HOLOWEDGE: QUANTUM GRAVITY INSPIRED LEARNING OF HIERARCHICAL EMBEDDINGS

Anonymous authors
Paper under double-blind review

ABSTRACT

Hyperbolic embeddings provide a natural representation of partially ordered data by encoding hierarchical relationships geometrically. However, existing learning methods suffer from undesirable features, including concentration of embedded data near the boundary, exclusion of the central region of hyperbolic space, and the necessity of pretraining using an external model. We introduce HoloWedge, a new framework for learning embeddings of hierarchical data inspired by theoretical physics, specifically the holographic approach to quantum gravity that relates the interior and the boundary of hyperbolic space. HoloWedge represents each data point by a subregion of hyperbolic space, a so-called wedge, which is bounded by a minimal-area hypersurface. We construct loss functions from areas of minimal surfaces, which encourage the containment relation between wedges to reflect the partial order within the underlying data. In the holographic model of theoretical physics, areas of minimal surfaces are identified with the entropies of the corresponding wedges. Our experiments demonstrate that our quantum gravity-inspired organizing principle outperforms existing approaches for reconstruction tasks in a moderate number of dimensions, while remaining competitive for link prediction tasks. The resulting embeddings are less concentrated at the boundary and utilize the full hyperbolic space. Moreover, the use of entropic ordering in our loss functions reduces sensitivity to initialization and removes the need for pretraining. Our results show that holography-inspired methods can effectively learn hierarchical representations.

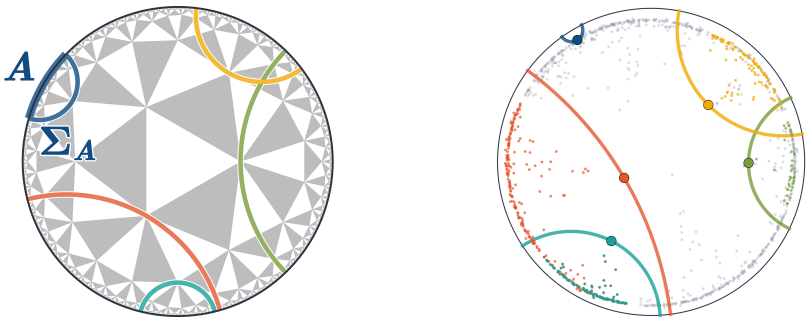
1 INTRODUCTION

Hierarchical structure is exhibited by data of many different origins such as lexical taxonomies (Miller et al., 1990), networks (Clauset et al., 2008) or particle physics (Woodward et al., 2024). Hierarchical representation learning aims at embedding such data in an auxiliary space in a way that reflects its structure. Taking d-dimensional hyperbolic space $\mathbb{H}^d$ as an embedding space seems particularly well suited, since it grows exponentially with distance, mirroring the exponential growth of descendants in tree-like hierarchies. This has led to the development of a variety of methods for learning hyperbolic embeddings of hierarchical data (Nickel & Kiela, 2017; Ganea et al., 2018a; Yu et al., 2024). In fact, hyperbolic geometry is now widely used beyond embeddings across different architectures (Ganea et al., 2018b; Gulcehre et al., 2019; Cetin et al., 2023; van Spengler et al., 2023; Klein et al., 2026) and applications (Mettes et al., 2023).

Some approaches, e.g. Poincaré embeddings (Nickel & Kiela, 2017) or Lorentzian models (Nickel & Kiela, 2018; Law et al., 2019), represent data as points, such that their hierarchical relationships have to be inferred indirectly through distances and norms. Other, subregion-based, methods such as entailment cones (Ganea et al., 2018a) and shadow cones (Yu et al., 2024), model data points as sets and represent the partial order through inclusion, outperforming point-based approaches.

However, existing methods for learning hyperbolic embeddings have practical drawbacks. Point-based methods tend to produce a concentration of embedded data near the boundary of hyperbolic space. Region based models such as Entailment and Poincaré-ball shadow cones need to exclude a ball-shaped region from the center of hyperbolic space, precisely where the most general concepts would be located. This makes initialization problematic for points near the origin.

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.



(a) (b)

Figure 1: **(a)** Organization of information in the holographic AdS/CFT correspondence: the information associated to a boundary region A is encoded in a minimal surface area Σ_A in the interior. The image was created using the Hypertiling library (Schrauth et al., 2025) **(b)** Learned representation of the mammals subtree of WordNet using HoloWedge. The bold circles are select nodes near the root node and arbitrarily colored. Their associated minimal surfaces are displayed. Their associated wedge is the region between the minimal surface and the boundary. The color of the remaining nodes (shown without their respective wedges) is chosen to be that of the closest colored ancestor.

We introduce a new framework which addresses these limitations leveraging a hierarchical encoding from physics, in particular from quantum gravity. According to the AdS/CFT (Anti-de Sitter space/conformal field theory) correspondence (Maldacena, 1999; Ammon & Erdmenger, 2015), gravity in a hyperbolic space $\mathbb{H}^d$ can equivalently be formulated as a quantum theory at the boundary of this space, see Figure 1a. This idea is also referred to as *holography*. The information contained in a subregion A on the boundary is encoded in an associated 'entanglement' wedge $\mathcal{W}_A$ in the gravitational description. According to Ryu & Takayanagi (2006), the amount of information contained in A, captured by the 'entanglement' entropy, can be computed from a minimal area surface Σ_A which is anchored on the boundary of A and separates $\mathcal{W}_A$ and its complement. More generally, areas of minimal surfaces quantify information-theoretic properties of subregions in a geometric and differentiable way, suitable to be used in machine learning.[1]

Building on these ideas we construct *HoloWedge*, a new approach to hyperbolic embeddings inspired by quantum gravity. Instead of embedding data as points, we represent it as a partially ordered set of wedges in hyperbolic space which correspond to ball-shaped subregions of the boundary, see Figure 1b. In this case, the minimal surfaces are simply hyperbolic planes and the wedges are hyperbolic half-spaces. We use areas of minimal surfaces as building blocks of appropriate loss functions, containing an 'entropy' term accounting for generality and radial ordering as well as a 'conditional mutual information' term which rewards containment of related wedges and penalizes proximity of unrelated wedges. In addition to yielding state-of-the-art performance, our construction does not suffer from the issues that affect existing approaches. Our main contributions are:

- We develop a new method for learning embeddings of data in hyperbolic space which is directly inspired from quantum gravity. Hierarchical relations are represented through *wedge inclusion*, making transitivity a direct consequence of set inclusion.
- We propose a natural and differentiable geometric function which quantifies relative generality and relation analogous to (quantum) information theoretic concepts of von Neumann entropy and conditional mutual information.
- The construction is well-defined at every positive radius and supports stable training from random initialization, without pretraining or excluding a neighborhood of the origin. Embeddings remain well within the hyperbolic interior, mitigating boundary crowding.
- Our method outperforms the evaluated baselines on reconstruction tasks in moderate embedding dimensions and achieves competitive performance on link prediction.

[1]For other works relating AdS/CFT and machine learning, see (Hashimoto et al., 2018; Hashimoto, 2019; Ahn et al., 2025; Deb & Sanghavi, 2025; Mirafzali et al., 2026).

2 RELATED WORK

Order Embeddings (Vendrov et al., 2016) models partial order by considering subset relations between cones in Euclidean space, anchored at points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N_+$. A relation $\mathbf{x} \preceq \mathbf{y}$ exists iff $x_i \geq y_i$ for all coordinates $i = 1, ..., N$, and the energy $E(\mathbf{x}, \mathbf{y}) = \| \max(0, \mathbf{y} - \mathbf{x})\|^2$ measures the degree to which this condition is violated for an ordered pair of points (x, y). This model yields strong results on predicting withheld relations of the WordNet hierarchy. However, representation capacity is limited by the fact that all cones defined in the above way eventually intersect. Other approaches for Euclidean models include Box Embeddings by Boratko et al. (2021) and Zhang et al. (2022), which represent entailment relations through containment between axis-aligned boxes. The performance of these models is competitive with approaches in hyperbolic space on reconstructing hierarchical relations and predicting unseen links.

The utility of hyperbolic space for embedding partially ordered data was first exploited in the Poincaré Embeddings model Nickel & Kiela (2017), where data points are mapped onto the Poincaré ball $\mathbb{B}^d$. Geometric proximity as measured by the hyperbolic distance $d_{\mathbb{B}}(\mathbf{x}, \mathbf{y})$ encodes semantic similarity. The coordinates of the embedded points are trained using approximate Riemannian Gradient Descent to find the optimal configuration. The model achieves strong results for the reconstruction task. However, since $d_{\mathbb{B}}(\mathbf{x}, \mathbf{y})$ is symmetric, the model has difficulty encoding ordered relations. Moreover, Poincaré Embeddings arrange most points very close to the boundary of hyperbolic space, obstructing visualization and limiting the expressivity. To address these shortcomings, in Ganea et al. (2018a) every datum is represented by a subregion of $\mathbb{B}^d$, the so-called Entailment Cone $\mathcal{S}_{\mathbf{x}}$ with apex $\mathbf{x}$. A hierarchical relation between two points $\mathbf{x}, \mathbf{y}$ corresponds to the inclusion $\mathbf{y} \in \mathcal{S}_{\mathbf{x}}$. Requiring certain symmetry properties as well as transitivity determines an optimal form for $\mathcal{S}_{\mathbf{x}}$. However, the resulting cones are only well-defined for $\|\mathbf{x}\| \geq \epsilon$, making initialization challenging. To address this issue, Ganea et al. initialize their model with Poincaré embeddings pretrained using the method of (Nickel & Kiela, 2017), which they report yields their best results.

Yu et al. (2024) generalized the entailment cone approach to the Shadow Cone framework, defined both for the Poincaré ball and the Poincaré half-space model of hyperbolic geometry. Their framework is physics inspired and models entailment relations by overlapping shadows of objects under a common light source. This approach yields very strong performances on link prediction tasks for several datasets. However, it remains unclear to what extent the improvement can be attributed to the unbounded Euclidean domain of the used Poincaré half-space model, which allows the embedded points to spread further apart than in the Poincaré ball. For the Poincaré ball geometry, again, points must not be initialized too close to the origin, such that this framework also requires an additional step of initialization. Like the entailment cones of Ganea et al., the Poincaré-ball formulations of Shadow Cones exclude a central region where the cones are undefined.

3 HOLOWEDGE EMBEDDINGS

3.1 GEOMETRIC SETUP

Hyperbolic space $\mathbb{H}^d$ is the unique simply connected Riemannian manifold of negative curvature in d dimensions. Our choice is motivated by the AdS/CFT correspondence (Maldacena, 1999; Ammon & Erdmenger, 2015), where $\mathbb{H}^d$ describes the spatial part of anti-de Sitter (AdS) spacetime. Different choices of coordinates on this manifold are known as different 'models' of hyperbolic space, even though they ultimately describe the same underlying geometry. In our work, we focus on the Poincaré ball model, although a formulation in the half-space is also possible.

For the purpose of the discussion below, we will find it useful to use coordinates $(\rho, \theta, \dots)$ which cover the full hyperbolic ball, such that the metric is given by the line-element[2]

$$ds^2 = L^2 \left(d\rho^2 + \sinh^2 \rho \, d\Omega^2_{d-1} \right). \tag{1}$$

Here, $d\Omega^2_{d-1} = d\theta^2 + \sin^2 \theta \, d\Omega^2_{d-2}$ is the line-element of a unit-sphere. We singled out a polar coordinate $\theta \in (0, \pi)$, which can be thought of as parameterizing lines of equal latitude. Moreover,

[2]The line-element measures the infinitesimal distance-squared between two points and thus defines the metric. For example, for two-dimensional Euclidean space it is $ds^2 = dx^2 + dy^2$.

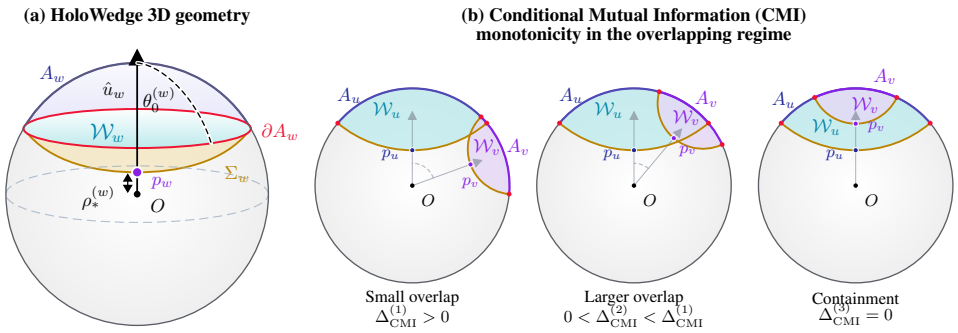


Figure 2: HoloWedge 3D geometry and monotonicity of conditional mutual information (CMI). Panel (a) shows the wedge $\mathcal{W}_w$, bounded by the boundary cap A_w and the minimal hypersurface Σ_w, with p_w at its deepest point. Panel (b) follows two wedges of fixed sizes within the overlapping regime. Decreasing γ_{uv} increases their overlap and lowers $\Delta_{\mathrm{CMI}}(u, v)$ until containment is reached, where $\Delta_{\mathrm{CMI}}(u, v) = 0$.

$\rho \in [0, \infty)$ is a radial coordinate and L is the curvature radius which we set to 1 in the following, such that the sectional curvature of our hyperbolic space is -1. Using our parametrization, the distance from the origin O to a point at radial coordinate ρ is simply ρ, while the boundary of $\mathbb{H}^d$ is approached as $\rho \to \infty$ and is (up to an overall scale factor) identified with the $(d-1)$-dimensional sphere S^{d-1}.

Note that this coordinate choice is not well suited for a numerical implementation due to the unboundedness of the ρ coordinate. Therefore, for all experiments reported in section 4, we express the same hyperbolic points in the Poincaré ball model $\mathbb{B}^d$ commonly used in hyperbolic representation learning (Nickel & Kiela, 2017; Ganea et al., 2018a; Yu et al., 2024), which can be obtained by the coordinate change $\rho = 2\,\mathrm{arctanh}(r)$ as discussed in Appendix C.

We embed all datapoints as subregions of $\mathbb{H}^d$, which are bounded by minimal area surfaces, i.e. hyperbolic planes, and which we call *wedges*. To define these subregions and introduce other useful quantities, consider a unit vector $\hat{u} \in S^{d-1}$ which selects a point on the boundary sphere, and a half-opening angle $\theta_0 \in (0, \pi/2]$. Together they define a boundary spherical cap

$$A(\hat{u}, \theta_0) = \left\{ \xi \in S^{d-1} \,\middle|\, \hat{u} \cdot \xi \geq \cos\theta_0 \right\}, \qquad \partial A = \left\{ \xi \in S^{d-1} \,\middle|\, \hat{u} \cdot \xi = \cos\theta_0 \right\} \cong S^{d-2}, \quad (2)$$

where ∂A denotes the boundary of the region A. The vector $\hat{u}$ points toward the center of the cap and defines its symmetry axis, while θ_0 determines its angular size.

Given A, there exists a minimal area surface Σ_A which is anchored on ∂A. Because the cap is rotationally symmetric around $\hat{u}$, the minimal surface has the same symmetry. Its shape is therefore completely determined by a curve in the two-dimensional (ρ, θ) meridian plane, while the full surface extends along the remaining $d-2$ dimensions. Choosing the polar axis along $\theta = 0$ we obtain the radial profile derived in Appendix D:

$$\rho(\theta) = \mathrm{arctanh}\left(\frac{\cos\theta_0}{\cos\theta} \right), \qquad 0 \leq \theta < \theta_0. \quad (3)$$

The surface reaches its minimum radial coordinate $\rho_* = \mathrm{arctanh}(\cos\theta_0)$ on the symmetry axis. For a cap with a different axis, the same profile holds in polar coordinates aligned with that axis. The cap A and the surface Σ_A enclose a subregion of $\mathbb{H}^d$ which we denote by $\mathcal{W}$ and refer to as the wedge associated with A.

To define our embedding, let w denote an abstract object in the dataset. We represent w as a wedge $\mathcal{W}_w$ with an associated cap region in direction $\hat{u}_w$ and angular size $\cos\theta_0^{(w)} = \tanh\rho_*^{(w)}$, parametrizing it by $p_w = (\rho_*^{(w)}, \hat{u}_w)$. This is illustrated in Figure 2a.

3.2 WEDGE ORDERING & ENTROPIC LEARNING

For two data points p_u and p_v, the angular separation between their axes is $\gamma_{uv} = d_{S^{d-1}}(\hat{u}_u, \hat{u}_v) = \arccos(\hat{u}_u \cdot \hat{u}_v)$. The relative geometry of the two wedges is therefore completely determined by $\theta_0^{(u)}$, $\theta_0^{(v)}$, and γ_{uv}. In particular, A_v is contained in A_u iff $\gamma_{uv} + \theta_0^{(v)} \leq \theta_0^{(u)}$. It can be shown that containment of the boundary caps implies a nesting of the corresponding bulk wedges (Wall, 2014),

$$A_u \subseteq A_v \quad \Rightarrow \quad \mathcal{W}_u \subseteq \mathcal{W}_v. \tag{4}$$

We model an ancestor–descendant relation as the inclusion of the wedge associated to the descendant in the wedge of the ancestor. Since the order is defined through set inclusion, transitivity follows automatically. In contrast, entailment cones (Ganea et al., 2018a) encode the order through point membership and guarantee transitivity by imposing the nested-cone condition $p_u \in \mathcal{C}_v \implies \mathcal{C}_u \subseteq \mathcal{C}_v$, with a similar condition used in (Yu et al., 2024). This condition results in an ϵ-ball at the center of hyperbolic space where no points can be located. In our approach transitivity imposes no additional conditions and thus wedges can be placed arbitrarily close to the center.[3]

Our containment rule is directly motivated by the holographic theory underlying our construction. In the AdS/CFT correspondence, equation 4 is known as *entanglement-wedge nesting* (Wall, 2014). In this geometric picture, detailed information is associated with the near-boundary region, whereas coarser information is represented deeper in the interior. Our method transfers this principle to hierarchical representation learning: specific objects correspond to smaller wedges near the boundary, whereas more general objects correspond to larger wedges extending farther into the interior, see Fig. 1b. The Ryu-Takayanagi (RT) formula (Ryu & Takayanagi, 2006) quantifies the information content of subregions on the boundary. It relates the area of Σ_A to the von Neumann entropy of A (the quantum analogue of Shannon entropy (Nielsen & Chuang, 2010)), quantifying the degree of randomness of the quantum state on A. Motivated by this relation, we associate with each embedded datum the following entropic measure:

$$S(w) \equiv S(A_w) = \log\left(\sin\theta_0^{(w)}\right). \tag{5}$$

In $d = 2$, where A_w is an interval, this has precisely the form of the von Neumann entropy.[4] Since the area of Σ_w is dimension-dependent, the functional form of equation 5 changes in holography for $d > 2$. For ease of evaluation, in our approach we use the $d = 2$ result rather than evaluating the (more complicated) exact dimension-dependent minimal surface area.

For $0 < \theta_0^{(w)} \leq \pi/2$, $S(w)$ increases monotonically with the opening angle, and hence with the size of $\mathcal{W}_w$. Using equation 5 we can also compare wedges pairwise. Given two wedges $\mathcal{W}_u$ and $\mathcal{W}_v$ we quantify their relative configuration through the conditional mutual information (CMI) (Fawzi & Renner, 2015) between their non-shared parts, conditioned on their intersection:

$$\Delta_{\mathrm{CMI}}(u, v) \equiv S(A_u) + S(A_v) - S(A_u \cup A_v) - S(A_u \cap A_v). \tag{6}$$

The entropies of the union $S(A_u \cup A_v)$ and intersection $S(A_u \cap A_v)$ are again defined as the areas of minimal surfaces which end on the boundaries of the respective regions. In general dimensions, there is no closed form expression for the corresponding area and thus again we use the $d = 2$ result in any dimension, where both cases reduce to the entropy of (unions) of intervals. In the case where $A_u \cup A_v$ consists of two disjoint intervals, there are two candidate minimal surfaces, the smaller of which is chosen

$$S(A_u \cup A_v) = \min\{S_{\mathrm{conn}}, S_{\mathrm{disc}}\} = \min\{S_{\mathrm{conn}}, S(A_u) + S(A_v)\}. \tag{7}$$

Figure 3 illustrates this exchange between the two configurations, while the complete construction is given in Appendix E. The connected branch gives $\Delta_{\mathrm{CMI}} > 0$, whereas the disconnected branch gives $\Delta_{\mathrm{CMI}} = 0$.

[3]The angular direction $\hat{u}_w$ is undefined only at the isolated origin. In our implementation, this coordinate singularity is handled numerically without imposing a finite lower bound on the radial coordinate.

[4]More precisely, the entropy has the form $S(A) = a\log(\sin\theta_0^{(w)}) + b_{\mathrm{cutoff}}$, where the scale $a > 0$ and offset b_{cutoff} are the same for every wedge. The offset arises because the boundary is at infinite hyperbolic distance, so the surface area is evaluated only up to a common cutoff. Removing the offset and dividing by a therefore gives equation 5 without affecting the geometric ordering.

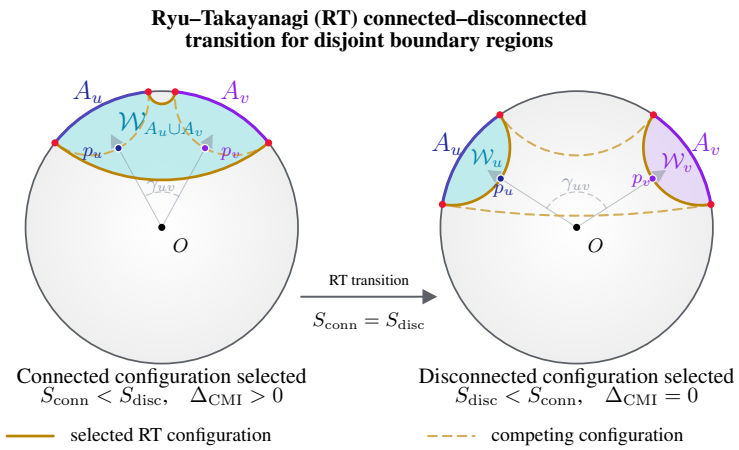


Figure 3: Competing Ryu–Takayanagi (RT) configurations for two disjoint boundary regions.

Intuitively, Δ_{CMI} measures information shared between the overlapping parts of the wedges after their common region has been taken into account. Strong subadditivity guarantees that $\Delta_{\text{CMI}} \geq 0$ (Nielsen & Chuang, 2010; Witten, 2020). If $A_u \cap A_v = \emptyset$, equation 6 reduces to the mutual information between the two disjoint boundary regions. If $A_u \subseteq A_v$, the four terms in equation 6 cancel, also yielding $\Delta_{\text{CMI}}(u, v) = 0$. Figure 2b summarizes the behavior throughout the overlapping regime until containment is reached.

As just explained, two geometrically distinct configurations — containment and sufficiently separated disjoint regions — can have the same CMI. To distinguish these two cases, we introduce the signed angular gap $q_{uv} = \gamma_{uv} - \theta_0^{(u)} - \theta_0^{(v)}$. Overlapping regions satisfy $q_{uv} < 0$, external tangency corresponds to $q_{uv} = 0$, and disjoint regions satisfy $q_{uv} > 0$. We turn this geometric transition into the smooth branch factor

$$b(q_{uv}) = \tanh\left(\frac{q_{uv}}{T_{\text{b}}}\right), \qquad T_{\text{b}} > 0, \tag{8}$$

where the branch temperature T_{b} controls the smoothness around $q_{uv} = 0$. For finite T_{b}, this factor regularizes the overlap–disjoint boundary at external tangency into a smooth crossover.

Combining this phase information with the CMI, we define the symmetric geometric score

$$z_{\text{CMI}}(u, v) = 1 + b(q_{uv}) \exp[-\Delta_{\text{CMI}}(u, v)]. \tag{9}$$

The exponential factor approaches one as the CMI vanishes, while $b(q_{uv})$ distinguishes whether this saturation occurs in the overlapping or disjoint branch, such that $z_{\text{CMI}}(u, v)$ vanishes if one wedge is contained in the other and continuously approaches two as they are moved further apart. Both q_{uv} and $\Delta_{\text{CMI}}(u, v)$ are invariant under the exchange $u \leftrightarrow v$. Hence $z_{\text{CMI}}(u, v) = z_{\text{CMI}}(v, u)$, and this score characterizes the relative configuration of the two wedges without determining the direction of their hierarchical relation.

To introduce directionality, we introduce a radial entropic-ordering penalty with margin m_{rad}

$$R(u, v) = \max\{0,\ S(u) + m_{\text{rad}} - S(v)\}, \qquad m_{\text{rad}} \geq 0, \tag{10}$$

where from this point onward we use (u, v) to denote an ordered candidate relation: u is the more specific descendant and v is a more general ancestor. Unlike $z_{\text{CMI}}(u, v)$, this term is generally not invariant under exchanging u and v. The penalty vanishes when $S(v) \geq S(u) + m_{\text{rad}}$. Since the entropy in equation 5 increases monotonically with the opening angle, this condition requires the candidate parent v to represent a larger boundary cap, a larger wedge, and a point closer to the origin than the child u.

We finally define the directional relation energy, which is minimal if $\mathcal{W}_v$ contains $\mathcal{W}_u$,

$$E(u, v) = z_{\text{CMI}}(u, v) + \lambda_{\text{rad}} R(u, v). \qquad \lambda_{\text{rad}} \geq 0. \tag{11}$$

3.3 Loss Functions

Point-based Poincaré embeddings use the symmetric hyperbolic distance as a pairwise compatibility measure (Nickel & Kiela, 2017). The symmetric score $z_{\mathrm{CMI}}(u, v)$ plays an analogous role in HoloWedge, but replaces point-to-point proximity with a richer comparison of wedge configurations. Hierarchical directionality is introduced by $R(u, v)$, which is analogous in purpose to the asymmetric partial order encoded by entailment cones (Ganea et al., 2018a).

We will optimize the directional relation energy, equation 11, which combines both quantities, using and comparing both cross-entropy and max-margin loss, defined below. For each observed relation $(u, v_+) \in \mathcal{D}$ in a given dataset, we construct the candidate set $\mathcal{C}(u) = \left\{v_+, v_1^-, \ldots, v_K^-\right\}$, containing the observed parent v_+ and K sampled candidate parents v_i^- that do not form observed relations with u. The cross-entropy loss (Nickel & Kiela, 2017) is:

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{|\mathcal{D}|} \sum_{(u,v_+)\in\mathcal{D}} \log \frac{\exp[-E(u, v_+)/T_{\mathrm{rank}}]}{\sum_{v\in\mathcal{C}(u)} \exp[-E(u, v)/T_{\mathrm{rank}}]}. \tag{12}$$

Here, $T_{\mathrm{rank}} > 0$ is the ranking temperature. Minimizing $\mathcal{L}_{\mathrm{CE}}$ encourages the observed parent v_+ to have lower energy than the sampled candidate parents. In HoloWedge, this means favoring a pair of compatible, overlapping wedges with the correct parent–child entropy ordering.

Alternatively, we optimize the same energy using a relative energy-gap hinge loss inspired by the max-margin loss functions used for Order Embeddings and hyperbolic entailment cones (Vendrov et al., 2016; Ganea et al., 2018a):

$$\mathcal{L}_{\mathrm{hinge}} = \frac{1}{|\mathcal{D}|} \sum_{(u,v_+)\in\mathcal{D}} \left[E(u, v_+) + \frac{\lambda_{\mathrm{neg}}}{K} \sum_{i=1}^{K} \max\left\{0,\, m_{\mathrm{energy}} + E(u, v_+) - E(u, v_i^-)\right\} \right]. \tag{13}$$

Here, $m_{\mathrm{energy}} > 0$ is the required energy separation and $\lambda_{\mathrm{neg}} \geq 0$ controls the contribution of the negative-ranking term. The first term directly lowers the energy of observed relations, while the hinge term penalizes a candidate parent v_i^- only when $E(u, v_i^-) < E(u, v_+) + m_{\mathrm{energy}}$.

4 Experiments

In this section, we compare the performance of our models on the tasks of **reconstruction** and **link prediction** against Poincaré embeddings (Nickel & Kiela, 2017), Entailment Cones (Ganea et al., 2018a), and Shadow Cones (Yu et al., 2024). The former two models are implemented within our code, so we reproduce the results for all datasets. For the shadow cone model, we cite available results achieved on the Poincaré ball. Tests are performed on four datasets, which exhibit a hierarchical structure: WordNet (Fellbaum, 1998) and its mammal subtree, Hearst (Hearst, 1992; Le et al., 2019), and a generated fully ternary tree.

4.1 Evaluation tasks and dataset preparation

For reconstruction, we train each model on the full transitive closure of the respective datasets. The original data is then reconstructed from this embedding via a ranking procedure. Each of the considered models has an energy function $E(u, v)$, which the model is trained to minimize for each correct link (u, v). We denote the set of all links in the original data as $\mathcal{D} = \{(u, v)\}$. To evaluate the embedding, the energy for each positive link (u, v) is ranked among the set of negative links involving u, i.e., the set $\{E(u, v') \mid (u, v) \notin \mathcal{D}\}$. We record the mean rank over all observed links, as well as the mean average precision of the ranking, and use these two parameters to evaluate our embeddings. Additionally, we record the mean squared norm of the embedding, in order to quantify how well the learned embeddings are distributed in hyperbolic space.

For link prediction we follow the strategy of Ganea et al. (2018a). We split the full transitive closure of our data into basic links, which are part of the transitive reduction, and non-basic links. Basic links are always included in the training data. The remaining, non-basic links are split in the following way: Between 0 and 90% of these links are added to the training set, while 5% of the non-basic links

Table 1: Reconstruction results on the mammal subtree of WordNet, the complete WordNet noun hierarchy, and a synthetically generated fully ternary tree with three root nodes and a depth of five. Lower mean rank (MR) and higher MAP are better; best numbers are **bolded**.

		Mammals			**Nouns**			**Ideal tree**		
	Dim.	5	10	20	5	10	20	5	10	20
Poincaré emb.	MR ↓	1.328	1.260	1.269	18.566	14.998	14.042	1.170	1.140	1.157
	MAP ↑	0.937	0.942	0.943	0.574	0.635	0.648	0.958	0.964	0.961
	$\langle \|\|x\|\|^2 \rangle$	0.997	0.997	0.996	1.000	1.000	1.000	0.996	0.996	0.995
Ent. cones	MR ↓	1.748	1.811	2.164	31.356	25.670	23.359	**1.053**	1.021	1.011
	MAP ↑	0.936	0.939	0.939	**0.819**	**0.861**	0.879	**0.990**	0.996	0.998
	$\langle \|\|x\|\|^2 \rangle$	0.673	0.670	0.669	0.670	0.669	0.669	0.675	0.672	0.668
HoloWedge CE loss	MR ↓	**1.241**	**1.037**	**1.050**	**17.564**	**4.968**	**2.716**	1.158	**1.005**	**1.008**
	MAP ↑	**0.965**	**0.991**	**0.988**	0.692	0.828	**0.906**	0.971	**0.999**	**0.999**
	$\langle \|\|x\|\|^2 \rangle$	0.841	0.794	0.766	0.858	0.802	0.777	0.820	0.771	0.750
HoloWedge Hinge loss	MR ↓	1.368	1.106	1.096	24.147	10.681	7.942	1.357	1.109	1.121
	MAP ↑	0.953	0.986	0.983	0.609	0.753	0.792	0.940	0.979	0.979
	$\langle \|\|x\|\|^2 \rangle$	0.732	0.697	0.669	0.751	0.706	0.691	0.668	0.648	0.629

form the validation set and another 5% form the test set. Additionally, for both reconstruction and link prediction, we generate a fixed number of negative examples for each positive link in the training set. We can generate these negatives either by keeping the child node and corrupting the parent node, or by keeping the parent and corrupting the child. For the entailment cone model, we generate ten child-corrupted negatives for each positive link, for all other models ten parent-corrupted negatives, because these choices yield the best results for the respective models.

The datasets which we use to train our models are directed acyclic graphs. For WordNet, its mammal subtree, and the ideal tree, this is true a priori, whereas for Hearst we can ensure this by following the procedure of Yu et al. (2024), i.e., we cut all cycles appearing in these datasets open by randomly removing one node. Moreover, for Hearst, we build the transitive closure only from the 50,000 links with the highest confidence score, such that the size of the dataset does not become too large.

4.2 Training details

Following Nickel & Kiela (2017), we initialize the embedded points for all models by sampling the coordinates of each point from a uniform distribution over the interval $[-\sigma_{\text{init}}, \sigma_{\text{init}}]$ in every dimension. For the Poincaré embeddings model, we use their initialization scale of $\sigma_{\text{init}} = 10^{-4}$. For our models, we obtained the best results for $\sigma_{\text{init}} = 0.01$. A more detailed discussion of the initialization scale can be found in Appendix A, where we demonstrate that our approach is less sensitive to this parameter compared to Poincaré embeddings. On the smaller datasets (mammals and ideal tree), we train our models for 500 epochs for the reconstruction task and 300 epochs for link prediction. The larger datasets (nouns and Hearst) require longer training, therefore we use 1500 epochs for reconstruction and 500 for link prediction. The first 100 epochs of training for the entailment cone model are done with the Poincaré embeddings algorithm, and the coordinates are subsequently rescaled by a factor of 0.7 following Ganea et al. (2018a). We found it advantageous to use a bigger learning rate on the larger datasets. For Poincaré embeddings, we use a learning rate of 0.3 for the smaller datasets and 1.0 for the larger ones, for the entailment cone model we use a learning rate of 0.003 for smaller and 0.01 for larger datasets. Our models use a learning rate of 0.075 for small and 0.3 for large datasets. Note that we use a burn-in strategy for all models, reducing the learning rate by a factor of 0.01 for the first 20 epochs. Additional choices for hyperparameters can be found in Appendix A.

4.3 Results and discussion

To illustrate our results, figure 1b shows the embedding of the mammals dataset obtained by our HoloWedge model with cross-entropy loss in two dimensions after 500 epochs of training.

The performance of our models on the reconstruction task can be found in Table 1. For the mammals dataset, our HoloWedge model with cross-entropy loss yields the best performance in 5, 10, and 20

Table 2: F1 score (%) on link prediction for WordNet Noun and Hearst, with best numbers **bolded**. Higher is better.

Noun	$d = 5$					$d = 10$				
Non-basic edge percentage	0%	10%	25%	50%	90%	0%	10%	25%	50%	90%
Poincaré embeddings	32.4	72.0	78.5	79.8	80.8	34.4	76.8	81.2	82.0	82.7
Entailment cones	21.6	**77.0**	**80.2**	**82.5**	**87.3**	24.8	**80.5**	**84.1**	**85.4**	**86.8**
HoloWedge - CE loss	30.0	71.3	75.0	75.9	76.1	33.3	78.4	81.4	82.2	81.7
HoloWedge - Hinge loss	**32.7**	68.0	71.6	70.6	69.0	**38.1**	73.1	75.0	75.9	76.1
Hearst	$d = 5$					$d = 10$				
Non-basic edges	0%	1%	2%	5%		0%	1%	2%	5%	
Poincaré embeddings	21.5	40.1	43.3	45.0		22.6	40.9	43.8	46.6	
Entailment cones	17.0	47.3	59.1	70.2		17.3	43.9	59.0	72.2	
HoloWedge - CE loss	**24.2**	59.4	**67.9**	**73.7**		**25.0**	61.1	**68.4**	**74.4**	
HoloWedge - Hinge loss	21.5	**63.3**	66.0	67.7		22.2	**66.6**	67.5	68.6	

dimensions, while the model with a hinge loss also outperforms previous approaches in 10 and 20 dimensions. The HoloWedge model with cross-entropy loss also performs best for an ideal tree in 10 and 20 dimensions, and it yields the best mean rank for the nouns dataset in 5 and 10 dimensions as well as the best overall performance in 20 dimensions. Note that in all cases, the values of the average squared norm $\langle \|x\|^2 \rangle$ show that our models arrange the points in a more central region of hyperbolic space compared to Poincaré embeddings, which heavily concentrate points near the boundary, where $\|x\|^2 = 1$. Together with Figure 1b, this demonstrates that our models utilize the full hyperbolic space. Entailment cones also arrange points in a more central region compared to Poincaré embeddings, but this can largely be attributed to the rescaling by a factor of 0.7 applied after pretraining. The results on link prediction for the larger datasets (full WordNet data and Hearst) are shown in table 2, while the results for the smaller datasets (mammals subtree of WordNet and an ideal tree) are shown in Table 3 in appendix B. On the ideal tree and the Hearst dataset, we outperform previous approaches in all but one setting, and for the WordNet hierarchy and its mammals subtree, performance remains competitive with the other models. Overall, the strong reconstruction performance of our HoloWedge approach indicates that partially ordered data are represented while preserving most of the information. Moreover, its competitive performance on link prediction shows that our models are able to infer previously unseen hierarchical relations.

5 Conclusion

In this work we have introduced HoloWedge, an entropy-driven embedding algorithm for hierarchical data in hyperbolic space. Data points are represented by wedges of hyperbolic space and their relations are quantified using entropic concepts inspired from the holographic approach to quantum gravity.

Since the wedges are hyperbolic half-spaces, our approach can be understood as a combination of embeddings and SVM in hyperbolic space (Cho et al., 2019; Weber et al., 2020; Chien et al., 2021), where to each data point we assign a hyperplane separating hierarchically subordinate data points from the rest.

Our method underlines the power of hyperbolic methods for hierarchical data by placing them into the context of holography. In addition, holography suggests that embedding data in hyperbolic space can equivalently be formulated as embedding associated caps or, correspondingly, ball-shaped regions on the lower-dimensional boundary of hyperbolic space, where hierarchical relations are again represented by nesting. Indeed, note that our loss function is formulated using only properties of the cap regions on the boundary of $\mathbb{H}^d$. Moreover, we can transform our model to the half-plane model of hyperbolic space by a coordinate transformation, such that each cap region on the boundary of $\mathbb{H}^d$ is equivalently described by a ball on $\mathbb{R}^{d-1}$. This provides a theoretical explanation for why region embeddings in Euclidean space can achieve a comparable performance to hyperbolic models, c.f. Boratko et al. (2021); Zhang et al. (2022), and suggests to reformulate other architectures utilizing hyperbolic space in the language of holography.

AI USE STATEMENT

In this work, we have not used generative AI tools for any tasks with required disclosure. Regarding tasks with recommended disclosure, we have used generative AI tools for creating and editing software code (including TikZ code for some of the figures), for identifying relevant literature, and for aiding and polishing writing. We have reviewed all AI-assisted work. In particular, all LLM-generated or modified code, including TikZ code for the images, was verified and tested for correctness by 2 authors. AI-assisted literature search was used only to supplement our manual search, and all suggested works were verified by 3 authors for relevance and correct citation. LLM-made suggestions for writing were proofread by all authors and checked for logical and linguistic errors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work raises no ethical concerns.

REPRODUCIBILITY STATEMENT

We have taken the following steps, in order to make our results reproducible. We provide an anonymized repository, which contains the source code for our models, as well as the baselines of Poincaré embeddings and entailment cones which we compare against, at https://anonymous.4open.science/r/HoloWedge-7198/. The parameters which we used to obtain our results are specified in Section 4.2 and Appendix A. The steps taken to process the used datasets are described in Section 4.1. Additional information on the geometric setup is given in Appendix C, where we derive the Poincaré-ball coordinates used in our implementation. Furthermore, we derive the Ryu–Takayanagi surface profile in Appendix D and discuss the connected and disconnected configurations in Appendix E.

REFERENCES

Byoungjoon Ahn, Hyun-Sik Jeong, Keun-Young Kim, and Kwan Yun. Holographic reconstruction of black hole spacetime: machine learning and entanglement entropy. *JHEP*, 01, 2025.

Martin Ammon and Johanna Erdmenger. *Gauge/Gravity Duality: Foundations and Applications*. Cambridge University Press, 2015.

Michael Boratko, Dongxu Zhang, Nicholas Monath, Luke Vilnis, Kenneth L Clarkson, and Andrew McCallum. Capacity and bias of learned geometric embeddings for directed graphs. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34. Curran Associates, Inc., 2021.

Edoardo Cetin, Benjamin Paul Chamberlain, Michael M. Bronstein, and Jonathan J Hunt. Hyperbolic deep reinforcement learning. In *The Eleventh International Conference on Learning Representations*, 2023.

Eli Chien, Chao Pan, Puoya Tabaghi, and Olgica Milenkovic. Highly scalable and provably accurate classification in poincaré balls. In James Bailey, Pauli Miettinen, {Yun Sing} Koh, Dacheng Tao, and Xindong Wu (eds.), *Proceedings - 21st IEEE International Conference on Data Mining, ICDM 2021*, Proceedings - IEEE International Conference on Data Mining, ICDM, United States, 2021. Institute of Electrical and Electronics Engineers Inc.

Hyunghoon Cho, Benjamin DeMeo, Jian Peng, and Bonnie Berger. Large-margin classification in hyperbolic space. In Kamalika Chaudhuri and Masashi Sugiyama (eds.), *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*. PMLR, 16–18 Apr 2019.

Aaron Clauset, Cristopher Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453:98–101, 2008.

Anirudh Deb and Yaman Sanghavi. Aspects of holographic entanglement using physics-informed-neural-networks. 9 2025.

Omar Fawzi and Renato Renner. Quantum conditional mutual information and approximate markov chains. *Communications in Mathematical Physics*, 340(2), 2015.

Christiane Fellbaum (ed.). *WordNet: An Electronic Lexical Database*. The MIT Press, 05 1998.

Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic entailment cones for learning hierarchical embeddings, 2018a.

Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks, 2018b.

Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro, and Nando de Freitas. Hyperbolic attention networks. In *International Conference on Learning Representations*, 2019.

Koji Hashimoto. AdS/CFT correspondence as a deep Boltzmann machine. *Phys. Rev. D*, 99(10), 2019.

Koji Hashimoto, Sotaro Sugishita, A. Tanaka, and Akio Tomiya. Deep learning and the ads/cft correspondence. *Physical Review D*, 2018.

Matthew Headrick. Entanglement rényi entropies in holographic theories. *Physical Review D*, 82 (12), December 2010.

Marti A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *COLING 1992 Volume 2: The 14th International Conference on Computational Linguistics*, 1992.

Timo Klein, Thomas Lang, Andrii Shkabrii, Alexander Sturm, Kevin Sidak, Lukas Miklautz, Claudia Plant, Yllka Velaj, and Sebastian Tschiatschek. Understanding and improving hyperbolic deep reinforcement learning. In *The Fourteenth International Conference on Learning Representations*, 2026.

Marc Law, Renjie Liao, Jake Snell, and Richard Zemel. Lorentzian distance learning for hyperbolic representations. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*. PMLR, 09–15 Jun 2019.

Matthew Le, Stephen Roller, Laetitia Papaxanthos, Douwe Kiela, and Maximilian Nickel. Inferring concept hierarchies from text corpora via hyperbolic embeddings. In Anna Korhonen, David Traum, and Lluís Màrquez (eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 2019. Association for Computational Linguistics.

Juan Maldacena. The large-N limit of superconformal field theories and supergravity. *International Journal of Theoretical Physics*, 38(4), April 1999.

Pascal Mettes, Mina Ghadimi Atigh, Martin Keller-Ressel, Jeffrey Gu, and Serena Yeung. Hyperbolic deep learning in computer vision: A survey, 2023.

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. Introduction to wordnet: An on-line lexical database*. *International Journal of Lexicography*, 3 (4), 12 1990.

Ehsan Mirafzali, Sanjit Shashi, Sanya Murdeshwar, Edgar Shaghoulian, Daniele Venturi, and Razvan Marinescu. Holographic generative flows with AdS/CFT. 1 2026.

Maximilian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations, 2017.

Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the Lorentz model of hyperbolic geometry. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*. PMLR, 10–15 Jul 2018.

Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, 2010.

Shinsei Ryu and Tadashi Takayanagi. Holographic derivation of entanglement entropy from AdS/CFT. *Physical Review Letters*, 96, 2006.

Manuel Schrauth, Felix Dusel, Yanick Thurn, Florian Goth, Dietmar Herdt, and Jefferson S. E. Portela. hypertiling, August 2025.

Max van Spengler, Erwin Berkhout, and Pascal Mettes. Poincaré resnet. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Ivan Vendrov, Ryan Kiros, Sanja Fidler, and Raquel Urtasun. Order-embeddings of images and language, 2016.

Aron C. Wall. Maximin surfaces, and the strong subadditivity of the covariant holographic entanglement entropy. *Classical and Quantum Gravity*, 31(22), 2014.

Melanie Weber, Manzil Zaheer, Ankit Singh Rawat, Aditya K Menon, and Sanjiv Kumar. Robust large-margin learning in hyperbolic space. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33. Curran Associates, Inc., 2020.

Edward Witten. A mini-introduction to information theory. *Rivista del Nuovo Cimento*, 43(4), 2020.

Nathaniel S. Woodward, Sang Eon Park, Gaia Grosso, Jeffrey Krupa, and Philip Harris. Product manifold machine learning for physics, 2024.

Tao Yu, Toni J. B. Liu, Albert Tseng, and Christopher De Sa. Shadow cones: A generalized framework for partial order embeddings, 2024.

Dongxu Zhang, Michael Boratko, Cameron Musco, and Andrew McCallum. Modeling transitivity and cyclicity in directed graphs via binary code box embeddings. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35. Curran Associates, Inc., 2022.

A TRAINING DETAILS

Sensitivity to the initialization scale. In the following, we investigate the performance of our model depending on the initialization scale σ_{init}, comparing with the model of Nickel & Kiela (2017). We restrict our numerical analysis to this distance-based model because it uses the same Poincaré-ball model and the same unmodified random initialization, making σ_{init} directly comparable. Standard entailment cones instead start from a Poincaré embedding pretrained for 100 epochs and exclude a finite neighborhood of the origin (Ganea et al., 2018a). Likewise, the Poincaré-ball formulations of Shadow Cones project initialized points outside the central light-source region (Yu et al., 2024). Their nominal initialization scales therefore do not define the same initial distribution and are not included in this controlled comparison. We have, however, run the same tests for Entailment cones, finding that our model is less sensitive to initialization, although only by a small margin.

For our analysis we initialize each Cartesian coordinate independently from $\mathcal{U}(-\sigma_{\text{init}}, \sigma_{\text{init}})$ and vary $\sigma_{\text{init}} \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$, while fixing the training budget to 300 epochs and the random seed to 43. To quantify sensitivity to the initialization scale, we measure the absolute relative change of each reconstruction metric M between two initialization scales:

$$\delta_M(\sigma_a, \sigma_b) = \frac{|M(\sigma_b) - M(\sigma_a)|}{M(\sigma_a)}. \tag{14}$$

Unless stated otherwise, we take $\sigma_a = 10^{-1}$ as the reference and $\sigma_b = 10^{-4}$, so that δ_M measures the change over three orders of magnitude in the initialization scale.

Relative to $\sigma_{\text{init}} = 10^{-1}$, the HoloWedge mean rank and MAP change by only 8.5% and 1.4%, respectively, on Mammals over the full three-order-of-magnitude range, compared to variations of

77.5% in mean rank and 13.3% in MAP between under distance-based training (Nickel & Kiela, 2017). On WordNet Nouns, the corresponding variations are 33.4% and 70.9% (ours), compared with 76.9% and 163.3% (theirs).

Thus, HoloWedge exhibits reduced sensitivity to the initialization scale, particularly in mean rank. This trend is consistent with our reconstruction results across the random seeds 42, 43, and 44.

Further hyperparameter choices. For training on the smaller datasets (mammals and the ideal tree), we use a batch size of 10 for training, whereas for the larger datasets we use a batch size of 50. We use the following values for the parameters in our energy and loss functions, which yield the best performance: a branch temperature of $T_{\text{branch}} = 0.2$, a ranking temperature of $T_{\text{rank}} = 0.3$ for the cross-entropy loss, as well as a radial margin of $m_{\text{rad}} = 0.2$. The radial term of our energy function is weighted with $\lambda_{\text{rad}} = 1.0$, and the negative term in the hinge loss is weighted with $\lambda_{\text{neg}} = 1.0$. Additionally, we use an energy separation of $m_{\text{energy}} = 0.5$ for the hinge loss.

B Additional experimental data

In this section, we record additional performance results for datasets not covered in the main part. In particular, Table 3 shows performance for the link prediction task on the mammals subtree of WordNet, and on an artificially generated ideal tree with three root nodes, a branching factor of three and a depth of five.

Table 3: F1 score (%) on link prediction for the mammal subtree and an ideal tree, with best numbers **bolded**. Higher is better.

Mammals	$d=2$					$d=5$				
Non-basic-edge percentage	0%	10%	25%	50%	90%	0%	10%	25%	50%	90%
Poincaré embeddings	25.1	47.4	50.7	54.6	57.3	42.9	64.5	67.9	65.2	70.5
Entailment cones	27.7	33.3	56.1	63.4	70.8	52.2	**76.7**	**82.0**	82.0	80.2
Shadow cones	**44.6**	**60.8**	**62.7**	**68.4**	67.9	**62.4**	69.5	81.4	**84.4**	**92.6**
HoloWedge - CE loss	28.7	44.4	56.7	66.3	**72.5**	48.3	72.3	77.1	79.5	81.5
HoloWedge - hinge loss	33.6	43.6	55.0	59.2	62.6	43.6	62.4	69.5	69.1	71.4
Ideal tree	$d=5$					$d=10$				
Non-basic-edge percentage	0%	10%	25%	50%	90%	0%	10%	25%	50%	90%
Poincaré embeddings	41.4	53.0	69.6	75.4	81.6	62.5	75.5	83.3	84.6	81.5
Entailment cones	31.3	33.2	52.3	77.9	**95.0**	56.7	73.5	84.3	89.8	96.0
HoloWedge - CE loss	**61.1**	**78.7**	**88.4**	**92.8**	93.2	**73.4**	**85.4**	**92.7**	**96.3**	**97.0**
HoloWedge - hinge loss	56.9	66.2	69.3	74.4	80.3	72.0	71.3	70.9	74.9	78.9

C Derivation of the Poincaré Ball Metric

The global metric of AdS_{d+1} can be written as

$$ds^2_{\mathrm{AdS}_{d+1}} = L^2 \left(-\cosh^2 \rho \, d\tau^2 + d\rho^2 + \sinh^2 \rho \, d\Omega^2_{d-1} \right), \tag{15}$$

where L is the curvature radius and $d\Omega^2_{d-1}$ is the metric of the unit $(d-1)$-sphere. On a constant-time slice, $d\tau = 0$, we obtain the induced metric $ds^2_{\mathbb{H}^d}$ present in the main text in equation 1. We now introduce coordinates on the open unit ball

$$\mathbb{B}^d = \left\{ \boldsymbol{x} \in \mathbb{R}^d \mid \|\boldsymbol{x}\| < 1 \right\}. \tag{16}$$

Writing $\boldsymbol{x} = r\,\widehat{\boldsymbol{u}}$, with $\widehat{\boldsymbol{u}} \in S^{d-1}$, we relate the two radial coordinates by

$$r = \|\boldsymbol{x}\| = \tanh\left(\frac{\rho}{2}\right), \tag{17}$$

while leaving the angular coordinates unchanged. This maps $\rho \in [0, \infty)$ to $r \in [0, 1)$. Differentiating the coordinate transformation gives

$$dr = \frac{1}{2}\left(1 - r^2\right) d\rho. \tag{18}$$

Therefore,

$$d\rho^2 = \frac{4}{(1-r^2)^2}\, dr^2. \tag{19}$$

The angular factor can be written as

$$\sinh\rho = \frac{2\tanh(\rho/2)}{1-\tanh^2(\rho/2)} = \frac{2r}{1-r^2}, \qquad \sinh^2\rho = \frac{4r^2}{(1-r^2)^2}. \tag{20}$$

Substituting these expressions into the hyperbolic metric, we obtain

$$ds^2_{\mathbb{H}^d} = L^2\left[\frac{4}{(1-r^2)^2}\, dr^2 + \frac{4r^2}{(1-r^2)^2}\, d\Omega^2_{d-1}\right] \tag{21}$$

$$= L^2\left(\frac{2}{1-r^2}\right)^2 \left(dr^2 + r^2 d\Omega^2_{d-1}\right). \tag{22}$$

The expression in parentheses is the Euclidean line element in spherical coordinates,

$$ds^2_E = dr^2 + r^2 d\Omega^2_{d-1}. \tag{23}$$

Thus, the constant-time slice of global AdS_{d+1} and the Poincaré ball are two coordinate descriptions of the same hyperbolic space. Note that the surface $r = 1$ is not part of the ball, but corresponds to $\rho \to \infty$ and lies at infinite hyperbolic distance.

D THE RYU-TAKAYANAGI SURFACES IN AdS_{d+1}

On a fixed-time slice $\tau = \tau^*$ of pure global AdS_{d+1}, the bulk geometry is $\mathbb{H}^d$. On this slice, the Ryu–Takayanagi (RT) formula (Ryu & Takayanagi, 2006) relates the entropy of a boundary region A to the area of the so-called RT surface Σ_A in the bulk of spacetime,

$$S(A) = \frac{\mathcal{A}(\Sigma_A)}{4G_N}. \tag{24}$$

Here G_N is Newton's gravitational constant in the $(d+1)$-dimensional bulk, which sets the overall conversion factor between area and entropy. The surface Σ_A has dimension $d-1$, i.e., it is codimension two in AdS_{d+1} and codimension one in $\mathbb{H}^d$. It is anchored on ∂A and homologous to A, meaning that A and Σ_A together bound a bulk region and can continuously be deformed into each other. Among all surfaces satisfying these conditions, the RT surface Σ_A by definition is the one with minimal area.

Area functional. To derive the shape of the RT surface, we use the metric in equation 1 and the cap $A(\hat{u}, \theta_0)$ in equation 2, choosing the polar axis along $\hat{u}$. Rotations and reflections of directions perpendicular to $\hat{u}$ form a symmetry of the configuration. They leave both the cap and the bulk metric unchanged while varying the remaining angular coordinates. Consequently, the RT surface also has this symmetry and extends along a full $(d-2)$-dimensional sphere S^{d-2} at each point of a curve in the (ρ, θ) meridian plane. For this radial profile, the induced metric on the surface is then

$$ds^2 = L^2\Big[\big(\rho'(\theta)^2 + \sinh^2\rho\big)d\theta^2 + \big(\sinh\rho\,\sin\theta\big)^2 d\Omega^2_{d-2}\Big], \tag{25}$$

and its area is computed as

$$\mathcal{A}[\rho] = L^{d-1}\Omega_{d-2}\int_0^{\theta_0} d\theta\, (\sinh\rho\,\sin\theta)^{d-2}\sqrt{\rho'(\theta)^2 + \sinh^2\rho}, \tag{26}$$

where $\rho'(\theta) = d\rho/d\theta$ and Ω_{d-2} is the line element of the unit sphere. Smoothness at the axis requires $\rho'(0) = 0$, while anchoring requires $\rho(\theta) \to \infty$ as $\theta \to \theta_0^-$. The area diverges at the asymptotic boundary, so area comparisons are performed using the same large-ρ cutoff for every surface.

Surface profile. One could now go on and find a function $\rho(\theta)$ which minimizes equation 26. However, the surface minimizing equation 26 can be found directly from the geometry, without numerically searching over candidate surfaces. In the hyperboloid model of $\mathbb{H}^d$, totally geodesic codimension–one submanifolds are obtained by intersecting the hyperboloid with linear hyperplanes through the origin. Here, totally geodesic means that geodesics lying within the intersection are also geodesics of $\mathbb{H}^d$. The extrinsic curvature of such a surface vanishes, and consequently its mean curvature also vanishes; the surface is therefore minimal. For a spherical cap, the rotational symmetry of the boundary condition selects a unique surface within this class.

We represent $\mathbb{H}^d$ as the hyperboloid in $\mathbb{R}^{1,d}$:

$$\mathbb{H}^d = \left\{ X \in \mathbb{R}^{1,d} : -(X^0)^2 + \sum_{i=1}^{d} (X^i)^2 = -L^2, \quad X^0 > 0 \right\}. \tag{27}$$

With the X^1 direction aligned with $\hat{u}$, its coordinates are

$$\begin{aligned} X^0 &= L\cosh\rho, & X^1 &= L\sinh\rho\cos\theta, \\ (X^2,\dots,X^d) &= L\sinh\rho\sin\theta\,\hat{n}, & \hat{n} &\in S^{d-2}. \end{aligned} \tag{28}$$

The symmetry about $\hat{u}$ selects hyperplanes of the form

$$\Pi_k : \quad X^1 - kX^0 = 0. \tag{29}$$

On $\mathbb{H}^d \cap \Pi_k$, the hyperplane equation becomes

$$\tanh\rho\,\cos\theta = k. \tag{30}$$

At the asymptotic boundary, $\rho \to \infty$ and $\theta \to \theta_0$, so the hyperplane equation gives $k = \cos\theta_0$. The anchoring condition therefore selects $\Sigma_A = \mathbb{H}^d \cap \Pi_{\cos\theta_0}$. For a spherical cap in pure AdS, this surface is the unique extremal surface and thus also the RT surface.

Solving for ρ gives the profile in equation 3:

$$\rho(\theta) = \operatorname{arctanh}\left(\frac{\cos\theta_0}{\cos\theta}\right), \qquad 0 \le \theta < \theta_0 < \frac{\pi}{2}. \tag{31}$$

Its minimum radial coordinate is attained at $\theta = 0$:

$$\rho_* = \operatorname{arctanh}(\cos\theta_0), \qquad \theta_0 = \arccos(\tanh\rho_*). \tag{32}$$

As $\theta \to \theta_0^-$, ρ diverges, so the surface reaches ∂A at the asymptotic boundary.

Hemisphere. The opening angle $\theta_0 = \pi/2$ remains part of our construction. In that case the implicit surface equation is $\tanh\rho\,\cos\theta = 0$: for $\rho > 0$ the surface lies at $\theta = \pi/2$, and it also contains the origin. Its minimum radial coordinate is $\rho_* = 0$, consistent with the limiting value of $\operatorname{arctanh}(\cos\theta_0)$. The surface is therefore an equatorial hyperplane, but it cannot be represented by the radial graph equation 31, because θ is fixed while ρ varies.

Use in HoloWedge. The area functional equation 26 defines which surface the RT formula selects. For spherical caps in pure AdS, we already know that surface analytically, so training does not require numerical area integration or a search over surfaces. Its full profile explains the wedge geometry; the implementation uses the cap axis $\hat{u}$ and the explicit relation between θ_0 and ρ_* to construct each wedge. Finding extremal surfaces for general boundary regions, including unions and intersections of cap regions on the sphere, is difficult and sometimes even impossible to do in closed form. Therefore, we have decided to use the area functional of two-dimensions as a proxy also in higher dimensions. This is possible, since we never compare more than two caps and can thus always project to the plane spanned by their vectors $\hat{u}_u, \hat{u}_v$.

E Connected and Disconnected Ryu–Takayanagi Configurations

This appendix details the Ryu–Takayanagi (RT) minimization in the case of several extremal surfaces, such as in equation 7. Such a case appears if the regions under consideration consist of several disconnected pieces. Then, there is more than one possible minimal surface. Depending on the exact configuration of the regions either class might be the smallest, as illustrated in Figure 3. Here, we will explain this in more detail in the case of two dimensions which is of relevance for us.

Two disjoint boundary regions. We recall briefly some definitions in the subsection 3.2. We consider the boundary regions A_u and A_v, with half-opening angles $\theta_u \equiv \theta_0^{(u)}, \theta_v \equiv \theta_0^{(v)}$, and angular separation γ_{uv} between their axes. They are disjoint whenever

$$q_{uv} = \gamma_{uv} - \theta_u - \theta_v > 0. \tag{33}$$

The endpoints of the two intervals divide the boundary circle into the intervals themselves and two complementary gaps. Their angular lengths are

$$g_{\text{near}} = \gamma_{uv} - \theta_u - \theta_v, \qquad g_{\text{far}} = 2\pi - \gamma_{uv} - \theta_u - \theta_v. \tag{34}$$

Thus, in the disjoint regime, $g_{\text{near}} = q_{uv} > 0$. The second expression in equation 34 accounts for the periodicity of the boundary circle.

For an interval of angular length ℓ, 5 reads

$$S(\ell) = \log\left[\sin\left(\frac{\ell}{2}\right)\right]. \tag{35}$$

In particular, $S(A_u) = S(2\theta_u)$ and $S(A_v) = S(2\theta_v)$.

The two candidate configurations. In the *disconnected configuration*, the two endpoints of each interval are paired with one another. The candidate surface is therefore the union of the individual surfaces Σ_u and Σ_v, and its entropy is

$$S_{\text{disc}} = S(A_u) + S(A_v) = S(2\theta_u) + S(2\theta_v). \tag{36}$$

Together with $A_u \cup A_v$, these surfaces bound two separate bulk regions. The corresponding entanglement wedge is therefore disconnected.

In the *connected configuration*, the endpoints are instead paired across the two boundary gaps. The two candidate surfaces span g_{near} and g_{far}, giving

$$S_{\text{conn}} = S(g_{\text{near}}) + S(g_{\text{far}}). \tag{37}$$

These surfaces bound a single bulk region together with $A_u \cup A_v$, and the corresponding entanglement wedge is connected. Here "connected" refers to this endpoint pairing and to the resulting wedge.

The RT formula selects the candidate with smaller total area, or equivalently smaller entropy. We recall equation 7 for completeness:

$$S(A_u \cup A_v) = \min\{S_{\text{conn}}, S_{\text{disc}}\}. \tag{38}$$

For fixed cap sizes, both candidates vary smoothly with γ_{uv}, but the minimizing candidate can change. The critical separation is determined by

$$S_{\text{conn}} = S_{\text{disc}}. \tag{39}$$

Using equation 35, this condition can be written explicitly as

$$\sin\theta_u \, \sin\theta_v = \sin\Big(\frac{g_{\text{near}}}{2}\Big)\sin\Big(\frac{g_{\text{far}}}{2}\Big). \tag{40}$$

At this point the two candidate surfaces have equal area and exchange dominance. This change of the minimizing RT configuration is the connected–disconnected phase transition (Headrick, 2010). The entropy in equation 38 remains continuous at the crossing.

Consequences for the CMI. Because the two boundary regions are disjoint, $A_u \cap A_v = \emptyset$ and $S(A_u \cap A_v) = 0$. Therefore, equation 6 reduces to

$$\begin{aligned} \Delta_{\text{CMI}}(u,v) &= S(A_u) + S(A_v) - S(A_u \cup A_v) \\ &= S_{\text{disc}} - \min\{S_{\text{conn}}, S_{\text{disc}}\} \\ &= \max\{0, S_{\text{disc}} - S_{\text{conn}}\}. \end{aligned} \tag{41}$$

Nearby disjoint regions can therefore lie in the connected branch and have positive CMI. After the transition, the disconnected candidate dominates and the RT prescription gives $S(A_u \cup A_v) = S(A_u) + S(A_v)$, so the CMI vanishes.