RR (Dettmers et al., 2018) is a subset of WN18 (Bordes et al., 2013), which only contains 11 relations. FB15k-237 is a subset of Freebase, which contains general knowledge facts and a few inverse relations. YouTube contains the interactions between its users, including contacts, shared friends and so on. In addition, we created pattern specific datasets accordingly to the relational patterns: symmetric, composition, and antisymmetric (Ali et al., 2021). We designed new test datasets based on the relational patterns present ín the WN18RR and FB15k-237 datasets. We first grouped all the relations into 9 categories based on the symmetric (S), antisymmetric (A), composition (C) patterns and their four mutual combinations (AC, AS, CS, ACS). Our categorization logic is straightforward as follows: A relation r is classified as an (i) X type: if it exclusively exhibits the X characteristic. (ii) XY type: if r displays both X and Y characteristics, and (iii) XYZ type: if r embodies X, Y, and Z

	FB15K-237				WN18RR				YouTube			
Model	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
ConvE	.325	.237	.356	.501	.430	.400	.440	.520	-	-	-	-
MuRE	.336	.245	.370	.521	.465	.436	.487	.554	-	-	-	-
DisMult	.241	.155	.263	.419	.430	.390	.440	.490	.040	.010	.030	.100
TransE	.294	-	-	.465	.226	-	-	.501	.180	-	.280	.470
ComplEx	.247	.158	.275	.428	.440	.410	.460	.510	**.320**	**.210**	.360	.540
RotatE	.338	.241	.375	.533	.476	.428	.492	.571	.250	.140	.300	.460
EllipsE	.344	.248	.382	.539	**.487**	**.441**	**.503**	**.574**	.288	.135	.380	.553
EllipsEs	**.346**	.248	**.386**	**.542**	.481	.436	.500	.568	.299	.149	.385	**.560**
Butterfly	.344	**.252**	.380	.531	.470	.413	.500	.569	.302	.169	.372	.536
ButtErflies	.344	.250	.381	.531	.456	.413	.478	.532	.318	.183	**.390**	.558

Table 1: Link prediction results on FB15h-237, WN18RR, and YouTube datasets in high dimension. The results of baseline models were reported from (Balazevic et al., 2019; Nayyeri et al., 2021).

Dataset	WN18RR	FB15k-237	YouTube
entities	41k	15k	2k
relations	11	237	5
Training	87k	272k	1114k
Validation	3k	18k	66k
Test	3k	20k	131k

Table 2: Statistics of the datasets.

Pattern	WN18RR	FB15k-237
Symmetric	3	3
Antisymmetric	7	205
Composition	1	147

Table 3: Statistics of relational patterns across WN18RR and Fb15k-237.

EllipsE						
Dataset	d	β	η	γ	α	λ
WN18RR	500	256	100	6	1	0.00005
FB15k-237	800	256	256	6	1	0.0001
YouTube	300	1024	256	12	1	0.0001
EllipsEs						
Dataset	d	β	η	γ	α	λ
WN18RR	800	1024	256	6	1	0.00005
FB15k-237	800	1024	256	12	0.5	0.0001
YouTube	300	1024	256	12	1	0.0001
ButtErfly						
Dataset	d	β	η	γ	α	λ
WN18RR	500	256	100	6	10	0.0001
FB15k-237	800	1024	256	9	1	0.00005
YouTube	300	1024	256	12	1	0.00005
ButtErflies						
Dataset	d	β	η	γ	α	λ
WN18RR	800	1024	256	12	0.5	0.0001
FB15k-237	800	512	256	12	1	0.00005
YouTube	300	1024	256	12	1	0.0001

Table 4: Optimal hyperparameter setting of our models.

characteristics. We use (iv) U_x for representing the union of relations with the specific property X. Thus, U_A which includes A, AC, AS, and ACS type relations, is the ordinal set of antisymmetric relations. Similarly, the set of symmetric and composition relations are U_S and U_C since both encompasses S, AS, CS, ACS types, and C, AC, CS, ACS types respectively. The statistics of the relational patterns U_A, U_S, and U_C in the benchmark datasets are shown in Table 3. We shall note that some relational categories can be empty given the characteristics of the dataset at hand. We compare our model with state-of-the-art models, including ConvE (Dettmers et al., 2018), MuRE (Balazevic et al., 2019), DisMult (Yang et al., 2015), TransE (Bordes et al., 2013), ComplEx (Trouillon et al., 2016), and RotatE (Sun et al., 2019b) on the original and the categorized datasets. More advanced KGE models (such as QuatE, ATTH, and GIE) partially or completely represent the interaction of relations with entities by rotation transformations. They either used Euclidean geometry (QuatE), non-Euclidean geometry (ATTH), or the combination of the two or more geometries (GIE) to define the rotation transformation. Consequently, they can be viewed as implicit extensions of the RotatE and TransE models, where their relational set, or at least a significant portion of it, forms a geometrical circle or line. Given this underlying similarity, we hypothesized that a more relevant comparison would be between these advanced models and their counterparts in KG GeoGebra variants.

6.2. Results and Analysis

Table 1 shows the performance comparison of our models and the baseline models on the three datasets. Overall, we majorly achieve significant improvement, or comparable results. Our models outperform all the baseline models for all the evaluation metrics on the FB15K-237 dataset, and for the H@3 and H@10 metrics on YouTube. On WN18RR, the two ellipse Knowledge GeoGebra models considerably out stand baselines. We no-

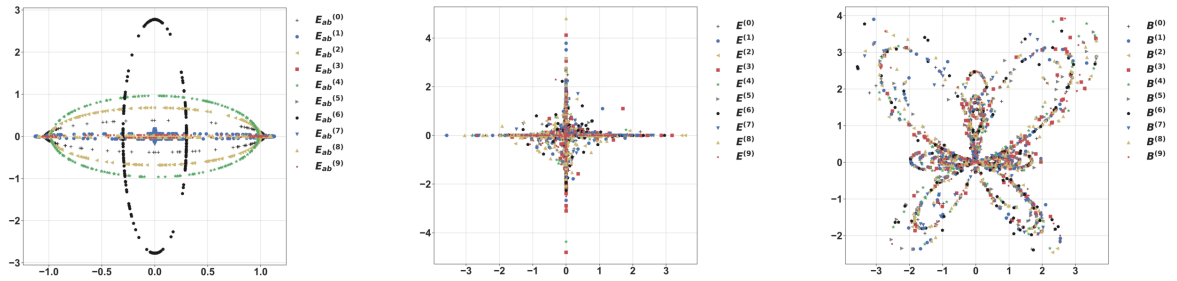

Figure 2: Distribution of relations in the first ten relational sets.

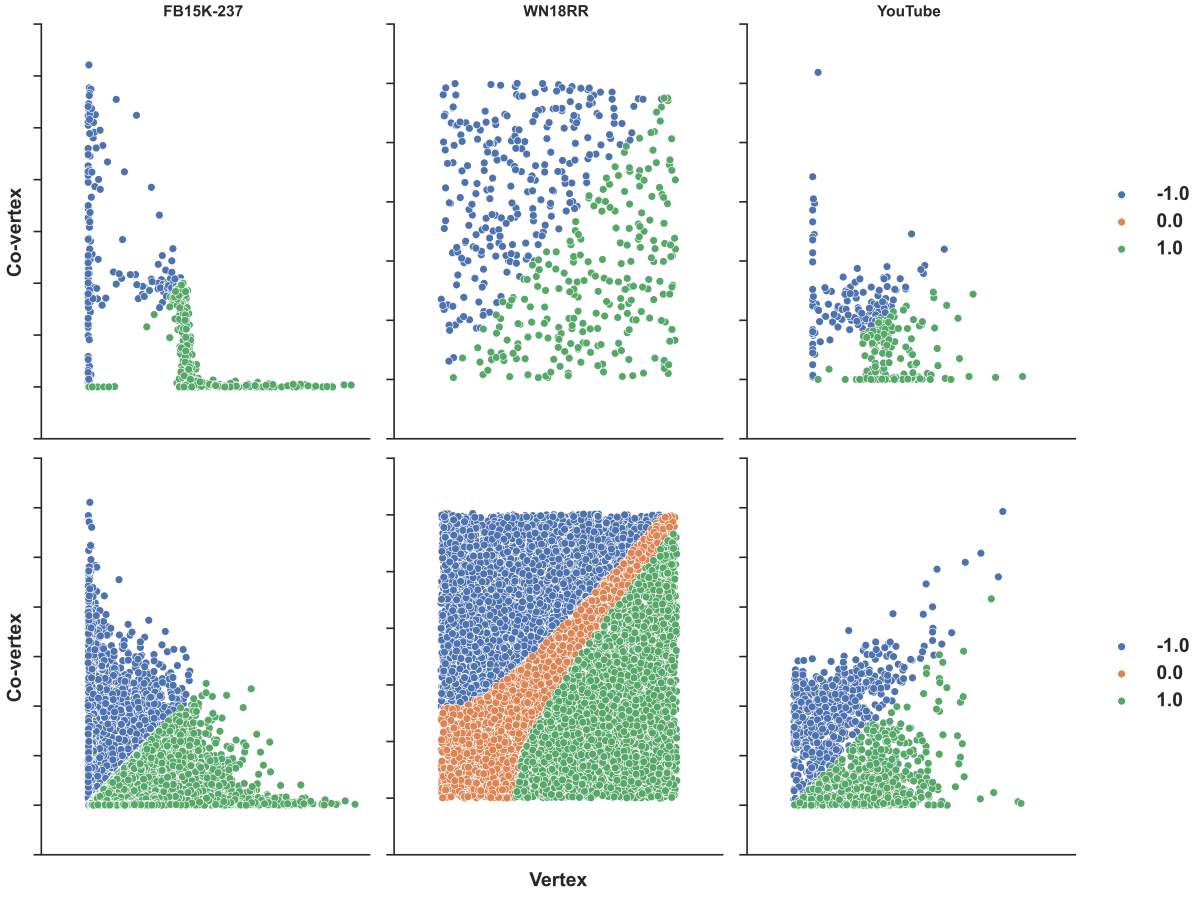

Figure 3: Distribution and classification of elliptical shapes based on the eccentricities $e = a^2 - b^2$. Eccentricity of relational points in class, -1, 0. and 1 satisfies $e \in (-\infty, -10^{-5}), e \in [-10^{-5}, 10^5]$, and $e \in (10^5, +\infty)$ respectively.

Task	Head		Tail	
	MRR	H@10	MRR	H@10
TransE	.251	.469	.365	.610
DistMult	.231	.422	.346	.560
ComplEx	.247	.453	.353	.588
RotatE	.258	.471	.373	.611
EllipsE	.269	**.481**	**.376**	.613
EllipsEs	.267	.480	.374	.616
ButtErfly	**.274**	.473	.373	.609
ButtErflies	.268	.464	**.376**	**.618**

Table 5: Head and tail prediction tasks on many-to-many test data set extracted from FB15k-237. The results of the baseline models were reported from (Li et al., 2022).

ticed a net superior performance of our models compare to TransE and RotatE. This proves that the non consideration of the relational set manifold led to suboptimal performance of KG completion.

Geometry of Relational Set Effective Learning. Fig. 2 contains the projection of the first ten relational sets into a single complex plane for EllipsE, EllipsEs, and ButtErfly models on the FB15K-237 dataset. It therefore shows the distribution of relations. We observed that EllipsE learned an elliptical distribution of relation embeddings in each dimension. ButttErfly mostly distributes relations on the lower part of the wings of the butterfly curve.

Type	A		C		AC		U_A		U_C		U_S	
Metric	MRR	H10	MRR	H10	MRR	H10	MRR	H10	MRR	H10	MRR	H10
TransE	.284	.490	.236	.702	.353	.537	.330	.522	.353	.545	**1.00**	**1.00**
DistMult	.248	.438	.265	.689	.340	.498	.310	.478	.338	.505	.341	.770
ComplEx	.269	.468	.262	.709	.350	.525	.323	.506	.349	.532	.714	.878
RotatE	.286	.493	.239	.694	.360	.542	.336	.526	.360	.548	**1.00**	**1.00**
EllipsE	**.341**	**.770**	.259	.709	.364	.547	.339	.527	.365	.554	**1.00**	**1.00**
EllipsEs	.295	.501	.236	.684	**.371**	**.552**	**.346**	**.535**	**.370**	**.559**	**1.00**	**1.00**
ButtErfly	.304	.498	.267	**.728**	.363	.535	.344	.523	.364	.543	**1.00**	**1.00**
ButtErflies	.300	.497	**.271**	.726	.361	.531	.341	.520	.362	.539	**1.00**	**1.00**

Table 6: Evaluation results on pattern specific datasets derived from FB15K-237. The test datasets S, AS and ACS are empty. Thus, U_S and CS are identical.

Type	A		U_A		U_C		U_S	
Metric	MRR	H10	MRR	H10	MRR	H10	MRR	H10
TransE	.101	.271	.103	.276	.122	.328	.438	.963
DistMult	.141	.275	.142	.279	.155	.320	.957	**.973**
ComplEx	.182	.308	.185	.311	**.217**	.337	**.960**	**.973**
RotatE	.196	.339	.197	**.380**	.200	**.392**	.957	.971
EllipsE	**.207**	**.342**	**.207**	.343	.205	.352	.958	.969
EllipsEs	**.207**	.332	.206	.334	.200	.358	.956	.969
ButtErfly	.200	.321	.199	.321	.199	.326	.935	.969
ButtErflies	.174	.279	.174	.280	.169	.288	.944	.969

Table 7: Evaluation results on pattern specific datasets derived from WN18RR. The test datasets C, AS, CS, and ACS are all empty sets. Thus, U_S and S, as well as U_C and AC, are identical sets.

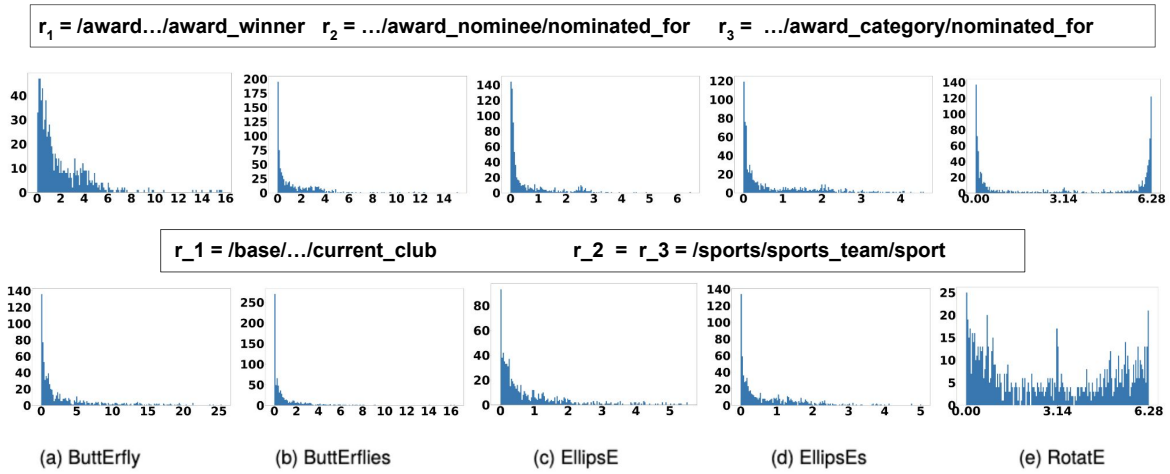

Figure 4: Histograms of composition of relation embeddings r_1, r_2, and $r_3 = r_2 \circ r_1$; where r_i are the names of relations in the FB15k-237 dataset. The evaluation of r_1 = /award/award_category/winners./award/award_honor/award_winner, r_2 = /award/award_nominee/award_nominations./award/award_nomination/nominated_for and r_3 = /award/award_category/nominees./award/award_nomination/nominated_for are displayed along the first row; while the second row reports on r_1 = /base/…/current_club, $r_2 = r_3$ =/sports/sports_team/sport. The results from each model are displayed column-wise.

We could not deduce an elliptical distribution learning from the EllipsEs model, since the distributions along each dimension are relation based. In other words, only a single relational point lies on the relation-specific curve in a given dimension. We only show the visualization of the results on the FB15K-237 dataset since WN18RR and YouTube datasets contains only 11 and 5 relations respectively; this results into a sparse distribution of relations. In order to reach a global appreciation

of the effective use of ellipses, we plotted vertex against co-vertex embeddings for all the relations in FB15K-237. We then classified the points into three categories based on the eccentricity $e = a^2 - b^2$. An eccentricity less than -1, bigger than 1, or equals 0 means the point lies on an ellipse with major axis the y-axis ($a < b$), the x-axis ($a > b$), or it lies on a circle ($a = b$). The top and bottom rows of Fig. 3 display the distribution of the vertex against co-vertex embeddings learned by EllipsE and EllipsEs respectively. It is noteworthy that vertex and co-vertex are relation-independent hyperparameters for the EllipsE model; therefore, both are learned dimension-wise. This explains the sparse distribution we observe in the top row plots. In the top row, points in class 0 are almost inexistent. This shows that EllipsE learned pure elliptical curves during training throughout the three datasets. EllipsEs also shows similar results, except the distribution on WN18RR dataset, where it can barely discriminate against circular shapes. We argue that this is the reason EllipsEs under-performs EllipsE on WN18RR, although it is the overall best performing model on the FB15K-237 dataset.

On Theoretical Evidence of Relation Composition. The evaluation of the ability of a model to infer relation composition patterns (r_1, r_2, r_3) requires comparing the embeddings of the composite relation (r_3) against the embeddings of the composition $(r_2 \circ r_1)$. For the RotatE model, this means $\theta^{r_3} \ominus (\theta^{r_2} \oplus \theta^{r_1})$ is a multiple of 2π; where θ^r are the relation embeddings. For our four models, we use the complex representation of the relation embeddings. That is,

$$\mathbf{z}^r = \begin{cases} \rho(\theta^r)e^{i\theta^r} & \text{for ButtErfly;} \\ \mathbf{b}^r\rho(\theta^r)e^{i\theta^r} & \text{for ButtErflies;} \\ \mathbf{a}\cos\theta^r \oplus i\mathbf{b}\sin\theta^r & \text{for EllipsE;} \\ \mathbf{a}^r\cos\theta^r \oplus i\mathbf{b}^r\sin\theta^r & \text{for EllipsEs.} \end{cases}$$

Our models infer composition by imposing $\mathbf{z}^{r_3} \ominus (\mathbf{z}^{r_2} \odot \mathbf{z}^{r_1}) = \mathbf{0}$. Fig. 4 shows evidence of implicit inference on two composition relational patterns extracted from FB15k-237. For comparison reasons, we also reported the evaluation of RotatE. The superiority of EllipsEs on EllipsE and ButtErflies on ButtErfly in preserving composition patterns can be apprehended from the columns of this figure.

Evaluation on Pattern Specific Datasets We observed from Tables 6 and 7 that antisymmetry is the dominant pattern in the WN18RR dataset, and in the FB15K-237 dataset, antisymmetry and composition patterns are dominants. WN18RR is richer in symmetric patterns than FB15K-237, however, FB15K-237 is much richer in antisymmetric and composition patterns. We are interested in these details since they help in understanding the interaction between model characteristics and dataset properties. The link prediction analyses on each one type shows that all of our proposed models perform better than the state of the art models RotatE, ComplEx, TransE, and DistMult on FB15K-237. EllipsE, RotatE and ComplEx are the best performing models on WN18RR, however, the difference in models' performances is not statistically significant. In particular, (i) our ellipse models are the best performing models in preserving all relational patterns except relations of type C on FB15K-237 and relations of type U_C and U_S on WN18RR. (ii) The butterfly models are superior in handling type C relations, and so are more convenient in embedding datasets rich in relations of type C.

Many-To-Many Pattern Modelling. As explained in Section 5.2, the H@3 and H@10 performances of ButtErflies on YouTube (Table 1) unveils the ability of the butterfly models to learn many-to-many patterns. We moreover experimentally showcased the ability of all our KGG models to learn many-to-many structural and topological patterns. We used the average number of head per tail and tail per head to group relations into four categories: 1-to-1, 1-to-M, N-to-1, and N-to-M. We therefore assess the MRR and H@10 average performances of each group on head and tail prediction tasks. The results on FB15K-237 are shown in Table 5. We noticed that our models consistently and significantly outperform the baselines for the N-to-M tail prediction, and this for both MRR and H@10. In particular, two of the proposed models, namely the ButtErfly or ButtErflies models showed outstanding performance for N-to-M relational subgraph link prediction.

7. Conclusion

In this paper, we propose a new perspective on the design of knowledge graph embedding models. To the best of our knowledge the GeoGebra framework is the first effort to consider the geometry of relation embeddings in KG representation learning. Compared to RotatE which distributes relation embeddings on a circle, our methods are obtained by replacing the circle with elliptical and butterfly curves. The evaluation of the performance of our models on link prediction tasks shows that our models can considerably improve upon the results of RotatE on three benchmark datasets. Furthermore, our analysis of the geometry of the relational set, and the ability of the models to preserve many-to-many and composition relational patterns demonstrate the effectiveness of the Knowledge GeoGebra approach.

Though we only considered two manifolds in this paper, we plan to extend this and increase the dimension of the manifolds. It is evident that geometric knowledge graph embedding models, which utilise rotation transformations, can be straightforwardly implemented in this framework.

Ethics Statement

In this study, the used datasets are publicly available and do not contain any personal data. In general, knowledge graph embeddings can reproduce and amplify bias that is present in the underlying knowledge graph (see (Fisher et al., 2020, 2019) for an overview over the issue and processes to mitigate bias). For further ethical considerations for knowledge graph embeddings across all stages of knowledge graph creation we refer to (Draschner et al., 2022).

Acknowledgements

We thank the reviewers and editors for their feedback on the submission. We also acknowledge the financial support by the Federal Ministry of Education and Research of Germany and by Sächsisches Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research „Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig" (ScaDS.AI). The work is also supported by the SECAI project (grant 57616814) funded by DAAD (German Academic Exchange Service), as well as the Leipzig University for providing GPU servers part of the evaluations.

8. Bibliographical References

Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

Mehdi Ali, Max Berrendorf, Charles Tapley Hoyt, Laurent Vermue, Mikhail Galkin, Sahand Sharifzadeh, Asja Fischer, Volker Tresp, and Jens Lehmann. 2021. Bringing light into the dark: A large-scale evaluation of knowledge graph embedding models under a unified framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):8825–8845.

American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.

Rie Kubota Ando and Tong Zhang. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853.

Galen Andrew and Jianfeng Gao. 2007. Scalable training of L_1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 33–40.

Ivana Balazevic, Carl Allen, and Timothy Hospedales. 2019. Multi-relational poincaré graph embeddings. *Advances in Neural Information Processing Systems*, 32.

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *ACM SIGMOD*, pages 1247–1250.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Zongsheng Cao, Qianqian Xu, Zhiyong Yang, Xiaochun Cao, and Qingming Huang. 2022. Geometry interaction knowledge graph embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5521–5529.

Yukuo Cen, Xu Zou, Jianwei Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. 2019. Representation learning for attributed multiplex heterogeneous network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1358–1368.

Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. 2020. Low-dimensional hyperbolic knowledge graph embeddings. *arXiv preprint arXiv:2005.00545*.

Payal Chandak, Kexin Huang, and Marinka Zitnik. 2023. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1):67.

Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

James W. Cooley and John W. Tukey. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Thirty-Second AAAI Conference*.

Carsten Felix Draschner, Hajira Jabeen, and Jens Lehmann. 2022. Ethical and sustainability considerations for knowledge graph based machine learning. In *2022 IEEE Fifth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 53–60. IEEE.

Joseph Fisher, Arpit Mittal, Dave Palfrey, and Christos Christodoulopoulos. 2020. Debiasing knowledge graph embeddings. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7332–7345.

Joseph Fisher, Dave Palfrey, Christos Christodoulopoulos, and Arpit Mittal. 2019. Measuring social bias in knowledge graph embeddings. *arXiv preprint arXiv:1912.02761*.

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

Matt Le, Stephen Roller, Laetitia Papaxanthos, Douwe Kiela, and Maximilian Nickel. 2019. Inferring concept hierarchies from text corpora via hyperbolic embeddings. *arXiv preprint arXiv:1902.00913*.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Chris Bizer. 2015. DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*, 6(2):167–195. Outstanding Paper Award (Best 2014 SWJ Paper).

Yizhi Li, Wei Fan, Chao Liu, Chenghua Lin, and Jiang Qian. 2022. Transher: translating knowledge graph embedding with hyper-ellipsoidal restriction. *arXiv preprint arXiv:2204.13221*.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29.

George A Miller. 1995. Wordnet: a lexical database for english. *CACM*, 38(11):39–41.

Mojtaba Nayyeri, Chengjin Xu, Franca Hoffmann, Mirza Mohtashim Alam, Jens Lehmann, and Sahar Vahdati. 2021. Knowledge graph representation learning using ordinary differential equations. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9529–9548.

Maximillian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30.

Mohammad Sadegh Rasooli and Joel R. Tetreault. 2015. Yara parser: A fast and accurate dependency parser. *Computing Research Repository*, arXiv:1503.06733. Version 2.

Frederic Sala, Chris De Sa, Albert Gu, and Christopher Ré. 2018. Representation tradeoffs for hyperbolic embeddings. In *International conference on machine learning*, pages 4460–4469. PMLR.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019a. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019b. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 57–66.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *CACM*, 57(10):78–85.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28.

Gerhard Weikum, Xin Luna Dong, Simon Razniewski, Fabian Suchanek, et al. 2021. Machine knowledge: Creation and curation of comprehensive knowledge bases. *Foundations and Trends® in Databases*, 10(2-4):108–490.

Han Xiao, Minlie Huang, Yu Hao, and Xiaoyan Zhu. 2015. Transa: An adaptive approach for knowledge graph embedding. *arXiv preprint arXiv:1509.05490*.

Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases.

Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. 2019. Quaternion knowledge graph embeddings. *Advances in neural information processing systems*, 32.