Proceedings of the 23rd International Conference on Principles of Knowledge Representation and Reasoning
KR meets Machine Learning and Explanation

RegD: Hierarchical Embeddings via Dissimilarity between Arbitrary Euclidean Regions

Hui Yang, **Jiaoyan Chen**
The University of Manchester
{hui.yang-2, jiaoyan.chen}@manchester.ac.uk

Abstract

Hierarchical data is common in many domains like life sciences and e-commerce, and its embeddings often play a critical role. While hyperbolic embeddings offer a theoretically grounded approach to representing hierarchies in low-dimensional spaces, current methods often rely on specific geometric constructs as embedding candidates. This reliance limits their generalizability and makes it difficult to integrate with techniques that model semantic relationships beyond pure hierarchies, such as ontology embeddings. In this paper, we present RegD, a flexible Euclidean framework that supports the use of arbitrary geometric regions—such as boxes and balls—as embedding representations. Although RegD operates entirely in Euclidean space, we formally prove that it achieves hyperbolic-like expressiveness by incorporating a depth-based dissimilarity between regions, enabling it to emulate key properties of hyperbolic geometry, including exponential growth. We establish the faithfulness of our approach. Furthermore, extensive empirical evaluations on diverse real-world datasets demonstrate consistent performance improvements over state-of-the-art methods, highlighting RegD's potential for broader applications, including ontology embedding tasks that extend beyond hierarchical structures.

1 Introduction

Embedding discrete data into low-dimensional vector spaces has become a cornerstone of modern machine learning. In Natural Language Processing (NLP), seminal works such as word2vec (Mikolov et al. 2013) and GloVe (Pennington, Socher, and Manning 2014) represent words as vectors to capture intricate linguistic relationships. Similarly, knowledge graph embedding methods (Bordes et al. 2013; Sun et al. 2019; Balazevic, Allen, and Hospedales 2019) encode entities and relations as vectors with their semantics concerned to facilitate reasoning and prediction.

Our work focuses on embedding hierarchical data into low-dimensional spaces. Such data represents *partial orders* over sets of elements, denoted as $u \prec v$, where u is a parent of v. These partial orders naturally manifest as trees or directed acyclic graphs (DAGs). The ability to effectively embed such hierarchical structures enables crucial operations like inferring sub- or superclasses of concepts and classifying nodes within graphs. These capabilities are essential for various tasks in knowledge management and discovery, particularly towards knowledge bases (Shi et al. 2024; Abboud et al. 2020), ontologies (He et al. 2024a; Chen et al. 2024) and taxonomies (Shen and Han 2022).

Current methods for embedding hierarchical data can be broadly categorized into two paradigms: *region-based* and *hyperbolic* approaches. The region-based approach usually represents entities as geometric regions in the Euclidean space, capturing hierarchical relationships through intuitive region-inclusion. However, these methods often experience degraded performance in low-dimensional settings due to the crowding effect inherent in Euclidean spaces.

In contrast, the hyperbolic approach takes advantage of the unique geometric properties of hyperbolic spaces (*e.g.,* their exponential growth in distance and volume) enabling more effective embeddings of tree-structured data in low dimensions. However, as shown in Table 1, existing hyperbolic methods often rely on specially constructed objects as embedding candidates (Ganea, Bécigneul, and Hofmann 2018; Yu et al. 2024), limiting their generalizability to data that encode richer semantics beyond hierarchy. For example, EntailmentCones (Ganea, Bécigneul, and Hofmann 2018) and ShadowCones (Yu et al. 2024) are not closed under intersection and thus cannot handle conjunction.

In this paper, we propose a flexible framework named RegD for modeling hierarchical data by embedding arbitrary regions in Euclidean spaces. Our framework relies on two novel dissimilarity metrics, *depth dissimilarity* d_{dep} and *boundary dissimilarity* d_{bd}, which combine the strengths of both region-based approaches in Euclidean spaces and hyperbolic methods. **The depth dissimilarity** (cf. Section 3.1) enables our model to achieve embedding expressiveness comparable to that of hyperbolic spaces by incorporating the "size" of the regions under consideration (cf. Theorem 1 and Propositions 1 and the general version in Section 4.1). **The boundary dissimilarity** (cf. Section 3.2) improves the representation of set-inclusion relationships among regions in Euclidean spaces. This allows for better identification of shallower and deeper descendants, thereby capturing hierarchical structures more effectively than traditional region-based approaches (cf. Proposition 2). Note that these dissimilarities may be negative or fail to satisfy the triangle inequality, and thus are generally not strict metric distances.

Notably, RegD can be applied to arbitrary regions[1], in-

[1]Here, "arbitrary regions" does not denote arbitrary subsets of

Method	Embeddings of a node u	Energy/score function for $u \prec v$
Poincaré (Nickel and Kiela 2017)	Points $\mathbf{u} \in \mathbb{H}^d$	$(1 + \alpha(\|\mathbf{u}\| - \|\mathbf{v}\|))d_k(\mathbf{u}, \mathbf{v})$
EntailmentCones (Ganea, Bécigneul, and Hofmann 2018)	Cones with apex $\mathbf{u} \in \mathbb{H}^d$ and angle $\theta_{\mathbf{u}} = \arcsin(\frac{(1-\|\mathbf{u}\|^2)}{\|\mathbf{u}\|})$	Angle-based function
ShadowCones (Yu et al. 2024)	Cones with apex $\mathbf{u} \in \mathbb{H}^d$ and angle $\theta_{\mathbf{u}} = \operatorname{arctan} \sinh(\sqrt{k}r)$	$d_{\mathbb{H}}(\mathbf{u}, \mathbf{v})$ **or** specific boundary distance
RegD (ours)	arbitrary regions in $\mathbb{R}^d$ (e.g., balls, boxes, detailed in Section 4.1)	$d_{\text{bd}}(reg_u, reg_v) + \lambda \cdot d_{\text{dep}}(reg_u, reg_v)$

Table 1: Comparison with hyperbolic methods in $\mathbb{H}^d$ (curvature $-k$, distance d_k). α, r and λ are predefined hyperparameters. Only the umbral half-space model of ShadowCones is shown, with boundary distance defined in Eq. 8.

cluding common geometric representations such as balls and boxes. This generality enables broad applicability across diverse geometric embeddings for various tasks, extending beyond purely hierarchical data to ontologies that include hierarchies and more complex relationships in Description Logic (Baader et al. 2017) (cf. Sections 5.2 and 5.3). Our main contributions are summarized as follows:

- We present a versatile framework that is able to embed hierarchical data as arbitrary regions in Euclidean space.
- We offer a rigorous theoretical analysis demonstrating that our framework retains the core embedding benefits of hyperbolic methods.
- Experiments on diverse real-world datasets demonstrate that our framework consistently outperforms existing approaches on embedding hierarchies and ontologies for reasoning and prediction.

Due to space limitations, implementation details, additional experimental results, as well as the code and datasets, are provided at the following anonymous link: https://github.com/HuiYang1997/RegD.

2 Preliminaries and Related Works

Manifold and Hyperbolic Space A d-dimensional *manifold* (Lee 2013), denoted $\mathcal{M}$, is a hypersurface embedded in an n-dimensional Euclidean space, $\mathbb{R}^n$, where $n \geq d$, and locally resembles $\mathbb{R}^d$.

A *Riemannian manifold* $\mathcal{M}$ is a manifold equipped with a Riemannian metric, enabling the definition of a distance function $d_{\mathcal{M}}(\mathbf{x}, \mathbf{y})$ for $\mathbf{x}, \mathbf{y} \in \mathcal{M}$. *Hyperbolic space*, denoted $\mathbb{H}^n$, is a Riemannian manifold with a constant negative curvature of $-k$, where $k > 0$ (Lee 2006). It can be represented using various isometric models, such as Poincaré half-space model, where the points are defined by the half-space: $U^n = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}_n > 0\}$, and the hyperbolic distance between $\mathbf{x}, \mathbf{y} \in U^n$ is given by

$$d_k(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{k}} \operatorname{arcosh}\left(1 + \frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{2\mathbf{x}_n \mathbf{y}_n}\right).$$

Euclidean space, but rather regular geometric domains of arbitrary shape that can be parameterized and that satisfy the general properties described in Section 4.1.

Region-based Methods Region-based methods embed the nodes of a directed acyclic graph (DAG) into regions on Euclidean space or Riemannian manifolds, such as boxes (Boratko et al. 2021; Zhang et al. 2022), balls (Suzuki, Takahama, and Onoda 2019), and cones (Vendrov et al. 2016), capturing hierarchical relationships through set-inclusion between these regions. Training is typically conducted using an inclusion loss, which is often defined in terms of the distance or volume of the regions (Vendrov et al. 2016). However, such loss functions may involve non-smooth operations, like maximization, which have been addressed through probabilistic (Vilnis et al. 2018; Dasgupta et al. 2020) or smooth (Boratko et al. 2021) approximations to improve performance. However, most of these methods suffer from the crowding effect in Euclidean space, which limits their expressiveness compared to our method and hyperbolic-based approaches.

Hyperbolic Methods Hyperbolic space embeddings (Nickel and Kiela 2017; Sonthalia and Gilbert 2020; Sala et al. 2018) have been introduced to model hierarchical structures by leveraging favorable properties of hyperbolic space, such as its exponential growth. However, these methods primarily rely on pointwise distances to encode hierarchies, which is less intuitive for representing transitivity properties. In contrast, *EntailmentCones* (Ganea, Bécigneul, and Hofmann 2018), and *ShadowCones* propose using regions in hyperbolic space that capture hierarchies through set inclusion, which can be regarded as region-based methods in hyperbolic spaces. Specifically, *EntailmentCones* introduces closed-form hyperbolic cones, defined by their apex coordinates, while *ShadowCones* uses different cones are inspired by the physical interplay of light and shadow. However, as shown in Table 1, these regions all take specific forms, limiting their generalizability to other tasks or methods.

Ontologies Ontologies use sets of statements (axioms) about concepts (unary predicates) and roles (binary predicates) for knowledge representation and reasoning. We focus on commonly used $\mathcal{EL}$-ontologies, defined as follows. Let $\mathsf{N_C} = \{A, B, \dots\}$, $\mathsf{N_R} = \{r, t, \dots\}$, and $\mathsf{N_I} = \{a, b, \dots\}$ be pairwise disjoint sets of *concept names* (also called *atomic concepts*), *role names*, and *individual names*, respectively. *$\mathcal{EL}$-concepts* are recursively defined from atomic concepts, roles, and individuals as follows: $\top \mid \bot \mid A \mid C \sqcap D \mid \exists r.C \mid \{a\}$ An *$\mathcal{EL}$-ontology* is a finite set of TBox axioms of the form $C \sqsubseteq D$. Note that here $\sqsubseteq$ denotes "subsumption",

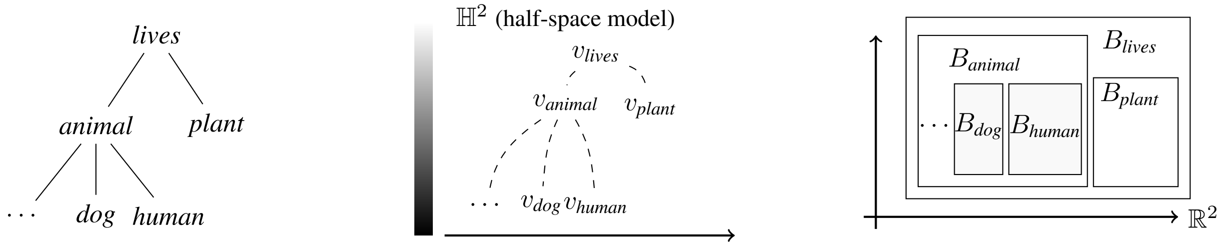

Figure 1: Demonstration of a taxonomy (left), its embeddings in the hyperbolic space (middle) and in the Euclidean space as boxes (right).

which is a transitive relation that can be considered a partial order $\prec$ by regarding $C \sqsubseteq D$ as $D \prec C$.

Ontology Embeddings Ontology embeddings aim to encode ontology entities into numerical vector representations to support a variety of downstream tasks, such as axiom prediction and logical inference. Existing approaches can be broadly categorized into two types: *geometric model-based* methods (Yang, Chen, and Sattler 2025; Jackermeier, Chen, and Horrocks 2024; Peng et al. 2022; Kulmanov et al. 2019), which represent ontology entities as geometric objects and construct a geometric model capturing the semantic structure of the target ontology; and *language model-based* methods (Yang et al. 2025; Chen et al. 2021), which learn embeddings from textual information, typically by leveraging pretrained large language models (LLMs).

3 Method

3.1 Depth Dissimilarity: A Similarity Measure Incorporating Depth

Unlike the Euclidean space, which is constrained by crowding effects that limit its embedding capacity, the hyperbolic space leverages exponential growth in distance and volume to offer superior embedding capabilities. This property makes the hyperbolic space particularly effective for representing tree-structured data in low-dimensional spaces. Notably, two key distinctions arise between region-based embeddings in the Euclidean space and hyperbolic embeddings:

1. *The hyperbolic space better discriminates different hierarchical layers than the Euclidean space.*

   Consider the taxonomy illustrated in Figure 1 (left). Intuitively, since *human* is a subcategory of *animal*, the semantic difference between *human* and *plant* should be greater than that between *animal* and *plant*. The hyperbolic space effectively captures this hierarchical relationship by permitting $d(v_{human}, v_{plant}) > d(v_{animal}, v_{plant}) + \Delta$, where Δ is an arbitrary gap. This arises from the property that the distance metric diverges to infinity near the boundary, as illustrated by the shadow density of the bar on the left-hand side of $\mathbb{H}^2$ in Figure 1 (middle). In contrast, region-based embeddings on Euclidean space may violate this hierarchical constraint, potentially placing the box B_{human} close to B_{plant}, resulting in a distance (*e.g.*, the Euclidean distance between box centers) similar to or even smaller than that between B_{animal} and B_{plant}.

2. *The hyperbolic space enables the distinct representation of an arbitrary number of child nodes.*

   As demonstrated in Figure 1, the box embeddings in the Euclidean space face inherent limitations when representing multiple children of a node, such as *animal*. As the number of children increases, their corresponding boxes must cluster within B_{animal}, leading to crowding. In contrast, the hyperbolic space can accommodate an arbitrary number of children while maintaining distinct separations between them. This capability arises from the exponential growth of distance near the boundary of the hyperbolic space, which allows unlimited child nodes to be positioned distinctly by placing them progressively closer to the boundary while preserving meaningful distances between them.

In the following sections, we introduce the notion of *depth dissimilarity* for regions in Euclidean space, which explicitly accounts for their "size." We show that this measure retains advantages analogous to those of hyperbolic spaces (Theorem 1), while offering a simpler structure that facilitates implementation (e.g., avoiding specialized techniques such as double-precision tensors and Riemannian-specific optimizers (Bécigneul and Ganea 2019)) and enabling potential extensions, e.g., integration with ontology embeddings; see Section 5.2). Moreover, we establish that depth dissimilarity encompasses hyperbolic distance as a special case (Proposition 1), while providing greater flexibility by preserving key properties of hyperbolic geometry through simple polynomial functions (Proposition 2).

Construction The depth dissimilarity serves as a similarity measure that quantifies the relationship between objects considering their hierarchical depth. As we use set-inclusion relations to model the hierarchy, this hierarchical depth can be represented through the size of the regions, such as their volumes or diameters. Formally, the depth dissimilarity is defined as follows, where the size of the regions is represented by a function $f(reg)$:

Definition 1 (Depth Dissimilarity). *Let $\mathcal{R}$ be a collection of* parameterized regions[2] *in the n-dimensional Euclidean space $\mathbb{R}^n$, i.e., each region $reg \in \mathcal{R}$ is characterized by an m-dim parameter $P(reg) \in \mathbb{R}^m$. The depth dissimilarity*

[2]The formal definition is deferred to Section 4.1 for simplicity. Examples using boxes and balls are shown in Example 1, 2.

between two regions $reg_1, reg_2 \in \mathcal{R}$, is defined as:

$$d_{dep}(reg_1, reg_2) = g\left(\frac{\|P(reg_1) - P(reg_2)\|_p^p}{f(reg_1)f(reg_2)}\right) \quad (1)$$

where $\|\cdot\|_p$ is the p-norm (i.e., $||\mathbf{x}||_p = (\sum_i \mathbf{x}_i^p)^{1/p}$), and:

- *$g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is an increasing function such that $g(x) = 0$ if and only if $x = 0$,*
- *$f : \mathcal{R} \to \mathbb{R}_{>0}$ is a function that measures the size of regions. It satisfies: $\lim_{reg \to \emptyset} f(reg) = 0$*[3].

We require f and g to have non-negative values to ensure the depth dissimilarity is non-negative. Additionally, we stipulate that $\lim_{reg\to\emptyset} f(reg) = 0$ to guarantee that as a region shrinks to an empty set, the dissimilarity between this object and others can approach infinity. This setting emulates the beneficial properties of the hyperbolic space, where the dissimilarity between two points can grow rapidly as they approach the boundary of the space (i.e., $\mathbf{x}_n = 0$ in the Poincaré half-space model). In our context, the boundary of the space $\mathcal{R}$ of a collection of (parametrized) regions in the Euclidean space is the empty set. By selecting an appropriate function f, we can control the rate at which the dissimilarity between two objects increases as they approach this boundary.

For clarity and tractability, we focus on the most representative region types, namely boxes and balls, to provide illustrative examples and empirical implementations. We defer the theoretical analysis of general parametrized region representations to Section 4.1.

Example 1. *Let $ball(\mathbf{c}, r) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{c}\| \leq r\}$ be a* ball *defined by a center $\mathbf{c} \in \mathbb{R}^n$ and a radius $r > 0$. By setting $g(x) = x$ and $f(ball) = r$, we obtain a depth dissimilarity of the form:*

$$d_{dep}(ball_1(\mathbf{c}_1, r_1), ball_2(\mathbf{c}_2, r_2)) = \frac{||\mathbf{c}_1 - \mathbf{c}_2||_p^p + |r_1 - r_2|^p}{r_1 \cdot r_2}. \quad (2)$$

Example 2. *Let $box(\mathbf{c}, \mathbf{o}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{c} - \mathbf{o} \leq \mathbf{x} \leq \mathbf{c} + \mathbf{o}\}$ be a* box *defined by a center $\mathbf{c} \in \mathbb{R}^n$ and an offset $\mathbf{o} \in \mathbb{R}^n_{>0}$. By setting $g(x) = x$ and $f(box) = \|\mathbf{o}\|$, we obtain a depth dissimilarity:*

$$d_{dep}(box_1(\mathbf{c}_1, \mathbf{o}_1), box_2(\mathbf{c}_2, \mathbf{o}_2)) = \frac{||\mathbf{c}_1 - \mathbf{c}_2||_p^p + ||\mathbf{o}_1 - \mathbf{o}_2||_p^p}{\|\mathbf{o}_1\| \cdot \|\mathbf{o}_2\|}. \quad (3)$$

The following result highlights that the depth dissimilarity exhibits properties analogous to those of the hyperbolic distance discussed earlier in this section. Specifically, the depth dissimilarity **(1)** effectively distinguishes concepts across different layers of the hierarchy with an arbitrary separation gap, denoted by Δ in item 1 of the following Theorem 1, and **(2)** distinctly represents an arbitrary number n of children within a single parent by a dissimilarity greater than M, as demonstrated in item 2 of the theorem.

[3] $reg \to \emptyset$ indicates that the region *reg* shrinks to the empty set (*e.g.*, its volume or diameter tends to zero).

Theorem 1. *Consider the region space $\mathcal{B}^n$ consisting of* ***balls*** *in $\mathbb{R}^n$, with the depth dissimilarity defined in Example 1. The following properties hold:*

1. *For any $ball_1, ball_2 \in \mathcal{B}^n$ and any $\Delta > 0$, there exists a positive constant r_0 such that for any $ball(\mathbf{c}', r') \subseteq ball_2$, if $r' \leq r_0$, then $d_{dep}(ball_1, ball(\mathbf{c}', r')) > d_{dep}(ball_1, ball_2) + \Delta$.*
2. *For any $ball \in \mathcal{B}^n$, any integer n and any $M > 0$, there exist subsets $ball_1, \ldots, ball_n \subseteq ball$ such that for any distinct $i, j \in \{1, \ldots, n\}$, we have $d_{dep}(ball_i, ball_j) > M$.*

The same conclusions hold for boxes, where $r_0 \in \mathbb{R}_{>0}$ is replaced with $\mathbf{o}_0 \in (\mathbb{R}_{>0})^n$, and the condition $r' \leq r_0$ is replaced by $\mathbf{o}' \leq \mathbf{o}_0$ element-wise.

Proof. **For item 1:** Let $reg_1 = ball(\mathbf{c}_1, r_1)$ and $reg_2 = ball(\mathbf{c}_2, r_2)$. For any ball $ball(\mathbf{c}', r') \subseteq ball(\mathbf{c}_2, r_2)$ with $r' \leq 0.5 \cdot r_1$, we have: $||\mathbf{c}_1 - \mathbf{c}'||_p^p + |r_1 - r'|^p > (0.5 \cdot r_1)^p$. Therefore:

$$d_{\text{dep}}(reg_1, ball(\mathbf{c}', r')) = \frac{||\mathbf{c}_1 - \mathbf{c}'||_p^p + |r_1 - r'|^p}{f(reg_1)f(ball(\mathbf{c}', r'))} > \frac{(0.5 \cdot r_1)^p}{f(reg_1)f(ball(\mathbf{c}', r'))}.$$

Thus, when $f(ball(\mathbf{c}', r')) < \epsilon$ with

$$\epsilon := \frac{(0.5 \cdot r_1)^p}{f(reg_1)[d_{\text{dep}}(reg_1, reg_2) + \Delta]},$$

we have $d_{\text{dep}}(reg_1, ball(\mathbf{c}', r')) > d_{\text{dep}}(reg_1, reg_2) + \Delta$. Since $\lim_{reg\to\emptyset} f(reg) = 0$ by assumption, there exists a $\delta > 0$ such that when $r < \delta$, we have $f(reg) \leq \epsilon$ for any region $reg = ball(\mathbf{c}, r)$. In conclusion, $r_0 = \min\{\delta, 0.5 \cdot r_1\}$ satisfies the required condition.

For item 2: Let $reg = ball(\mathbf{c}, r)$. For any $n, M > 0$, we can select n distinct vectors $\mathbf{c}_1, \ldots, \mathbf{c}_n \in ball(\mathbf{c}, 0.5 \cdot r)$ such that: $\|\mathbf{c}_i - \mathbf{c}_j\|_p^p > \delta > 0$ for some $\delta > 0$.

Similarly to item 1, we can choose r_i small enough such that for any ball $ball(\mathbf{c}_i, r_i)$, we have: $f(ball(\mathbf{c}_i, r_i)) < (\delta/M)^{0.5}$. Thus, $d_{\text{dep}}(ball(\mathbf{c}_i, r_i), ball(\mathbf{c}_j, r_j))$ is bigger than:

$$\frac{\delta}{f(ball(\mathbf{c}_i, r_i))f(ball(\mathbf{c}_j, r_j))} \geq M.$$

The proof for the case of boxes follows the same reasoning, with the radius r replaced by the offset $\mathbf{o}$ or its norm $||\mathbf{o}||$. □

Comparison with the Hyperbolic Distance The following theorem shows that our depth dissimilarity encompasses hyperbolic distance as a special case, obtained with specific choices of functions f, g in Definition 1.

Proposition 1. *Let $\mathbb{H}^{n+1}$ be the hyperbolic space with curvature -1. Assume the ball space $\mathcal{B}^n$ is parameterized as in Example 1, and equipped with the depth dissimilarity defined in Equation (1). Then the map $F\colon \mathcal{B}^n \to \mathbb{H}^{n+1}$ defined by $F(ball(\mathbf{c}, r)) = [\mathbf{c} : r]$ is a* bijective isometry *when $p = 2$, $g(x) = \operatorname{arcosh}(x + 1)$, and $f(ball(\mathbf{c}, r)) = \sqrt{2}\, r$.*

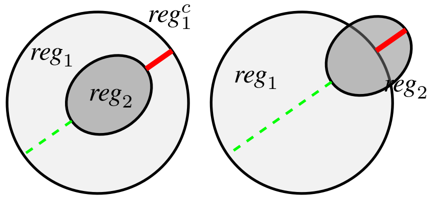

Figure 2: Illustration of $d_{\text{bd}}(reg_1, reg_2)$ (red) for $reg_2 \subseteq reg_1$ (left) or $reg_2 \not\subseteq reg_1$ (right). Green lines show the inverse: $d_{\text{bd}}(reg_2, reg_1)$.

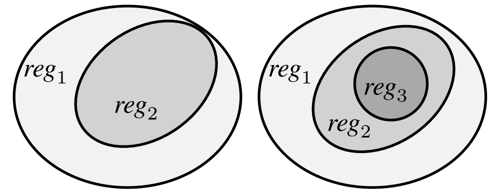

Figure 3: Illustration of internally tangent of item 1 (left) and item 2 (right) in Proposition 3.

Proof. Recall that when the curvature is -1, the distance in the hyperbolic space (half-plane model) is given by:

$$d_k(\mathbf{x}, \mathbf{y}) = \operatorname{arcosh}\left(1 + \frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{2\mathbf{x}_n \mathbf{y}_n}\right).$$

Then, the distance induced by the function F is of the form:

$$d^{\#}_{\mathcal{B}^n}(ball(\mathbf{c}, r), ball'(\mathbf{c}'.r')) = \tag{4}$$

$$\operatorname{arcosh}\left(1 + \frac{\|\mathbf{c} - \mathbf{c}'\|_2^2 + (r - r')^2}{2rr'}\right). \tag{5}$$

This coincides with the depth dissimilarity in Example 1 when $p = 2$, $g(x) = \operatorname{arcosh}(x + 1)$, and $f(ball(\mathbf{c}, r)) = \sqrt{2}\, r$. This completes the proof. □

Moreover, the following result demonstrates that even with a simple linear function $g(\cdot)$, our depth dissimilarity retains the ability to capture the hyperbolic structure. Specifically, there exists a map from the region space to the hyperbolic space that preserves the order of hyperbolic distances for all pairs of points, as stated below.

Proposition 2. *Following Proposition 1, let the depth dissimilarity d_{dep} be redefined by replacing g with $g(x) = k \cdot x\,(k > 0)$. Then the map F retains the following monotonicity property: for any points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \in \mathbb{H}^n$, $d_{\mathbb{H}^n}(\mathbf{x}_1, \mathbf{x}_2) < d_{\mathbb{H}^n}(\mathbf{x}_3, \mathbf{x}_4)$ if and only if*

$$d_{dep}(F^{-1}(\mathbf{x}_1), F^{-1}(\mathbf{x}_2)) < d_{dep}(F^{-1}(\mathbf{x}_3), F^{-1}(\mathbf{x}_4)).$$

Proof. This proposition follows directly from Proposition 1. The function $h(x) = \operatorname{arcosh}(x+1)$ is an increasing bijection from $\mathbb{R}_{\geq 0}$ to $\mathbb{R}_{\geq 0}$. Thus, for any $x, x' \geq 0$, we have:

$$x \leq x' \iff h^{-1}(x) \leq h^{-1}(x').$$

By assumption, $d_{\text{dep}}(\cdot,\cdot) = k \cdot h^{-1}(d_{\mathbb{H}^n}(\cdot,\cdot))$ for some $k > 0$. Therefore, for any points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \in \mathbb{H}^n$, we have: $d_{\mathbb{H}^n}(\mathbf{x}_1, \mathbf{x}_2) < d_{\mathbb{H}^n}(\mathbf{x}_3, \mathbf{x}_4)$ if and only if

$$d_{\text{dep}}(F^{-1}(\mathbf{x}_1), F^{-1}(\mathbf{x}_2)) < d_{\text{dep}}(F^{-1}(\mathbf{x}_3), F^{-1}(\mathbf{x}_4)).$$

□

It is important to recognize that evaluation metrics such as F1 scores and Hits@k rely solely on the ranking induced by the scoring function. As a result, only the relative order of the scores matters, while their absolute values are less important. Therefore, preserving the same ranking as that produced by hyperbolic distances is sufficient to attain comparable performance. Consequently, the depth-based dissimilarity defined in Equations 2 and 3, with the function $g(x) = x$, is well-suited for our implementation, as justified by Proposition 3.

3.2 Boundary Dissimilarity: A Non-Symmetric Measure of Inclusion

Although the depth dissimilarity d_{dep} introduced above has been shown to have great power for embedding hierarchical data, it is a symmetric metric and therefore inadequate for fully capturing the inherently non-symmetric hierarchical relationships between objects. To address this limitation, we introduce the boundary dissimilarity, specifically designed to reflect the partial order of regions defined by set inclusion.

Our boundary dissimilarity generalizes the signed-distance-to-boundary in ShadowCone (Theorem 4.2, (Yu et al. 2024)) by extending its applicability from specialized hyperbolic cone geometries to arbitrary Euclidean regions. This generalization is formally defined in Definition 2. As a result, our boundary dissimilarity can be computed in a much simpler form (Example 3) compared to the formulation used in ShadowCone (Equation 8).

Construction The boundary dissimilarity is defined to measure the minimal cost associated with transforming the spatial relationship between two regions reg_1 and reg_2. Specifically, it quantifies the cost of moving reg_2 out of reg_1 when $reg_2 \subseteq reg_1$, or moving reg_2 into reg_1 otherwise (when $reg_2 \not\subseteq reg_1$). This cost can be defined for arbitrary geometric objects based on either distance or volume within Euclidean or other spaces. Below, we introduce a boundary dissimilarity based on the Euclidean distance for two regions $reg_1, reg_2 \subseteq \mathbb{R}^n$, which consists of two cases:

1. *Containment ($reg_2 \subseteq reg_1$):* As illustrated on the left of Figure 2, when reg_2 is fully contained within reg_1, the boundary dissimilarity is defined by the minimum Euclidean distance between the complementary region reg_1^c and the points in reg_2 (*i.e.*, length of the red line). This distance quantifies the minimum translation cost required to move at least a part of reg_2 out of reg_1.

2. *Non-Containment ($reg_2 \not\subseteq reg_1$):* As shown in the right of Figure 2, when reg_2 is not fully contained within reg_1, the boundary dissimilarity is defined as the maximum Euclidean distance from the points in $reg_2 \setminus reg_1$ to reg_1 (*i.e.*, length of the red line). This distance quantifies the minimum translation cost for moving reg_2 into reg_1.

Let $d(reg, \mathbf{x}) := \min\{||\mathbf{x} - \mathbf{y}||_2 \mid \mathbf{y} \in reg\}$ be the distance of a point $\mathbf{x}$ to a region reg defined by the minimal distance from $\mathbf{x}$ to $\mathbf{y} \in reg$. The formal definition of the boundary dissimilarity is as follows:

Definition 2 (Boundary Dissimilarity)**.** *Given a region space $\mathcal{R}$, we define the boundary dissimilarity over $reg_1, reg_2 \in \mathcal{R}$ by:*

$$d_{bd}(reg_1, reg_2) = \begin{cases} -\min\limits_{\mathbf{x}_2 \in reg_2} \{d(reg_1^c, \mathbf{x}_2)\} & \text{if } reg_2 \subseteq reg_1 \\ \max\limits_{\mathbf{x}_2 \in reg_2 \setminus reg_1} \{d(reg_1, \mathbf{x}_2)\} & \text{else }. \end{cases} \tag{6}$$

Note that a negative sign is added to $d_{\text{bd}}(reg_1, reg_2)$ in the containment case ($reg_2 \subseteq reg_1$) to clearly distinguish it from other cases. Moreover, the boundary dissimilarity is inherently asymmetric, that is, $d_{\text{bd}}(reg_1, reg_2) \neq d_{\text{bd}}(reg_2, reg_1)$ in general. For example, as illustrated in Figure 2, the boundary dissimilarity in the reverse order, $d_{\text{bd}}(reg_2, reg_1)$, corresponds to the length of the green dashed line, which differs from the red line representing $d_{\text{bd}}(reg_1, reg_2)$.

Moreover, for widely used geometric objects such as balls and boxes, the boundary dissimilarity can be computed efficiently using simple arithmetic operations.

Example 3. *If $reg_1 = ball(\mathbf{c}_1, r_1)$ and $reg_2 = ball(\mathbf{c}_2, r_2)$, the boundary dissimilarity has the form (here the two cases can be unified into a single formula):*

$$d_{bd}(reg_1, reg_2) = \|\mathbf{c}_1 - \mathbf{c}_2\|_2 + r_2 - r_1$$

For the case of boxes, where $reg_1 = box(\mathbf{c}_1, \mathbf{o}_1)$ and $reg_2 = box(\mathbf{c}_2, \mathbf{o}_2)$, the boundary dissimilarity has the form:

$$d_{bd}(reg_1, reg_2) = \begin{cases} \max(|\mathbf{c}_1 - \mathbf{c}_2| + \mathbf{o}_2 - \mathbf{o}_1) & \text{if } reg_2 \subseteq reg_1, \\ || \max\{|\mathbf{c}_1 - \mathbf{c}_2| + \mathbf{o}_2 - \mathbf{o}_1, \mathbf{0}\}||_2 & \text{else}. \end{cases}$$

Here, $\max(\cdot)$ in the first line denotes the maximal value along all dimensions, while $\max\{\cdot, \cdot\}$ in the second line applies element-wise to the two vectors or values.

The following proposition demonstrates that our definition of the boundary dissimilarity effectively captures the inclusion relationship between two regions in two key aspects, as illustrated in Figure 3: (i) it identifies whether one region is (exactly) contained within another, and (ii) it enhances discrimination in inclusion chains, as smaller regions tend to have larger boundary dissimilarities. This property is useful for distinguishing shallow children from deeper ones. It is worth noting that the proposition below applies to any regions, not only boxes or balls.

Proposition 3. *For the boundary dissimilarity d_{bd} in Definition 2, the following properties hold:*

1. *$d_{bd}(reg_1, reg_2) \leq 0$ if and only if $reg_2 \subseteq reg_1$. Moreover, $d_{bd}(reg_1, reg_2) = 0$ if and only if reg_1 is* internally tangent *to reg_2. That is, $reg_1 \subseteq reg_2$, and their boundaries intersect at some point.*
2. *$d_{bd}(reg_1, reg_3) \leq d_{bd}(reg_1, reg_2)$ if $reg_3 \subseteq reg_2 \subseteq reg_1$.*

Proof. We prove each item one by one:

1. By definition, we have $d_{\text{bd}}(reg_1, reg_2) \leq 0$ if and only if $reg_2 \subseteq reg_1$. Next, we focus on the case $d_{\text{bd}}(reg_1, reg_2) = 0$.

   Note that if $d_{\text{bd}}(reg_1, reg_2) = 0$, we must have $reg_2 \subseteq reg_1$. Otherwise, we have

   $$d_{\text{bd}}(reg_1, reg_2) = \max_{\mathbf{x}_2 \in reg_2 \setminus reg_1} \{d(reg_1, \mathbf{x}_2)\} = 0$$

   Therefore, for any $\mathbf{x}_2 \in reg_2$, we have $d(reg_1, \mathbf{x}_2) = 0$, therefore $\mathbf{x}_2 \subseteq reg_1$ (assuming reg_1 is a closed set). Contradiction!
   Since $reg_1 \subseteq reg_2$, we have

   $$d_{\text{bd}}(reg_1, reg_2) = \max_{\mathbf{x}_2 \in reg_2} \{-d(reg_1^c, \mathbf{x}_2)\} = 0.$$

   Therefore, there must exist $\mathbf{x}_2 \in reg_2$ such that $d(reg_1^c, \mathbf{x}_2) = 0$, and thus $\mathbf{x}_2 \in reg_1^c$. Since we have $\mathbf{x}_2 \subseteq reg_2 \subseteq reg_1$, therefore, $\mathbf{x}_2 \in \partial(reg_1)$. Similarly, since $\mathbf{x}_2 \subseteq reg_1^c \subseteq reg_2^c$ and $\mathbf{x}_2 \in reg_2$, we also have $\mathbf{x}_2 \in \partial(reg_2)$. This finishes the proof of the first case.
2. By assumption, we have

   $$d_{\text{bd}}(reg_1, reg_2) = \max_{\mathbf{x}_2 \in reg_2} \{-d(reg_1^c, \mathbf{x}_2)\},$$

   $$d_{\text{bd}}(reg_1, reg_3) = \max_{\mathbf{x}_2 \in reg_3} \{-d(reg_1^c, \mathbf{x}_2)\}.$$

   Since $reg_3 \subseteq reg_2$, of course we have $d_{\text{bd}}(reg_1, reg_3) \leq d_{\text{bd}}(reg_1, reg_2)$. This finishes the proof of the second case.

□

Specific Constructions for Boxes or Balls For specific geometric regions like balls and boxes, we can create specialized distance functions to measure the set-inclusion relationship based on their intrinsic geometric properties or established methods. Our framework accommodates these specialized metrics by allowing them to replace the general boundary dissimilarity function.

1. *Volume-based dissimilarity for boxes:* Since the volume of a box can be computed as the product of its offsets along different dimensions, we can define a partial distance based on volume:

$$d_{\text{vol}}(reg_1, reg_2) = -\ln\left(\frac{\text{vol}(reg_1 \cap reg_2)}{\text{vol}(reg_2)}\right). \tag{7}$$

2. *Hyperbolic dissimilarity for balls:* (Yu et al. 2024) introduced a series of circular cones in hyperbolic space and defined a boundary dissimilarity based on the hyperbolic distance between the apex of these cones. By utilizing the natural mapping from balls to circular cones, we can derive a new boundary dissimilarity for balls as follows:

$$d_{\text{bd}}^{\text{cone}}(reg_1, reg_2) = \operatorname{arcsinh}\left(\frac{||\mathbf{c}_1 - \mathbf{c}_2||_2 - r_1}{r_2}\right) + \operatorname{arcsinh}(1) \tag{8}$$

3.3 Training

For a given pair (u, v), we define their energy as a weighted sum with weight λ that balances the contributions of the hyperbolic-like depth dissimilarity:

$$E(u, v) = d_{\text{bd}}(reg_u, reg_v) + \lambda \cdot d_{\text{dep}}(reg_u, reg_v), \tag{9}$$

We say that u is considered a parent of v (i.e., $u \prec v$) if $E(u,v) \le t$, where t is the threshold that achieved the best performance on the evaluation set.

For model training, we use d_{dep} from Equations (2) or (3) with the contrastive loss from (Yu et al. 2024):

$$\mathcal{L}(\gamma_1,\gamma_2) = \sum_{(u,v)\in P} \Bigg(\max\{E(u,v), \gamma_1\} \tag{10}$$

$$+ \log\Bigg(\sum_{(u,v')\in N} e^{\max\{\gamma_2 - d_{\text{bd}}(\mathit{reg}_u, \mathit{reg}_{v'}),0\}} \Bigg)\Bigg),$$

where P and N denote positive and negative sample pairs, respectively. For **positive pairs** $(u,v) \in P$, the loss based on $E(u,v)$ promotes both the containment of reg_v within reg_u (via the d_{bd} term) and their geometric similarity (via the d_{dep} term), whose contributions are controlled by the weight λ and the threshold γ_1. For **negative pairs** $(u,v') \in N$, since our primary goal is to push $\mathit{reg}_{v'}$ outside of reg_u, it is sufficient to use the boundary dissimilarity $d_{\text{bd}}(\mathit{reg}_u, \mathit{reg}_{v'})$ rather than the energy $E(u,v')$. A threshold γ_2 is used to regulate how far $\mathit{reg}_{v'}$ is pushed from reg_u.

4 On Parameterized Regions

4.1 General Parameterized Regions

In this section, we extend the results of Section 3.1 to general region case. A set of **parameterized regions** over Euclidean Space $\mathbb{R}^n$ can be defined as all regions *reg* of the following form:

$$\mathit{reg}(\boldsymbol{\theta}) = \{\mathbf{x} \in \mathbb{R}^n : f_i(\mathbf{x}, \boldsymbol{\theta}) \le 0, \quad \text{for } i = 1,\dots,n\}, \tag{11}$$

where $f_1,\dots,f_n$ are given (differentiable) functions from the given space to $\mathbb{R}$ and $\boldsymbol{\theta} = (\theta_1,\dots,\theta_m) \in \mathbb{R}^m$ is the m-dimensional parameter.

Let $\mathcal{R}$ denote a space of parameterized regions $\mathit{reg}(\boldsymbol{\theta})$ as defined in Equation 11, with $\boldsymbol{\theta} = (\theta_1,\dots,\theta_m) \in \mathbb{R}^m$. For simplicity, and without significant loss of generality, in this section, we assume that all parameterized regions spaces $\mathcal{R}$ satisfy the following properties:

- *Uniqueness:* Two regions coincide if and only if their parameters are equal, i.e., $\mathit{reg}(\boldsymbol{\theta}) = \mathit{reg}(\boldsymbol{\theta}') \iff \boldsymbol{\theta} = \boldsymbol{\theta}'$.
- *Volume-like parameter.* The last coordinate θ_m of $\boldsymbol{\theta}$ serves as a volume parameter such that $\theta_m > 0$ and $\lim_{\mathit{reg}(\boldsymbol{\theta})\to\emptyset} \theta_m = 0$
- *Non-empty regions:* For any $\boldsymbol{\theta} \in \mathbb{R}^m$ with $\theta_m > 0$, the corresponding region $\mathit{reg}(\boldsymbol{\theta})$ is non-empty.
- *Contractible:* Any region can be continuously shrunk to the empty set while remaining in the space $\mathcal{R}$. Formally, for any parameter vector $\boldsymbol{\theta}$, there exists a sequence $\{\boldsymbol{\theta}^k\}_{k=0}^{\infty}$ with $\boldsymbol{\theta}^0 = \boldsymbol{\theta}$ and $\theta_m^k > 0$ for all k, such that $\mathit{reg}(\boldsymbol{\theta}^0) \supset \mathit{reg}(\boldsymbol{\theta}^1) \supset \mathit{reg}(\boldsymbol{\theta}^2) \supset \cdots \to \emptyset$.

It is worth noting that the spaces of balls and boxes in Examples 1 and 2 satisfy all of the above properties, with the volume parameter given by the radius r or the offsets $\mathbf{o}$, respectively.

Moreover, since we assume a positive volume parameter $\theta_m > 0$ under the "volume-like parameter" property, a single point usually does not qualify as a valid region. Consequently, entities such as nominals should be modeled as small regions rather than as points.

We extend Theorem 1 to parameterized regions above as follows.

Theorem 2. *Consider the parameterized region space $\mathcal{R}$ that satisfies assumptions above, and equipped with the depth dissimilarity as defined in Definition 1. The following properties hold:*

1. *For any $\mathit{reg}_1, \mathit{reg}_2 \in \mathcal{R}$ and any $\Delta > 0$, there exists $\mathit{reg}' \subseteq \mathit{reg}_2$ such that*
$$d_{dep}(\mathit{reg}_1, \mathit{reg}) > d_{dep}(\mathit{reg}_1, \mathit{reg}_2) + \Delta, \quad \forall \mathit{reg} \subseteq \mathit{reg}'.$$
2. *For any $\mathit{reg} \in \mathcal{R}$, any integer n, and any $M > 0$, there exist subsets $\mathit{reg}_1,\dots,\mathit{reg}_n \subseteq \mathit{reg}$ such that for any distinct $i,j \in \{1,\dots,n\}$, $d_{dep}(\mathit{reg}_i, \mathit{reg}_j) > M$.*

Outline of proof. We sketch the main ideas, which closely follow the argument in the proof of Theorem 1.

For the first statement, by selecting a sufficiently small subset $\mathit{reg}' \subseteq \mathit{reg}_2$, we can ensure that

$$\|P(\mathit{reg}) - P(\mathit{reg}_1)\| > \delta$$

for some $\delta > 0$. Additionally, since $f(\mathit{reg}')$ can be made arbitrarily small as $\mathit{reg}' \to \emptyset$ (i.e., $\lim_{\mathit{reg}\to\emptyset} f(\mathit{reg}) = 0$), we can make the numerator of the depth dissimilarity arbitrarily small while keeping the denominator bounded below by a positive constant. Consequently, the depth dissimilarity $d_{\text{dep}}(\mathit{reg}_1, \mathit{reg}')$ can be made arbitrarily large.

For the second statement, a similar argument applies: by choosing the subsets $\mathit{reg}_1,\dots,\mathit{reg}_n$ to be sufficiently small and mutually separated within *reg*, we can ensure that each $f(\mathit{reg}_i)$ is small enough so that the depth dissimilarity between any pair $(\mathit{reg}_i, \mathit{reg}_j)$ exceeds M.

□

Propositions 1 and 3 are extended to arbitrary regions as follows.

Proposition 4. *Let $\mathbb{H}^m$ denote the hyperbolic space with curvature -1. Assume that $\mathcal{R}$ is a collection of parameterized regions satisfying the above assumptions, equipped with the depth dissimilarity defined in Equation equation 1. Then the map $F\colon \mathcal{R} \to \mathbb{H}^m$ defined by $F(\boldsymbol{\theta}) = \boldsymbol{\theta}$ is a* bijective isometry *between $\mathcal{R}$ and $\mathbb{H}^m$ when $p = 2$, $g(x) = \operatorname{arcosh}(x+1)$, and $f(\mathit{reg}(\boldsymbol{\theta})) = \sqrt{2}\,\theta_m$.*

Proof. Under the given condition, the distance in the $\mathbb{H}^{n+1}$ is given by:

$$d_k(\mathbf{x},\mathbf{y}) = \operatorname{arcosh}\left(1 + \frac{\|\mathbf{x}-\mathbf{y}\|_2^2}{2\mathbf{x}_n\mathbf{y}_n}\right).$$

Then, the distance induced by the function F is of the form:

$$d^{\#}_{\mathcal{R}}(\mathit{reg}(\boldsymbol{\theta}), \mathit{reg}(\boldsymbol{\theta'})) = \operatorname{arcosh}\left(1 + \frac{\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|_2^2}{2\theta_m\theta'_m}\right).$$

This coincides with the depth dissimilarity in Equation (1) when $p = 2$, $g(x) = \operatorname{arcosh}(x+1)$, and $f(\mathit{reg}(\boldsymbol{\theta})) = \sqrt{2}\,\theta_m$.

□

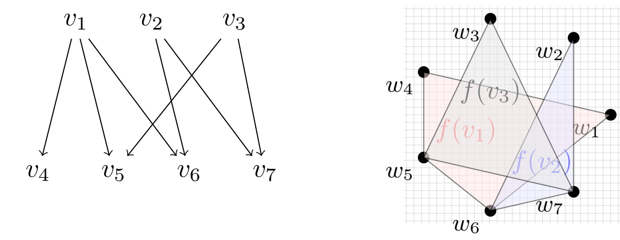

Figure 4: Example of a partially ordered set (left, $v \to v'$ means $v \prec v'$) and its embedding (right), with $S_{v_1} = \{w_1, w_4, w_5, w_6\}$, $S_{v_2} = \{w_2, w_6, w_7\}$, and $S_{v_3} = \{w_3, w_5, w_7\}$.

Proposition 5. *Following Proposition 4, let the depth dissimilarity d_{dep} be redefined by replacing g to $g(x) = k \cdot x$ ($k > 0$). Then the map F retains the following monotonicity property: for any points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \in \mathbb{H}^n$, $d_{\mathbb{H}^n}(\mathbf{x}_1, \mathbf{x}_2) < d_{\mathbb{H}^n}(\mathbf{x}_3, \mathbf{x}_4)$ if and only if*

$$d_{dep}(F^{-1}(\mathbf{x}_1), F^{-1}(\mathbf{x}_2)) < d_{dep}(F^{-1}(\mathbf{x}_3), F^{-1}(\mathbf{x}_4)).$$

Proof. The proof is the same as that of Proposition 2 because the function $h(x) = \operatorname{arcosh}(x+1)$ is an increasing bijection from $\mathbb{R}_{\geq 0}$ to $\mathbb{R}_{\geq 0}$. $\square$

4.2 Faithfulness

In this section, we show that embeddings based on arbitrary regions in the plane (i.e., $\mathbb{R}^2$) are *faithful*, that is, they capture the hierarchy accurately, as detailed in Definition 4.1. This result significantly strengthens existing theoretical guarantees. In particular, prior work shows that box embeddings can represent any DAG using $O((\Delta + 2) \log n)$ dimensions, where Δ denotes the maximum degree and n the number of nodes (Proposition 3 in (Boratko et al. 2021)). In contrast, we prove that two dimensions already suffice when arbitrary regions are allowed.

Definition 4.1 (Faithful embedding)**.** Let $(V, \prec)$ be a partially ordered set. An *n-dimensional embedding* f of V that assigns to each element $v \in V$ a region $f(v) \subseteq \mathbb{R}^n$ is said to be *faithful* if, for all $v \neq v' \in V$,

$$v \prec v' \iff f(v) \subseteq f(v').$$

Theorem 4.2. *For any partially ordered set $(V, \prec)$, there exists a 2-dimensional faithful embedding of V.*

Proof. Let $V = \{v_1, \ldots, v_m\}$. We construct a two-dimensional embedding as follows (see Figure 4):

1. Embed the elements v_i as distinct points w_i placed at the vertices of a regular m-gon in $\mathbb{R}^2$.
2. For each $v_i \in V$, let $S_{v_i} := \{w_i\} \cup \{\, w_j \mid v_j \prec v_i \,\}$. We define $f(v_i)$ to be the convex hull of the points in S_{v_i}.

The embedding is faithful as follows: For any $v_1 \neq v_2 \in V$: **(⇒)** if $v_1 \prec v_2$, then by transitivity of the partial order, $S_{v_1} \subseteq S_{v_2}$. Since the convex hull operator is monotone with respect to set inclusion, we obtain $f(v_1) \subseteq f(v_2)$. **(⇐)** if $f(v_1) \subseteq f(v_2)$. By our construction based on convex hull of n-gons, $w_1 \in f(v_2)$ implies $w_1 \in S_{v_2}$. By definition of S_{v_2}, this yields $v_1 \prec v_2$. $\square$

The construction above embeds each node as a polyhedral region with a variable number of vertices, which is generally impractical for real-world applications. In practice, one usually learns embeddings through a training procedure that restricts regions to a fixed parametric family as in Section 3.3.

5 Evaluation

Our experiments aim to address two questions: 1. How effectively do our methods capture hierarchical relationships? 2. Can they generalize to tasks involving more than hierarchies?

We evaluate two types of reasoning behavior. First, in Section 5.1, we examine transitivity reasoning over directed acyclic graphs (DAGs), focusing on transitive relations such as the “is-a” relation in the Noun dataset. Second, in Sections 5.2 and 5.3, we study entailment reasoning over ontologies, specifically for inferring concept subsumptions. For both tasks, we use F1-score as the primary metric for logical reasoning performance, assessing the accuracy of inferred results under the embedding setting.

5.1 Inferences over DAG

Benchmark Following (Yu et al. 2024), we evaluate our method on four real-world datasets consisting of Is-A relations: MCG (Wang et al. 2015; Wu et al. 2012), Hearst patterns (Hearst 1992), the WordNet (Fellbaum 1998) noun taxonomy, and its mammal subgraph. All models are trained exclusively on *basic edges*, which are edges not implied transitively by other edges in the graph. For validation and testing, we use the same sets as in (Yu et al. 2024), consisting of 5% of *non-basic (inferred)* edges, ensuring a fair comparison. We exclude non-basic edges from training since they can be transitively derived from basic edges. Including them would artificially inflate performance metrics without properly evaluating the embeddings' ability to capture hierarchical structures.

Baselines We compare our method RegD with (i) hyperbolic approaches such as EntailmentCone (Ganea, Bécigneul, and Hofmann 2018) and ShadowCone (Yu et al. 2024), which is the most recent state-of-the-art method; (ii) region-based methods like OrderEmbedding (Bordes et al. 2013), and τBox (Boratko et al. 2021). We also compare with the ontology embedding methods, ELBE (Peng et al. 2022) and ELEM (Kulmanov et al. 2019), which can be considered baseline approaches that embed the DAG as boxes or balls, respectively. We use F1-scores as in previous studies.

It is worth noting that some prior works, such as RotL and Rot2L (Wang et al. 2021), also simplify operations in hyperbolic space. However, they focus on Möbius addition and address link prediction in knowledge graphs. Therefore, we do not compare with these methods.

Results The performance comparison across different DAGs is shown in Table 2. RegD achieved the best performance on all four datasets. Notably, the box variant consistently outperformed the ball variant in most cases, which might be because boxes contain more parameters than balls when embedded in the same dimensional space. Interestingly,

Method		Mammal		Noun		MCG		Hearst	
		d=2	d=5	d=5	d=10	d=5	d=10	d=5	d=10
τBox		29.0	33.5	30.5	31.5	43.9	50.3	39.7	43.7
OE		25.4	31.0	28.8	30.8	36.3	46.6	34.6	40.7
ELBE (box baseline)		30.3	36.8	30.7	31.8	48.4	55.5	41.6	46.8
ELEM (ball baseline)		27.7	28.8	28.6	29.5	35.7	38.6	34.6	36.7
EntailmentCone*		54.4	56.3	29.2	32.1	25.3	25.5	22.6	23.7
ShadowCone*	(Umbral-half)	57.7	69.4	45.2	52.2	36.8	40.1	32.8	32.6
	(Penumbral-half)	52.8	67.8	44.6	51.7	35.0	37.6	26.8	28.4
RegD	(box)	**64.9**	71.6	53.8	51.3	**50.7**	**58.5**	**42.8**	**49.6**
	(ball)	62.7	**71.8**	**58.4**	**59.1**	44.9	46.8	37.7	37.7

Table 2: F1 score (%) on Mammal, WordNet noun, MCG, and Hearst. Results with * are coming from ShadowCones.

Method	GALEN		GO		ANATOMY	
	d=5	d=10	d=5	d=10	d=5	d=10
ELBE	20.7	21.2	36.9	42.4	43.1	43.0
+ τBox	20.8	20.7	32.2	34.7	42.2	47.2
+ RegD	**25.2**	**25.8**	**50.0**	**50.5**	**58.7**	**62.5**
ELEM	16.9	17.3	23.5	27.4	34.6	38.7
+ RegD	**19.2**	**18.8**	**36.4**	**40.0**	**52.5**	**55.5**

Table 3: F1 score (%) for the inference task.

Method	GALEN		GO		ANATOMY	
	d=5	d=10	d=5	d=10	d=5	d=10
ELBE	**21.0**	**24.2**	32.8	37.9	25.1	25.5
+ τBox	18.4	19.5	24.0	29.3	25.6	22.9
+ RegD	20.6	21.0	**37.1**	**44.1**	**41.4**	**45.3**
ELEM	16.7	16.6	54.5	54.2	**23.1**	**26.0**
+ RegD	**16.8**	**18.0**	**60.3**	**61.4**	21.7	21.9

Table 4: F1 score (%) for the prediction task.

region-based methods outperformed hyperbolic methods on the MCG and Hearst datasets. However, on the Noun and Mammal datasets, hyperbolic methods performed better. Nevertheless, our method performed consistently well in both cases, as it can adjust the hyperbolic component by setting different λ values in Equation (9).

5.2 Inference and Prediction over Ontologies

Benchmark We utilize three normalized biomedical ontologies: GALEN (Rector, Rogers, and Pole 1996), Gene Ontology (GO) (Ashburner et al. 2000), and Anatomy (Uberon) (Mungall et al. 2012). As in (Jackermeier, Chen, and Horrocks 2024), we use the entire ontology for training, and the complete set of inferred class subsumptions for testing. Those subsumptions can be regarded as partial order pairs $u \prec v$. Evaluation is performed using 1,000 subsumptions randomly sampled from the test set. Similar to inference over DAG, negative samples are generated by randomly replacing the child of each positive pair 10 times.

Baselines We consider two representative geometric, model-based ontology embedding methods: ELBE (Peng et al. 2022) and ELEM (Kulmanov et al. 2019), which embed ontology concepts as balls and boxes, respectively. Those methods can be enhanced by RegD (or τBox in the box case) by incorporating their corresponding energy functions to model inclusion constraints between geometric objects. Other hierarchy embedding methods are excluded from our evaluation due to their incompatibility with ontology embedding frameworks. For instance, OE represents concepts using cones, which cannot be directly integrated with ELBE or ELEM for ontology-specific reasoning tasks.

Results The results are summarized in Tables 3. We can see that RegD yields consistent improvements across all ontologies for inference tasks. Specifically, it gains an F1 score increase of more than 45% with ELBE method on the ANATOMY ontology. Conversely, the τBox plugin consistently reduces performance across nearly all test cases, underscoring its limited applicability to tasks involving more than hierarchies.

5.3 Prediction over Ontologies

We use the same baselines and datasets as described in Section 5.2. However, in this prediction task, we partition the original ontologies directly into 80% for training, 10% for validation, and 10% for testing as in (Jackermeier, Chen, and Horrocks 2024). For the link prediction task, we focus on specific parts of the validation and testing sets, represented as $\exists r.B \prec ?A$, where A and B are concept names and r is a role. This setup is equivalent to link prediction tasks $(?A, r, B)$ in knowledge graphs if we regard A, B, and r as the head entity, tail entity, and relation, respectively.

Results Table 3 summarizes the results. RegD shows mixed results: while it generally improves performance, it decreases scores on the GALEN ontology and ANATOMY ontology with ELEM. This degradation likely occurs in challenging prediction cases where all methods perform poorly. Nevertheless, RegD achieves significant improvements in other cases, notably increasing F1 score by 77.6% with ELBE on the ANATOMY ontology. In contrast, the τBox plugin consistently reduces performance across almost all test cases.

6 Conclusion

We introduced a framework RegD for low-dimensional embeddings of hierarchies, leveraging two dissimilarity metrics between regions. Our method, applicable to regions in the Euclidean space, demonstrates versatility and has the potential for a wide range of tasks involving data beyond hierarchies. Additionally, we showed that our approach achieves comparable embedding performance to hyperbolic methods while being significantly simpler to implement.

For future work, we are interested in generalizing our framework to support a wider variety of region types beyond the balls and boxes considered here, so that it can be better adapted to the requirements of diverse downstream tasks. Another promising direction for future work is to investigate replacing existing hyperbolic embedding components in related methods with our approach, including recent studies that couple large language models with hyperbolic embeddings to learn semantic hierarchies (He et al. 2024b), which can be extended from atomic concept to complex ones by integrating the ontology embeddings as in Section 5.2.

AI Declaration

During the preparation of this work, the authors used AI to improve the work's readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgements

This work is funded by EPSRC projects OntoEm (EP/Y017706/1).

References

Abboud, R.; Ceylan, İ. İ.; Lukasiewicz, T.; and Salvatori, T. 2020. Boxe: A box embedding model for knowledge base completion. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; et al. 2000. Gene ontology: tool for the unification of biology. *Nature genetics* 25(1):25–29.

Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.

Balazevic, I.; Allen, C.; and Hospedales, T. M. 2019. Multi-relational poincaré graph embeddings. In Wallach, H. M.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E. B.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, 4465–4475.

Bécigneul, G., and Ganea, O. 2019. Riemannian adaptive optimization methods. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Boratko, M.; Zhang, D.; Monath, N.; Vilnis, L.; Clarkson, K. L.; and McCallum, A. 2021. Capacity and bias of learned geometric embeddings for directed graphs. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y. N.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, 16423–16436.

Bordes, A.; Usunier, N.; García-Durán, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In Burges, C. J. C.; Bottou, L.; Ghahramani, Z.; and Weinberger, K. Q., eds., *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*, 2787–2795.

Chen, J.; Hu, P.; Jiménez-Ruiz, E.; Holter, O. M.; Antonyrajah, D.; and Horrocks, I. 2021. Owl2vec*: embedding of OWL ontologies. *Mach. Learn.* 110(7):1813–1845.

Chen, J.; Mashkova, O.; Zhapa-Camacho, F.; Hoehndorf, R.; He, Y.; and Horrocks, I. 2024. Ontology embedding: A survey of methods, applications and resources. *CoRR* abs/2406.10964.

Dasgupta, S. S.; Boratko, M.; Zhang, D.; Vilnis, L.; Li, X.; and McCallum, A. 2020. Improving local identifiability in probabilistic box embeddings. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Fellbaum, C. 1998. Wordnet: An electronic lexical database. *MIT Press google schola* 2:678–686.

Ganea, O.; Bécigneul, G.; and Hofmann, T. 2018. Hyperbolic entailment cones for learning hierarchical embeddings. In Dy, J. G., and Krause, A., eds., *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*, 1632–1641. PMLR.

He, Y.; Chen, J.; Dong, H.; Horrocks, I.; Allocca, C.; Kim, T.; and Sapkota, B. 2024a. Deeponto: A python package for ontology engineering with deep learning. *Semantic Web* 15(5):1991–2004.

He, Y.; Yuan, Z.; Chen, J.; and Horrocks, I. 2024b. Language models as hierarchy encoders. In *Advances in Neural Information Processing Systems (NeurIPS)*. Accepted.

Hearst, M. A. 1992. Automatic acquisition of hyponyms from large text corpora. In *14th International Conference on Computational Linguistics, COLING 1992, Nantes, France, August 23-28, 1992*, 539–545.

Jackermeier, M.; Chen, J.; and Horrocks, I. 2024. Dual box embeddings for the description logic el^{++}. In Chua, T.; Ngo, C.; Kumar, R.; Lauw, H. W.; and Lee, R. K., eds., *Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024*, 2250–2258. ACM.

Kulmanov, M.; Liu-Wei, W.; Yan, Y.; and Hoehndorf, R. 2019. EL embeddings: Geometric construction of models for the description logic EL++. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 6103–6109. International Joint Conferences on Artificial Intelligence Organization.

Lee, J. M. 2006. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer Science & Business Media.

Lee, J. 2013. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer Science & Business Media.

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. In Bengio, Y., and LeCun, Y., eds., *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*.

Mungall, C. J.; Torniai, C.; Gkoutos, G. V.; Lewis, S. E.; and Haendel, M. A. 2012. Uberon, an integrative multi-species anatomy ontology. *Genome biology* 13:1–20.

Nickel, M., and Kiela, D. 2017. Poincaré embeddings for learning hierarchical representations. In Guyon, I.; von Luxburg, U.; Bengio, S.; Wallach, H. M.; Fergus, R.; Vishwanathan, S. V. N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 6338–6347.

Peng, X.; Tang, Z.; Kulmanov, M.; Niu, K.; and Hoehndorf, R. 2022. Description logic EL++ embeddings with intersectional closure.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In Moschitti, A.; Pang, B.; and Daelemans, W., eds., *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, 1532–1543. ACL.

Rector, A. L.; Rogers, J. E.; and Pole, P. 1996. The galen high level ontology. In *Medical Informatics Europe'96*. IOS Press. 174–178.

Sala, F.; Sa, C. D.; Gu, A.; and Ré, C. 2018. Representation tradeoffs for hyperbolic embeddings. In Dy, J. G., and Krause, A., eds., *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*, 4457–4466. PMLR.

Shen, J., and Han, J. 2022. *Automated Taxonomy Discovery and Exploration*. Synthesis Lectures on Data Mining and Knowledge Discovery. Springer.

Shi, J.; Dong, H.; Chen, J.; Wu, Z.; and Horrocks, I. 2024. Taxonomy completion via implicit concept insertion. In Chua, T.; Ngo, C.; Kumar, R.; Lauw, H. W.; and Lee, R. K., eds., *Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024*, 2159–2169. ACM.

Sonthalia, R., and Gilbert, A. C. 2020. Tree! I am no tree! I am a low dimensional hyperbolic embedding. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Sun, Z.; Deng, Z.; Nie, J.; and Tang, J. 2019. 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.

Suzuki, R.; Takahama, R.; and Onoda, S. 2019. Hyperbolic disk embeddings for directed acyclic graphs. In Chaudhuri, K., and Salakhutdinov, R., eds., *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, 6066–6075. PMLR.

Vendrov, I.; Kiros, R.; Fidler, S.; and Urtasun, R. 2016. Order-embeddings of images and language. In Bengio, Y., and LeCun, Y., eds., *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

Vilnis, L.; Li, X.; Murty, S.; and McCallum, A. 2018. Probabilistic embedding of knowledge graphs with box lattice measures. In Gurevych, I., and Miyao, Y., eds., *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, 263–272. Association for Computational Linguistics.

Wang, Z.; Wang, H.; Wen, J.; and Xiao, Y. 2015. An inference approach to basic level of categorization. In Bailey, J.; Moffat, A.; Aggarwal, C. C.; de Rijke, M.; Kumar, R.; Murdock, V.; Sellis, T. K.; and Yu, J. X., eds., *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*, 653–662. ACM.

Wang, K.; Liu, Y.; Lin, D.; and Sheng, M. 2021. Hyperbolic geometry is not necessary: Lightweight euclidean-based models for low-dimensional knowledge graph embeddings. In Moens, M.; Huang, X.; Specia, L.; and Yih, S. W., eds., *Findings of the Association for Computational Linguistics: EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 16-20 November, 2021*, 464–474. Association for Computational Linguistics.

Wu, W.; Li, H.; Wang, H.; and Zhu, K. Q. 2012. Probase: a probabilistic taxonomy for text understanding. In Candan, K. S.; Chen, Y.; Snodgrass, R. T.; Gravano, L.; and Fuxman, A., eds., *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2012, Scottsdale, AZ, USA, May 20-24, 2012*, 481–492. ACM.

Yang, H.; Chen, J.; He, Y.; Gao, Y.; and Horrocks, I. 2025. Language models as ontology encoders. In *International Semantic Web Conference*, 443–461. Springer.

Yang, H.; Chen, J.; and Sattler, U. 2025. Transbox: EL^{++}-closed ontology embedding. In *THE WEB CONFERENCE 2025*.

Yu, T.; Liu, T. J. B.; Tseng, A.; and Sa, C. D. 2024. Shadow cones: A generalized framework for partial order embeddings. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.

Zhang, D.; Boratko, M.; Musco, C.; and McCallum, A. 2022. Modeling transitivity and cyclicity in directed graphs via binary code box embeddings. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.