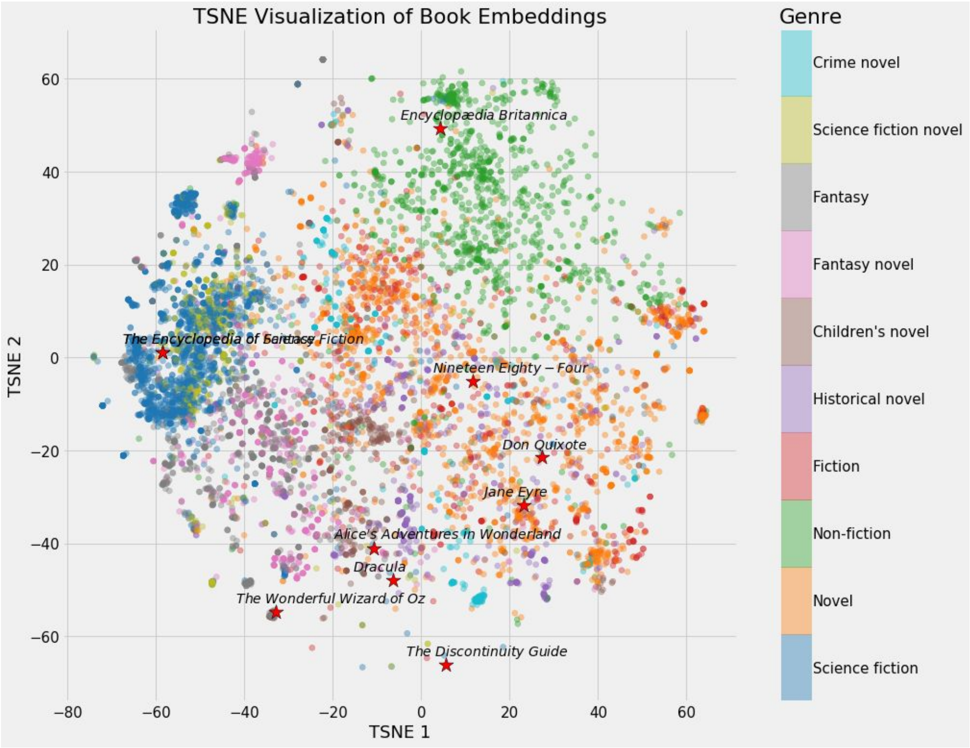
Poincaré Embeddings for Learning Hierarchical Representations

KJ and Zane

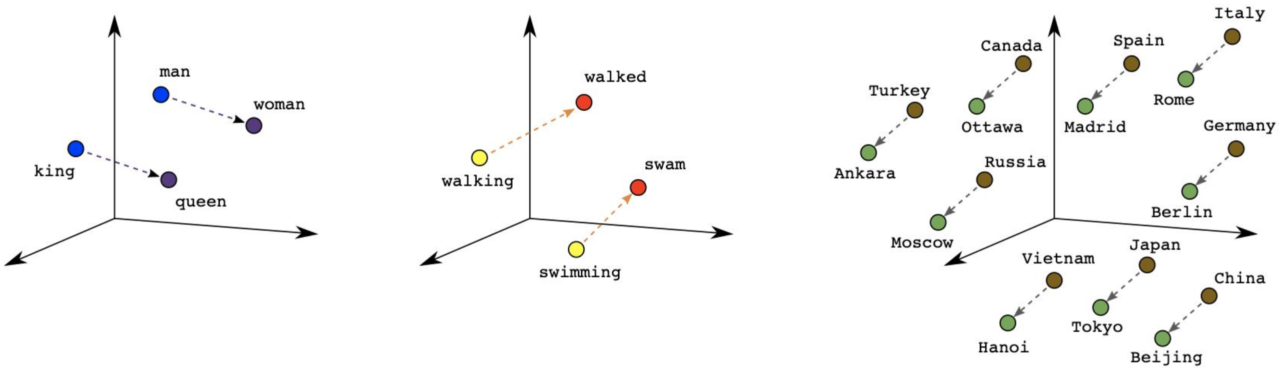
What are “embeddings”?

A mapping of a discrete, categorical variable to a vector of continuous numbers



Objective of embeddings...

The goal of embeddings is to organize symbolic objects (e.g., words, entities, concepts) in a way such that their similarity in the embedding space reflects their semantic or functional similarity.



Male-Female   Verb Tense   Country-Capital

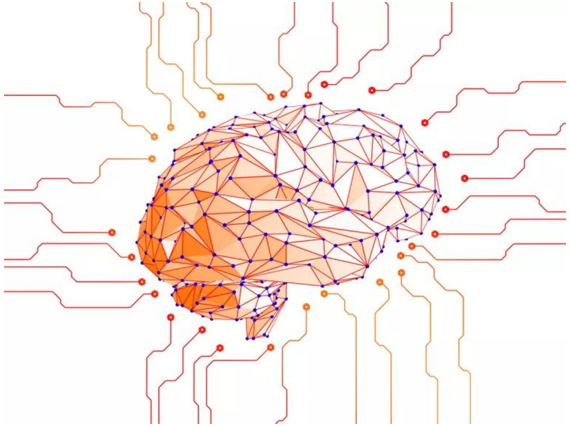
What are the different types of embeddings?

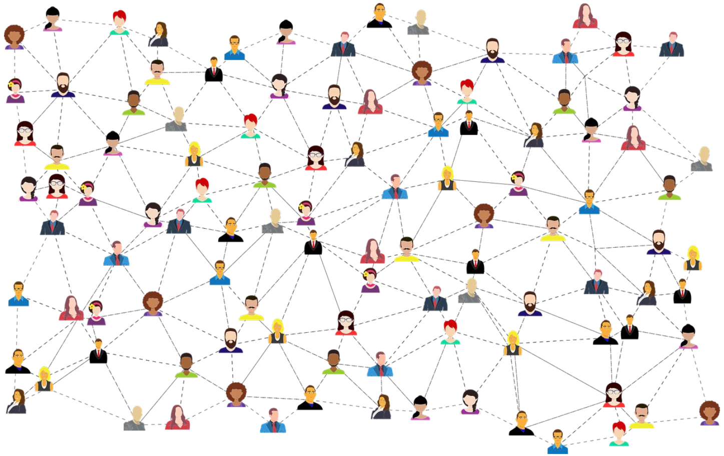
- Embeddings of words
  - Word2Vec [17], GloVe [23], FastText [4]
  - For machine translation and sentiment analysis

- Embeddings of graphs
  - Latent space embeddings [13], Node2Vec [11], DeepWalk [24]
  - For community detection and link prediction in social networks

- Embeddings of multi-relational data
  - Rescal [19], TransE [6], Universal Schema [27]
  - For knowledge graph completion and information extraction

Relational Data





3,000 G$ | 7,000 G$ | 20,000 G$ | 60,000 G$ | 100,000 G$ | 500,000 G$ 1000 largest flows worldwide in 2011

Understanding Relationships

Link Prediction

- Predict truth-value of a missing link based on observed truths
- If I am from Ohio, I am probably a US citizen

Entity Resolution

- Find nodes that refer to the same underlying identity
- Both nodes are from Toledo, Ohio and both are in the MSAI program, these nodes probably refer to the same entity

Community Detection

- Clustering based on relationships (an extension of graph clustering)
- Identifying functional parts of the brain that are connected in similar ways

Issue with the current state of embeddings...

The previously mentioned embeddings are learned in Euclidean vector spaces, which do not account for a latent hierarchical structure that exists inherently within the data.

Euclidean space is flat. Why doesn't that work for hierarchical structures?

- What properties do we need to embed a tree (or any graph) well?

Issue with the current state of embeddings...

Although non-linear embeddings can encode such hierarchies, they have such a large dimensionality that they are infeasible computationally.

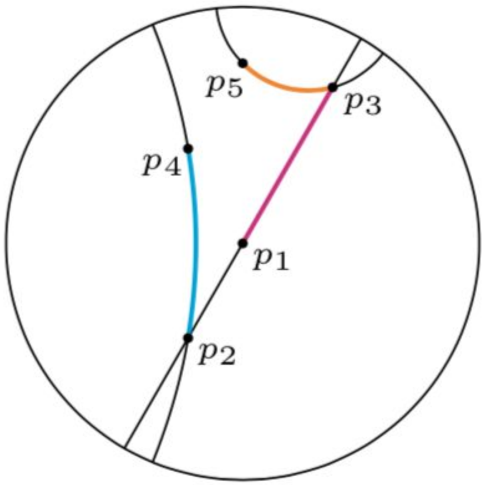
Solution proposed...

Use hyperbolic space instead of Euclidean space

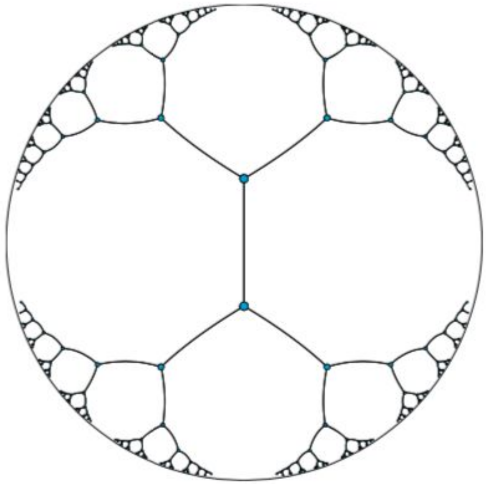
- Natural hierarchical structure
- Continuous tree-like structure

Model of hyperbolic space that we would like to use: Poincaré Disk

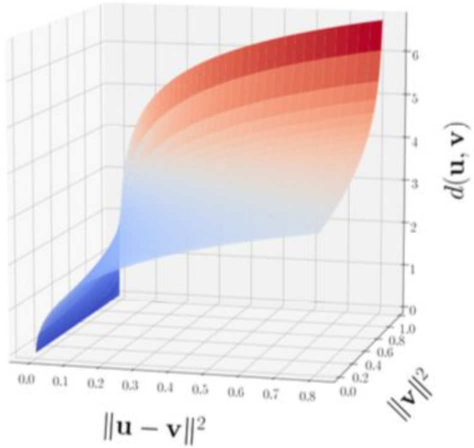
Embeddings and Hyperbolic Geometry



(a) Geodesics of the Poincaré disk



(b) Embedding of a tree in $\mathcal{B}^2$



(c) Growth of Poincaré distance

Poincaré Embeddings

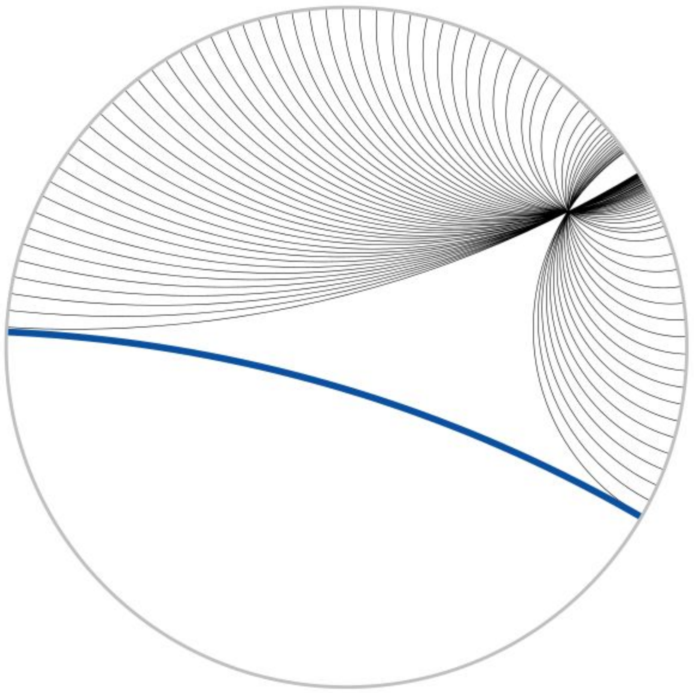
Will improve upon existing methods by:

1. Learning more parsimonious embeddings for superior generalization performance and decreased runtime and memory complexity
2. Gaining additional insights about the relationships between symbols and the important of individual symbols

Poincaré Ball Model

- Distance function is differentiable making it well-suited for gradient-based optimization
- Extracts the multiple latent hierarchies that can co-exist in the data

Fully unsupervised.



Riemannian Manifold

let $\mathcal{B}^d = \{\boldsymbol{x} \in \mathbb{R}^d \mid \|\boldsymbol{x}\| < 1\}$ be the *open* d-dimensional unit ball, where $\|\cdot\|$ denotes the Euclidean norm. The Poincaré ball model of hyperbolic space corresponds then to the Riemannian manifold $(\mathcal{B}^d, g_{\boldsymbol{x}})$, i.e., the open unit ball equipped with the Riemannian metric tensor

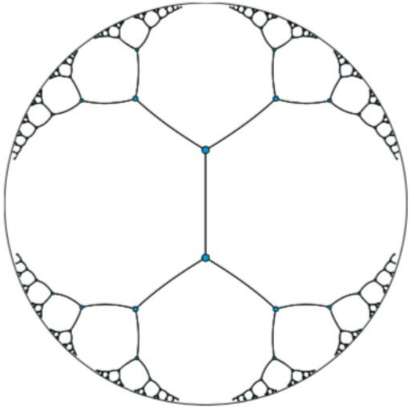
$$g_{\boldsymbol{x}} = \left(\frac{2}{1 - \|\boldsymbol{x}\|^2}\right)^2 g^E,$$

where $\boldsymbol{x} \in \mathcal{B}^d$ and g^E denotes the Euclidean metric tensor.

Poincaré Distance

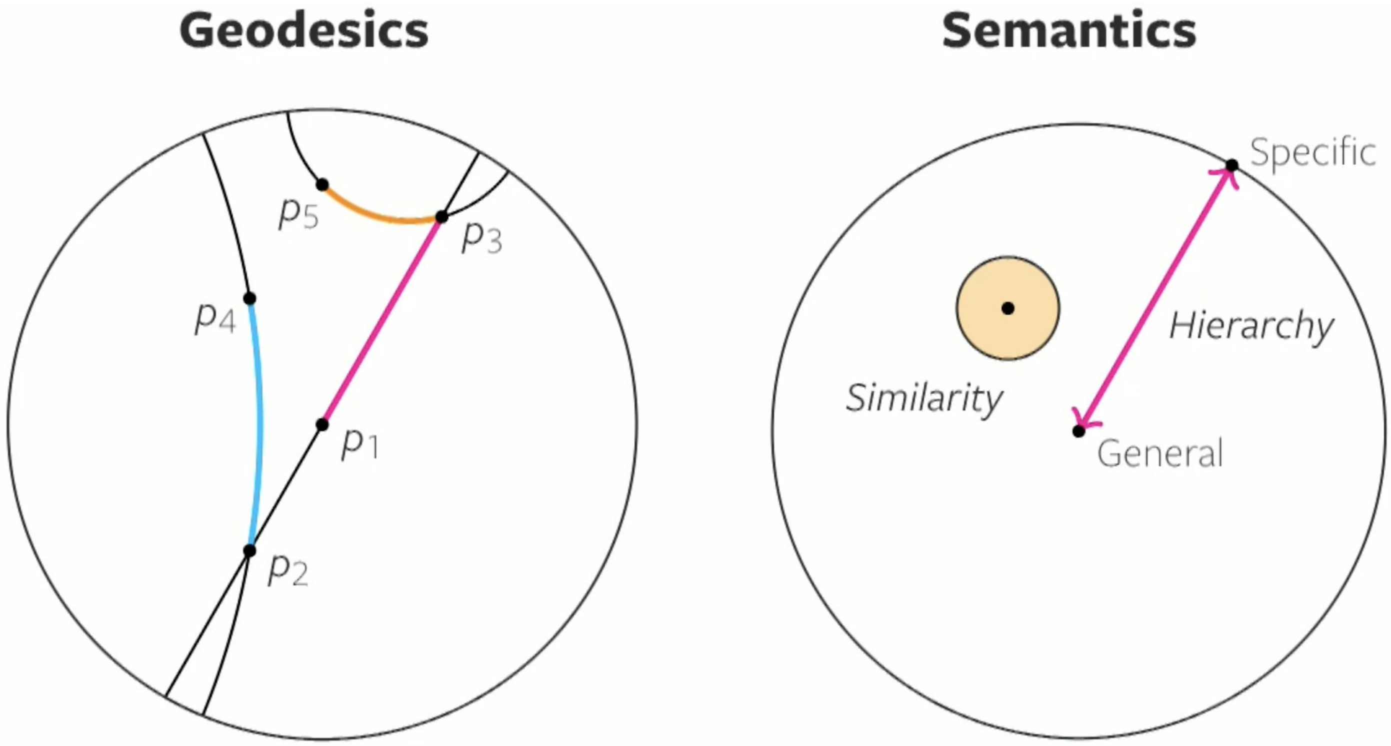
Furthermore, the distance between points $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{B}^d$ is given as

$$d(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{arcosh}\left(1 + 2\frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{(1 - \|\boldsymbol{u}\|^2)(1 - \|\boldsymbol{v}\|^2)}\right). \quad (1)$$



(b) Embedding of a tree in $\mathcal{B}^2$

Poincaré Distance (Visualized)



Poincaré Objective Function

$$\Theta' \leftarrow \arg\min_{\Theta} \mathcal{L}(\Theta) \qquad \text{s.t. } \forall\, \boldsymbol{\theta}_i \in \Theta : \|\boldsymbol{\theta}_i\| < 1.$$

Optimization

- Aim to minimize computational and memory complexity of updates
- Algorithm can be parallelized for further runtime improvements

Riemannian Optimization

- **Riemannian Stochastic Gradient Descent** (Bonnabel, 2013)

$$\boldsymbol{\theta}_{t+1} = \mathfrak{R}_{\boldsymbol{\theta}_t}\left(-\eta_t \nabla_R \mathcal{L}(\boldsymbol{\theta}_t)\right)$$

- **Riemannian Gradient**: The Poincaré ball is a *conformal* model of hyperbolic space $\Rightarrow$ angles are identical, vector length can differ

$$\nabla_R = g_\theta^{-1} \nabla_E = \frac{(1 - \|\boldsymbol{\theta}_t\|^2)^2}{4} \nabla_E$$

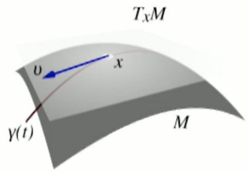
- **Retraction**: We simply use the Euclidean retraction

$$\mathfrak{R}_{\theta_t}(\boldsymbol{x}) = \boldsymbol{\theta}_t + \boldsymbol{x}$$

- **Parameter Update**

$$\boldsymbol{\theta}_{t+1} \leftarrow \text{proj}\left(\boldsymbol{\theta}_t - \eta_t \frac{(1 - \|\boldsymbol{\theta}_t\|^2)^2}{4} \nabla_E\right)$$

$$\text{where } \text{proj}(\boldsymbol{\theta}) = \begin{cases} \boldsymbol{\theta}/\|\boldsymbol{\theta}\| - \varepsilon & \text{if } \|\boldsymbol{\theta}\| \geq 1 \\ \boldsymbol{\theta} & \text{otherwise}. \end{cases}$$



Training Details

- Embeddings are randomly generated from the uniform distribution (-0.001, 0.001)

- “Burn-in” phase with reduced learning rate

Evaluation

The Poincaré embeddings are evaluated on a variety of tasks:

- Embedding of taxonomies
- Link prediction in networks
- Modeling lexical entailment

And compared to 2 distance metrics:

- Euclidean
- Translational (when applicable)

Embedding Taxonomies

Goal: Evaluating the ability of Poincaré embeddings to embed data that exhibits a clear latent hierarchical structure

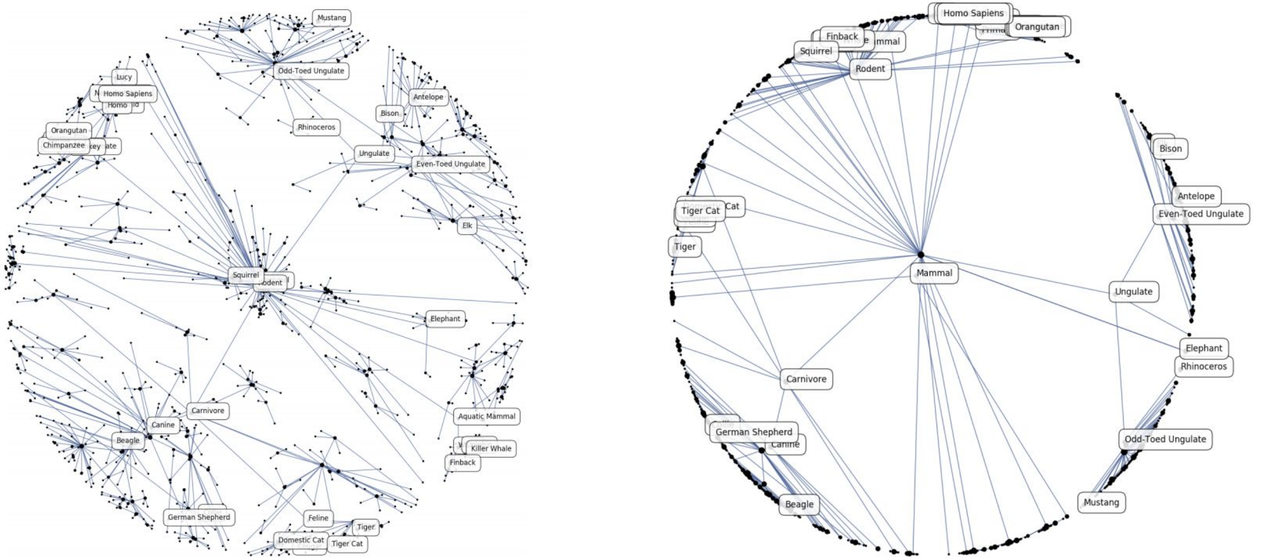
Experiments on the transitive closure of the WORDNET noun hierarchy in two settings:

- Reconstruction
- Link Prediction

Embedding Taxonomies

			Dimensionality					
			5	10	20	50	100	200
WordNet Reconstruction	Euclidean	Rank	3542.3	2286.9	1685.9	1281.7	1187.3	1157.3
		MAP	0.024	0.059	0.087	0.140	0.162	0.168
	Translational	Rank	205.9	179.4	95.3	92.8	92.7	91.0
		MAP	0.517	0.503	0.563	0.566	0.562	0.565
	Poincaré	Rank	4.9	4.02	3.84	3.98	3.9	**3.83**
		MAP	0.823	0.851	0.855	0.86	0.857	**0.87**
WordNet Link Pred.	Euclidean	Rank	3311.1	2199.5	952.3	351.4	190.7	81.5
		MAP	0.024	0.059	0.176	0.286	0.428	0.490
	Translational	Rank	65.7	56.6	52.1	47.2	43.2	40.4
		MAP	0.545	0.554	0.554	0.56	0.562	0.559
	Poincaré	Rank	5.7	**4.3**	4.9	4.6	4.6	4.6
		MAP	0.825	0.852	0.861	**0.863**	0.856	0.855

Embedding Taxonomies



(a) Intermediate embedding after 20 epochs

(b) Embedding after convergence

Network Embeddings

Goal: Evaluating the ability of Poincaré embeddings in of terms representation size and generalization performance.

Experiments on four social networks:

- ASTROPH
- CONDMAT
- GROQ
- HEPPH

Network Embeddings

		Dimensionality							
		Reconstruction				Link Prediction			
		10	20	50	100	10	20	50	100
AstroPh N=18,772; E=198,110	**Euclidean**	0.376	0.788	0.969	0.989	0.508	0.815	0.946	0.960
	Poincaré	0.703	0.897	0.982	0.990	0.671	0.860	0.977	0.988
CondMat N=23,133; E=93,497	**Euclidean**	0.356	0.860	0.991	0.998	0.308	0.617	0.725	0.736
	Poincaré	0.799	0.963	0.996	0.998	0.539	0.718	0.756	0.758
GrQc N=5,242; E=14,496	**Euclidean**	0.522	0.931	0.994	0.998	0.438	0.584	0.673	0.683
	Poincaré	0.990	0.999	0.999	0.999	0.660	0.691	0.695	0.697
HepPh N=12,008; E=118,521	**Euclidean**	0.434	0.742	0.937	0.966	0.642	0.749	0.779	0.783
	Poincaré	0.811	0.960	0.994	0.997	0.683	0.743	0.770	0.774

Lexical Entailment

Goal: Evaluating the ability of Poincaré embeddings ability to capture graded lexical entailment

Tested with HyperLex

- Quantifies to what degree X is a type of Y via ratings on a scale of [0, 10]

Lexical Entailment

Table 3: Spearman's ρ for Lexical Entailment on HYPERLEX.

	FR	**SLQS-Sim**	**WN-Basic**	**WN-WuP**	**WN-LCh**	**Vis-ID**	**Euclidean**	**Poincaré**
ρ	0.283	0.229	0.240	0.214	0.214	0.253	0.389	0.512

Findings

Introduced Poincaré embeddings for symbolic data and that they can learn similarity and hierarchy of objects simultaneously

Proposed an efficient algorithm to computer embeddings

Showed that Poincaré embeddings are more advantageous than Euclidean for hierarchical data

Future Work

- Expand the applications of Poincaré embeddings
- Derive models that are tailored to specific applications
- Use full Riemannian optimization approach to further increase the quality of the embeddings and lead to faster convergence