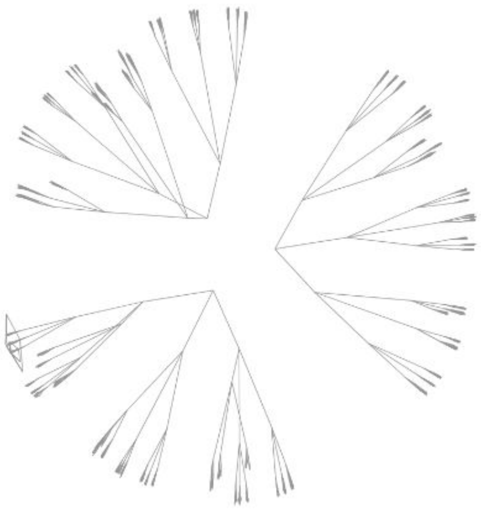
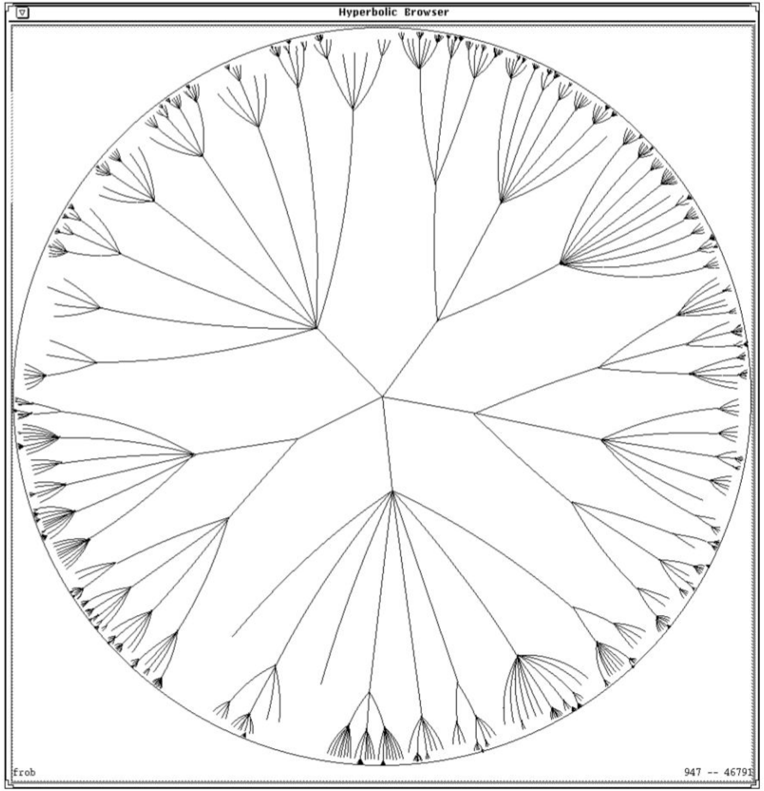


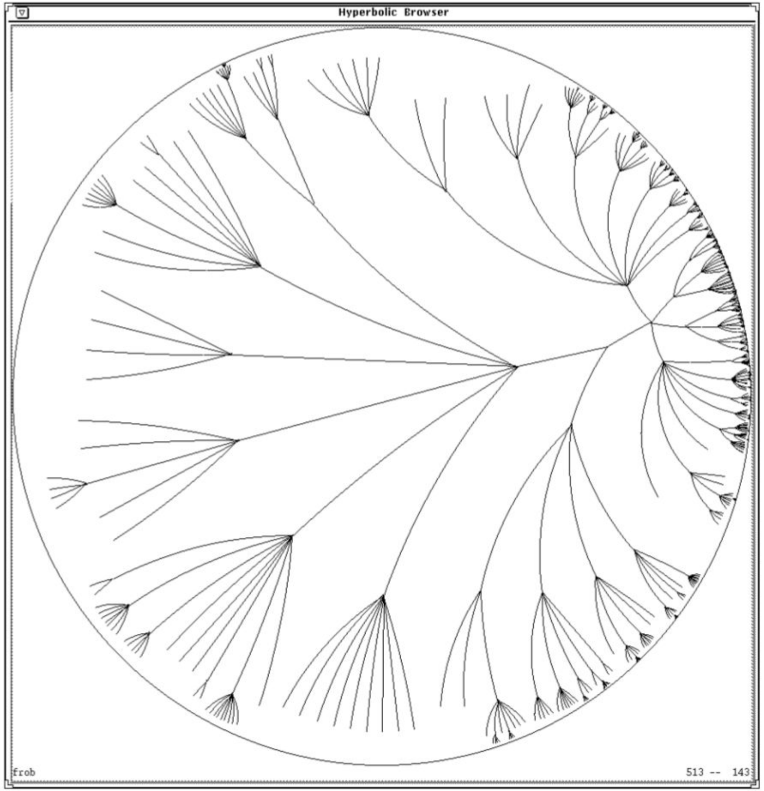
Hyperbolic Entailment Cones for Learning Hierarchical Embeddings

Octavian-Eugen Ganea, Gary Becigneul, Thomas Hofmann

ETH Zurich, Switzerland

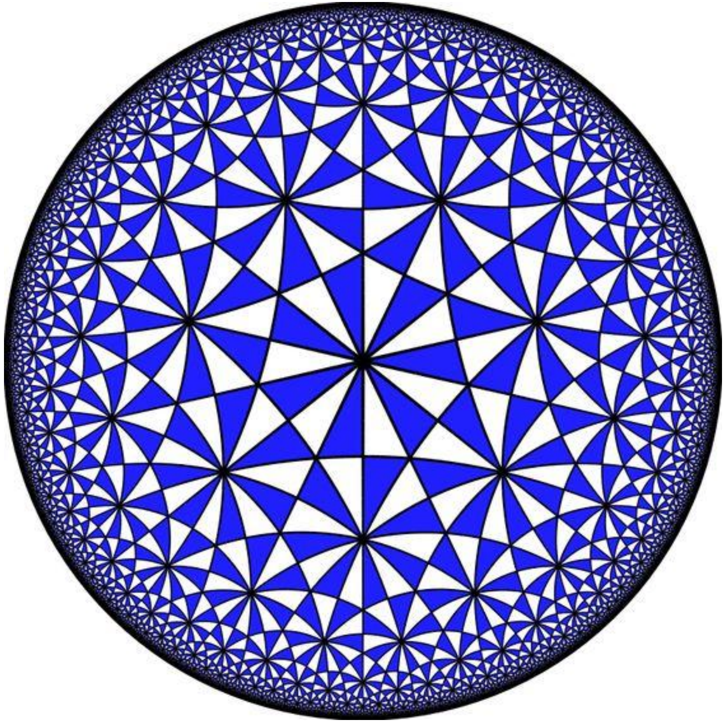
Hyperbolic Geometry for Embedding Hierarchies





[1] J. Lamping et al. "A focus+ context technique based on hyperbolic geometry for visualizing large hierarchies." *SIGCHI* 1995.

Poincaré Ball: distances dilated towards the border



$$g_x(u, v) = \frac{2}{1 - \|x\|^2} \langle u, v \rangle$$

Problem Definition & Goals

Entailment:

- *v* is a *subconcept* of *u*
- *Directed* edges in a directed acyclic graph

We want to:

1. Geometrically model entailment in the embedding space
2. Exploit the power of hyperbolic geometry

Poincaré Embeddings [2]

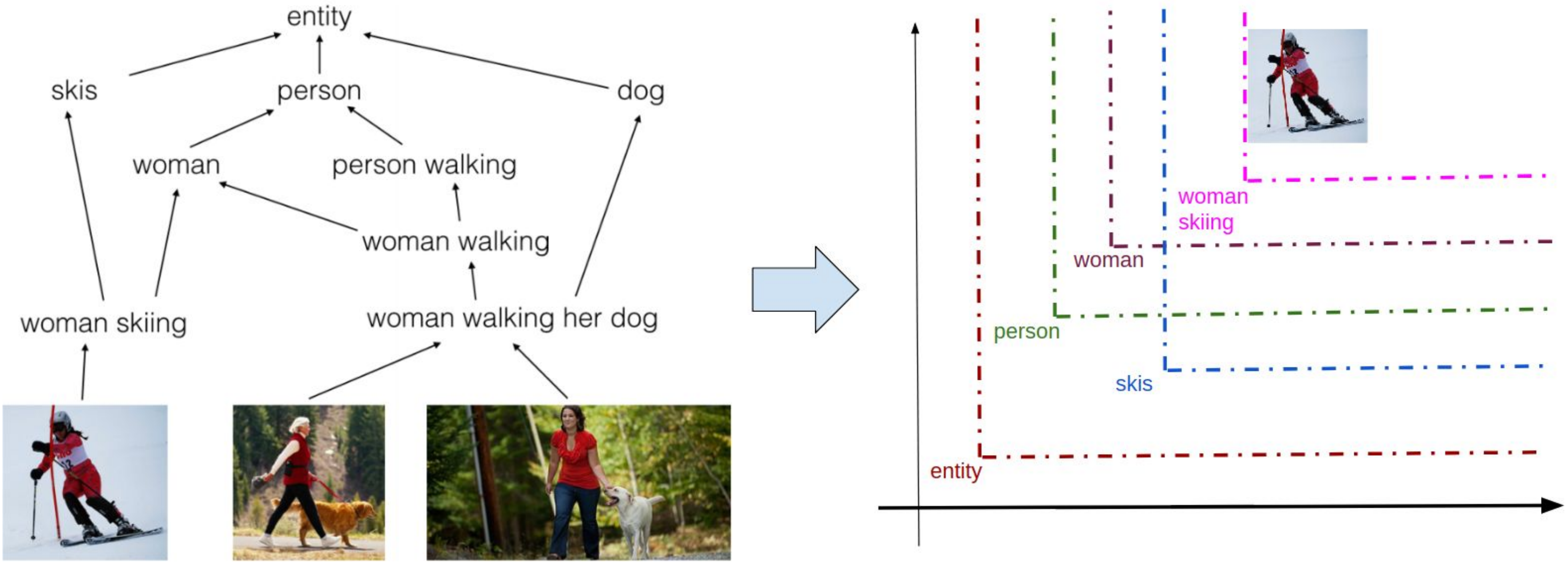
- Train loss: $\mathcal{L}(\theta) = \sum_{(u,v)\in\mathcal{E}} \log \frac{e^{-d(u,v)}}{\sum_{v'\in\mathcal{N}(u)} e^{-d(u,v')}}$

- Predict entailment at test time via heuristic

$$score(is-a(u,v)) = (1+\alpha(\|u\|-\|v\|))\cdot d(u,v)$$

[2] Nickel, Maximillian, and Douwe Kiela. "Poincaré embeddings for learning hierarchical representations." *NIPS* 2017.

Order Embeddings [3]



- Geometric modeling of partial order relations induced by a tree/DAG via simple **entailment cones** in $\mathbb{R}^d$
- Drawbacks

[3] Vendrov, Ivan, et al. "Order-embeddings of images and language." *ICLR 2016*

Riemannian Entailment Cones

- Generalize **convex entailment cones** to any Riemannian manifold:

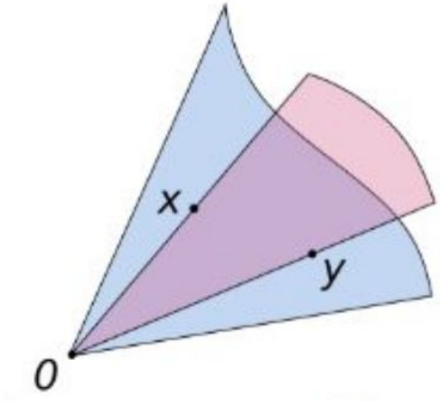
$$(u, v) \in \mathcal{E} \Longleftrightarrow v \in \mathfrak{S}_u$$

- Induce a **partial order** in the embedding space

Riemannian Convex Cones

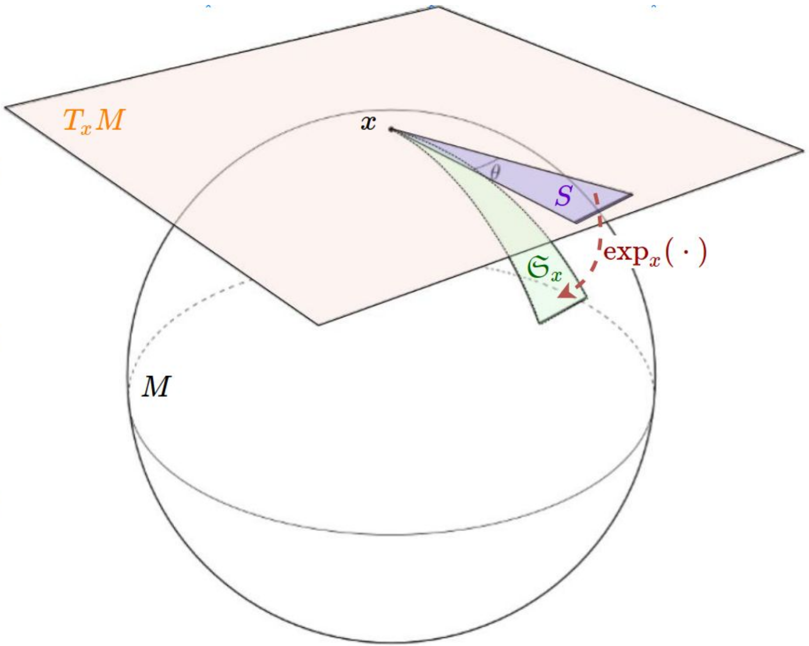
EUCLIDEAN CONVEX CONES

$v_1, v_2 \in S, \forall \alpha, \beta \geq 0$
$\Rightarrow \alpha v_1 + \beta v_2 \in S$



RIEMANNIAN CONVEX CONES

$S \subseteq T_x\mathcal{M},\ \mathfrak{S}_x := \exp_x(S) \subseteq \mathcal{M}$

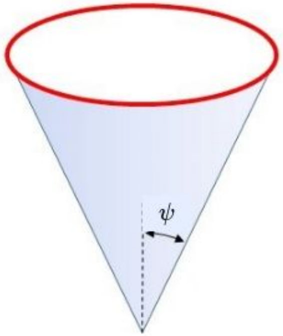


Angular Entailment Cones in the Poincaré Ball

Satisfy four intuitive properties:

1) axial symmetry (depends on aperture angle ψ)

$$S_x^{\psi} := \{v : \angle(v, x) \leq \psi(x)\}, \quad \mathfrak{S}_x^{\psi} := \exp_x(S_x^{\psi})$$

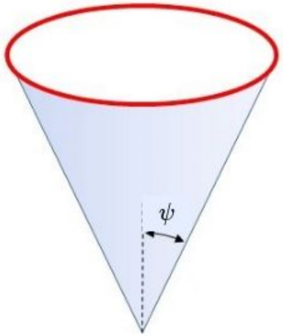


Angular Entailment Cones in the Poincaré Ball

Satisfy four intuitive properties:

1) axial symmetry (depends on aperture angle ψ)

$$S_x^{\psi} := \{v : \angle(v, x) \leq \psi(x)\}, \quad \mathfrak{S}_x^{\psi} := \exp_x(S_x^{\psi})$$



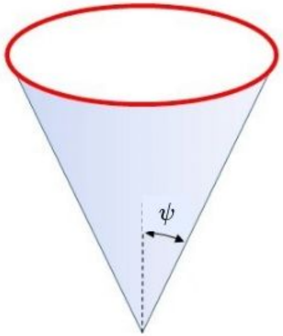
2) continuous cone aperture function ψ

Angular Entailment Cones in the Poincaré Ball

Satisfy four intuitive properties:

1) axial symmetry (depends on aperture angle ψ)

$$S_x^{\psi} := \{v : \angle(v, x) \leq \psi(x)\}, \quad \mathfrak{S}_x^{\psi} := \exp_x(S_x^{\psi})$$



2) continuous cone aperture function ψ

3) rotation invariance (only depends on norm of x)

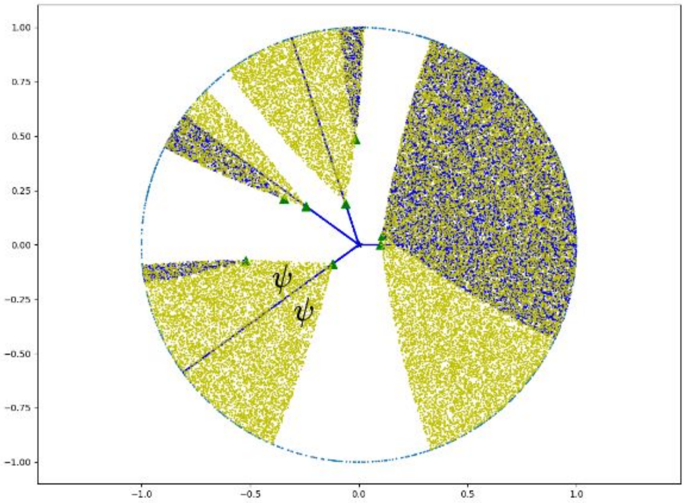
$$\exists \tilde{\psi} \quad \text{s.t.} \quad \psi(x) = \tilde{\psi}(\|x\|)$$

Angular Entailment Cones in the Poincaré Ball

Satisfy four intuitive properties:

4) (hardest) transitivity of nested angular cones

$$\forall x, x' \in \mathbb{D}^n : \quad x' \in \mathfrak{S}_x^{\psi(x)} \implies \mathfrak{S}_{x'}^{\psi(x')} \subseteq \mathfrak{S}_x^{\psi(x)}$$



Angular Entailment Cones in the Poincaré Ball

Four properties $\Longrightarrow$ closed form expression of ψ

$$\mathfrak{S}_x^{\psi(x)} = \left\{ y \in \mathbb{D}^n \middle| \quad \Xi(x, y) \leq \arcsin\left(K \frac{1 - \|x\|^2}{\|x\|} \right) \right\}$$

Angular Entailment Cones in the Poincaré Ball

Four properties $\Longrightarrow$ closed form expression of ψ

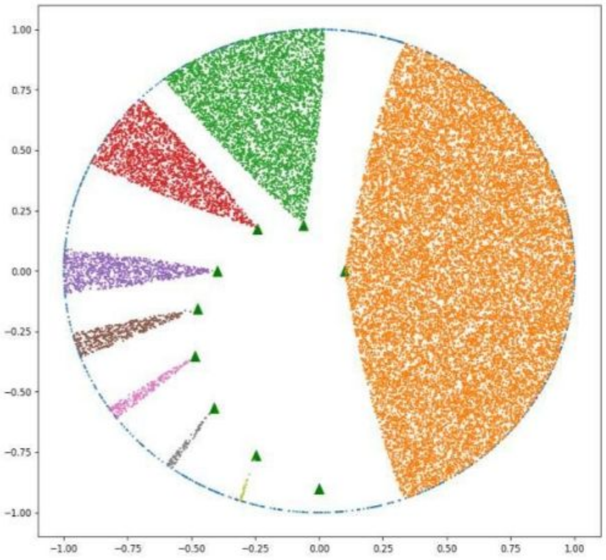
$$\mathfrak{S}_x^{\psi(x)} = \left\{ y \in \mathbb{D}^n \middle| \quad \Xi(x,y) \leq \mathbf{arcsin}\left(K \frac{1 - \|x\|^2}{\|x\|} \right) \right\}$$

where

$$\Xi(x,y) = \mathbf{arccos}\left(\frac{< x, y > (1 + \|x\|^2) - \|x\|^2 (1 + \|y\|^2)}{\|x\| \cdot \|x - y\| \sqrt{1 + \|x\|^2 \|y\|^2 - 2 < x, y >}} \right)$$

Angular Entailment Cones in the Poincaré Ball

Four properties $\Longrightarrow$ closed form expression of ψ



Similar derivation can be done for Euclidean angular cones

Learning with Entailment Cones

- Learning from positive and negative pairs of nodes in a DAG
- loss:

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

- penalty: $E(u,v) := \max(0, \Xi(u,v) - \psi(u)),$

Riemannian Optimization

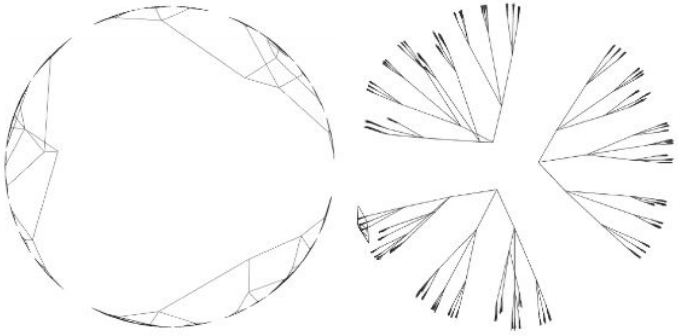
- SGD becomes:

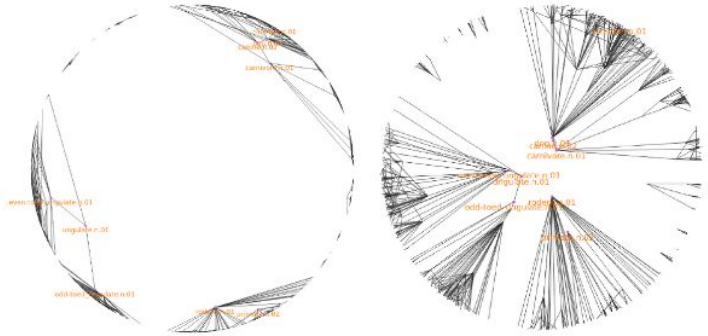
$$u \leftarrow \exp_u(-\eta \nabla_u^R \mathcal{L}), \quad u \in \mathbb{D}$$

- Riemannian gradient:

$$\nabla_u^R \mathcal{L} = (1/\lambda_u)^2 \nabla_u \mathcal{L}, \quad \text{conformal factor } \lambda_u = \frac{2}{1 - \|u\|^2}$$

Experiments - Qualitative





- Left deck: uniform tree of depth 7 and branching factor 3
- Right deck: WordNet mammal subtree - 4230 edges, 1165 nodes
- Each left: Poincare embeddings
- Each right: our hyperbolic cones

Quantitative Experiments - Link Prediction in DAGs

- Transitive closure of WordNet hierarchy: 82K nodes, 660K edges

	Embedding Dimension = 5				Embedding Dimension = 10			
	Percentage of Transitive Closure Edges used during Training							
	0%	10%	25%	50%	0%	10%	25%	50%
Simple Euclidean Emb	26.8%	71.3%	73.8%	72.8%	29.4%	75.4%	78.4%	78.1%
Poincaré Emb	29.4%	70.2%	78.2%	83.6%	28.9%	71.4%	82.0%	85.3%
Order Emb	**34.4%**	70.2%	75.9%	81.7%	**43.0%**	69.7%	79.4%	84.1%
Our Euclidean Cones	28.5%	69.7%	75.0%	77.4%	31.3%	81.5%	84.5%	81.6%
Our Hyperbolic Cones	29.2%	**80.1%**	**86.0%**	**92.8%**	32.2%	**85.9%**	**91.0%**	**94.4%**

Table 1. Test F1 results for various models. Simple Euclidean Emb and Poincaré Emb are the Euclidean and hyperbolic methods proposed by (Nickel & Kiela, 2017), Order Emb is proposed by (Vendrov et al., 2015).

HDL HYPERBOLIC DEEP LEARNING

A nascent and promising field

Website: hyperbolicdeeplearning.com

Code: github.com/dalab/hyperbolic_cones