HYPERBOLIC GRAVITATIONAL CONE: A PHYSICALLY-INSPIRED APPROACH FOR PARTIAL ORDER EMBEDDINGS

Anonymous authors
Paper under double-blind review

ABSTRACT

Partial-order modeling is essential for capturing asymmetric entailment semantics in knowledge graph reasoning and multimodal hierarchical alignment, which necessitates an inductive bias capable of strictly preserving transitivity. While hyperbolic entailment cones provide rigorous geometric guarantees for this property, their efficacy is limited by two constraints: geometric rigidity and optimization stagnation. Specifically, isotropic apertures mask semantic heterogeneity, and binary objectives sever gradient flow for entailed pairs. To address these issues, we propose Hyperbolic Gravitational Cone (HGC), inspired by the physical principle that mass governs semantic influence. The HGC formulates the entailment scope as a dynamic gravitational field, yielding a closed-form cone aperture modulated jointly by a learnable anisotropic mass and radial depth. To ensure hierarchical consistency, this mass function is strictly governed by positivity, radial monotonicity, convexity, and directional anisotropy. Furthermore, a continuous gravitational energy is derived to replace rigid binary constraints. By employing a soft-gated mechanism to evaluate angular deviation and coupling it with parental mass and axial distance, this energy function provides informative gradients even within the cone interior, driving precise structural optimization. Crucially, radial transitivity is mathematically guaranteed by imposing strict monotonicity on mass decay, preserving logical nesting along any given radial direction. Empirical evaluations demonstrate that HGC achieves superior performance on standard partial-order benchmarks, where the learned mass distributions reveal the underlying structure of semantic hierarchies and quantify the intrinsic breadth of concepts.

1 INTRODUCTION

Partial order relations constitute the structural backbone of knowledge representation, providing the essential logical scaffolding that governs the asymmetric entailment semantics ubiquitous in ontologies, taxonomies, and directed acyclic graphs (Ren et al., 2020; Ma et al., 2021; Dhall et al., 2020; Song et al., 2025; Kulmanov et al., 2019; Vilnis et al., 2018; Wang et al., 2023; Berg et al., 2025; Sastry et al., 2025; Chen et al., 2025; Li et al., 2023; Wang et al., 2021; Sinha et al., 2024). Fundamentally distinct from proximity-based association relationships, partial orders impose a rigorous hierarchical topology that dictates the unidirectional flow of semantic generality. Encoding partial-order constraints establishes the intrinsically correct problem formulation that aligns with the data's true logic, rather than merely serving as an auxiliary tool to artificially boost model expressiveness. To achieve this structural alignment, geometric relational embeddings (Xiong et al., 2023; Peng et al., 2022) inject the necessary inductive bias by utilizing spatially extended structures such as cones or boxes. By explicitly enforcing containment logic, these models enable representation spaces that are naturally isomorphic to the partial-order structures they aim to capture.

However, the efficacy of geometric embeddings is inextricably linked to the underlying manifold (Nickel & Kiela, 2017; Bronstein et al., 2017; Gu et al., 2019). Data governed by partial-order relations inherently exhibits exponential expansion, creating a fundamental curvature mismatch with flat Euclidean space. As illustrated in Fig. 1(a), the number of entailed entities grows exponentially with depth, whereas Euclidean volume expands only polynomially. This capacity bottleneck precipitates a catastrophic representational collapse: deep nodes are forced to crowd at the periphery (Fig:1(b)),

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

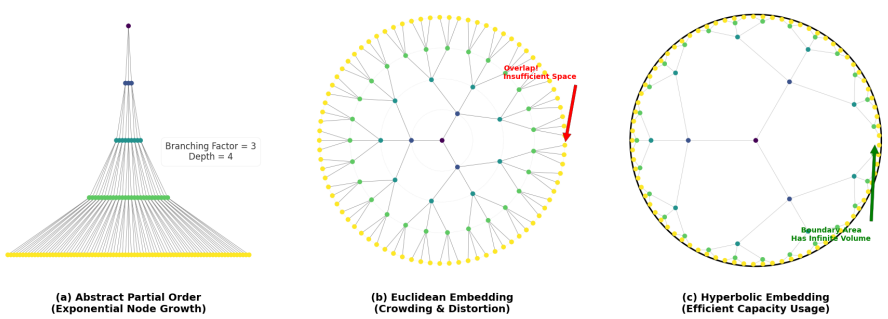


Figure 1: Geometric Capacity Comparison. (a) Intrinsic Structure: Partial-order data exhibits an exponentially growing tree structure. (b) Euclidean Distortion: In Euclidean space, volume grows polynomially (r^d). To accommodate exponential nodes, the distance between nodes at deeper levels must shrink drastically, leading to crowding and overlap (marked in red). (c) Hyperbolic Fit: In the Poincaré disk, volume grows exponentially (e^r). Although nodes appear crowded near the boundary (Euclidean view), their intrinsic hyperbolic distances remain large and constant.

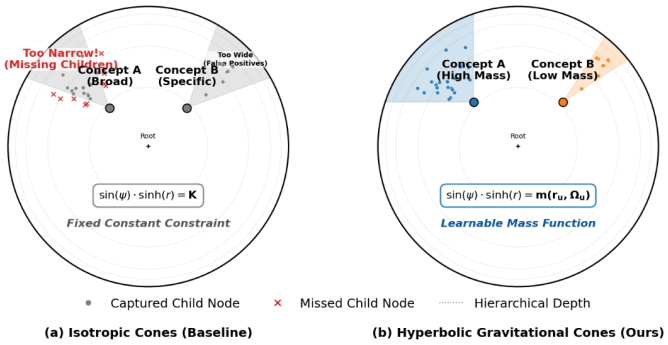


Figure 2: The Geometric Rigidity Bottleneck and Our Solution. (a) Isotropic Entailment Cones: Existing methods rely on a rigid geometric constraint $\sin(\psi)\sinh(r) = K$, where the aperture ψ is strictly determined by radial depth r. (b) Hyperbolic Gravitational Cone (Ours): We introduce a physics-inspired formulation $\sin(\psi)\sinh(r) = m(r_u, \Omega_u)$, where the aperture is modulated by a learnable anisotropic mass $m(r_u, \Omega_u)$. This allows high-mass concepts to generate wide gravitational fields capturing all descendants, while low-mass concepts contract their apertures for precision.

obliterating the separability required for fine-grained concepts. Consequently, hyperbolic manifolds emerge as the optimal geometric solution (Fig. 1(c)). Their exponential volume expansion aligns perfectly with the intrinsic growth of hierarchical trees, leveraging the vast capacity near the boundary to preserve partial-order integrity with arbitrarily low distortion (Sarkar, 2012).

To fully exploit this manifold advantage, recent advances have shifted from point-wise representations to region-based hyperbolic modeling. The Hyperbolic Entailment Cones (HEC) (Ganea et al., 2018; Le et al., 2019) pioneered this direction by formalizing entailment through angular containment, where a parent's conical aperture is strictly determined by its radial depth. Generalizing this formalism through a physical lens, Shadow Cones (Yu et al., 2024; Tseng et al., 2023) reinterpreted the structure as shadows cast by opaque, fixed-radius objects from a light source. The primary success of these methods lies in their capacity to translate logical partial orders into strict geometric containment constraints within a negatively curved space.

Despite their ability to capture basic hierarchical depth, these approaches are hindered by dual limitations: geometric rigidity and optimization stagnation. First, the representations are structurally rigid. Existing methods impose rotational invariance, where cone apertures are defined solely by radial depth. Even physics-inspired approaches like Shadow Cones (Yu et al., 2024) oversimplify concepts into fixed-size proxies, lacking an intrinsic definition of "mass" to faithfully capture heterogeneous semantic influence. As illustrated in Fig. 2(a), this rigidity forces concepts at comparable hierarchical

depths (e.g., the broad *"Animal"* and specific *"Artifact"*) to adopt identical entailment scopes. Ignoring intrinsic directional variability fails to allocate sufficient geometric volume to broad semantic families, creating a semantic resolution bottleneck: the aperture is either too narrow to encompass rich entailments (causing false negatives) or too wide for specific ones (inviting false positives). Second, the optimization landscape is ill-posed. Current objectives rely on binary containment constraints, assigning zero energy to any valid entailment regardless of its interior position. This mechanism severs the gradient flow inside the cone, preventing the model from distinguishing between marginal and robust entailments, thereby hindering the fine-grained refinement of the hierarchy.

To overcome these geometric and optimization barriers, we propose **Hyperbolic Gravitational Cone (HGC)**, a physically-inspired approach that reimagines entailment as a continuous gravitational potential field. In HGC, each node exerts a semantic influence determined by a learnable, anisotropic mass. As illustrated in Fig. 2(b), this mass function dynamically modulates the cone aperture based jointly on radial depth and semantic direction, allowing the model to adaptively expand for broad concepts or contract for specific ones. This capability enables HGC to capture heterogeneous semantic breadth that rigid geometric priors cannot represent. Furthermore, replacing the hard zero-energy interior, our formulation introduces a graded potential well. This design guarantees informative, non-zero gradients everywhere, including for entailed pairs, driving the adaptive refinement of the cone structure while strictly preserving radial transitivity.

Our main contributions are summarized as follows:

- **Anisotropic Mass Framework:** We introduce a direction-aware mass function governed by Möbius transformations. This mechanism allows cone apertures to vary continuously across semantic directions, overcoming isotropy limitations while satisfying the critical structural properties of monotonicity and convexity to guarantee radial transitivity.
- **Gravitational Potential Optimization:** We derive a differentiable energy function where the potential magnitude is scaled by the hypernym's mass. Replacing rigid binary constraints with a soft-gated potential well eliminates vanishing gradients within the cone interior. This objective enforces semantic centrality, actively pulling hyponyms towards the gravitational axis of their hypernyms to ensure robust and precise entailment.
- **State-of-the-Art Performance:** Extensive experiments demonstrate that HGC achieves state-of-the-art results. Notably, it exhibits exceptional sample efficiency in low-resource regimes and strong numerical stability in low-dimensional spaces.

2 Background: Hyperbolic Geometry and Entailment Cones

In this section, we establish the essential geometric concepts of hyperbolic space and review the Hyperbolic Entailment Cone (Ganea et al., 2018), which serves as the architectural foundation for our dynamic gravitational extension.

Hyperbolic Space $\mathbb{H}_c^n$ is the unique simply connected Riemannian manifold with constant negative curvature $-c(c > 0)$ (Cannon et al., 1997). In this work, we operate within the Poincaré ball model due to its conformal property, which preserves Euclidean angles and thus facilitates intuitive cone aperture definitions.

The Poincaré Ball Model. The n-dimensional Poincaré ball $(\mathbb{D}_c^n, g^{\mathbb{D}})$ is defined as the manifold $\mathbb{D}_c^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < 1/\sqrt{c}\}$ equipped with the Riemannian metric tensor $g_x^{\mathbb{D}} = \lambda_x^2 g^{\mathbb{E}}$, where the conformal factor $\lambda_x := \frac{2}{1-c\|\mathbf{x}\|^2}$ and $g^{\mathbb{E}}$ is the Euclidean metric tensor. The induced geodesic distance between two points $\mathbf{x}, \mathbf{y} \in \mathbb{D}_c^n$ is defined as:

$$d_{\mathbb{D}}(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{c}} \cosh^{-1}\left(1 + \frac{2c\|\mathbf{x} - \mathbf{y}\|^2}{(1 - c\|\mathbf{x}\|^2)(1 - c\|\mathbf{y}\|^2)}\right), \tag{1}$$

where $\|\cdot\|$ denotes the standard Euclidean norm. A comprehensive overview of foundational hyperbolic geometry concepts is provided in Appendix B.

Hyperbolic Entailment Cones. Ganea et al.Ganea et al. (2018) model partial orders $\mathbf{v} \preceq \mathbf{u}$ by associating each node u with a conical region $\mathcal{C}(\mathbf{u}) = \{\mathbf{v} \in \mathbb{D}_c^n | \Xi(\mathbf{u}, \mathbf{v}) \leq \psi(\mathbf{u})\}$. As illustrated in Fig. 5, the containment condition relies on comparing the deviation angle $\Xi(\mathbf{u}, \mathbf{v})$ with the cone half-

aperture $\psi(\mathbf{u})$. Detailed derivations of deviation angular and optimal aperture bounds are deferred to Appendix C.

3 HYPERBOLIC GRAVITATIONAL CONE

The Hyperbolic Gravitational Cone projects discrete partial-order relations into a continuous hyperbolic manifold. Fundamentally, it reimagines hierarchical entailment not as static geometric containment, but as a dynamic interactions within a gravitational potential field.

3.1 OVERALL FRAMEWORK

As illustrated in Fig. 3, the proposed model orchestrates this physical-geometric interaction through a pipeline comprising three stages: **(1) Hyperbolic Encoding:** Discrete concepts are projected into the Poincaré ball via a hyperbolic encoder, transforming symbolic data into a continuous, curvature-aware manifold. **(2) Anisotropic Mass Modeling:** A core innovation of HGC is the generation of a dynamic entailment boundary. As detailed in Secs. 3.2 and 3.3, each node is assigned a learnable anisotropic mass that physically dictates its cone aperture, thereby explicitly encoding the concept's semantic breadth. **(3) Gravitational Energy Optimization:** The system is optimized end-to-end by minimizing a gravitational potential energy (Sec. 3.4). This objective pulls child nodes into the gravitational cone of their parents while strictly enforcing physical constraints through regularization.

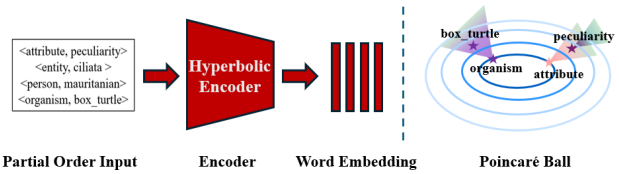


Figure 3: The Overall Framework of Hyperbolic Gravitational Cone.

3.2 GEOMETRIC DEFINITION OF HYPERBOLIC GRAVITATIONAL CONE

For a given partial-order pair $v \preceq u$, the discrete concepts are first projected into continuous coordinates $\mathbf{u}, \mathbf{v} \in \mathbb{D}_c^n$ via a hyperbolic embedding layer. To mathematically formulate the entailment region $\mathcal{C}_{grav}(\mathbf{u})$ for the concept u, we draw inspiration from the ***Universal Law of Gravitation***, where the potential field generated by a massive object is jointly governed by its intrinsic mass and relative spatial distance. Inspired by this mechanism, the HGC dynamically modulated the semantic influence of each node according to its learnable mass and radial depth. This physical analogy yields a well-defined entailment cone that is strictly axially symmetric around the geodesic ray $\overrightarrow{\mathbf{ou}}$, as visualized in Fig. 4.

Definition 1 (Hyperbolic Gravitational Cone.) Let $\mathbf{u} \in \mathbb{D}_c^n \setminus \{\mathbf{o}\}$ be a node with hyperbolic norm $r_u = \frac{2}{\sqrt{c}} \tanh^{-1}(\sqrt{c}\|\mathbf{u}\|)$ and angular direction $\Omega_u = \frac{\mathbf{u}}{\|\mathbf{u}\|}$. The Hyperbolic Gravitational Cone of $\mathbf{u}$ is defined as the set of points $\mathbf{v}$ encompassed within a mass-modulated half-aperture $\psi_{grav}(\mathbf{u})$:

$$\mathcal{C}_{grav}(\mathbf{u}) = \{\mathbf{v} \in \mathbb{D}_c^n \mid \Xi(\mathbf{u}, \mathbf{v}) \leq \psi_{grav}(\mathbf{u})\}, \tag{2}$$

where the half-aperture $\psi_{grav}(\mathbf{u})$ is dynamically determined by the mass function $m(r_u, \Omega_u)$ via the geometric condition:

$$\sin\left(\psi_{grav}(\mathbf{u})\right) \cdot \sinh(\sqrt{c} r_u) = m(r_u, \Omega_u). \tag{3}$$

Given that the volume of hyperbolic space expands exponentially, the half-aperture ψ must undergo inverse scaling to preserve a consistent semantic scope. In the standard Hyperbolic Entailment Cones formulation (Eq. 21), this contraction is governed by a rigid, isotropic global constant K. To overcome this fundamental limitation, we supersede the static constant K with a dynamic, highly expressive proportionality factor: the anisotropic semantic mass $m(r_u, \Omega_u)$. Crucially, by enforcing strict radial monotonicity on this anisotropic mass field, we establish a rigorous mathematical guarantee: the hyperbolic gravitational cone of any descendant node is unconditionally contained within the geometric boundaries of its ancestor.

Proposition 1 (Ray-wise Transitivity Guarantee). *Consider two nodes* $\mathbf{u}$ *and* $\mathbf{v}$ *aligned on the same geodesic ray from the origin, with* $r_v > r_u$ *(i.e.,* v *is a descendant of* u*). The nesting condition* $\mathcal{C}_{grav}(\mathbf{v}) \subseteq \mathcal{C}_{grav}(\mathbf{u})$ *holds if and only if the mass function satisfies* $m(r_v, \Omega) \leq m(r_u, \Omega)$.

Proposition 1 encapsulates the critical boundary condition for transitive containment along any corresponding geodesic ray. The formal proof of this proposition is provided in Appendix D.1.

3.3 Construction of the Anisotropic Mass Function

The mass function $m(r_u, \Omega_u)$ plays a pivotal role in governing the semantic influence of each node. To faithfully model the heterogeneous nature of hierarchical data, characterized by entailment strengths that vary across different semantic domains, we introduce a learnable anisotropic mass function. To mathematically guarantee that the induced geometric cones constitute a valid partial order, the mass function is rigorously constrained by four foundational properties:

- **Positivity and Asymptotic Stability:** $m(r_u, \Omega_u) > 0$ and $\lim_{r_u \to +\infty} m(r_u, \Omega_u) = m_0$. Strict positivity ensures the formation of geometrically valid, non-degenerate apertures. Furthermore, the asymptotic limit establishes m_0 as the intrinsic base mass, guaranteeing that even infinitely specific concepts (i.e., extreme leaf nodes) retain a minimal semantic footprint to maintain numerical stability and prevent aperture collapse in deep subtrees.
- **Radial Monotonicity:** $\partial m / \partial r \leq 0$. Broad, generalized concepts situated near the origin (small r_u) must possess a larger mass, whereas specific, fine-grained concepts situated near the boundary (large r_u) require a diminished mass. Crucially, this monotonicity serves as a sufficient condition to strictly guarantee the transitivity of the induced partial order.
- **Radial Convexity:** $\partial^2 m / \partial r^2 \geq 0$. Strict convexity is enforced to ensure that the decay of influence attenuates with depth. This mathematical property formalizes the intuition that semantic transitions become more subtle at deeper hierarchical levels (e.g., $husky \preceq dog \preceq animal$), ensuring the model retains sufficient capacity for fine-grained inference.
- **Directional Anisotropy:** The mass function exhibits an explicit dependency on the angular coordinate Ω_u, breaking the rotational symmetry. Specifically, for any two nodes $\mathbf{u}, \mathbf{v} \in \mathbb{D}_c^n$ situated at identical radial depth ($r_u = r_v$) but occupying distinct semantic orientations ($\Omega_u \neq \Omega_v$), the framework permits $m(r_u, \Omega_u) \neq m(r_v, \Omega_v)$. This flexibility empowers concepts like *animal* and *artifact* to possess distinct entailment capacities despite residing at equivalent hierarchical levels.

Guided by these principles, the mass function is formulated as:

$$m(r_u, \Omega_u) = m_0 \left(1 + \alpha\left(\Omega_u\right) \cdot e^{-\beta r_u}\right), \tag{4}$$

where $m_0 > 0$ acts as a base mass constant, $\alpha\left(\Omega_u\right) > 0$ is a learnable directional factor, and $\beta > 0$ controls the radial decay rate. This formulation is strictly positive, monotonically decreasing, and convex with respect to r_u. By substituting this mass formulation into the transitivity boundary condition defined in Eq. 3, the closed-form geometric expression for the cone half-aperture $\psi_{grav}(\mathbf{u})$ is obtained:

$$\psi_{grav}(\mathbf{u}) = \arcsin\left(\frac{m_0\left(1 + \alpha\left(\Omega_u\right) e^{-\beta r_u}\right)}{\sinh(\sqrt{c} r_u)}\right). \tag{5}$$

This explicit formulation bridges the physical and geometric aspects of our framework. It shows that the angular span of the entailment region is not merely a result of the geodesic distance from the origin, but is also heavily influenced by its intrinsic semantic mass.

Learning Anisotropy via Möbius Transformation. To model the complex, non-linear dependency of the entailment scope on semantic direction, the anisotropic factor $\alpha(\Omega_u)$ is parameterized via a shallow neural network. Specifically, a Möbius linear layer is first employed to project the directional vector Ω_u into a high-dimensional feature space. This is subsequently followed by a linear aggregation layer that maps the hidden representation to a scalar value. Finally, to strictly guarantee the essential positivity constraint required by the underlying mass function (Eq. 4), the $Softplus$ activation is applied at the terminal output:

$$\mathbf{h} = \tanh(M_1 \otimes_c \Omega_u), \tag{6}$$

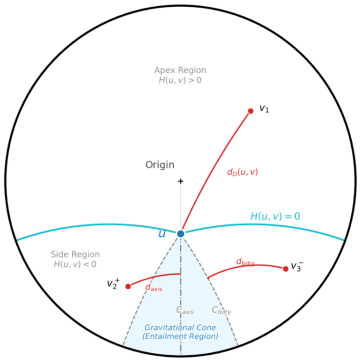


Figure 4: Illustration of Distance Calculations in HGC. The space is divided by the altitude plane $H(u, v) = 0$ (teal). In the apex region (v_1), distance is measured to the node u. In the side region, distance is defined orthogonally: to the central axis for entailed nodes (v_2^+, red line d_{axis}) and to the boundary for non-entailed nodes (v_3^-, red line d_{bdry}).

$$\alpha(\Omega_u) = \text{Softplus}(w_2^\top \mathbf{h} + b_2), \tag{7}$$

where $M_1 \in \mathbb{R}^{m \times n}$ denotes the learnable Möbius weight matrix, and $w_2 \in \mathbb{R}^{d_{hidden}}$ is the aggregation vector. This architecture empowers the model to capture intricate directional semantics. Coupling this learnable parameter with radial decay enables the proposed mass function to generalize prior isotropic formulations and create node-specific, direction-dependent apertures.

3.4 Gravitational Potential Optimization

To resolve the optimization stagnation caused by zero-energy where interior gradients vanish, the entailment relationship is modeled as a continuous gravitational field. Within this physics-inspired framework, entailment strength is quantified by the "work" required to position a hyponym node $\mathbf{v}$ within the gravitational influence of its hypernym $\mathbf{u}$. Unlike prior approaches that assign zero energy to all points inside the cone, our formulation establishes a graded potential. This energy actively pulls entailed nodes towards the gravitational core (i.e., the cone axis) while propelling non-entailed nodes beyond the event horizon (i.e., the cone boundary).

Gravitational Energy Formulation. The gravitational potential energy quantifies the degree to which a given entailment constraint is violated. It is composed of three physically coupled components: **(1) Gravitational Source Strength**, where the field intensity is exclusively governed by the hypernym's mass. This asymmetric design ensures robust optimization signals even for deep, light-weight leaves. **(2) Angular Potential**, which imposes a soft-gated penalty based on the deviation from the cone axis. **(3) Radial Depth**, which scales the energy by the orthogonal hyperbolic distance, ensuring that angular violations at greater hierarchical depths incur higher geometric penalties. Formally, the energy terms for positive ($\mathcal{P}$) and negative ($\mathcal{N}$) pairs are defined as:

$$E^+_{grav}(\mathbf{u}, \mathbf{v}) = \underbrace{\text{sg}[m(\mathbf{u})]}_{\text{Source Strength}} \cdot \underbrace{\zeta\left(\Xi(\mathbf{u}, \mathbf{v}) - \psi(\mathbf{u})\right)}_{\text{Angular Potential}} \cdot \underbrace{d_{axis}(\mathbf{u}, \mathbf{v})}_{\text{Radial Depth}}, \tag{8}$$

$$E^-_{grav}\left(\mathbf{u}, \mathbf{v}'\right) = \text{sg}[m(\mathbf{u})] \cdot \zeta\left(\Xi\left(\mathbf{u}, \mathbf{v}'\right) - \psi(\mathbf{u})\right) \cdot d_{bdry}\left(\mathbf{u}, \mathbf{v}'\right), \tag{9}$$

where $m(\mathbf{u})$ serves as a notational shorthand for the anisotropic mass function $m(r_u, \Omega_u)$. Crucially, a stop-gradient operation, denoted as $\text{sg}[\cdot]$, is applied to this source term. This operation ensures that the mass acts solely as an independent weighting coefficient for the field intensity, preventing the optimization process from trivially minimizing the energy by eroding the mass $m(\mathbf{u})$ to zero. The term $\zeta(\cdot)$ denotes a temperature-scaled soft-gating function, defined as:

$$\zeta(x) = \text{Softplus}(\tau x) = \ln(1 + e^{\tau x}), \tag{10}$$

where $\tau > 0$ acts as a hyperparameter controlling the steepness of potential barrier. This smooth approximation imposes a linear penalty for violations ($x > 0$) while maintaining a continuous, non-zero gradient for entailment constraints ($x \leq 0$).

Geometric Distance Metric. The radial depth terms in the energy function strictly depend on the relative position of a given node $\mathbf{v}$ with respect to the gravitational cone $\mathcal{C}_{grav}(\mathbf{u})$. To systematically quantify this spatial relationship, we define two distinct metrics: Distance to Cone Axis (d_{axis}), designed to refine semantic alignment, and Distance to Cone Boundary (d_{bdry}) formulated to enforce geometric separation.

To rigorously accommodate the hyperbolic geometry, a relative altitude function $H(\mathbf{u}, \mathbf{v}) = \Xi(\mathbf{u}, \mathbf{v}) - \psi(\mathbf{u}) - \pi/2$ is introduced. As illustrated in Fig. 4, this function partitions the manifold into two distinct regions: the "Apex Region" ($H \geq 0$, situated above the cone structure) and the "Side Region" ($H < 0$, situated alongside the cone). Consequently, the effective distances are formulated on this partition.

When node $\mathbf{v}$ resides in the apex region, the shortest path to the cone structure degenerates to the direct hyperbolic geodesic distance, $d_{\mathbb{D}}(\mathbf{u}, \mathbf{v})$. Conversely, within the side region, the distance is determined via orthogonal projection onto the respective geometric structures. Applying the hyperbolic Law of Sines, these distances are formalized as:

$$d_{axis}(\mathbf{u}, \mathbf{v}) = \frac{1}{\sqrt{c}} \operatorname{arsinh}\left(\sinh(\sqrt{c} d_{\mathbb{D}}(\mathbf{u}, \mathbf{v})) \sin(\Xi(\mathbf{u}, \mathbf{v}))\right), \tag{11}$$

$$d_{bdry}(\mathbf{u}, \mathbf{v}') = \frac{1}{\sqrt{c}} \operatorname{arsinh}\left(\sinh\left(\sqrt{c} d_{\mathbb{D}}(\mathbf{u}, \mathbf{v}')\right) \sin\left(\Xi(\mathbf{u}, \mathbf{v}') - \psi_{grav}(\mathbf{u})\right)\right). \tag{12}$$

Comprehensive derivations for these distance metrics and the gradient analysis of the gravitational energy are provided in Appendices D.2 and D.3, respectively.

3.5 Learning Objective and Physical Regularization

The proposed framework is trained end-to-end by minimizing a joint objective function that simultaneously enforces partial-order task constraints and the geometric consistency of the mass function.

Max-Margin Task Loss. To circumvent the boundary collapse problem frequently observed when applying contrastive loss in hyperbolic space (Zhang et al., 2025), we employ a max-margin loss. This formulation explicitly encourages positive pairs to descend into the potential well, while ensuring negative pairs are propelled sufficiently far from the decision boundary, exceeding a predefined margin γ:

$$\mathcal{L}_{task} = \sum_{(\mathbf{u}, \mathbf{v}) \in \mathcal{P}} E^{+}_{grav}(\mathbf{u}, \mathbf{v}) + \sum_{(\mathbf{u}, \mathbf{v}') \in \mathcal{N}} \max\left(0, \gamma - E^{-}_{grav}(\mathbf{u}, \mathbf{v}')\right). \tag{13}$$

Physics-Informed Regularization. A core theoretical contribution of HGC is the mathematical guarantee of transitivity, which is governed by the properties of the mass function. To strictly enforce radial monotonicity ($\partial m / \partial r \leq 0$) and convexity ($\partial^2 m / \partial r^2 \geq 0$) during optimization, explicit regularization terms are introduced.

To circumvent computationally prohibitive Hessian evaluations, these derivative are efficiently approximated via central finite differences. Specifically, for a given node $\mathbf{u}$ situated at a radial norm $r = \|\mathbf{u}\|_{\mathbb{D}}$, the localized mass variations are sampled at infinitesimally perturbed radii $r \pm \delta$:

$$\Delta_1 = \frac{\partial m}{\partial r} \approx \frac{m(r + \delta) - m(r - \delta)}{2\delta}, \tag{14}$$

$$\Delta_2 = \frac{\partial^2 m}{\partial r^2} \approx \frac{m(r + \delta) - 2m(r) + m(r - \delta)}{\delta^2}. \tag{15}$$

The total physical regularization objective is then formulated as:

$$\mathcal{L}_{reg} = \underbrace{ReLU(\Delta_1)}_{\text{Monotonicity}} + \underbrace{ReLU(-\Delta_2)}_{\text{Convexity}} + \underbrace{\Delta_1^2}_{\text{Magnitude}}. \tag{16}$$

Here, the magnitude penalty bounds extreme gradients within the mass function, preventing pathological oscillations and ensuring a stable, gradual evolution of semantic influence across varying hierarchical depths.

Necessity of the Regularization. While the analytical structure of mass function (Eq. 4) inherently incorporates radial decay, it remains theoretically susceptible to topological violations during highly non-linear optimization. Specifically, the anisotropic component $\alpha(\Omega_u)$ is parameterized by a highly expressive neural network. As node embeddings traverse the hyperbolic space, localized fluctuations or aggressive optimization steps could theoretically overpower the radial decay. This would lead to localized inversions where a child node illicitly gains a larger mass than its ancestor.

By integrating these finite-difference regularization terms $\mathcal{L}_{reg}$, we establish rigorous geometric guardrails. These constraints unconditionally guarantee that, irrespective of the neural network's parameter configuration, the resulting mass field strictly complies with the fundamental physical laws of monotonic decay and convexity. This mechanism structurally safeguards hierarchical transitivity throughout the entire training process.

Finally, the overall optimization objective of the HGC framework is formulated as:

$$\mathcal{L} = \mathcal{L}_{task} + \lambda \cdot \mathcal{L}_{reg}, \tag{17}$$

where λ governs the strength of physical constraints.

4 Experiments

4.1 Datasets

We evaluate the representation and inference capabilities of Hyperbolic Gravitational Cone across three benchmark datasets: WordNet Mammals, WordNet Noun (Fellbaum, 1998), and the Microsoft Concept Graph (MCG) (Wang et al., 2015; Li et al., 2017). These datasets exhibit diverse scales and topological complexities, spanning from the highly structured Mammal subtree to the large-scale MCG ontology. Following standard protocols (Yu et al., 2024), we exclusively consider "Is-A" (hypernym-hyponym) relations to focus on strict partial-order structures. Since HGC operates as a fundamental representation learning framework independent of task-specific architectures, we focus on pure hierarchical reconstruction to validate its intrinsic geometric expressiveness. Comprehensive statistics and data pre-processing details are provided in Appendix E.

4.2 Baselines and Evaluation Metrics

Baselines. We compare HGC against representative baselines spanning both Euclidean and hyperbolic spaces. Within the Euclidean domain, we consider prominent Box Embedding models: GBC-box (Zhang et al., 2022) and VBC-box (Boratko et al., 2021). For the hyperbolic domain, we compare against state-of-the-art cone-based representation methods: Hyperbolic Entailment Cones (Ganea et al., 2018) and Shadow Cones (Yu et al., 2024). Notably, HEC is constrained by isotropic apertures, whereas Shadow Cones rely on fixed-radius physical proxies, providing a structural contrast to our dynamic gravitational formulation. HGC implementation details and hyperparameter configuration are provided in Appendix F.

Evaluation Metrics. In alignment with standard protocols for hierarchical reconstruction (Ganea et al., 2018; Yu et al., 2024), we frame the task as a binary classification problem. Specifically, the model predicts whether a directed pair (u, v) constitutes a valid entailment relation $v \preceq u$. We report the *F1 Score*, which provides a balanced assessment of *Precision* and *Recall*.

4.3 Main Results

The hierarchical reconstruction performance of HGC and competitive baselines is presented in Table 1 (Mammal sub-graph) and Table 2 (WordNet Noun and MCG). Across diverse scales and topological complexities, HGC demonstrates consistent state-of-the-art performance. The empirical results highlight three critical advantages of our framework: (1) Robustness in Ultra-Low Dimensions (e.g., given 50% training data, HGC achieves an F1 score of 80.5%, outperforming the best baseline Umbral-half-space by an 11.1% absolute margin at $d = 2$ on Mammal); (2) Sample Efficiency in Low-Resource Regimes (e.g., achieving superior zero-shot inference on Noun); and (3) Scalability on Complex Topologies (e.g., establishing a new peak *F1* of 91.6% at $d = 10$ on the densely branched MCG dataset). A detailed quantitative analysis of these three observations is deferred to Appendix G.

Table 1: F1 Score (%) on the Mammal Sub-graph. Best results are highlighted in bold. The symbol "–" indicates that the model failed to converge (e.g., encountering NaN values during training).

Method	**Dimension = 2**					**Dimension = 5**				
	0%	10%	25%	50%	90%	0%	10%	25%	50%	90%
GBC-box	23.4	25.0	23.7	43.1	48.2	35.8	60.1	66.8	83.8	**97.6**
VBC-box	20.1	26.1	31.0	33.3	34.7	30.9	43.1	58.6	74.9	69.3
Hyperbolic Entailment Cone	**54.4**	61.0	71.0	66.5	73.1	56.3	81.0	84.1	83.6	82.9
Umbral-half-space	47.8	64.3	68.7	69.4	70.1	67.6	**84.3**	86.1	**89.4**	89.4
Penumbral-half-space	40.1	69.4	67.7	66.2	–	62.9	79.0	77.5	81.0	85.1
Penumbral-Poincaré-ball	27.2	34.9	38.4	55.6	71.4	35.9	46.9	68.4	79.5	88.0
HGC (Ours)	**54.4**	**71.9**	**74.3**	**80.5**	**77.2**	**70.4**	79.3	**88.6**	89.3	89.8

Table 2: F1 score (%) on WordNet Noun and MCG datasets.

Dataset		Noun				MCG			
Non-basic-edge Percentage		0%	10%	25%	50%	0%	10%	25%	50%
Entailment Cone	d=5	29.2	78.1	84.6	92.1	25.3	56.1	52.1	60.2
	d=10	32.1	82.9	91.0	95.2	25.5	58.9	55.5	63.8
Umbral-half-space	d=5	42.2	84.6	86.5	90.4	**38.6**	**76.8**	81.1	**84.2**
	d=10	44.5	86.6	87.8	91.8	**42.0**	**81.5**	85.8	88.9
Penumbral-half-space	d=5	**50.5**	83.6	87.2	89.4	34.8	76.6	80.8	83.6
	d=10	52.6	85.8	88.0	90.7	39.6	81.2	85.7	89.1
HGC (Ours)	d=5	48.2	**88.9**	**92.6**	**92.6**	35.0	76.3	**81.3**	83.9
	d=10	**55.3**	**89.7**	**94.7**	**95.3**	40.7	79.8	**87.7**	**91.6**

Beyond pure performance metrics, the results underscore a fundamental architectural superiority. Prior state-of-the-art methods, such as Shadow Cones, necessitate heuristic, dataset-specific selection between distinct physical assumptions (Umbral vs. Penumbral) and manifold models (Half-space vs. Poincar'e Ball) to avoid performance degradation. In contrast, HGC operates as a unified, end-to-end framework. By autonomously learning semantic influence via anisotropic gravitational mass, HGC eliminates the reliance on rigid physical proxies, delivering a highly adaptable solution for partial-order modeling. Further qualitative visualizations of HGC and its continuous potential field from multiple perspectives are provided in Appendix H.

5 Conclusion and Future Work

This paper introduces the Hyperbolic Gravitational Cone, fundamentally shifting partial-order representation learning from static geometric constraints to a dynamic, physically-grounded paradigm. By conceptualizing entailment as a continuous potential field governed by anisotropic mass, HGC successfully overcomes the semantic resolution bottlenecks and optimization stagnation inherent in prior rigid cone models. Consequently, this formulation guarantees informative gradients throughout the embedding space, systematically enforcing strict asymmetry and radial transitivity to reconstruct complex hierarchies with high fidelity.

While highly effective, the current framework assumes a singular gravitational center per entity, limiting its applicability to polysemous concepts residing in overlapping taxonomies. To address this, future work will investigate multi-center gravitational fields to disentangle polysemy accurately. Furthermore, extending this dynamic physical paradigm to temporal knowledge graphs and cross-modal entailment presents a promising avenue for advancing complex semantic representation.

AI USE STATEMENT

During the preparation of this manuscript, we used ChatGPT solely to assist with language refinement and document formatting. Specifically, its application was limited to improving language clarity and readability, organizing the text structure, formatting LaTeX code, and refining figure captions. We explicitly declare that all core scientific contributions, including the proposed methodologies, theoretical analyses, mathematical proofs and experimental designs, are strictly our original work and were developed without any assistance from generative AI. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Ralph Abboud, Ismail Ilkan Ceylan, Thomas Lukasiewicz, and Tommaso Salvatori. BoxE: A box embedding model for knowledge base completion. *Advances in Neural Information Processing Systems*, 33:9649–9661, 2020.

Gary Becigneul and Octavian-Eugen Ganea. Riemannian adaptive optimization methods. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=r1eiqi09K7.

Paul Berg, Léo Buecher, Björn Michele, Minh-Tan Pham, Laetitia Chapel, and Nicolas Courty. Multi-prototype hyperbolic learning guided by class hierarchy. *International Journal of Computer Vision*, 133(11):7969–7984, 2025.

Michael Boratko, Dongxu Zhang, Nicholas Monath, Luke Vilnis, Kenneth L Clarkson, and Andrew McCallum. Capacity and bias of learned geometric embeddings for directed graphs. *Advances in Neural Information Processing Systems*, 34:16423–16436, 2021.

Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

James W Cannon, William J Floyd, Richard Kenyon, and Walter R Parry. Hyperbolic geometry. *Flavors of geometry*, 31(59-115):2, 1997.

Ying Chen, Lianming Zhang, Jiusheng Li, and Pingping Dong. Hyperbolic graph representation learning: methods, applications and challenges—a survey. *Neurocomputing*, 652:131044, 2025.

Karan Desai, Maximilian Nickel, Tanmay Rajpurohit, Justin Johnson, and Shanmukha Ramakrishna Vedantam. Hyperbolic image-text representations. In *International Conference on Machine Learning*, volume 202, pp. 7694–7731, 2023.

Ankit Dhall, Anastasia Makarova, Octavian Ganea, Dario Pavllo, Michael Greeff, and Andreas Krause. Hierarchical image classification using entailment cone embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020.

Yu Duan, Feiping Nie, Huimin Chen, Zhanxuan Hu, Rong Wang, and Xuelong Li. Hyperbolic hierarchical representation learning for generalized category discovery. *IEEE Transactions on Neural Networks and Learning Systems*, 36(12):20050–20064, 2025.

Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, 1998.

Octavian Ganea, Gary Becigneul, and Thomas Hofmann. Hyperbolic entailment cones for learning hierarchical embeddings. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 1646–1655, 2018.

Albert Gu, Frederic Sala, Beliz Gunel, and Christopher Ré. Learning mixed-curvature representations in product spaces. In *International Conference on Learning Representations*, 2019.

Maxat Kulmanov, Wang Liu-Wei, Yuan Yan, and Robert Hoehndorf. EL embeddings: Geometric construction of models for the description logic EL++. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 6103–6109, 2019.

Matthew Le, Stephen Roller, Laetitia Papaxanthos, Douwe Kiela, and Maximilian Nickel. Inferring concept hierarchies from text corpora via hyperbolic embeddings. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 3231–3241, 2019.

Anchen Li, Bo Yang, Huan Huo, Hongxu Chen, Guandong Xu, and Zhen Wang. Hyperbolic neural collaborative recommender. *IEEE Transactions on Knowledge and Data Engineering*, 35(9): 9114–9127, 2023.

Lin Li, Jiahui Li, Jiaming Lei, Jun Xiao, Feifei Shao, and Long Chen. H^2EM: Learning hierarchical hyperbolic embeddings for compositional zero-shot learning, 2025. URL `https://arxiv.org/abs/2512.20029`.

Xiang Li, Luke Vilnis, and Andrew McCallum. Improved representation learning for predicting commonsense ontologies. *arXiv preprint arXiv:1708.00549*, 2017.

Xiang Li, Luke Vilnis, Dongxu Zhang, Michael Boratko, and Andrew McCallum. Smoothing the geometry of probabilistic box embeddings. In *International Conference on Learning Representations*, 2019.

Mingyu Derek Ma, Muhao Chen, Te-Lin Wu, and Nanyun Peng. HyperExpan: Taxonomy expansion with hyperbolic representation learning. In *Findings of the Association for Computational Linguistics: EMNLP*, pp. 4182–4194, 2021.

Johannes Messner, Ralph Abboud, and Ismail Ilkan Ceylan. Temporal knowledge graph completion using box embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 7779–7787, 2022.

Maximilian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in Neural Information Processing Systems*, 30:6338–6347, 2017.

Avik Pal, Max van Spengler, Guido Maria D'Amely di Melendugno, Alessandro Flaborea, Fabio Galasso, and Pascal Mettes. Compositional entailment learning for hyperbolic vision-language models. In *International Conference on Learning Representations*, 2025.

Wei Peng, Tuomas Varanka, Abdelrahman Mostafa, Henglin Shi, and Guoying Zhao. Hyperbolic deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10023–10044, 2022.

Samuele Poppi, Tobia Poppi, Federico Cocchi, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. Safe-CLIP: Removing NSFW concepts from vision-and-language models. In *European Conference on Computer Vision*, pp. 340–356, 2024.

Tobia Poppi, Tejaswi Kasarla, Pascal Mettes, Lorenzo Baraldi, and Rita Cucchiara. Hyperbolic safety-aware vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4222–4232, 2025.

Vaibhav Rathore, Divyam Gupta, and Biplab Banerjee. Hidisc: A hyperbolic framework for domain generalization with generalized category discovery. In *Advances in Neural Information Processing Systems*, 2025.

Hongyu Ren, Weihua Hu, and Jure Leskovec. Query2box: Reasoning over knowledge graphs in vector space using box embeddings. In *International Conference on Learning Representations*, 2020.

Rik Sarkar. Low distortion delaunay embedding of trees in hyperbolic plane. In *Graph Drawing*, pp. 355–366, 2012.

Srikumar Sastry, Aayush Dhakal, Eric Xing, Subash Khanal, and Nathan Jacobs. Global and local entailment learning for natural world imagery. In *International Conference on Computer Vision*, pp. 15770–15780, 2025.

Aditya Sinha, Siqi Zeng, Makoto Yamada, and Han Zhao. Learning structured representations with hyperbolic embeddings. *Advances in Neural Information Processing Systems*, 37:91220–91259, 2024.

Junru Song, Tianlei Chen, Yang Yang, and Feifei Wang. HeteroHTC: Enhancing hierarchical text classification via heterogeneity encoding of label hierarchy. *Expert Systems with Applications*, 271: 126558, 2025.

Albert Tseng, Tao Yu, Toni Liu, and Christopher De Sa. ConeHeads: Hierarchy aware attention. *Advances in Neural Information Processing Systems*, 36:51421–51433, 2023.

Ivan Vendrov, Ryan Kiros, Sanja Fidler, and Raquel Urtasun. Order-embeddings of images and language. In *International Conference on Learning Representations*, 2016.

Àlex Pujol Vidal, Kamal Nasrollahi, Thomas B. Moeslund, and Sergio Escalera. Machine unlearning in hyperbolic vs. euclidean multimodal contrastive learning: Adapting alignment calibration to MERU. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1635–1644, 2025.

Luke Vilnis, Xiang Li, Shikhar Murty, and Andrew McCallum. Probabilistic embedding of knowledge graphs with box lattice measures. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 263–272, 2018.

Shen Wang, Xiaokai Wei, Cicero Nogueira Dos Santos, Zhiguo Wang, Ramesh Nallapati, Andrew Arnold, and Philip S. Yu. Knowledge graph representation via hierarchical hyperbolic neural graph embedding. In *IEEE International Conference on Big Data*, pp. 540–549, 2021.

Xudong Wang, Shufan Li, Konstantinos Kallidromitis, Yusuke Kato, Kazuki Kozuka, and Trevor Darrell. Hierarchical open-vocabulary universal image segmentation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 21429–21453. Curran Associates, Inc., 2023.

Zhongyuan Wang, Haixun Wang, Ji-Rong Wen, and Yanghua Xiao. An inference approach to basic level of categorization. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, pp. 653–662, 2015.

Wei Wu, Xiaomeng Fan, Yuwei Wu, Zhi Gao, Pengxiang Li, Yunde Jia, and Mehrtash Harandi. Modality alignment across trees on heterogeneous hyperbolic manifolds. In *International Conference on Learning Representations*, pp. 11519–11554, 2026.

Bo Xiong, Mojtaba Nayyeri, Ming Jin, Yunjie He, Michael Cochez, Shirui Pan, and Steffen Staab. Geometric relational embeddings: A survey. *arXiv preprint arXiv:2304.11949*, 2023.

Yi.shi Xu, Dongsheng Wang, Bo Chen, Ruiying Lu, Zhibin Duan, and Mingyuan Zhou. Hyperminer: Topic taxonomy mining with hyperbolic embedding. In *Advances in Neural Information Processing Systems*, volume 35, pp. 31557–31570, 2022.

Menglin Yang, Jiahong Liu, Irwin King, and Rex Ying. UHCone: Universal hyperbolic cone for implicit hierarchical learning. In *ICML Workshop on Geometry-grounded Representation Learning and Generative Modeling*, 2024.

Tao Yu, Toni J. B. Liu, Albert Tseng, and Christopher De Sa. Shadow cones: A generalized framework for partial order embeddings. In *International Conference on Learning Representations*, 2024.

Dongxu Zhang, Michael Boratko, Cameron Musco, and Andrew McCallum. Modeling transitivity and cyclicity in directed graphs via binary code box embeddings. *Advances in Neural Information Processing Systems*, 35:10587–10599, 2022.

Yifei Zhang, Hao Zhu, Menglin Yang, Jiahong Liu, Rex Ying, Irwin King, and Piotr Koniusz. Understanding and mitigating hyperbolic dimensional collapse in graph contrastive learning. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1984–1995, 2025.

APPENDIX

This appendix provides comprehensive technical and supplementary materials to support the main text. Specifically, it is logically organized into three parts:

Part I: Theoretical Foundations

Appendix A: Related Work

Appendix B: Additional Key Concepts in Hyperbolic Geometry

Appendix C: Hyperbolic Entailment Cones

Appendix D: Theoretical Analysis

Part II: Experimental Setup

Appendix E: Data Pre-processing

Appendix F: Implementation Details and Hyperparameter Configuration

Part III: Supplementary Evaluations

Appendix G: Main Results and Discussion

Appendix H: Visualization of the Hyperbolic Gravitational Cone

Appendix I: Ablation Study

A RELATED WORK

We now review approaches that utilize deterministic geometric structures to embed partial order relations in Euclidean and hyperbolic spaces, as well as their applications.

A.1 PARTIAL ORDER MODELING IN EUCLIDEAN SPACE

Early approaches to embedding partial orders were grounded in Euclidean space, shifting the focus from symmetric distances to asymmetric geometric constraints. Order Embeddings (Vendrov et al., 2016) first model entailment via coordinate-wise inequalities ($x \preceq y \Leftrightarrow \bigwedge_i x_i \geq y_i$, with y as the hypernym), establishing a cone-based partial order. To capture uncertainty and disjoint concepts, this framework was extended to Box Embeddings (Ren et al., 2020; Vilnis et al., 2018; Abboud et al., 2020; Messner et al., 2022; Li et al., 2019), representing entities as hyper-rectangles where entailment is defined by geometric inclusion. While algorithmically innovative, these Euclidean approaches face a fundamentally topological mismatch: Euclidean volume grows only polynomially with radius ($V \propto r^d$), whereas hierarchical data expands exponentially. This capacity bottleneck inevitably leads to high distortion when embedding deep taxonomies.

A.2 HYPERBOLIC PARTIAL ORDER EMBEDDINGS AND APPLICATIONS

To overcome the structural bottlenecks inherent in Euclidean space, recent research has increasingly shifted toward hyperbolic geometry. Characterized by an exponential expansion of volume with respect to the radius ($V \propto e^r$). Pioneering theoretical work by Sarkar Sarkar (2012) rigorously proved that trees can be embedded into the Poincaré disk with arbitrarily low distortion. Building upon this, Poincaré Embeddings (Nickel & Kiela, 2017) demonstrated the superior inductive bias of hyperbolic manifolds for hierarchical data. However, these foundational approaches rely strictly on symmetric distance that struggle to capture asymmetric entailment relations.

To resolve this symmetry limitation, the field has transitioned toward region-based formalisms, specifically utilizing cone structures to model partial orders. The Hyperbolic Entailment Cone (Ganea et al., 2018) pioneered this direction by defining optimal entailment regions, while Shadow Cones (Yu et al., 2024) generalized this via a physics-inspired "light-shadow" framework. Fundamentally,

these methods optimize a geometric energy function governed by binary containment constraints, penalizing nodes only when they fall outside the cone boundary. Due to their rigorous geometric guarantees, such cone-based priors have achieved broad utility across diverse domains, ranging from implicit hierarchy discovery (Yang et al., 2024) and taxonomy induction (Sastry et al., 2025; Duan et al., 2025; Rathore et al., 2025; Xu et al., 2022), to cross-modal semantic alignment (Desai et al., 2023; Li et al., 2025; Pal et al., 2025; Wu et al., 2026) and emerging frontiers like machine unlearning (Vidal et al., 2025; Poppi et al., 2024; 2025).

B Additional Key Concepts in Hyperbolic Geometry

Hyperbolic Space $\mathbb{H}_c^n$ is the unique simply connected Riemannian manifold with constant negative curvature $-c(c > 0)$ (Cannon et al., 1997). As established in Fig. 1(c), its intrinsic exponential volume expansion allows for the embedding of tree-like structures (Sarkar, 2012). Geometrically, this capacity arises from the negative curvature, where non-intersecting geodesics diverge exponentially. This continuous divergence creates an exponentially expanding continuous space toward the periphery, perfectly accommodating the increasing number of nodes at deeper hierarchical levels. In this work, we operate within the Poincaré ball model due to its conformal property, which preserves Euclidean angles and thus facilitates intuitive cone aperture definitions.

The Poincaré norm, representing the geodesic distance from the origin, is defined as:

$$\|\mathbf{x}\|_{\mathbb{D}} := d_{\mathbb{D}}(\mathbf{0}, \mathbf{x}) = \frac{2}{\sqrt{c}} tanh^{-1}(\sqrt{c}\|\mathbf{x}\|). \tag{18}$$

Mobius Matrix Multiplication serves as the canonical generalization of Euclidean linear transformations to hyperbolic geometry. By enabling the continuous distortion, rotation, and scaling of vectors directly within the Poincaré ball, it allows for rich geometric transformations without violating the manifold's boundary constraints. For a linear map $\mathbf{M} \in \mathbb{R}^{m\times n}$ and a point $\mathbf{x} \in \mathbb{D}_c^n$ (assuming $\mathbf{Mx} \neq 0$), the Möbius matrix-vector multiplication is defined as:

$$\mathbf{M} \otimes_c \mathbf{x} = \frac{1}{\sqrt{c}} tanh\left(\frac{\|\mathbf{Mx}\|}{\|\mathbf{x}\|} \tanh^{-1}(\sqrt{c}\|\mathbf{x}\|)\right) \frac{\mathbf{Mx}}{\|\mathbf{Mx}\|}, \tag{19}$$

where $\otimes_c$ denotes the Möbius matrix-vector multiplication.

C Hyperbolic Entailment Cones

As illustrated in Fig. 5, the containment condition relies on comparing the deviation angle $\Xi(\mathbf{u}, \mathbf{v})$ with the cone half-aperture $\psi(\mathbf{u})$.

Deviation Angle Calculation. Due to the conformality of the Poincaré ball, the hyperbolic angle $\Xi(\mathbf{u}, \mathbf{v})$ coincides with the Euclidean angle between tangent vectors. As illustrated in Fig. 5, $\Xi(\mathbf{u}, \mathbf{v})$ is defined as the deviation angle between the cone axis (the geodesic ray extending from the origin $\mathbf{o}$ through $\mathbf{u}$) and the geodesic segment connecting $\mathbf{u}$ to $\mathbf{v}$. It can be explicitly computed as:

$$\Xi(\mathbf{u}, \mathbf{v}) = \pi - \angle ouv = \cos^{-1} \frac{\langle \mathbf{u}, \mathbf{v} \rangle (1 + c\|\mathbf{u}\|^2) - \|\mathbf{u}\|^2(1 + c\|\mathbf{v}\|^2)}{\|\mathbf{u}\| \cdot \|\mathbf{u} - \mathbf{v}\| \sqrt{1 + c^2\|\mathbf{u}\|^2\|\mathbf{v}\|^2 - 2c \langle \mathbf{u}, \mathbf{v} \rangle}}, \tag{20}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard Euclidean inner product.

Transitivity and Aperture Formulation. To rigorously guarantee the transitivity of the induced partial order (i.e., $w \preceq v \wedge v \preceq u \Rightarrow w \preceq u$), the half-aperture of entailment cone, denoted as $\psi(\mathbf{u})$, must contract as the node $\mathbf{u}$ moves away from the origin. The optimal half-aperture to satisfy this geometric constraint is derived as:

$$\psi(\mathbf{u}) = arcsin\left(K \cdot \frac{1 - c\|\mathbf{u}\|^2}{2\sqrt{c}\|\mathbf{u}\|}\right) \approx arcsin\left(\frac{K}{sinh(\sqrt{c}d_{\mathbb{D}}(\mathbf{o}, \mathbf{u}))}\right), \tag{21}$$

where K acts as a global scaling constant controlling the hierarchy width. The model optimizes a max-margin loss function formulated as:

$$\mathcal{L} = \sum_{(\mathbf{u},\mathbf{v})\in\mathcal{P}} E(\mathbf{u}, \mathbf{v}) + \sum_{(\mathbf{u},\mathbf{v}')\in\mathcal{N}} \max\left(0, \gamma - E(\mathbf{u}, \mathbf{v}')\right), \tag{22}$$

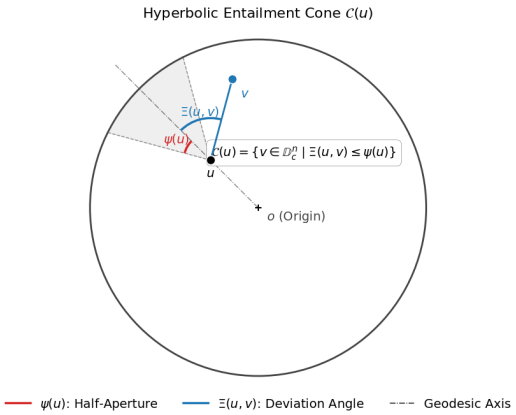


Figure 5: Geometric Construction of the Hyperbolic Entailment Cone. The cone $\mathcal{C}(u)$ (shaded region) is defined in the Poincaré ball with the parent node u as the apex. The cone's axis is the geodesic ray originating from the origin o and passing through u (dashed line). The containment condition is determined by comparing the actual deviation angle $\Xi(u, v)$ (blue arc) with the cone half-aperture $\psi(u)$ (red arc). In this illustration, the child node v lies outside the cone $\Xi(u, v) > \psi(u)$, indicating that the entailment relation $v \preceq u$ does not hold.

for margin $\gamma > 0$, where $\mathcal{P}$ and $\mathcal{N}$ define samples of positive and negative edges, $E(\mathbf{u}, \mathbf{v}) = \max(0, \Xi(\mathbf{u}, \mathbf{v}) - \psi(\mathbf{u}))$. A critical limitation of this paradigm is its reliance on a binary objective. It assigns strictly zero energy to all valid entailment relations $\Xi \leq \psi$, resulting in vanishing gradients inside the cone—a bottleneck of our framework resolves.

D Theoretical Analysis

To substantiate the theoretical rigor of HGC, this section provides formal derivations for ray-wise transitivity, explicit distance metrics, and continuous gradient behavior. Furthermore, we analyze the topological robustness of our mass function in mitigating the ϵ-hole singularity at the origin, and formally establish HGC's relationship with prior models.

D.1 Proof of Ray-wise Transitivity

This section provides the formal proof for Proposition 1, establishing that the radial monotonicity of the mass function strictly guarantees geometric containment along any geodesic ray.

Proof. Let $\mathbf{u}$ and $\mathbf{v}$ be two node embeddings aligned on the same geodesic ray extending from the origin, such that they share an identical angular direction Ω. Let their hyperbolic radii be r_u and r_v, respectively, with $r_u < r_v$ (implying that $\mathbf{v}$ is hierarchy deeper than $\mathbf{u}$). We aim to prove that if $m(r_v, \Omega) \leq m(r_u, \Omega)$, then the gravitational cone of $\mathbf{v}$ is unconditionally nested within the cone of $\mathbf{u}$, denoted as $\mathcal{C}_{grav}(\mathbf{v}) \subseteq \mathcal{C}_{grav}(\mathbf{u})$.

According to the geometric axioms of the Poincaré ball, the sufficient condition for a rotationally symmetric cone to be geometrically contained by an ancestor cone located on the same axis is governed by the following inequality (Theorem 3, (Ganea et al., 2018)):

$$\sin(\psi_v)\sinh(\sqrt{c}r_v) \leq \sin(\psi_u)\sinh(\sqrt{c}r_u). \tag{23}$$

Recall the geometric definition of the Hyperbolic Gravitational Cone Eq. 3:

$$\sin(\psi_{\text{grav}}(x)) \cdot \sinh(\sqrt{c}r_x) = m(r_x, \Omega_x). \tag{24}$$

By substituting this mass-based aperture definition into the containment inequality (Eq.23), the geometric condition simplifies directly to:

$$m(r_v, \Omega) \leq m(r_u, \Omega). \tag{25}$$

Because our physical regularization strictly enforces $\frac{\partial m}{\partial r} \leq 0$ for any given Ω, the mass function is monotonically non-increasing with respect to radial depth. Therefore, the condition $r_u < r_v \Rightarrow$

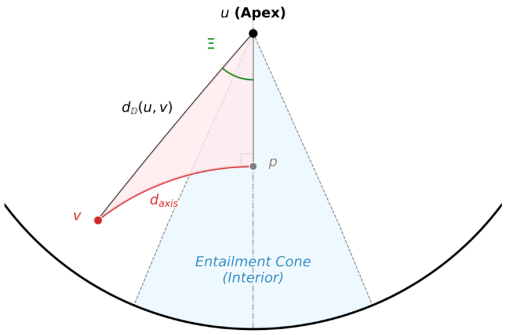


Figure 6: Geometric Derivation of the Distance to Cone Axis (d_{axis}) via Hyperbolic Trigonometry.

$m(r_v, \Omega) \leq m(r_u, \Omega)$ is universally satisfied. This structurally guarantees strict geometric transitivity along any specific geodesic ray.

Discussion: The Trade-off Between Anisotropy and Global Transitivity. While Proposition 1 guarantees strict geometric containment along specific radial directions, it is crucial to acknowledge that absolute global transitivity—where entailment holds across all angular directions—is mathematically incompatible with a fully anisotropic framework. In an anisotropic setting, distinct mass decay across different directions could theoretically permit a deep node in one semantic branch to exhibit a wider aperture than a shallow node in another, potentially violating cross-branch transitivity.

However, we argue that enforcing absolute global transitivity imposes an overly rigid geometric prior that hinders semantic resolution. Real-world taxonomies are inherently heterogeneous; an abstract concept in a dense domain should not be geometrically constrained by the properties of an unrelated sparse domain. By relaxing global transitivity to strict ray-wise transitivity, HGC acquires the necessary geometric flexibility to model this directional variability.

D.2 Derivation of Distance Metrics

This section formalizes the closed-form expressions for the geodesic distances d_{axis} and d_{bdry}.

Geometric Setup. Consider a target node $\mathbf{v}$ and its hypernym node $\mathbf{u}$. We define the cone axis as the geodesic ray originating from the origin $\mathbf{o}$ and passing through $\mathbf{u}$. When the relative altitude $H(\mathbf{u}, \mathbf{v}) \leq 0$, the orthogonal projection of $\mathbf{v}$ onto the axis $\overrightarrow{\mathbf{ou}}$ falls on the ray extending outwards from $\mathbf{u}$. Let $\mathbf{p}$ denote this orthogonal projection point. As illustrated in Fig. 6, we construct a hyperbolic right-angled triangle $\triangle uvp$, characterized by the following properties: the interior angle at $\mathbf{u}$ corresponds to the angular deviation $\Xi(\mathbf{u}, \mathbf{v})$; the segment $\mathbf{vp}$ is orthogonal to the axis $\overrightarrow{\mathbf{ou}}$, ensuring the angle at projection $\mathbf{p}$ is $\pi/2$.

We apply the Hyperbolic Law of Sines, which relates the interior angles (A, B, C) of a hyperbolic triangle to its opposite geodesic edge lengths (a, b, c):

$$\frac{\sin A}{\sinh a} = \frac{\sin B}{\sinh b} = \frac{\sin C}{\sinh c}. \tag{26}$$

Within $\triangle uvp$, the hypotenuse connecting $\mathbf{u}$ and $\mathbf{v}$ has length $d_{\mathbb{D}}(\mathbf{u}, \mathbf{v})$. The opposite side to angle $\Xi(\mathbf{u}, \mathbf{v})$ is the segment $\mathbf{vp}$, representing the perpendicular distance to the axis, $d_{axis}(\mathbf{u}, \mathbf{v})$. Substituting these geometric entities into Eq.26 yields:

$$\sinh(\sqrt{c} \cdot d_{\text{axis}}) = \sinh(\sqrt{c} \cdot d_{\mathbb{D}}(\mathbf{u}, \mathbf{v})) \cdot \sin(\Xi(\mathbf{u}, \mathbf{v})). \tag{27}$$

By isolating d_{axis} and applying the inverse hyperbolic sine, we obtain the exact closed-form solution:

$$\sqrt{c} \cdot d_{\text{axis}} = \operatorname{arsinh}\left(\sinh(\sqrt{c} d_{\mathbb{D}}(\mathbf{u}, \mathbf{v})) \sin(\Xi(\mathbf{u}, \mathbf{v}))\right), \tag{28}$$

$$d_{\text{axis}}(\mathbf{u}, \mathbf{v}) = \frac{1}{\sqrt{c}} \operatorname{arsinh}\left(\sinh(\sqrt{c} d_{\mathbb{D}}(\mathbf{u}, \mathbf{v})) \sin(\Xi(\mathbf{u}, \mathbf{v}))\right). \tag{29}$$

The derivation for the distance to the cone boundary d_{bdry}, proceeds analogously. By substituting the central axis with the cone boundary ray, the internal angle simply becomes the residual angular deviation $\Xi(\mathbf{u}, \mathbf{v}') - \psi_{grav}(\mathbf{u})$.

D.3 Gradient Analysis of the Gravitational Energy

To mathematically demonstrate how HGC resolves the optimization stagnation inherent in prior models (e.g., HEC (Ganea et al., 2018) & Shadow Cones (Yu et al., 2024)), we analyze the gradient behavior of the proposed energy function. The angular component of the positive energy term (Eq. 8) is defined via a temperature-scaled Softplus activation:

$$E_{\text{ang}}(\Delta\theta) = \text{Softplus}(\tau \cdot \Delta\theta) = \ln\left(1 + e^{\tau\Delta\theta}\right), \tag{30}$$

where $\Delta\theta = \Xi(\mathbf{u}, \mathbf{v}) - \psi(\mathbf{u})$ denotes the angular deviation, and $\tau > 0$ regulates the steepness. Crucially, $\Delta\theta > 0$ indicates the node lies outside the cone, whereas $\Delta\theta < 0$ implies interior containment.

Applying the chain rule, the gradient of this energy with respect to the angular violation $\Delta\theta$ resolves to a scaled sigmoid function $\sigma(\cdot)$:

$$\begin{aligned} \frac{\partial E_{\text{ang}}}{\partial \Delta\theta} &= \frac{\partial}{\partial \Delta\theta}\left[\ln(1 + e^{\tau\Delta\theta})\right] = \frac{1}{1 + e^{\tau\Delta\theta}} \cdot \tau e^{\tau\Delta\theta} \\ &= \tau \cdot \frac{e^{\tau\Delta\theta}}{1 + e^{\tau\Delta\theta}} = \tau \cdot \sigma(\tau\Delta\theta). \end{aligned} \tag{31}$$

We evaluate this gradient magnitude across two distinct geometric regimes:

Case 1: Boundary Violation ($\Delta\theta > 0$, Outside Cone.) When the node resides outside the entailment region, the gradient asymptotically approaches:

$$\lim_{\Delta\theta \to +\infty} \frac{\partial E_{\text{ang}}}{\partial \Delta\theta} = \tau. \tag{32}$$

This provides a robust, constant restoring force proportional to τ, actively pulling the node back toward the cone boundary without causing numerical explosion.

Case 2: Constraint Satisfaction ($\Delta\theta < 0$, Inside Cone.) This represents the critical regime where prior binary objectives, utilizing $\max(0, \Delta\theta)$, yield strictly zero gradients. In contrast, because the exponential function is strictly positive, the HGC gradient becomes:

$$\frac{\partial E_{\text{ang}}}{\partial \Delta\theta} = \tau \cdot \sigma(\tau\Delta\theta) = \frac{\tau}{1 + e^{-\tau\Delta\theta}} > 0. \tag{33}$$

This guarantees a continuous, non-zero gradient flow throughout the entire interior domain. Even deep inside the cone, the gradient remains strictly positive, continuously driving hyponym nodes to align with the hypernym's central gravitational axis. This structurally prevents optimization stagnation and enforces semantic centrality.

D.4 Discussion on the Impact of the ϵ-Hole Singularity

A fundamental topological artifact of origin-centered hyperbolic cones is the "ϵ-hole" singularity. As the radial norm of a node approaches zero ($r_u \to 0$), both its angular direction Ω_u and aperture $\psi_{grav}(u)$ became mathematically undefined.

In standard HEC framework (Ganea et al., 2018), this singularity precipitates severe optimization hazards. If a root node is initialized or pushed too close to the origin, the geometric constraints structurally collapse, triggering numerical instability (e.g., NaN gradients) and destroying the learned hierarchy. Consequently, prior methods rely heavily on multi-stage pre-training heuristics to artificially distance embeddings from the origin prior to enforcing cone constraints.

The proposed HGC framework intrinsically mitigates the "ϵ-hole" singularity via its continuous physics-informed formulations. Firstly, the soft-gated potential function acts as a numerical buffer during gradient computation near the origin. As described in Sec. D.3, regardless of the magnitude of

numerical perturbations introduced by angle computations Ξ near singular points, the gradient $\sigma(\tau x)$ is strictly bounded within the interval [0,1], which ensures stable optimization. More fundamentally, from a geometric perspective, the base mass m_0 within the anisotropic mass function explicitly regulates near-origin behavior. As $r_u \to 0$, the ratio $m(r_u, \Omega_u)/\sinh(\sqrt{c}r_u)$ rapidly diverges. Through the bounded geometric formulation in Eq. 5, this seamlessly translates the aperture to $\psi_{grav} \to \pi/2$, effectively flattening the cone into a half-space. Therefore, HGC interprets embeddings near the singularity not as numerical anomalies, but as absolute root concepts commanding infinite gravitational reach. This profound geometric consistency enables HGC to be trained robustly from scratch utilizing simple uniform initialization, elegantly eliminating the necessity for heuristic pre-training phases.

D.5 Mathematical Relationship to Previous Models

The HGC framework functions as a rigorous geometric generalization of prior cone-based models. Specifically, we demonstrate that both HEC and Shadow Cones can be mathematically recovered as specialized configurations of HGC.

D.5.1 Relationship to Hyperbolic Entailment Cones

The original HEC model (Ganea et al., 2018) defines the cone aperture strictly as a function of radial depth: $\sin(\psi) \approx K/\sinh(\sqrt{c}r)$, where K is a global constant. Within the HGC framework, if directional anisotropy is disabled by zeroing the learnable factor $\alpha(\Omega) = 0$, and the base mass m_0 is fixed to K, the mass function reduces to:

$$m(r_u, \Omega_u) = m_0 \left(1 + 0 \cdot e^{-\beta r_u}\right) = K. \tag{34}$$

Substituting this static mass into the HGC aperture definition recovers the exact HEC formulation. Consequently, HEC is revealed to be an isotropic special case of HGC, constrained by a constant, direction-independent semantic mass.

D.5.2 Relationship to Shadow Cones

Shadow Cones (Yu et al., 2024) formulate entailment via optical shadows cast by physical proxies possessing a fixed radius r_{obj}. The resulting aperture is determined entirely by the geometric projection of this static proxy.

In HGC, the mass term serves as the functional equivalent of the proxy's "effective size". By enforcing the mass to decay at a specific rate that mimics the hyperbolic visual angle of a fixed-radius ball, HGC reduces to the Shadow Cone formulation. Fundamentally, Shadow Cones assume that a concept's "blocking capability" (mass) is governed exclusively by static physical dimensions r_{obj}. HGC inherently relaxes this rigid constraint by rendering the mass dynamically learnable and fully anisotropic. This paradigm shift allows nodes to function as semantic entities with variable densities, establishing Shadow Cones as a special case of the generalized HGC framework.

E Data Pre-processing

Following the standard evaluation protocol established by Shadow Cones (Yu et al., 2024), we compute the transitive closure of the hierarchy graphs. The relation edges are subsequently partitioned into a transitive reduction set utilized during training to preserve structural connectivity, and the remaining transitive closure set. This latter set is reserved for validation and testing to systematically evaluate the model's capacity for hierarchical inference. Comprehensive statistics for each dataset are summarized in Table 3.

F Implementation Details and Hyperparameter Configuration

We implement HGC utilizing the *HTorch* library[1], a PyTorch-based framework optimized for hyperbolic deep learning, enabling efficient end-to-end training within the Poincaré ball. To ensure reproducibility, we detail our experimental protocols and hyperparameter configurations below.

[1] https://github.com/ydtydr/HTorch

Table 3: Detailed Statistics of the Datasets.

Statistic	Mammal	Noun	MCG
Max Depth	8	18	31
No. of Nodes	1,179	82,114	22,665
No. of Components	4	2	657
No. of Basic Relations	1,176	84,363	38,288
No. of Total Relations	5,361	661,127	1,134,348

Initialization and Curvature. Prior research has underscored the profound sensitivity of hyperbolic optimization to initial conditions (Ganea et al., 2018). In alignment with established practices (Nickel & Kiela, 2017), we initialize node embeddings using a uniform distribution $\mathcal{U}[-\epsilon, \epsilon]$ centered at the origin. Across all experiments, the manifold curvature is strictly fixed at $-c = -1.0$. Instead of tuning a global curvature parameter, we rely on the learnable anisotropic mass to dynamically capture varying hierarchical scales.

Training Protocol and Negative Sampling. We train the HGC framework utilizing the transitive reduction (i.e., the minimal set of basic edges) of the hierarchy DAG as the foundation for positive examples. To rigorously evaluate the model's capacity for hierarchical inference, we construct training sets with varying levels of difficulty by progressively injecting 0% to 90% of the non-basic edges (the transitive closure). The remaining non-basic edges are evenly divided between the validation and test sets. For negative sampling, consistent with standard methodologies (Ganea et al., 2018; Nickel & Kiela, 2017), we generate $N_{neg} = 10$ negative samples for each positive pair (u, v) by randomly corrupting either the hypernym u or the hyponym v. To comprehensively assess HGC's geometric capacity, we configure the embedding dimensions according to the scale and topological complexity of each dataset. For the fine-grained Mammal sub-graph, we evaluate the model in extremely constrained low-dimensional spaces (d=2 and d=5) to test its robustness against crowding. For the massive and deeply branched WordNet Noun and MCG ontologies, we scale the embedding dimension to d=5 and d=10, providing sufficient geometric volume to disentangle complex, heterogeneous semantic clusters.

Hyperparameter Configuration. To fully exploit the physical dynamics of the HGC framework, we conducted a systematic grid search over the key hyperparameters governing the gravitational field. The search space is selected as follows: base mass $m_0 \in \{0.001, 0.002, 0.01, 0.02\}$, radial decay rate $\beta \in \{0.1, 0.2, 0.5, 1.0, 2.0\}$, energy margin $\gamma \in \{0.1, 0.2, 0.5\}$, potential temperature $\tau \in \{1, 10, 20, 50\}$, and physical constraint $\lambda \in \{0.1, 0.01, 0.001\}$. The anisotropic mass network is configured with a hidden dimension of 64. Optimal configurations for all main experiments are selected based on validation set performance. Additionally, we introduce a perturbation step $\delta = 10^{-3}$ to guarantee numerical stability during the finite-difference regularization of the mass function. Finally, to perform optimization on the manifold, we utilize Riemannian Adam (RAdam) (Becigneul & Ganea, 2019), which adapts adaptive moment estimation to the hyperbolic geometry via the exponential map.

G Main Results and Discussion

The hierarchical reconstruction performance of HGC and the competitive baselines is presented in Table 1 for the Mammal sub-graph, while the results for the large-scale Noun and MCG datasets are summarized in Table 2. Across diverse scales and topological complexities, HGC demonstrates highly consistent performance. We highlight three key observations from the empirical results.

1) Robustness in Ultra-Low Dimension Spaces.

A notable performance gain for HGC is observed under the extremely constrained $d = 2$ setting (Table 1). While Euclidean methods (GBC-box, VBC-box) exhibit sub-optimal capacity and certain hyperbolic variants (e.g., Penumbral-half-space) encounter numerical instability (denoted by "–"), HGC maintains robust optimization across almost all levels. Specifically, given 50% training data, HGC achieves an F1 score of 80.5%, outperforming the best baseline (Umbral-half-space at 69.4%) by an absolute margin of 11.1%. This indicates that the proposed anisotropic mass formulation

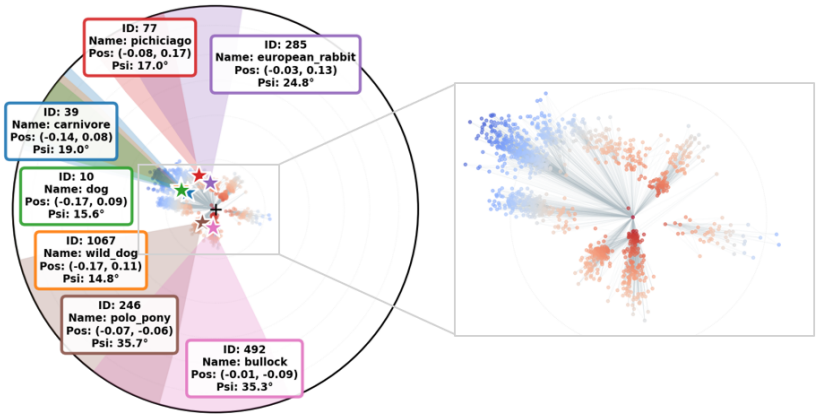


Figure 7: Visualization of Learned HGC Embeddings and Gravitational Cones on the Mammal Dataset. Macro View (Left): We highlight two key geometric properties: (1) Hierarchical Transitivity *(IDs 39, 10, 1067)*: The chain $wild_dog \preceq dog \preceq carnivore$ demonstrates strict cone containment, where the ancestor's cone (blue) encapsulates the descendant's cone (green), which in turn contains the leaf node (orange). (2) Directional Anisotropy *(IDs 246, 496)*: Nodes such as $polo_pony$ (gray) and $bullock$ (pink) occupy distinct angular sectors with unique aperture shapes, confirming that entailment scope is modulated by semantic direction rather than radial depth alone. The text boxes display node metadata and the learned half-aperture (ψ). Micro View (Right): A zoomed-in visualization of the central region reveals the dense, tree-like topological structure captured by HGC, showing how child nodes cluster within the gravitational influence of their parents.

effectively mitigates the geometric bottleneck inherent to isotropic cones, allowing nodes to adaptively organize within highly congested, low-dimensional manifolds without structural collapse.

2) Sample Efficiency in Low-Resource Regimes.

A critical measure of a model's inductive bias is its ability to infer global hierarchies with minimal structural supervision (i.e., the 0% and 10% non-basic-edge settings). As detailed in Table 2, HGC exhibits superior sample efficiency on the WordNet Noun dataset. At $d = 10$ with 0% additional transitive edges (zero-shot inference), HGC achieves 55.3%, surpassing the strongest Shadow Cone variant (52.6%). Similarly, at the 10% training data, HGC reaches 88.9% ($d = 5$) and 89.7% ($d = 10$). These results confirm that the continuous gravitational energy provides informative, non-zero gradients that successfully guide the formation of coherent taxonomies even when explicit transitive links are severely limited.

3) Scalability and Expressiveness on Complex Topologies.

The MCG dataset presents a rigorous challenge due to its deep hierarchy and massive branching factor. While physics-inspired Shadow Cones (Umbral-half-space) show strong performance in lower dimensions ($d = 5$), HGC surpasses them as the embedding capacity increases. At $d = 10$ on MCG, HGC achieves state-of-the-art results of 87.7% (25% data) and 91.6% (50% data). This trend demonstrates that, given sufficient manifold capacity, HGC's learnable mass function effectively disentangles dense semantic clusters, dynamically adjusting cone apertures to accommodate highly heterogeneous semantic branches.

H Visualization of the Hyperbolic Gravitational Cone

To intuitively validate the structural properties and geometric expressiveness of HGC, we visualize the learned embeddings and their corresponding gravitational fields on the Mammal dataset (evaluated in $\mathbb{D}_c^2$ using a 90% training split). Note that the shaded regions in the visualization depict the full entailment cones (spanning an angle of $2\psi_{\text{grav}}$), to explicitly illustrate the complete semantic scope of each concept.

As shown in the macro view of Fig. 7(Left), the visualization highlights two fundamental characteristics: **(1) Hierarchical Transitivity.** The logical nesting of entailment scopes is explicitly preserved. Observing the evolutionary chain from *carnivore (ID 39)* to *dog (ID 10)* and subsequently to *wild-dog (ID 1067)*, we see a clear structural containment: the broader gravitational cone of the ancestor fully encapsulates the progressively narrower cones of its descendants. This confirms that our physics-informed monotonicity regularization successfully enforces transitive containment across multiple hierarchy levels. **(2) Directional Anisotropy.** The heterogeneous nature of semantic influence is clearly manifested. Concepts such as *polo-pony (ID 246)* and *bullock (ID 492)* are mapped to distinct angular sectors, generating cone apertures uniquely adapted to their specific semantic directions. Unlike isotropic models where the aperture is rigidly determined by the radial depth, HGC empowers nodes at comparable distances from the origin to develop heterogeneous influence scopes tailored to their intrinsic semantic breadth.

Furthermore, the zoomed-in panel in Fig. 7 (Right) demonstrates that HGC effectively disentangles the dense, heavily populated regions near the origin. By mapping concepts into coherent semantic branches governed by gravitational potential, the model maintains high structural fidelity even in highly congested hyperbolic neighborhoods.

Empirical Validation of the Potential Field. To visually substantiate the advantages of our continuous energy formulation over zero-energy rigid boundaries, the local energy landscape of a representative concept is plotted in Fig. 8, a high-resolution sub-region (spanning a normalized spatial extent of 0.25×0.25) centered at the apex node *dog* is extracted. The background contour map explicitly visualizes the computed positive energy E^+ across a dense discrete grid. The superimposed white arrows constitute the background gradient vector field ($-\nabla E^+$), indicating the trajectory of optimization towards the local energy minima.

To demonstrate the optimization behaviors, several true descendant nodes are sampled, with their corresponding analytical gradient vectors (red arrows) scaled proportionally by magnitude. The distinct optimization behaviors of these nodes provide critical insights into the underlying mechanics of the HGC framework: **(1) Gradient Response to Boundary Violations.** Consider the node "dalmatian", which is located on the fringe of the entailment region, slightly exceeding the theoretical angular boundary defined by the dashed red line ($\Xi > \psi$). Consequently, this node is subjected to a substantial restoring gradient, indicated by the pronounced magnitude of its corresponding red arrow, directed orthogonally towards the semantic axis. This observation confirms that the soft-gated potential function imposes a severe, distance-scaled penalty for boundary violations, effectively forcing outlier nodes back into the valid entailment scope. **(2) Mitigation of Optimization Stagnation.** Conversely, examine the node *puppy*, which is situated deep within the entailment cone, strictly satisfying the geometric boundary condition ($\Xi < \psi$). Under conventional isotropic models (e.g., HEC), the gradient for such a node would evaluate to identically zero, precipitating optimization stagnation. In the HGC formulation, however, the gradient vector for *puppy* clearly exhibits a non-zero magnitude directed inward toward the central semantic axis (white dotted line). Although its gradient norm is quantitatively smaller than that of the boundary-violating node *dalmatian*, this persistent, non-zero restorative force ensures continuous structural refinement. This mechanism empirically validates that HGC actively enforces semantic centrality, optimizing the internal hierarchical arrangement rather than merely satisfying a binary geometric constraint.

I Ablation Study

Detailed ablation studies are provided in this section, including an in-depth analysis of the performance gains attributable to the anisotropic mass formulation and the continuous gravitational energy field.

I.1 Impact of Anisotropic Mass Formulation

To isolate and quantify the necessity of the proposed anisotropic mass function (Eq. 4), the full HGC architecture is compared against two isotropic baselines lacking directional awareness. The evaluated configurations are specified as follows: **(1) HGC-Isotropic (constant mass):** The directional factor is eliminated by setting $\alpha(\Omega_u) = 0$ and $\beta = 0$, reducing the semantic mass to a global constant $m(r_u, \Omega_u) = m_0$. This variant is geometrically equivalent to the HEC framework while being

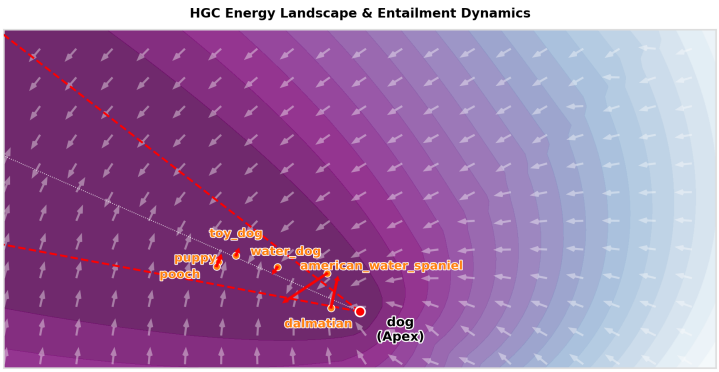


Figure 8: Energy Landscape and Gradient Vector Field of the Gravitational Energy. The visualization depicts the local potential well around the apex node *dog*. The background color represents the magnitude of the gravitational energy E^+ (darker purple indicates lower energy). White arrows constitute the background gradient field, indicating the direction of optimization. The red dashed lines define the theoretical cone boundary $\psi_{grav}(u)$, while the white dotted line represents the semantic axis. We sample key descendant nodes and plot their actual gradient vectors (red arrows), with lengths proportional to the gradient magnitude.

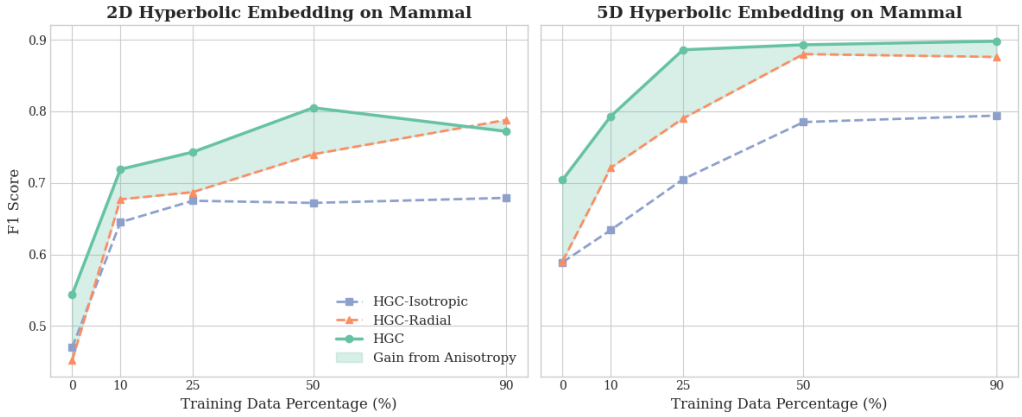
optimized via our gravitational energy. **(2) HGC-Radial (depth-only):** We restrict $\alpha(\Omega_u) = 1$ but retain the radial decay β, yielding $m(r_u, \Omega_u) = m_0 \cdot e^{-\beta r_u}$. This isolates the performance variance attributable exclusively to depth-dependent contraction. **(3) HGC:** Our proposed architecture utilizing the fully learnable anisotropic mass.

The F1 scores across three hierarchical datasets under varying training sparsity are plotted in Fig. 9. The experimental results demonstrate three critical findings: **(1) Radial Decay is Insufficient for Dense Topologies.** While HGC-Radial occasionally yields marginal improvements over the constant-mass baseline (e.g., Mammal 5D), both isotropic variants fail to surpass an F1 score of 55% on large-scale datasets. This quantitatively demonstrates that relying solely on radial depth is structurally inadequate for resolving semantic overlap in dense taxonomies. **(2) Anisotropy Solves the Semantic Resolution Bottleneck.** The introduction of directional awareness via the fully learnable anisotropic mass yields substantial performance gains, visually highlighted by the green shaded areas ("Gain from Anisotropy"). This confirms that the anisotropic inductive bias successfully disentangles heterogeneous entailment scopes, particularly in data-scarce scenarios (0% and 10% training data). **(3) Anisotropic Gains Scale with Dataset Complexity.** The performance differential between the full HGC and its isotropic counterparts widens dramatically on the massive Noun and MCG hierarchies, achieving an absolute improvement of exceeding 35% (e.g., MCG 10D, 50% data). This distinct scaling trend provides compeling empirical evidence that dynamic, direction-dependent apertures are an absolute mathematical necessity for high-fidelity representation learning in complex networks.

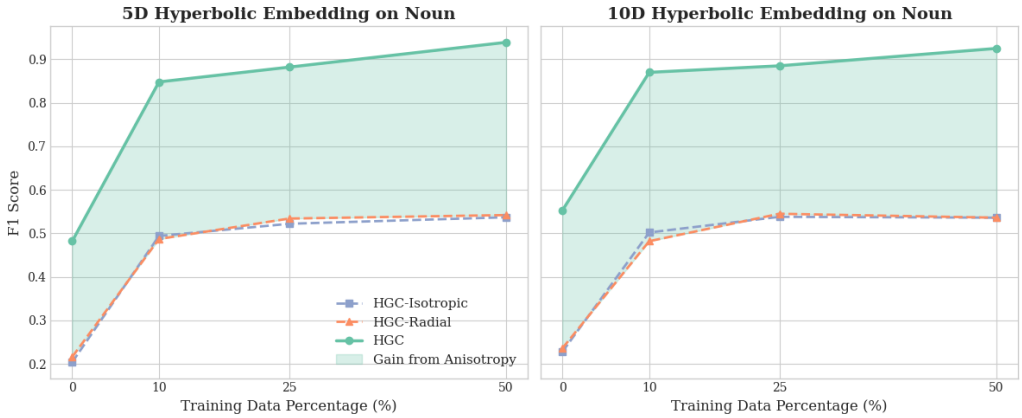
I.2 Impact of Continuous Gravitational Energy

To validate the contribution of the continuous gravitational potential, the HGC energy formulation is contrasted with traditional binary geometric objectives. This comparison empirically validates that maintaining informative, non-zero gradients within the entailment cone is imperative for fine-grained structural alignment. The evaluated variants are: **(1) HGC-Max-Margin:** The graded potential is replaced with a angle energy in Eq.22 (Ganea et al., 2018). Consequently, any entailed node $\mathbf{v}$ falling inside the cone is assigned strictly zero energy. **(2) HGC-Contrastive:** The graded potential is substituted with an InfoNCE contrastive loss paradigm (Yu et al., 2024), utilizing the distance to the cone boundary as the energy function. **(3) HGC:** Our proposed formulation employs the soft-gated potential $\zeta(\cdot)$ coupled with the geodesic distance.

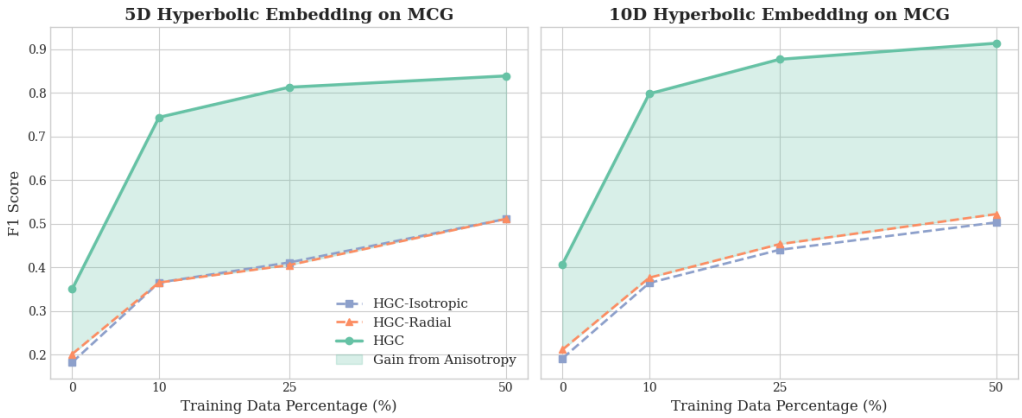
Empirical validation of these variants (Fig. 10) yields the following insights: **(1) The Pitfalls of Optimization Stagnation.** The HGC-Max-Margin variant consistently exhibits severe performance degradation across all settings, struggling to surpass 60% F1 on Noun and MCG. **(2) Boundary**



(a) The effect of anisotropy on Mammal dataset.

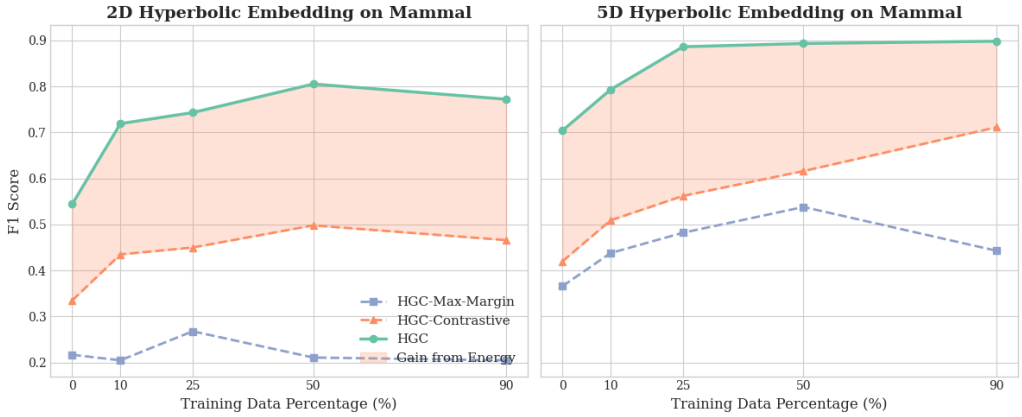


(b) The effect of anisotropy on Noun dataset.

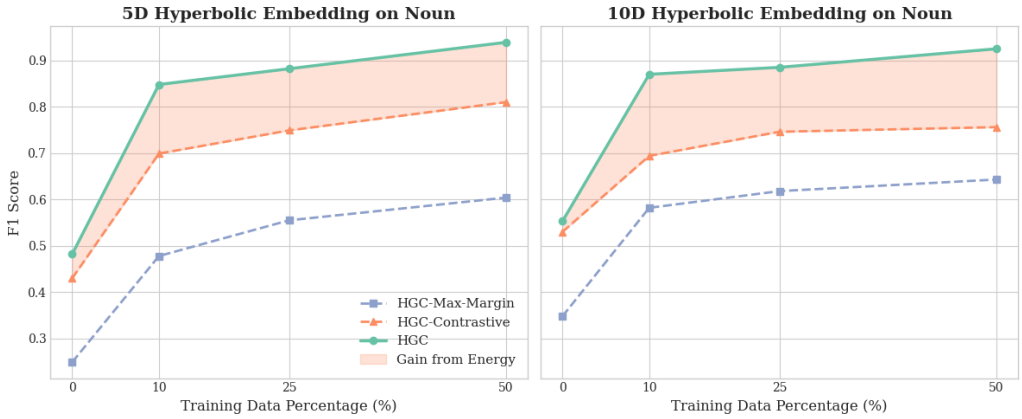


(c) The effect of anisotropy on MCG dataset.

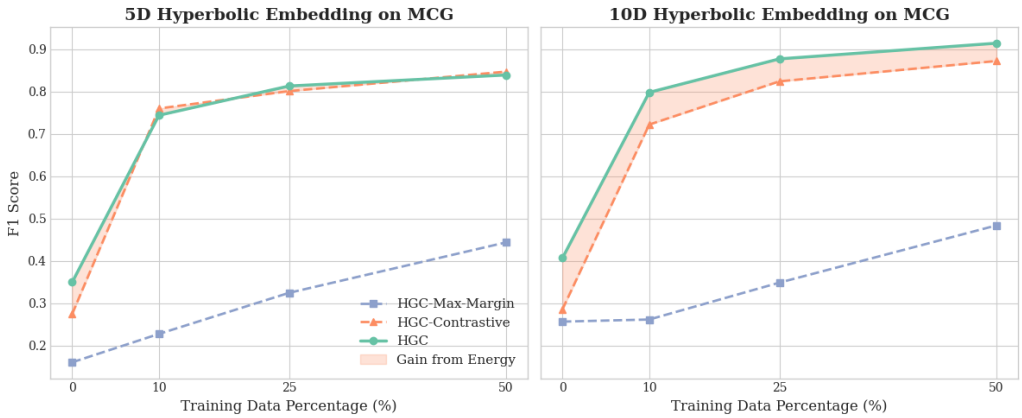
Figure 9: Ablation study on the effect of anisotropy and radial decay. Shaded areas indicate the performance gain attributed to anisotropic modeling across varying training data percentages.



(a) The effect of gravitational energy on Mammal dataset.



(b) The effect of gravitational energy on Noun dataset.



(c) The effect of gravitational energy on MCG dataset.

Figure 10: Ablation study on the effect of continuous gravitational energy. Shaded areas indicate the performance gain attributed to energy function across varying training data percentages.

Collapse in High-Density Spaces. Although the HGC-Contrastive variant generally outperforms the max-margin approach by providing relative ranking signals, a significant performance gap remains relative to the full HGC model (e.g., Noun 10D). We attribute this to the "boundary collapse" phenomenon: unconstrained contrastive objectives tend to greedily push negative samples towards the Poincaré infinity, destabilizing the geometric integrity of the manifold. **(3) Superiority of the Continuous Potential Field.** The full HGC formulation consistently establishes the upper bound of performance (orange shaded areas). By coupling parental mass with a soft-gated potential and axial distance, HGC maintains robust, inward-directed gradients even deep within the cone interior. This mechanism actively enforces "semantic centrality", pulling descendant nodes toward the parent's core axis. Geometrically, this central positioning maximizes the available hyperbolic volume for the child's own entailment cone, ensuring sufficient spatial capacity to accommodate subsequent generations and thereby reconstructing complex, deep taxonomies with high fidelity.