Capturing Knowledge Graphs and Rules with Octagon Embeddings

Victor Charpenay[1] , **Steven Schockaert**[2]
[1]Mines Saint-Etienne, UMR 6158 LIMOS, Saint-Etienne, France
[2]Cardiff University, Cardiff, UK
victor.charpenay@emse.fr, schockaerts1@cardiff.ac.uk

Abstract

Region based knowledge graph embeddings represent relations as geometric regions. This has the advantage that the rules which are captured by the model are made explicit, making it straightforward to incorporate prior knowledge and to inspect learned models. Unfortunately, existing approaches are severely restricted in their ability to model relational composition, and hence also their ability to model rules, thus failing to deliver on the main promise of region based models. With the aim of addressing these limitations, we investigate regions which are composed of axis-aligned octagons. Such octagons are particularly easy to work with, as intersections and compositions can be straightforwardly computed, while they are still sufficiently expressive to model arbitrary knowledge graphs. Among others, we also show that our octagon embeddings can properly capture a non-trivial class of rule bases. Finally, we show that our model achieves competitive experimental results.

1 Introduction

Knowledge graphs, i.e. sets of (entity, relation, entity) triples, have become one of the most popular frameworks for knowledge representation, with applications ranging from search [Reinanda *et al.*, 2020] and recommendation [Guo *et al.*, 2022] to natural language processing [Schneider *et al.*, 2022] and computer vision [Marino *et al.*, 2017]. Their popularity has spurred an extensive line of work dedicated to representation learning on knowledge graphs. Most works in this area focus on the paradigm of knowledge graph embedding [Bordes *et al.*, 2011], aiming to learn a vector representation $\mathbf{e}$ for each entity e and a scoring function s_r for each relation r such that $s_r(\mathbf{e}, \mathbf{f})$ reflects the likelihood that (e, r, f) is a valid triple, i.e. that entity e is in relation r with entity f.

Knowledge graph embeddings are intended to capture semantic regularities, making it possible to predict plausible triples that were missing from the original knowledge graph. Consider, for instance, the seminal TransE model [Bordes *et al.*, 2013], which uses scoring functions of the form $s_r(\mathbf{e}, \mathbf{f}) = -d(\mathbf{e} + \mathbf{r}, \mathbf{f})$, where $\mathbf{r} \in \mathbb{R}^n$ is an embedding of the relation r, $\mathbf{e}, \mathbf{f} \in \mathbb{R}^n$ are entity embeddings of the same dimension and d is a distance metric. Note that $s_r(\mathbf{e}, \mathbf{f})$ achieves its maximal value of 0 iff $\mathbf{e} + \mathbf{r} = \mathbf{f}$. We can thus say that the embedding fully supports the triple (e, r, f) when $s_r(\mathbf{e}, \mathbf{f}) = 0$. If $\mathbf{r_1} + \mathbf{r_2} = \mathbf{r_3}$ we have that $s_{r_1}(\mathbf{e}, \mathbf{f}) = 0$ and $s_{r_2}(\mathbf{f}, \mathbf{g}) = 0$ together imply $s_{r_3}(\mathbf{e}, \mathbf{g}) = 0$. In this sense, we can say that the embedding captures the following rule:

$$r_1(X, Y) \wedge r_2(Y, Z) \rightarrow r_3(X, Z) \tag{1}$$

This correspondence between knowledge graph embeddings and symbolic rules is appealing. For instance, if we already know that (1) is valid, we can impose $\mathbf{r_1} + \mathbf{r_2} = \mathbf{r_3}$ when learning the embedding. We can also use this correspondence to inspect which semantic dependencies a learned embedding is capturing. Unfortunately, TransE has some inherent limitations, which mean that certain knowledge graphs cannot be faithfully captured [Wang *et al.*, 2018]. This has been addressed in more recent models such as ComplEx [Trouillon *et al.*, 2017], ConvE [Dettmers *et al.*, 2018] and TuckER [Balazevic *et al.*, 2019], to name just a few, but while these models often perform better on the task of link prediction, the connection between their parameters and the captured semantic dependencies is considerably more opaque.

Region based knowledge graph embedding models [Gutiérrez-Basulto and Schockaert, 2018; Abboud *et al.*, 2020; Pavlovic and Sallinger, 2023] aim to get the best of both worlds, increasing the expressivity of TransE while maintaining an explicit correspondence between model parameters and semantic dependencies. They represent each relation r as a geometric region $X_r \subseteq \mathbb{R}^{2n}$. We say that a triple (e, r, f) is captured by the embedding if $\mathbf{e} \oplus \mathbf{f} \in X_r$, where $\oplus$ denotes vector concatenation. To characterise semantic dependencies, we can then exploit the fact that the intersection, subsumption and composition of relations can be naturally modelled in terms of their embeddings X_r. For instance, we say that the embedding captures the rule $r_1(X, Y) \wedge r_2(X, Y) \rightarrow r_3(X, Y)$ iff $X_{r_1} \cap X_{r_2} \subseteq X_{r_3}$. To model rules of the form (1), we can characterise relational composition as follows:

$$X_r \diamond X_s = \{\mathbf{e} \oplus \mathbf{g} \mid \exists \mathbf{f} \in \mathbb{R}^n \,.\, \mathbf{e} \oplus \mathbf{f} \in X_r \wedge \mathbf{f} \oplus \mathbf{g} \in X_s\} \tag{2}$$

We then say that (1) is captured iff $X_{r_1} \diamond X_{r_2} \subseteq X_{r_3}$. Rules of the form (1), which are sometimes referred to as closed path rules, play an important role in link prediction [Meilicke *et al.*, 2019]. When designing a region based model, it is thus important that the composition of regions can be straightforwardly characterised. However, existing approaches are severely limited in this respect. For instance, BoxE [Abboud *et al.*, 2020] cannot model relational composition at all, while ExpressivE [Pavlovic and Sallinger, 2023] uses parallelograms, which are not closed under composition.

To address these concerns, our aim in this paper is to develop a region based model which is as simple as possible, while (i) still being expressive enough to capture arbitrary knowledge graphs and (ii) using regions which are closed under intersection and composition. Based on these desiderata, we arrive at a model which relies on axis-aligned octagons (with all angles fixed at 45 degrees). We show that despite their simplicity, the proposed octagon embeddings are sufficiently expressive to properly capture a large class of rule bases. This is an important property, among others because it means that we can inject prior knowledge, in the form of a given rule base, without unintended consequences. Our result is considerably more general than what is possible with BoxE, which is not able to capture any closed path rules, and more general than what is known about ExpressivE. Moreover, because compositions and intersections of octagons can be straightforwardly computed, octagon embeddings are considerably more practical than existing alternatives. While our main focus is on better understanding the expressivity of region based models, we have also empirically evaluated the proposed octagon embeddings. We found octagons to achieve results close to the current state-of-the-art. This demonstrates that learning octagon embeddings is a promising strategy, especially in contexts where both knowledge graphs and rules need to be modelled.[1]

2 Preliminaries

Let $\mathcal{E}$ and $\mathcal{R}$ be sets of entities and relations respectively. We consider knowledge graph embeddings in which each entity $e \in \mathcal{E}$ is represented by a vector $\mathbf{e} \in \mathbb{R}^n$ and each relation $r \in \mathcal{R}$ is represented by a region $X_r \subseteq \mathbb{R}^{2n}$. We refer to such representations as region-based knowledge graph embeddings. We say that a triple (e, r, f) is captured or supported by a given embedding if $\mathbf{e} \oplus \mathbf{f} \subseteq X_r$. This notion of support allows us to unambiguously associate a knowledge graph with a given geometric embedding, which in turn allows us to study the expressivity of knowledge graph embedding models. In practice, we typically use regions with soft boundaries. This makes learning easier and it is better aligned with the fact that link prediction is typically treated as a ranking problem rather than a classification problem. We will return to the issue of learning regions with soft boundaries in Section 5.

Coordinate-wise region embeddings. Region based embeddings were first studied from a theoretical point of view in [Gutiérrez-Basulto and Schockaert, 2018]. Their setting allowed relations to be represented by arbitrary convex polytopes. Using such regions is not feasible in practice, however, as they require exponentially many parameters. A natural solution is to use regions X_r which can be described using n

[1]An extended version of this paper with supplementary materials can be found at https://arxiv.org/abs/2401.16270.

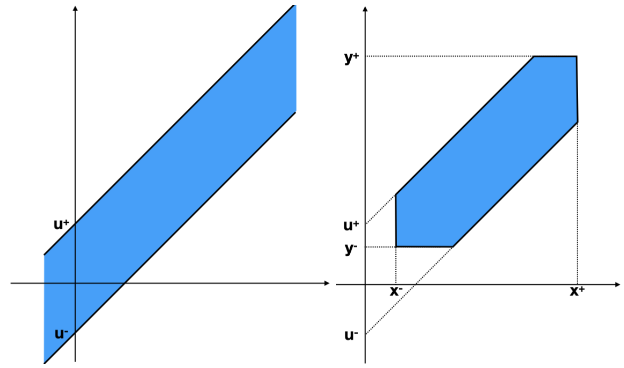

Figure 1: Basic region based embeddings: TransE bands (left) and hexagons (right).

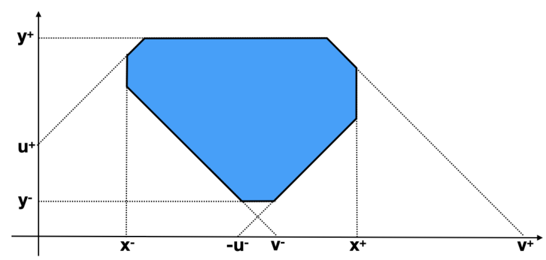

Figure 2: Region-based embeddings with octagons.

two-dimensional regions X_r^i:

$$X_r = \{(x_1, ..., x_n, y_1, ..., y_n) \mid \forall i \in \{1, ..., n\}. (x_i, y_i) \in X_r^i\} \quad (3)$$

We call such region based embeddings *coordinate-wise*. For example, we can view TransE as a coordinate-wise region based embedding model, where

$$X_r^i = \{(x, y) \mid u_i^- \leq y - x \leq u_i^+\} \quad (4)$$

In this case, the two-dimensional regions correspond to unbounded bands, as illustrated in Figure 1. ExpressivE [Pavlovic and Sallinger, 2023] is also a coordinate-wise model, where the regions X_r^i are parallelograms. An important drawback of parallelograms is that they are not closed under composition: if $X_{r_1}^i$ and $X_{r_2}^i$ are parallelograms, then it may be that $X_{r_1}^i \diamond X_{r_2}^i$ is not a parallelogram, with $\diamond$ defined as in (2). Similarly, parallelograms are not closed under intersection. To address these limitations, we can consider some alternatives. Starting from the TransE bands defined in (4), a natural generalisation is to add domain and range constraints, i.e. to use hexagonal regions of the following form, as also illustrated in Figure 1:

$$\begin{aligned} X_r^i = \{(x, y) \mid\, & u_i^- \leq y - x \leq u_i^+, \\ & x_i^- \leq x \leq x_i^+, y_i^- \leq y \leq y_i^+\} \end{aligned} \quad (5)$$

Unfortunately, as we will see in Section 4, such hexagons are not sufficiently expressive to capture arbitrary knowledge graphs. As a minimal extension, we consider octagons of the following form, as illustrated in Figure 2:

$$X_r^i = \{(x, y) \mid u_i^- \leq y - x \leq u_i^+, v_i^- \leq x + y \leq v_i^+, \quad (6)$$

$$x_i^- \leq x \leq x_i^+, y_i^- \leq y \leq y_i^+\}$$

Throughout the paper, we will refer to regions of the form (6) as octagons. However, note that some of these regions are degenerate, having in fact fewer than eight vertices.

Cross-coordinate region embeddings. Some region based models go beyond coordinate-wise embeddings. We will refer to them as *cross-coordinate* models. For instance, BoxE [Abboud *et al.*, 2020] uses regions of the following form:

$$X_r = \{(x_1, \ldots, x_{2n}, y_1, \ldots, y_{2n}) \mid \forall i \in \{1, \ldots, n\} \,.\, (x_i, y_{n+i}) \in A_r^i \wedge (x_{n+i}, y_i) \in B_r^i\} \tag{7}$$

An advantage of cross-coordinate models is that they can rely on simpler two-dimensional regions. For instance, the regions A_r^i and B_r^i in the BoxE model essentially correspond to TransE bands. While a coordinate-wise model with TransE bands is not sufficiently expressive to capture arbitrary knowledge graphs, by using cross-coordinate comparisons, BoxE can circumvent this limitation.

Capturing rules. The main promise of region based models is that the spatial configuration of the regions reveals the semantic dependencies that are captured. Let us first consider a closed-path rule of the following form:

$$r_1(X_1, X_2) \wedge \ldots \wedge r_k(X_k, X_{k+1}) \rightarrow r(X_1, X_{k+1}) \tag{8}$$

Such a rule can be encoded using relational composition and set inclusion, as follows:

$$r_1 \circ \ldots \circ r_k \subseteq r$$

Accordingly, we say that a region based embedding captures this rule iff the following inclusion holds:

$$X_{r_1} \diamond \ldots \diamond X_{r_k} \subseteq X_r$$

Other types of rules can be modelled in a similar way. For instance, a rule such as $r_1(X, Y) \wedge r_2(X, Y) \rightarrow r(X, Y)$ is captured by a region based embedding if $X_{r_1} \cap X_{r_2} \subseteq X_r$.

To illustrate how rules can be captured by region based embeddings, first consider a coordinate-wise model with TransE bands, i.e. regions of the form (4). We have that $X_{r_1} \diamond \ldots \diamond X_{r_k} \subseteq X_r$ holds iff $\sum_{i=1}^k u_{r_i}^- \geq u_r^-$ and $\sum_{i=1}^k u_{r_i}^+ \leq u_r^+$. This means that every rule of the form (8) can be satisfied. However, the model is not sensitive to the order in which the relations in the body appear. For instance, whenever $r_1(X, Y) \wedge r_2(Y, Z) \rightarrow r(X, Z)$ is satisfied, we also have that $r_2(X, Y) \wedge r_1(Y, Z) \rightarrow r(X, Z)$ is satisfied, which is clearly undesirable. Coordinate-wise models with TransE bands are thus not sufficiently expressive to properly capture rules.

Now consider a cross-coordinate model such as BoxE, using regions of the form (7). It holds that $(x_1, \ldots, x_{2n}, z_1, \ldots, z_{2n}) \in X_{r_1} \diamond X_{r_2}$ iff there are $y_1, \ldots, y_{2n} \in \mathbb{R}$ such that for every $i \in \{1, \ldots, n\}$ we have:

$$(x_i, y_{n+i}) \in A_{r_1}^i \qquad (x_{n+i}, y_i) \in B_{r_1}^i$$
$$(y_i, z_{n+i}) \in A_{r_2}^i \qquad (y_{n+i}, z_i) \in B_{r_2}^i$$

This is the case iff for every $i \in \{1, \ldots, n\}$ we have:

$$(x_i, z_i) \in A_{r_1}^i \diamond B_{r_2}^i \qquad (x_{n+i}, z_{n+i}) \in B_{r_1}^i \diamond A_{r_2}^i$$

While $X_{r_1} \diamond X_{r_2}$ constrains the pairs (x_i, z_i) and (x_{n+i}, z_{n+i}), any region X_{r_3} rather constrains the pairs (x_i, z_{n+i}) and (x_{n+i}, z_i). It is thus not possible to have $X_{r_1} \diamond X_{r_2} \subseteq X_{x_3}$, unless in trivial cases where $X_{r_1} = \emptyset$, $X_{r_2} = \emptyset$, or (if infinite bounds are allowed) $X_{r_3} = \mathbb{R}^{4n}$. In all these trivial cases, the embedding also captures other rules, e.g. if $X_{r_1} = \emptyset$ then all rules of the form $r_1(X, Y) \wedge r_4(Y, Z) \rightarrow r_5(X, Z)$ are also satisfied.

ExpressivE is able to capture rules of the form (8), while avoiding unintended consequences [Pavlovic and Sallinger, 2023]. However, in practice we are typically interested in capturing *sets* of such rules, and it is unclear under which conditions this is possible with ExpressivE. Another drawback of ExpressivE is that compositions and intersections of regions are difficult to compute, which, among others, makes checking whether a given rule is satisfied computationally expensive. Moreover, it is not clear how injecting rules into the learning process can then be done in a practical way.

3 Modelling Relations with Octagons

In this paper, we focus on coordinate-wise models where the two-dimensional regions are octagons. Let us write $X_r = [O_1^r, \ldots, O_n^r]$ to denote that relation r is defined using the octagons $O_1^r, \ldots, O_n^r$, in the sense that $(x_1, \ldots, x_n, y_1, \ldots, y_n) \in X_r$ iff $(x_i, y_i) \in O_i^r$ for each $i \in \{1, \ldots, n\}$. Clearly, if $X_r = [O_1^r, \ldots, O_n^r]$ and $X_s = [O_1^s, \ldots, O_n^s]$ then we have $X_r \cap X_s = [O_1^r \cap O_1^s, \ldots, O_n^r \cap O_n^s]$ and $X_r \diamond X_s = [O_1^r \diamond O_1^s, \ldots, O_n^r \diamond O_n^s]$. To study how properties such as reflexivity, symmetry and transitivity can be satisfied in octagon embeddings, it is thus sufficient to study these properties for individual octagons, which is what we focus on in this section.

Parameterisation. Let us write $\mathsf{Octa}(x_i^-, x_i^+, y_i^-, y_i^+, u_i^-, u_i^+, v_i^-, v_i^+)$ to denote the octagon defined in (6). Note that the eight parameters are not independent. For instance, we have that $\mathsf{Octa}(x^-, x^+, y^-, y^+, u^-, u^+, v^-, v^+) = \mathsf{Octa}(x^-, x^+, y^-, y^+, u^-, u^+, \max(v^-, x^- + y^-), v^+)$. Indeed, if $x \geq x^-$ and $y \geq y^-$, we also have $x + y \geq x^- + y^-$, meaning that the bound $x + y \geq v^-$ can be strengthened to $x + y \geq \max(x^- + y^-, v^-)$. We call the parameters $(x^-, x^+, y^-, y^+, u^-, u^+, v^-, v^+)$ *normalised* if they cannot be strengthened, i.e. if increasing any of the lower bounds or decreasing any of the upper bounds always leads to a different octagon. As the following proposition reveals, we can easily normalise any set of parameters.[2]

Proposition 1. *Consider the following set of parameters:*

$$\begin{aligned}
x_1^- &= \max(x^-, v^- - y^+, y^- - u^+, 0.5 \cdot (v^- - u^+)) \\
x_1^+ &= \min(x^+, v^+ - y^-, y^+ - u^-, 0.5 \cdot (v^+ - u^-)) \\
y_1^- &= \max(y^-, u^- + x^-, v^- - x^+, 0.5 \cdot (u^- + v^-)) \\
y_1^+ &= \min(y^+, u^+ + x^+, v^+ - x^-, 0.5 \cdot (u^+ + v^+)) \\
u_1^- &= \max(u^-, y^- - x^+, v^- - 2x^+, 2y^- - v^+) \\
u_1^+ &= \min(u^+, y^+ - x^-, v^+ - 2x^-, 2y^+ - v^-) \\
v_1^- &= \max(v^-, x^- + y^-, u^- + 2x^-, 2y^- - u^+)
\end{aligned}$$

[2] All proofs can be found in the supplementary materials.

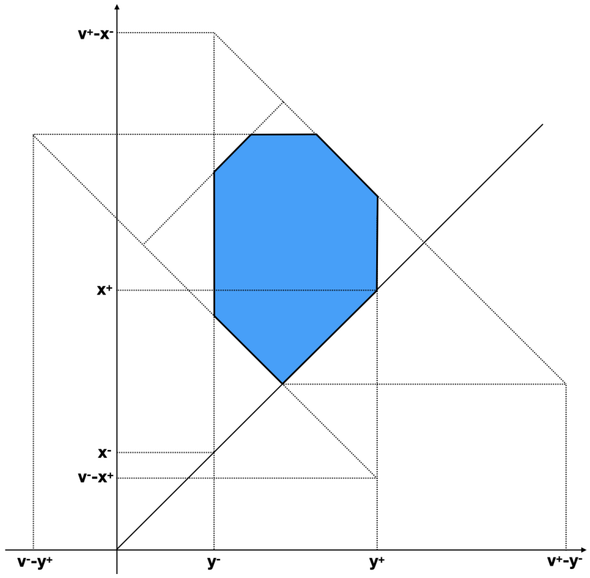

Figure 3: Example of an octagon encoding a transitive relation.

$$v_1^+ = \min(v^+, x^+ + y^+, u^+ + 2x^+, 2y^+ - u^-)$$

Then $\mathsf{Octa}(x^-, x^+, y^-, y^+, u^-, u^+, v^-, v^+) = \mathsf{Octa}(x_1^-, x_1^+, y_1^-, y_1^+, u_1^-, u_1^+, v_1^-, v_1^+)$*. Furthermore, if we have* $x_1^- > x_1^+$*,* $y_1^- > y_1^+$*,* $u_1^- > u_1^+$ *or* $v_1^- > v_1^+$ *then* $\mathsf{Octa}(x^-, x^+, y^-, y^+, u^-, u^+, v^-, v^+) = \emptyset$*. Otherwise, the set of parameters* $(x_1^-, x_1^+, y_1^-, y_1^+, u_1^-, u_1^+, v_1^-, v_1^+)$ *is normalised.*

Capturing relational properties. As we saw in the previous section, two operations are of particular importance for capturing rules: intersection and composition. The *intersection* of $\mathsf{Octa}(x_1^-, x_1^+, y_1^-, y_1^+, u_1^-, u_1^+, v_1^-, v_1^+)$ and $\mathsf{Octa}(x_2^-, x_2^+, y_2^-, y_2^+, u_2^-, u_2^+, v_2^-, v_2^+)$ is the following octagon:

$$\begin{aligned}\mathsf{Octa}(&\max(x_1^-, x_2^-), \min(x_1^+, x_2^+), \max(y_1^-, y_2^-),\\ &\min(y_1^+, y_2^+), \max(u_1^-, u_2^-), \min(u_1^+ u_2^+),\\ &\max(v_1^-, v_2^-), \min(v_1^+, v_2^+))\end{aligned}$$

Octagons are also closed under *composition*, as the following proposition reveals.

Proposition 2. *Let* $O_1 = \mathsf{Octa}(x_1^-, x_1^+, y_1^-, y_1^+, u_1^-, u_1^+, v_1^-, v_1^+)$ *and* $O_2 = \mathsf{Octa}(x_2^-, x_2^+, y_2^-, y_2^+, u_2^-, u_2^+, v_2^-, v_2^+)$ *be non-empty octagons with normalised parameters. Then* $O_1 \diamond O_2 = \mathsf{Octa}(x_3^-, x_3^+, y_3^-, y_3^+, u_3^-, u_3^+, v_3^-, v_3^+)$*, where:*

$$\begin{aligned}
x_3^- &= \max(x_1^-, x_2^- - u_1^+, v_1^- - x_2^+)\\
x_3^+ &= \min(x_1^+, x_2^+ - u_1^-, v_1^+ - x_2^-)\\
y_3^- &= \max(y_2^-, u_2^- + y_1^-, v_2^- - y_1^+)\\
y_3^+ &= \min(y_2^+, u_2^+ + y_1^+, v_2^+ - y_1^-)\\
u_3^- &= \max(y_2^- - x_1^+, u_2^- + u_1^-, v_2^- - v_1^+)\\
u_3^+ &= \min(y_2^+ - x_1^-, u_2^+ + u_1^+, v_2^+ - v_1^-)\\
v_3^- &= \max(x_1^- + y_2^-, u_2^- + v_1^-, v_2^- - u_1^+)\\
v_3^+ &= \min(x_1^+ + y_2^+, u_2^+ + v_1^+, v_2^+ - u_1^-)
\end{aligned}$$

The *inverse* of a relation can also be straightforwardly characterised. Let us define the inverse of an octagon $O = \mathsf{Octa}(x^-, x^+, y^-, y^+, u^-, u^+, v^-, v^+)$ as:

$$O^{\text{inv}} = \mathsf{Octa}(y^-, y^+, x^-, x^+, -u^+, -u^-, v^-, v^+) \qquad (9)$$

Then clearly we have $(x, y) \in O$ iff $(y, x) \in O^{\text{inv}}$, hence the relation encoded by O^{inv} is indeed the inverse of the relation encoded by O. It also follows that O is *symmetric* if the following conditions are satisfied: $x^- = y^-$, $x^+ = y^+$ and $u^- = -u^+$. Similarly, assuming the parameters are normalised, the relation captured by the octagon O is *reflexive* over its domain $[x^-, x^+]$, i.e. $\forall x \in [x^-, x^+] \,.\, (x, x) \in O$, iff the following conditions are satisfied:

- $u^- \leq 0 \leq u^+$;
- $v^- \leq 2x^- \leq 2x^+ \leq v^+$.

From the definition of composition, it straightforwardly follows that a non-empty octagon with normalised coordinates is *transitive*, i.e. satisfies $O \diamond O \subseteq O$, iff the following conditions are satisfied:

- $u^- \leq \max(y^- - x^+, 2u^-, v^- - v^+)$;
- $u^+ \geq \min(y^+ - x^-, 2u^+, v^+ - v^-)$;
- $v^- \leq \max(x^- + y^-, u^- + v^-, v^- - u^+)$;
- $v^+ \geq \min(x^+ + y^+, u^+ + v^+, v^+ - u^-)$.

We can also show the following.

Proposition 3. *Let* $O = \mathsf{Octa}(x^-, x^+, y^-, y^+, u^-, u^+, v^-, v^+)$ *and assume that these parameters are normalised. It holds that* $O \diamond O = O$ *iff one of the following conditions is satisfied:*

- $O = \emptyset$*;*
- $O = [x^-, x^+] \times [y^-, y^+]$ *with* $x^+ \geq y^-$ *and* $x^- \leq y^+$*;*
- $O = \{(x, x) \mid x \in [x^-, x^+]\}$*;*
- $u^- = 0$*,* $u^+ > 0$*,* $v^- < v^+$*,* $v^- - x^+ \leq x^-$*,* $x^+ \leq v^+ - x^-$*,* $v^- - y^+ \leq y^-$*,* $y^+ \leq v^+ - y^-$ *and* $u^+ = \min(y^+ - x^-, v^+ - v^-)$*;*
- $u^- < 0$*,* $u^+ = 0$*,* $v^- < v^+$*,* $v^- - x^+ \leq x^-$*,* $x^+ \leq v^+ - x^-$*,* $v^- - y^+ \leq y^-$*,* $y^+ \leq v^+ - y^-$ *and* $u^- = \max(y^- - x^+, v^- - v^+)$*.*

Figure 3 shows an octagon which satisfies the conditions from the fourth case. If we compose an octagon with itself a sufficient number of times, we always end up with one of the octagon types from Proposition 3. In particular, let us define $O^{(m)}$ as $O^{(m-1)} \diamond O$ for $m \geq 2$ and $O^{(1)} = O$. We then have the following result.

Proposition 4. *If* $v^- < v^+$ *then there exists some* $m \geq 1$ *such that* $O^{(m)} \diamond O = O^{(m)}$*.*

4 Expressivity of Octagon Embeddings

We now study the ability of octagon embeddings to capture knowledge graphs and rules. We will denote a given knowledge graph embedding as γ. Such an embedding represents each entity $e \in \mathcal{E}$ as the vector $\gamma(e) \in \mathbb{R}^n$ and each relation $r \in \mathcal{R}$ as the region $\gamma(r) \subseteq \mathbb{R}^{2n}$. We say that γ is an octagon embedding if it is a coordinate-wise model in which $\gamma(r)$ is defined in terms of octagons, for each $r \in \mathcal{R}$.

4.1 Capturing Knowledge Graphs

With each region based embedding γ we can associate the knowledge graph $\mathcal{G}_\gamma = \{(e, r, f) \mid \gamma(e) \oplus \gamma(f) \in \gamma(r)\}$. An important question is whether a given region based embedding model is capable of modelling any knowledge graph, i.e. whether for any knowledge graph $\mathcal{G} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ there exists an embedding γ such that $\mathcal{G} = \mathcal{G}_\gamma$. When using coordinate-wise embeddings with TransE bands, this is not possible [Wang *et al.*, 2018; Kazemi and Poole, 2018]. In fact, for a coordinate-wise model with hexagons of the form (5) this is still not possible as the following proposition reveals.

Proposition 5. *Let X_r be a hexagon region of the form* (5) *and let $\mathbf{e}, \mathbf{f} \in \mathbb{R}^n$. If $\mathbf{e} \oplus \mathbf{f} \in X_r$ and $\mathbf{f} \oplus \mathbf{e} \in X_r$ then we also have $\mathbf{e} \oplus \mathbf{e} \in X_r$.*

From this proposition, it follows that the hexagon model cannot correctly capture a knowledge graph containing the triples (e, r, f) and (f, r, e) but not (e, r, e). However, if we use octagons instead of hexagons, we can correctly capture any knowledge graph.

Proposition 6. *Let $\mathcal{G} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ be a knowledge graph. There exists an octagon embedding γ such that $\mathcal{G} = \mathcal{G}_\gamma$.*

4.2 Capturing Rules

We now analyse which sets of rules can be correctly captured by octagon embeddings. Previous work [Abboud *et al.*, 2020] has focused in particular on the following types of rules[3]:

Symmetry:	$r(X, Y) \rightarrow r(Y, X)$	(10)
Asymmetry:	$r(X, Y) \rightarrow \neg r(Y, X)$	(11)
Inversion:	$r_1(X, Y) \rightarrow r_2(Y, X)$	(12)
Hierarchy:	$r_1(X, Y) \rightarrow r_2(X, Z)$	(13)
Intersection:	$r_1(X, Y) \wedge r_2(X, Y) \rightarrow r_3(X, Y)$	(14)
Mutual exclusion:	$r_1(X, Y) \wedge r_2(X, Y) \rightarrow \bot$	(15)
Composition:	$r_1(X, Y) \wedge r_2(Y, Z) \rightarrow r_3(X, Z)$	(16)

Let us define the inverse region X_r^{inv} as follows:

$$X_r^{inv} = \{(x_1, ..., x_n, y_1, ..., y_n) \mid (y_1, ..., y_n, x_1, ..., x_n) \in X_r\}$$

Note that when $X_r = [O_1, \ldots, O_n]$ then $X_r^{inv} = [O_1^{inv}, \ldots, O_n^{inv}]$ with O_i^{inv} as defined in (9). Each of the aforementioned rule types can be straightforwardly modelled using regions. In particular, we say that a region based embedding γ captures or satisfies (10) iff $\gamma(r) \subseteq \gamma(r)^{inv}$; (11) iff $\gamma(r) \cap \gamma(r)^{inv} = \emptyset$; (12) iff $\gamma(r_1) \subseteq \gamma(r_2)^{inv}$; (13) iff $\gamma(r_1) \subseteq \gamma(r_2)$; (14) iff $\gamma(r_1) \cap \gamma(r_2) \subseteq \gamma(r_3)$; (15) iff $\gamma(r_1) \cap \gamma(r_2) = \emptyset$; and (16) iff $\gamma(r_1) \diamond \gamma(r_2) \subseteq \gamma(r_3)$. For a set of rules $\mathcal{K}$ and a rule ρ, we write $\mathcal{K} \models \rho$ to denote that ρ is entailed by $\mathcal{K}$ in the usual sense. We also write $\gamma \models \rho$ to denote that the embedding γ captures the rule ρ.

[3] *Asymmetry* rule was called *anti-symmetry* in [Abboud *et al.*, 2020]. We use the former to avoid any confusion with the standard notion of anti-symmetric relations. Furthermore, [Abboud *et al.*, 2020] defined *inversion* as $r_1(X, Y) \equiv r_2(Y, X)$. We consider inversion as a rule, since the equivalence can be straightforwardly recovered by two inversion rules.

We are interested in analysing whether a given region based embedding model can capture a set of rules $\mathcal{K}$, without capturing any other rules, i.e. such that for any rule ρ it holds that $\gamma \models \rho$ iff $\mathcal{K} \models \rho$. Whether this is possible in general depends on the kinds of rules which $\mathcal{K}$ is permitted to contain. We first show a negative result. We call a region based embedding γ convex if $\gamma(r)$ is convex for every $r \in \mathcal{R}$.

Proposition 7. *There exists a consistent set of hierarchy, intersection and mutual exclusion rules $\mathcal{K}$ such that every convex embedding of the form* (3) *or* (7) *which satisfies $\mathcal{K}$ also satisfies some hierarchy rule which is not entailed by $\mathcal{K}$.*

We can show the same result when only hierarchy, intersection and asymmetry rules are permitted. Note that Proposition 7 also applies to BoxE, and thus contradicts the claim from [Abboud *et al.*, 2020] that BoxE can capture arbitrary consistent sets of symmetry, asymmetry, inversion, hierarchy, intersection and mutual exclusion rules. The underlying issue relates to the fact that the considered embeddings are defined from two-dimensional regions, and when these regions are convex, Helly's theorem imposes restrictions on when intersections of such regions can be disjoint. Furthermore, note that [Abboud *et al.*, 2020] did not consider composition rules, as such rules cannot be captured by BoxE. When we omit asymmetry, mutual exclusion and composition rules, we can capture any rule base, as the following result shows.

Proposition 8. *Let $\mathcal{K}$ be a set of symmetry, inversion, hierarchy and intersection rules. There exists a coordinate-wise octagon embedding γ which satisfies $\mathcal{K}$, and which only satisfies those symmetry, inversion, hierarchy and intersection rules which are entailed by $\mathcal{K}$, and which does not satisfy any asymmetry, mutual exclusion and composition rules.*

Let us now shift our focus to composition rules. First, we have the following negative result, showing that octagon embeddings cannot capture arbitrary sets of composition rules.

Proposition 9. *Let $\mathcal{K} = \{r_1(X, Y) \wedge r_1(Y, Z) \rightarrow r_2(X, Z), r_2(X, Y) \wedge r_2(Y, Z) \rightarrow r_3(X, Z)\}$. It holds that every octagon embedding γ which satisfies $\mathcal{K}$ also satisfies some composition rule which is not entailed by $\mathcal{K}$.*

The underlying issue relates to composition rules where the same relation appears more than once. If we exclude such cases, we can correctly embed sets of composition rules.

Definition 1. *An* extended composition rule *over a set of relations $\mathcal{R}$ is an expression of the form:*

$$r_1(X_1, X_2) \wedge \ldots \wedge r_k(X_k, X_{k+1}) \rightarrow r_{k+1}(X_1, X_{k+1})$$

where $k \geq 1$ and $r_1, \ldots, r_{k+1} \in \mathcal{R}$. Such a rule is called regular *if we have $r_i \neq r_j$ for $i \neq j$.*

Regular composition rules generalise the hierarchy and composition rules from [Abboud *et al.*, 2020], noting that the latter work also required the three relations appearing in composition rules to be distinct.

Proposition 10. *Let $\mathcal{K}$ be a set of regular composition rules. Assume that any extended composition rule entailed by $\mathcal{K}$ is either a trivial rule of the form $r \subseteq r$ or a regular rule. There exists an octagon embedding γ which satisfies $\mathcal{K}$, and which only satisfies those extended composition rules which are entailed by $\mathcal{K}$.*

This result is significant given the importance of composition rules for link prediction [Meilicke *et al.*, 2019], among others. Moreover, a similar result was not yet shown in previous work on region based embeddings, to the best of our knowledge. At the same time, however, the result is limited by the fact that only regular rules are considered and no inverse relations. Knowledge graph completion often relies (explicitly or implicitly) on rules such as $\textit{playsForTeam}(X,Y) \wedge \textit{playsForTeam}^{-1}(Y,Z) \wedge \textit{playsSport}(Z,U) \rightarrow \textit{playsSport}(X,U)$. Our analysis suggests that capturing sets of such rules would require an octagon model with cross-coordinate comparisons, which we leave as a topic for future work.

5 Learning Octagon Embeddings

In our theoretical analysis, determining whether a triple (e,r,f) is supported by an embedding was based on a discrete criterion: either $\mathbf{e} \oplus \mathbf{f}$ lies in X_r or it does not. In practice, however, we need to learn regions with fuzzy boundaries, given that link prediction is typically treated as a ranking problem, rather than a classification problem. Continuous representations are typically also easier to learn than discrete structures. Let us consider a region $X_r = [O_1^r, \ldots, O_n^r]$, where $O_i^r = \mathsf{Octa}(x_i^-, x_i^+, , y_i^-, y_i^+, , u_i^-, u_i^+, , v_i^-, v_i^+,)$. Let us write $\mathbf{x}^- = (x_1^-, \ldots, x_n^-)$ and similar for $\mathbf{x}^+$, $\mathbf{y}^-$, $\mathbf{y}^+$, $\mathbf{u}^-$, $\mathbf{u}^+$, $\mathbf{v}^-$ and $\mathbf{v}^+$. Octagon embeddings impose the following four constraints on an entity pair (e,f):

$$\mathbf{x}^- \leq \mathbf{e} \leq \mathbf{x}^+ \qquad \mathbf{y}^- \leq \mathbf{f} \leq \mathbf{y}^+$$
$$\mathbf{u}^- \leq \mathbf{f} - \mathbf{e} \leq \mathbf{u}^+ \qquad \mathbf{v}^- \leq \mathbf{e} + \mathbf{f} \leq \mathbf{v}^+$$

We will refer to the regions defined by these four constraints as *bands*. We can straightforwardly use the sigmoid function σ to convert the constraints into soft scores. For the u-constraint we can use the following "distance" to the u-band:

$$\text{dist}_u(e,r,f) = \sigma\big(|(\mathbf{f} - \mathbf{e}) - \mathbf{u}_c| - \mathbf{u}_w\big) \tag{17}$$

where we write $\mathbf{u}_c = (\mathbf{u}^+ + \mathbf{u}^-)/2$ and $\mathbf{u}_w = (\mathbf{u}^+ - \mathbf{u}^-)/2$. The functions dist_x, dist_y and dist_v are defined analogously. Note that $\mathbf{e} \oplus \mathbf{f} \in X_r$ iff $\text{dist}_x(e,r,f) \geq 0.5$, $\text{dist}_y(e,r,f) \geq 0.5$, $\text{dist}_u(e,r,f) \geq 0.5$ and $\text{dist}_v(e,r,f) \geq 0.5$. Furthermore note that BoxE and ExpressivE also convert interval constraints into soft scores. However, rather than using scores of the form (17), these models use a piecewise linear function of the distance to the centre of the interval. Although a different slope is used for points inside and outside the interval, we may question to what extent the resulting embeddings still define regions. While the regions defined by scoring functions of the form (17) also have fuzzy boundaries, this "fuzziness" is mostly limited to the immediate vicinity of the boundary.

We also consider a variant of the distance function (17) which uses learnable attention weights $\mathbf{u}_{\text{att}}$, as follows:

$$\text{dist}'_u(e,r,f) = \mathbf{u}_{\text{att}} \odot \text{dist}_u(e,r,f) \tag{18}$$

where we write $\odot$ for component-wise multiplication. Weighted variants of dist_x, dist_y and dist_v are defined analogously. The attention weights intuitively allow the model to "forget" certain constraints, e.g. by focusing only (or primarily) on the u band in some coordinates.

	FB15k-237				WN18RR			
	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR
TransE	22.3	37.2	53.1	33.2	01.3	40.1	52.9	22.3
BoxE	23.8	**37.4**	**53.8**	**33.7**	40.0	47.2	54.1	45.1
ExpressivE	**24.3**	36.6	51.2	33.3	**46.4**	**52.2**	**59.7**	**50.8**
u	23.1	<u>37.3</u>	<u>53.2</u>	33.2	01.6	39.9	51.5	22.0
ux	23.3	37.1	52.5	33.1	01.9	39.0	51.6	21.8
uxy	23.2	37.2	53.1	33.2	01.8	40.8	52.8	22.8
uv^*	<u>24.1</u>	36.9	52.8	<u>33.6</u>	43.6	48.5	52.9	46.9
$uvxy^*$	<u>24.1</u>	36.7	51.7	33.2	<u>43.6</u>	<u>49.2</u>	<u>56.1</u>	<u>47.9</u>

Table 1: Link prediction performances of region based embedding models. Configurations with * use the variant with attention weights.

As in BoxE and ExpressivE, coordinate-wise scores are aggregated to a scalar by taking their negated norm. We define $\text{dist}(e,r,f) = \text{dist}_x(e,r,f) \oplus \text{dist}_y(e,r,f) \oplus \text{dist}_u(e,r,f) \oplus \text{dist}_v(e,r,f)$ and score triples as follows (and similar for the weighted variant):

$$s(e,r,f) = -\| \,\text{dist}(e,r,f)\, \|_p \tag{19}$$

In the above expression, p is a hyperparameter. Learning occurs by minimizing the following margin loss function with self-adversarial negative sampling [Sun *et al.*, 2019]:

$$-\log\big(\sigma(\lambda - s(e,r,f))\big) - \sum_{i=1}^{|N|} \alpha_i \log\big(\sigma(s(e_i,r,f_i) - \lambda)\big)$$

where $\alpha = \text{softmax}(s(e_1,r,f_1) \oplus \ldots \oplus s(e_n,r,f_n)$, the margin λ is a hyperparameter, and $N = \{(e_1,r,f_1), \ldots, (e_n,r,f_n)\}$ is a set of randomly sampled negative examples.

6 Experimental Results

We consider several variants of our model. The full model is denoted by *uvxy*, whereas *ux* refers to a model in which dist_v and dist_y are not used, and similar for the other variants. For uv and $uvxy$, we have used the variant with attention weights (18). For the other configurations, attention weights were not used, as they were found to make little difference[4]. Note that our u model is almost identical to TransE, except that an explicit width parameter is used, while *uv* can be seen as a variant of ExpressivE without learnable scale factors.

Table 1 compares the performance of our model with TransE [Bordes *et al.*, 2013], BoxE [Abboud *et al.*, 2020] and ExpressivE [Pavlovic and Sallinger, 2023] for link prediction, using the two most common benchmarks [Toutanova and Chen, 2015; Dettmers *et al.*, 2018]. As usual, performance is measured using Hits@k and Mean Reciprocal Rank (MRR). All results are for 1000-dimensional embeddings for FB15k-237 and 500-dimensional embeddings for WN18RR, following BoxE and ExpressivE's experimental setup. Other hyperparameters are specified in the supplementary materials.

The performance of u is almost identical to TransE, which suggests that the addition of an explicit width parameter plays

[4]An analysis of the impact of attention weights can be found in the supplementary materials.

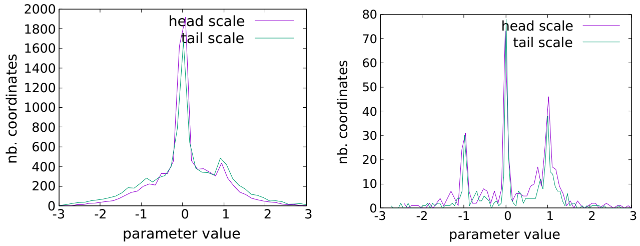

Figure 4: Distribution of scale factors in 40-dimensional ExpressivE embeddings for FB15k-237 (left) and WN18RR (right).

no significant role. Adding the x, y and v constraints does not bring any benefits on FB15k-237. On WN18RR, however, introducing the v constraint has a big impact on H@1, which goes from less than 1% for u, ux and uxy to more than 40% for uv and $uvxy$. This result is in line with Proposition 5, which states that hexagons cannot capture relations that are both symmetric and irreflexive: for most WN18RR relations, models without v constraints incorrectly rank (e, r, e) triples first[5]. Including the x and y constraints also improves performances in the case of WN18RR. Introducing attention weights, as in (18), has a significant positive effect on the overall performance on WN18RR but not for FB15k-237. As Table 1 shows, $uvxy$ outperforms BoxE on all metrics. Further analysis can be found in the supplementary materials.

Octagon embeddings slightly underperform ExpressivE in most cases. We may wonder whether this reflects the inherent limitations of using fixed angles. In fact, when inspecting the scale factors in learned ExpressivE embedddings, we noticed that most are close to -1, 0 or 1 (see Figure 4). In other words, most of the learned parallelograms are diamonds (uv), bands (u, v, x, y) or squares (xy), all of which can be represented in our octagon model. This suggests that the underperformance of the octagon model is rather the result of overparameterisation. On the one hand, we have shown theoretically that all four constraints are needed to capture semantic dependencies, and this is also confirmed by our empirical results on WN18RR. On the other hand, not all constraints might be needed for all coordinates. The use of attention weights was inspired by this view, but the weights themselves of course also introduce further parameters.

7 Related Work

Knowledge graph embeddings have generally been classified in three broad categories: linear models, tensor decomposition models and neural models [Hogan *et al.*, 2021]. Linear models use a distance to score triples, after applying a linear transformation to entity or relation embeddings, e.g. a translation [Bordes *et al.*, 2013] or rotation [Sun *et al.*, 2019]. Region based models can essentially be viewed under the same umbrella, where the idea of computing distances between points is generalised to evaluating relationships between points and regions. Tensor decomposition models, such as ComplEx [Trouillon *et al.*, 2017; Lacroix *et al.*, 2018], TuckER [Balazevic *et al.*, 2019] and QuatE [Zhang *et al.*, 2019], view knowledge graphs as a three-dimensional tensor, which can be factorised into (low-rank) entity and relation embeddings. While these models tend to perform well, they often come with a significantly higher computational cost. Another important limitation is that they are far less interpretable in terms of how they capture semantic dependencies between relations. Tensor decomposition models also have important theoretical limitations when it comes to capturing rules [Gutiérrez-Basulto and Schockaert, 2018]. Neural models, including ConvE [Dettmers *et al.*, 2018] and graph neural network based approaches [Zhu *et al.*, 2021], are even less interpretable than tensor decomposition models.

Knowledge graph embeddings have been combined with rules in various ways. For instance, it has been proposed to incorporate learned rules [Guo *et al.*, 2018] or hard ontological constraints [Abboud *et al.*, 2020] when learning knowledge graph embeddings. Conversely, learned embeddings have been exploited for implementing an efficient rule mining system [Omran *et al.*, 2018]. The ability of models to capture rules has also been evaluated indirectly, for instance by analysing their inductive generalisation capabilities [Trouillon *et al.*, 2019] or their ability to capture intersections and compositions of relations for evaluating complex queries [Arakelyan *et al.*, 2021; Wang *et al.*, 2023].

Region based approaches are closely related to ontology embeddings. In particular, several models have recently been introduced for the Description Logic $\mathcal{EL}^{++}$, which features subsumption and composition of relations, as well as union and intersection of concepts. Concepts are typically embedded as hyperballs [Mondal *et al.*, 2021] or hypercubes [Xiong *et al.*, 2022]. The Box^{2}EL model [Jackermeier *et al.*, 2023] also represents relations as regions, based on BoxE, which was found to improve performance. Given the limitations of BoxE on modeling composition, an evaluation of more expressive region based models is likely to yield even better results, although such an evaluation is yet to be performed.

8 Conclusions

The study of region based knowledge graph embedding essentially aims to identify the simplest models that are capable of capturing particular types of reasoning patterns. Despite the extensive work on knowledge graph embedding, there are still surprisingly many open questions on this front. Octagons where shown to be capable of capturing arbitrary knowledge graphs, while a simpler model based on hexagons is not. Octagons also have the desired property of being closed under intersection and composition, which is not the case for closely related models such as BoxE and ExpressivE. We have furthermore shown that the octagon model is capable of modelling (particular) sets of composition rules, which is an important property that was not yet shown for existing region based models. Our empirical results show that octagon embeddings slightly underperform ExpressivE and further work is needed to better understand whether this performance gap can be closed by changing how the octagons are learned.

[5]This result still holds without attention weights: on WN18RR, $uvxy$ achieves a H@1 of 39.6, H@3 of 45.9, H@10 of 50.7 and MRR of 43.8 without trainable attention weights.

Acknowledgments

Computations have been performed on the supercomputer facilities of the Mésocentre Clermont-Auvergne of the Université Clermont Auvergne. Steven Schockaert was supported by grants from EPSRC (EP/W003309/1) and the Leverhulme Trust (RPG-2021-140).

References

[Abboud *et al.*, 2020] Ralph Abboud, İsmail İlkan Ceylan, Thomas Lukasiewicz, and Tommaso Salvatori. BoxE: A box embedding model for knowledge base completion. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

[Arakelyan *et al.*, 2021] Erik Arakelyan, Daniel Daza, Pasquale Minervini, and Michael Cochez. Complex query answering with neural link predictors. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

[Balazevic *et al.*, 2019] Ivana Balazevic, Carl Allen, and Timothy M. Hospedales. TuckER: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 5184–5193. Association for Computational Linguistics, 2019.

[Bordes *et al.*, 2011] Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. Learning structured embeddings of knowledge bases. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011*, pages 301–306. AAAI Press, 2011.

[Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795, 2013.

[Dettmers *et al.*, 2018] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 1811–1818. AAAI Press, 2018.

[Guo *et al.*, 2018] Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Knowledge graph embedding with iterative guidance from soft rules. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 4816–4823. AAAI Press, 2018.

[Guo *et al.*, 2022] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. A survey on knowledge graph-based recommender systems. *IEEE Trans. Knowl. Data Eng.*, 34(8):3549–3568, 2022.

[Gutiérrez-Basulto and Schockaert, 2018] Víctor Gutiérrez-Basulto and Steven Schockaert. From knowledge graph embedding to ontology embedding? an analysis of the compatibility between vector space representations and rules. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018*, pages 379–388. AAAI Press, 2018.

[Hogan *et al.*, 2021] 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 Sequeda, Steffen Staab, and Antoine Zimmermann. *Knowledge Graphs*. Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool Publishers, 2021.

[Jackermeier *et al.*, 2023] Mathias Jackermeier, Jiaoyan Chen, and Ian Horrocks. Box^{2}EL: Concept and role box embeddings for the description logic EL++. *CoRR*, abs/2301.11118, 2023.

[Kazemi and Poole, 2018] Seyed Mehran Kazemi and David Poole. Simple embedding for link prediction in knowledge graphs. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 4289–4300, 2018.

[Lacroix *et al.*, 2018] Timothée Lacroix, Nicolas Usunier, and Guillaume Obozinski. Canonical tensor decomposition for knowledge base completion. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 2869–2878. PMLR, 2018.

[Marino *et al.*, 2017] Kenneth Marino, Ruslan Salakhutdinov, and Abhinav Gupta. The more you know: Using knowledge graphs for image classification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 20–28. IEEE Computer Society, 2017.

[Meilicke *et al.*, 2019] Christian Meilicke, Melisachew Wudage Chekol, Daniel Ruffinelli, and Heiner Stuckenschmidt. Anytime bottom-up rule learning for knowledge graph completion. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 3137–3143. ijcai.org, 2019.

[Mondal *et al.*, 2021] Sutapa Mondal, Sumit Bhatia, and Raghava Mutharaju. EmEL++: Embeddings for EL++ description logic. In *Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering (AAAI-MAKE 2021), Stanford University, Palo Alto, California, USA, March 22-24, 2021*, volume 2846 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021.

[Omran *et al.*, 2018] Pouya Ghiasnezhad Omran, Kewen Wang, and Zhe Wang. Scalable rule learning via learning representation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 2149–2155. ijcai.org, 2018.

[Pavlovic and Sallinger, 2023] Aleksandar Pavlovic and Emanuel Sallinger. ExpressivE: A spatio-functional embedding for knowledge graph completion. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.

[Reinanda *et al.*, 2020] Ridho Reinanda, Edgar Meij, and Maarten de Rijke. Knowledge graphs: An information retrieval perspective. *Found. Trends Inf. Retr.*, 14(4):289–444, 2020.

[Schneider *et al.*, 2022] Phillip Schneider, Tim Schopf, Juraj Vladika, Michael Galkin, Elena Simperl, and Florian Matthes. A decade of knowledge graphs in natural language processing: A survey. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, AACL/IJCNLP 2022 - Volume 1: Long Papers, Online Only, November 20-23, 2022*, pages 601–614. Association for Computational Linguistics, 2022.

[Sun *et al.*, 2019] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

[Toutanova and Chen, 2015] Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality, CVSC 2015, Beijing, China, July 26-31, 2015*, pages 57–66. Association for Computational Linguistics, 2015.

[Trouillon *et al.*, 2017] Théo Trouillon, Christopher R. Dance, Éric Gaussier, Johannes Welbl, Sebastian Riedel, and Guillaume Bouchard. Knowledge graph completion via complex tensor factorization. *J. Mach. Learn. Res.*, 18:130:1–130:38, 2017.

[Trouillon *et al.*, 2019] Théo Trouillon, Éric Gaussier, Christopher R. Dance, and Guillaume Bouchard. On inductive abilities of latent factor models for relational learning. *J. Artif. Intell. Res.*, 64:21–53, 2019.

[Wang *et al.*, 2018] Yanjie Wang, Rainer Gemulla, and Hui Li. On multi-relational link prediction with bilinear models. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 4227–4234. AAAI Press, 2018.

[Wang *et al.*, 2023] Dingmin Wang, Yeyuan Chen, and Bernardo Cuenca Grau. Efficient embeddings of logical variables for query answering over incomplete knowledge graphs. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 4652–4659. AAAI Press, 2023.

[Xiong *et al.*, 2022] Bo Xiong, Nico Potyka, Trung-Kien Tran, Mojtaba Nayyeri, and Steffen Staab. Faithful embeddings for EL++ knowledge bases. In *The Semantic Web - ISWC 2022 - 21st International Semantic Web Conference, Virtual Event, October 23-27, 2022, Proceedings*, volume 13489 of *Lecture Notes in Computer Science*, pages 22–38. Springer, 2022.

[Zhang *et al.*, 2019] Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. Quaternion knowledge graph embeddings. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 2731–2741, 2019.

[Zhu *et al.*, 2021] Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal A. C. Xhonneux, and Jian Tang. Neural bellman-ford networks: A general graph neural network framework for link prediction. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 29476–29490, 2021.