Hyperbolic Embeddings

刘闯
chuangliu@whu.edu.cn
2019.03.08

Reference

1. Poincaré Embeddings for Learning Hierarchical Representations , Facebook, 2017 NIPS , Maximilian Nickel, Douwe Kiela
2. Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry, Facebook, 2018 PMLR, Maximilian Nickel, Douwe Kiela
3. Representation Tradeoffs for Hyperbolic Embeddings, Stanford , 2018 PMLR, Christopher De Sa, Christopher Re
4. Hyperbolic entailment cones for learning hierarchical embeddings, 2018 PMLR , Octavian-Eugen Ganea, Thomas Hofmann

庞加莱嵌入

Poincaré Embeddings for Learning Hierarchical Representations

摘要:

1. 为了学习层次化的表征， 将图数据嵌入到双曲空间中(n 维庞加莱球：well-suited for gradient-based optimization)
2. Riemannian optimization(并行化)

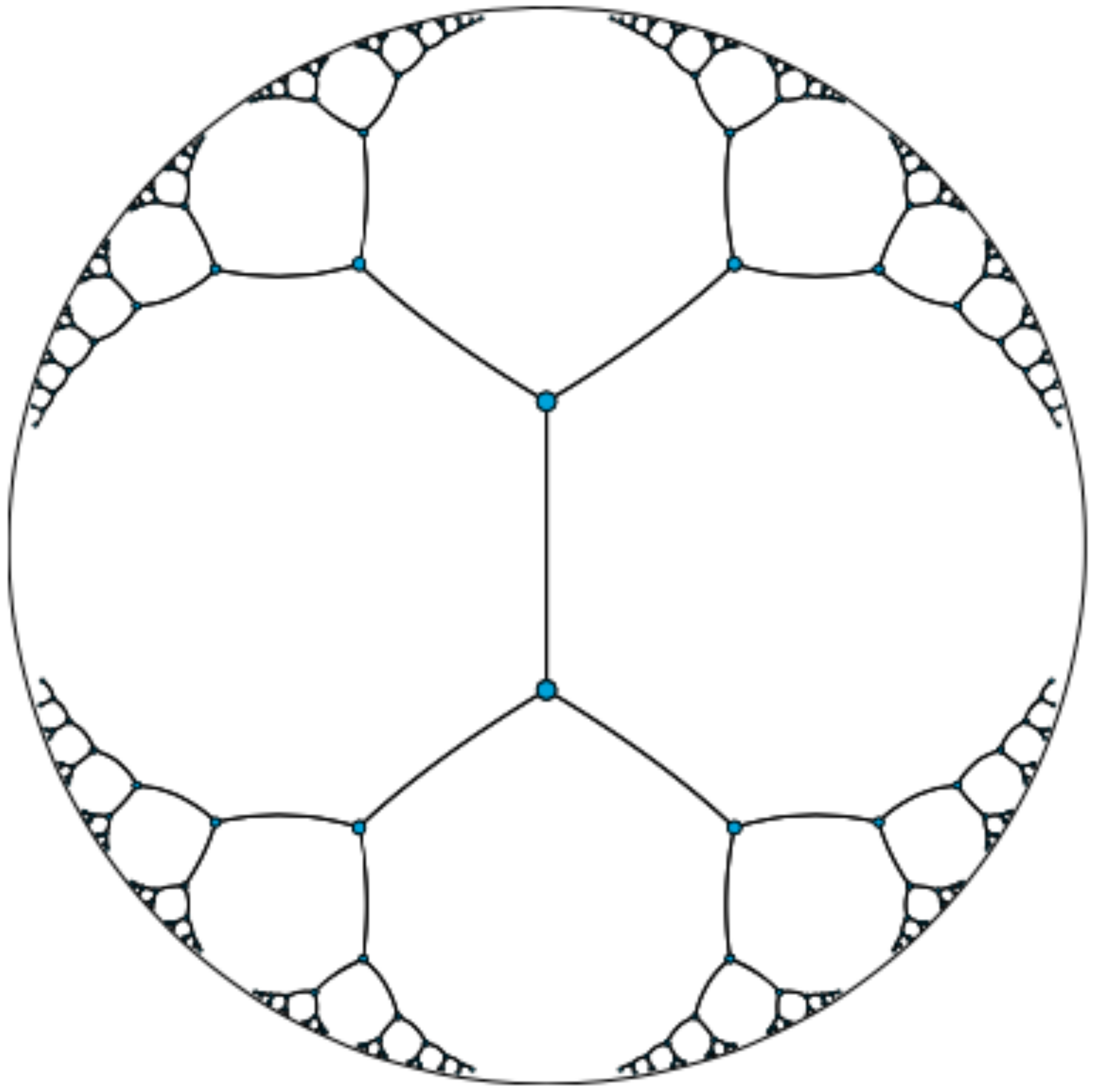
主要解决存在潜在层级结构的数据:

1. 存在幂律分布的数据往往存在层级结构[1]
2. 树结构的数据[2]

1. *Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. Physical Review E,*
2. *Tree-like structure in large social and information networks , ICDM , 2013*

双曲空间

简单来说： 可以将树看作离散的双曲空间, 双曲空间看作连续的树结构



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

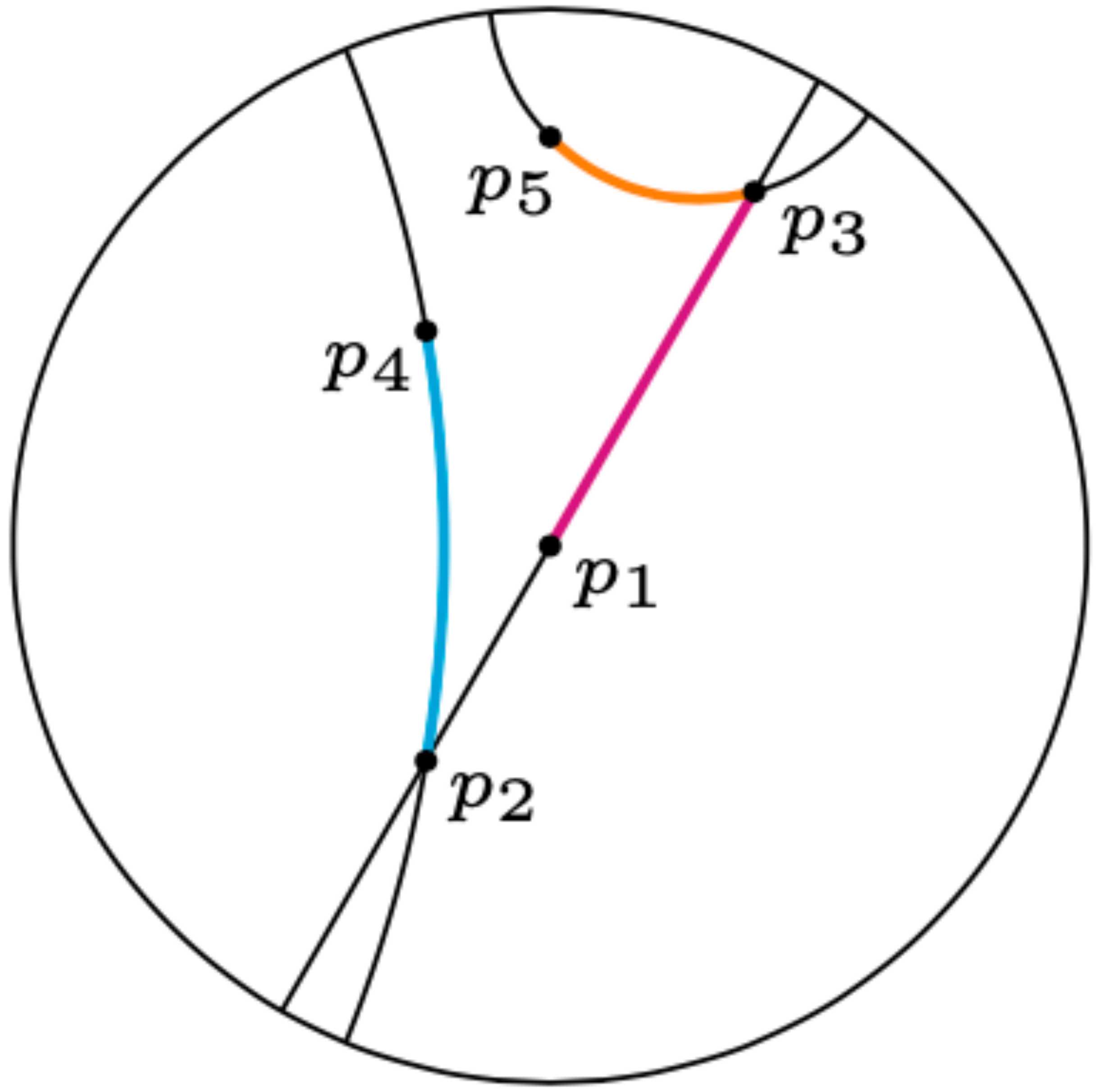
所有的黑色线段等长

庞加莱嵌入

距离

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

1. 距离平滑改变， 这是能够找到分层连续嵌入的关键特性
2. 方程是对称的， 空间的分层管理仅仅依赖于距离中心的距离，由于这种自控制性， 可以用于无监督网络分层学习
3. 不仅可以学习分层结构，还可以学习相似性



(a) Geodesics of the Poincaré disk

测地线：最短双曲距离

文中用的是庞加莱球，而不是庞加莱圆盘：

1. 很多数据有多个隐层级结构共存， 二维嵌入不太好表达
2. 高维嵌入降低优化难度

庞加莱嵌入

损失函数

$$\mathscr{L}(\Theta) = \sum_{(u,v)\in\mathscr{D}} \log \frac{e^{-d(\boldsymbol{u},\boldsymbol{v})}}{\sum_{\boldsymbol{v}'\in\mathscr{N}(u)} e^{-d(\boldsymbol{u},\boldsymbol{v}')}}$$

欧氏空间嵌入： word2vec

d 是距离

双曲空间嵌入

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

庞加莱嵌入

To compute Poincaré embeddings for a set of symbols $\mathcal{S} = \{x_i\}_{i=1}^n$, we are then interested in finding embeddings $\Theta = \{\boldsymbol{\theta}_i\}_{i=1}^n$, where $\boldsymbol{\theta}_i \in \mathcal{B}^d$. We assume we are given a problem-specific loss function $\mathcal{L}(\Theta)$ which encourages semantically similar objects to be close in the embedding space according to their Poincaré distance. To estimate Θ, we then solve the optimization problem

$$\Theta' \leftarrow \underset{\Theta}{\arg\min}\, \mathcal{L}(\Theta) \qquad \text{s.t. } \forall\, \boldsymbol{\theta}_i \in \Theta : \|\boldsymbol{\theta}_i\| < 1. \tag{2}$$

庞加莱球有黎曼流形结构， 通过随机黎曼优化方法： RSGD[1]，RSVRG

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

根据保角变换，将黎曼梯度下降转换为欧氏梯度：

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

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

$$\nabla_E = \frac{\partial \mathscr{L}(\boldsymbol{\theta})}{\partial d(\boldsymbol{\theta}, \boldsymbol{x})} \frac{\partial d(\boldsymbol{\theta}, \boldsymbol{x})}{\partial \boldsymbol{\theta}}$$

$$\frac{\partial d(\boldsymbol{\theta}, \boldsymbol{x})}{\partial \boldsymbol{\theta}} = \frac{4}{\beta\sqrt{\gamma^2 - 1}}\left(\frac{\|\boldsymbol{x}\|^2 - 2\langle \boldsymbol{\theta}, \boldsymbol{x}\rangle + 1}{\alpha^2}\boldsymbol{\theta} - \frac{\boldsymbol{x}}{\alpha}\right)$$

1. Silvere Bonnabel. Stochastic gradient descent on Riemannian manifolds. IEEE Trans. Automat. Contr., 58(9):2217–2229, 2013.

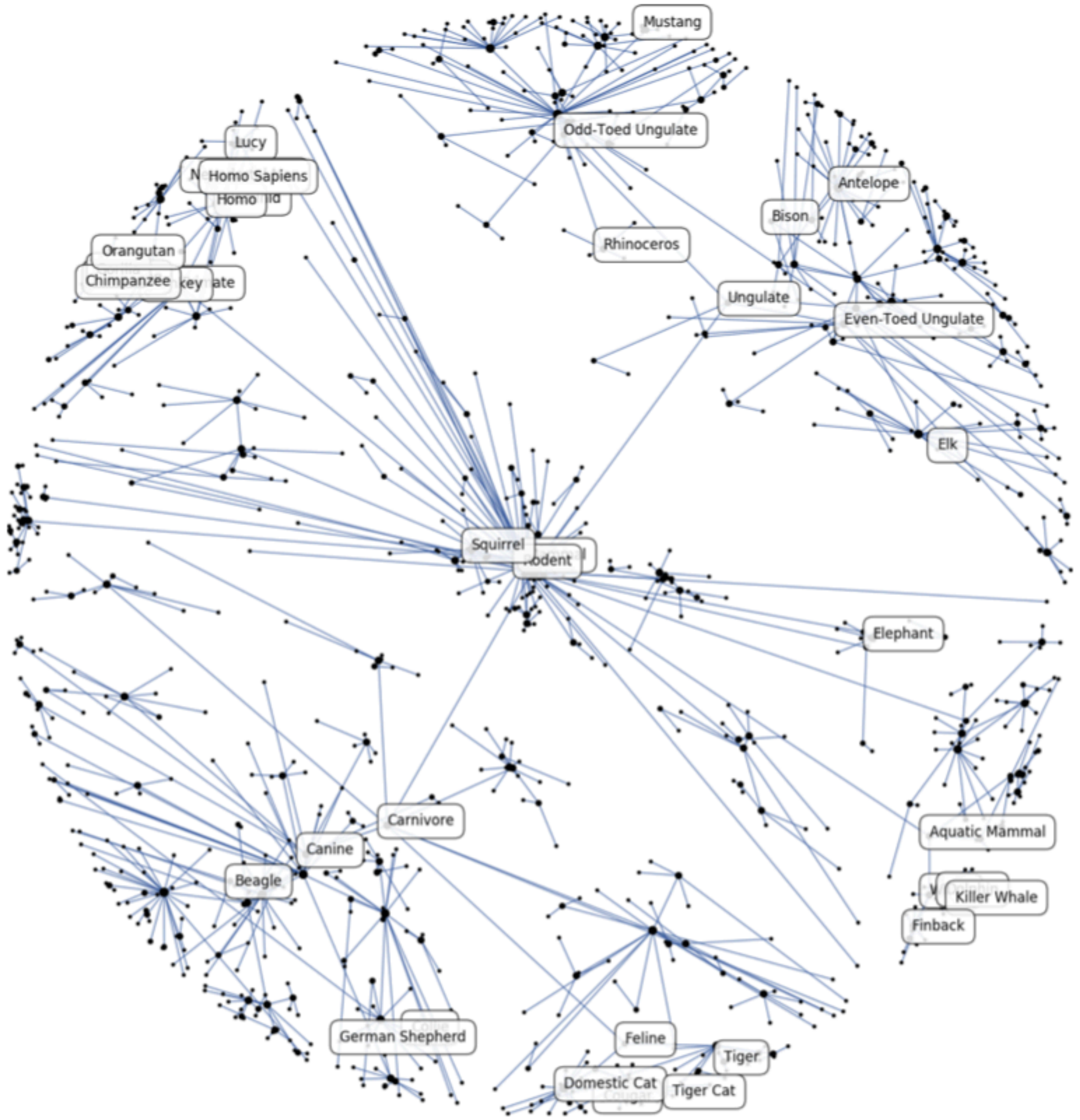
Results

Table 1: Experimental results on the transitive closure of the WordNet noun hierarchy. Highlighted cells indicate the best Euclidean embeddings as well as the Poincaré embeddings which achieve equal or better results. Bold numbers indicate absolute best results.

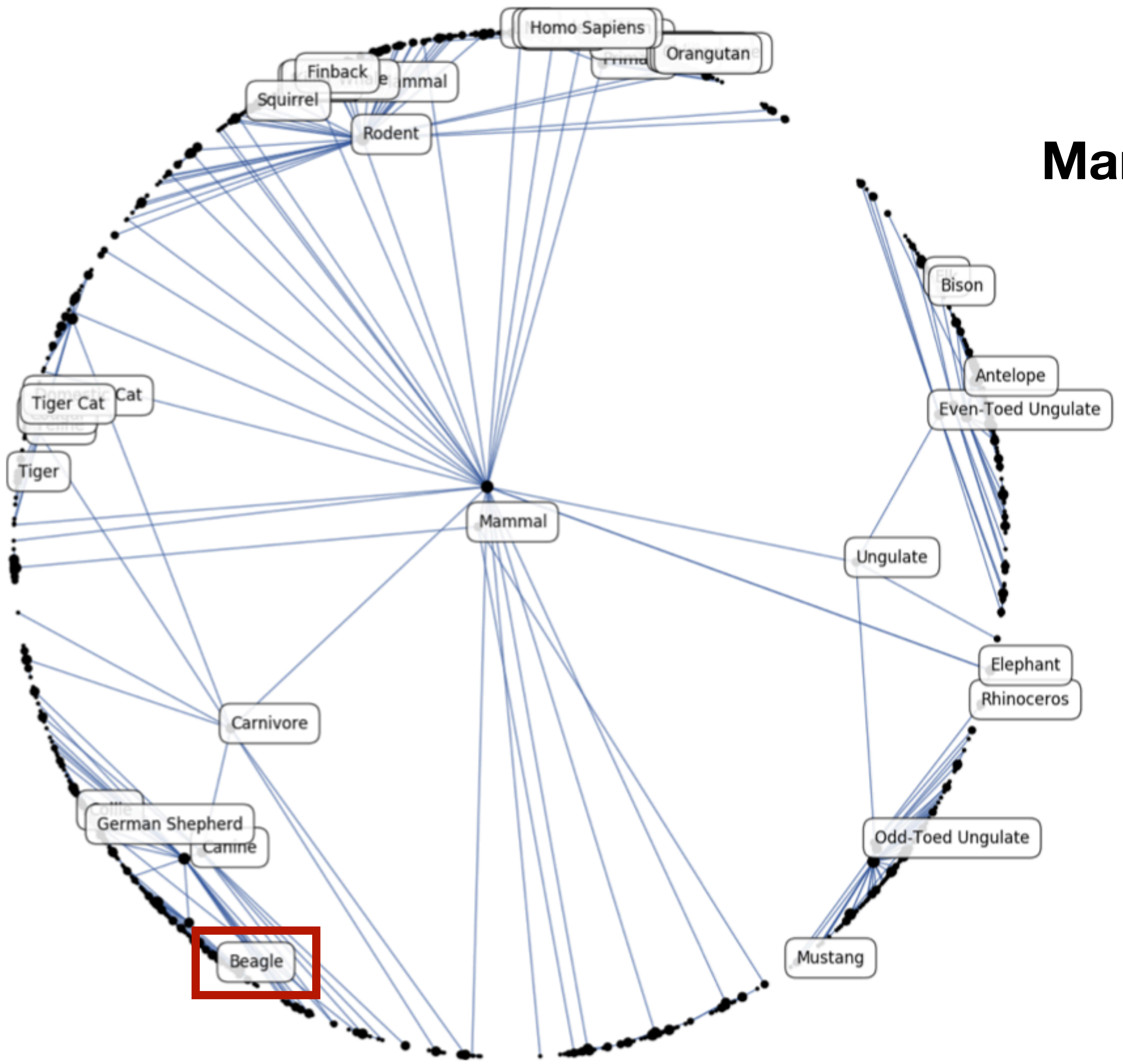
			Dimensionality					
			5	10	20	50	100	200
WordNet **Reconstruction**	**Euclidean**	Rank	3542.3	2286.9	1685.9	1281.7	1187.3	1157.3
		MAP	0.024	0.059	0.087	0.140	0.162	0.168
	Translational	Rank	205.9	179.4	95.3	92.8	92.7	91.0
		MAP	0.517	0.503	0.563	0.566	0.562	0.565
	Poincaré	Rank	4.9	4.02	3.84	3.98	3.9	**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

Reconstruction and Link Prediction

Results



(a) Intermediate embedding after 20 epochs



(b) Embedding after convergence

Mammal : 哺乳动物

carnivore : 食肉动物

canine： 犬类

Beagle: 比格犬

Rodent ： 啮齿动物

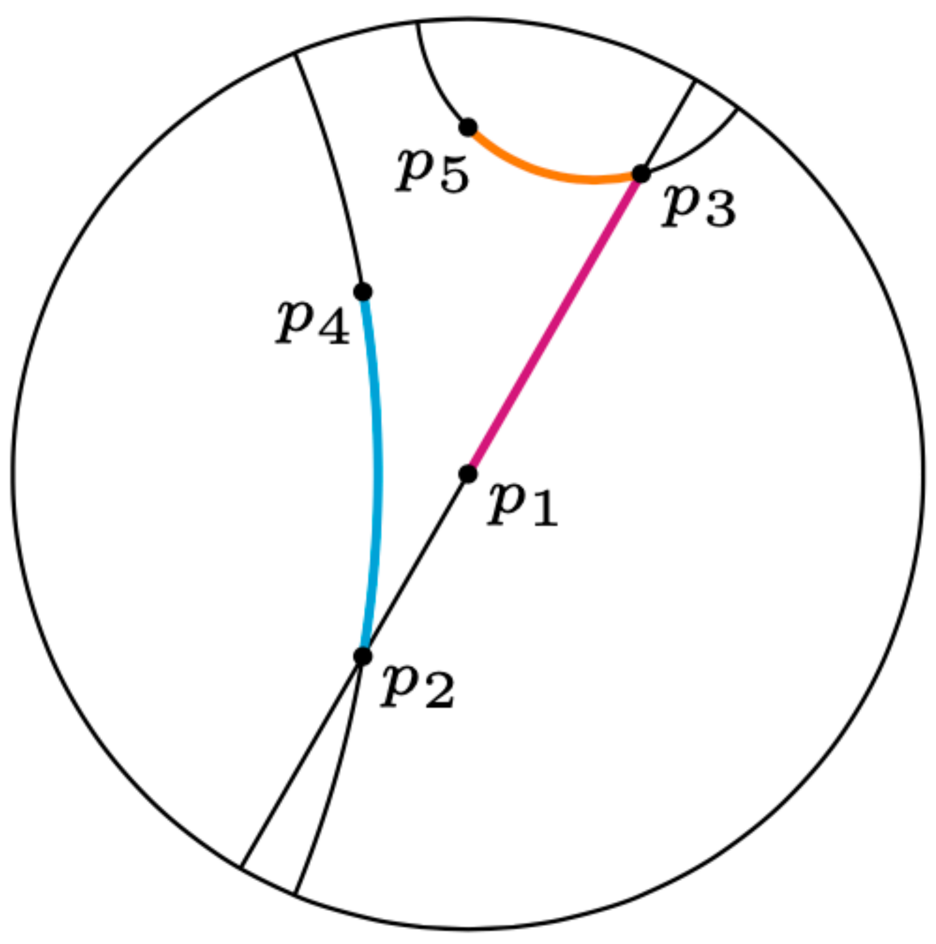
洛伦兹嵌入

摘要：

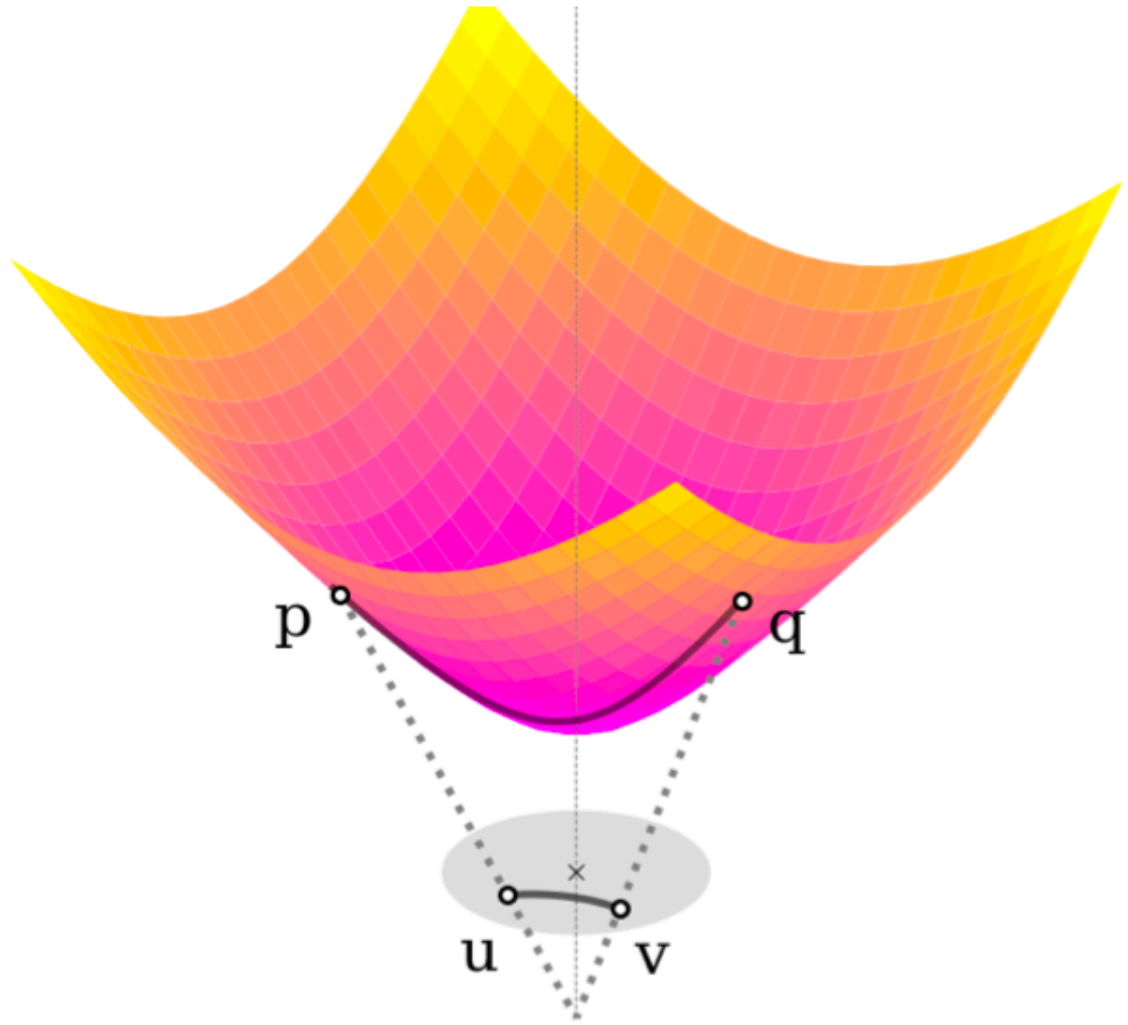
Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry

网络嵌入到另一种双曲模型： Lorentz model

1. **避免了数值不稳定性， 更适合优化过程**
2. **更加高效，准确**



(a) Geodesics in the Poincaré disk.



(b) Lorentz model of hyperbolic geometry.

uv 是pq 在庞加莱盘上映射点

Results

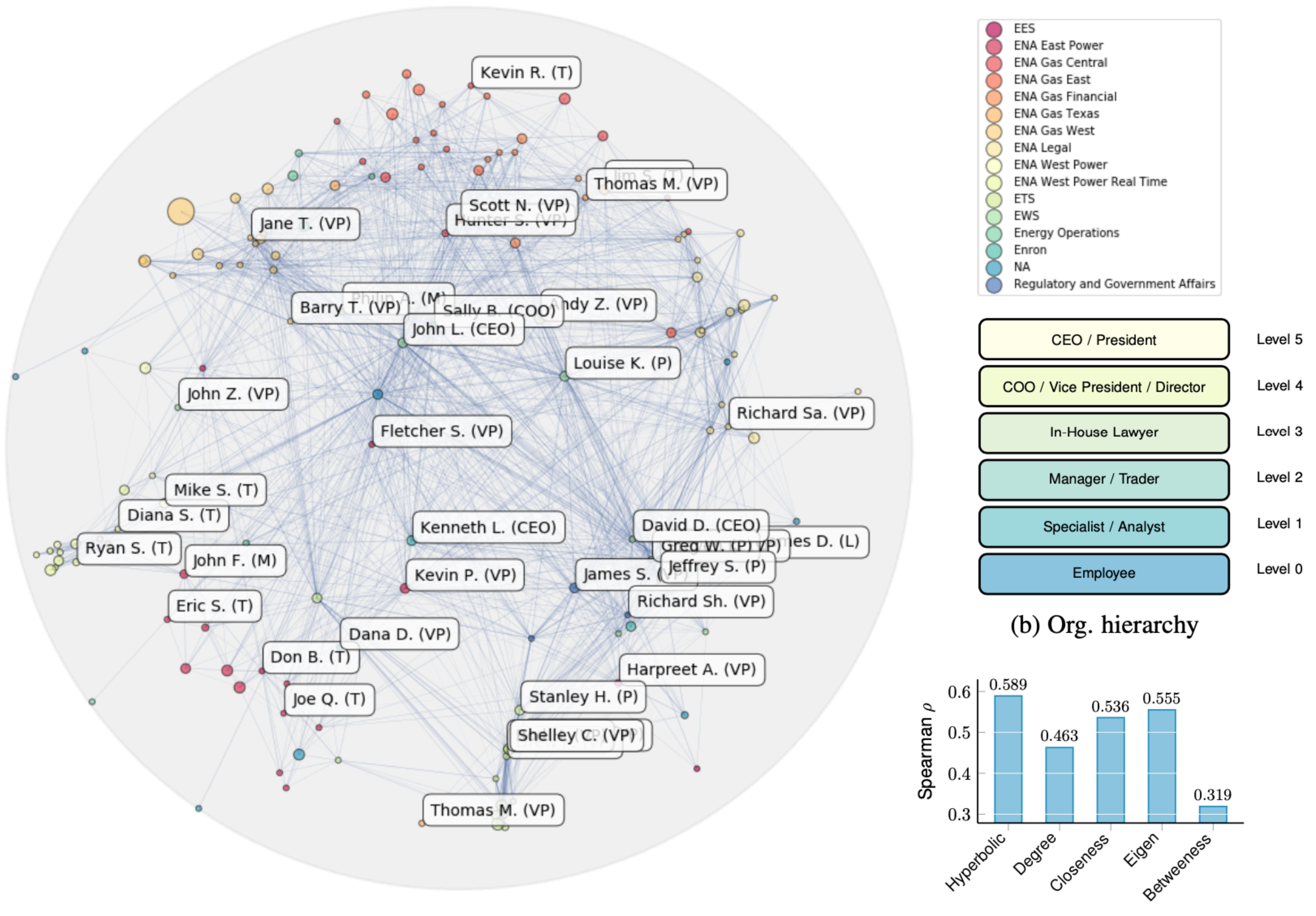
Table 2: Evaluation of Taxonomy Embeddings. MR = Mean Rank, MAP = Mean Average Precision ρ = Spearman rank-order correlation. $\Delta\%$ indicates the relative improvement of optimization in the Lorentz model.

		WordNet Nouns			WordNet Verbs			EuroVoc			ACM			MeSH		
		2	5	10	2	5	10	2	5	10	2	5	10	2	5	10
MR	Poincaré	90.7	4.9	4.02	10.71	1.39	1.35	2.83	1.25	1.23	4.14	1.8	1.71	61.11	14.05	12.8
	Lorentz	22.8	3.18	2.95	3.64	1.26	1.23	1.63	1.24	1.17	3.05	1.67	1.63	38.99	14.13	12.42
	$\Delta\%$	74.8	35.1	36.2	66.0	9.6	8.9	42.4	6.1	3.4	26.3	7.2	4.8	36.2	-0.5	2.9
MAP	Poincaré	11.8	82.8	86.5	36.5	91.0	91.2	64.3	94.0	94.4	69.3	94.1	94.8	19.5	76.3	79.4
	Lorentz	30.5	92.3	92.8	57.9	93.5	93.3	87.1	95.8	96.5	82.9	96.6	97.0	34.8	77.7	79.9
	$\Delta\%$	61.3	10.3	6.8	58.6	2.7	2.3	35.6	1.6	2.0	19.6	2.7	2.3	43.9	1.8	0.6
ρ	Poincaré	13.8	57.2	58.5	11.0	54.1	55.1	37.5	57.5	61.4	59.8	63.5	62.9	42.2	69.9	74.9
	Lorentz	41.0	58.9	59.5	47.9	55.5	56.6	54.5	61.7	67.5	65.9	65.9	65.9	64.5	71.4	76.3

低维情况下表现更好

Results

不仅有分层， 还有相似性



(a) Embedding of the Enron communication graph

(b) Org. hierarchy

(c) Rank-order correlation

Figure 2: Embedding of the Enron email corpus. Abbreviations in parentheses indicate organizational role: CEO = Chief Executive Officer, COO = Chief Operating Officer, P = President, VP = Vice President, D = Director, M = Manager, T = Trader. Blue lines indicate edges in the graph. Node size indicates node degree.

双曲嵌入理论分析

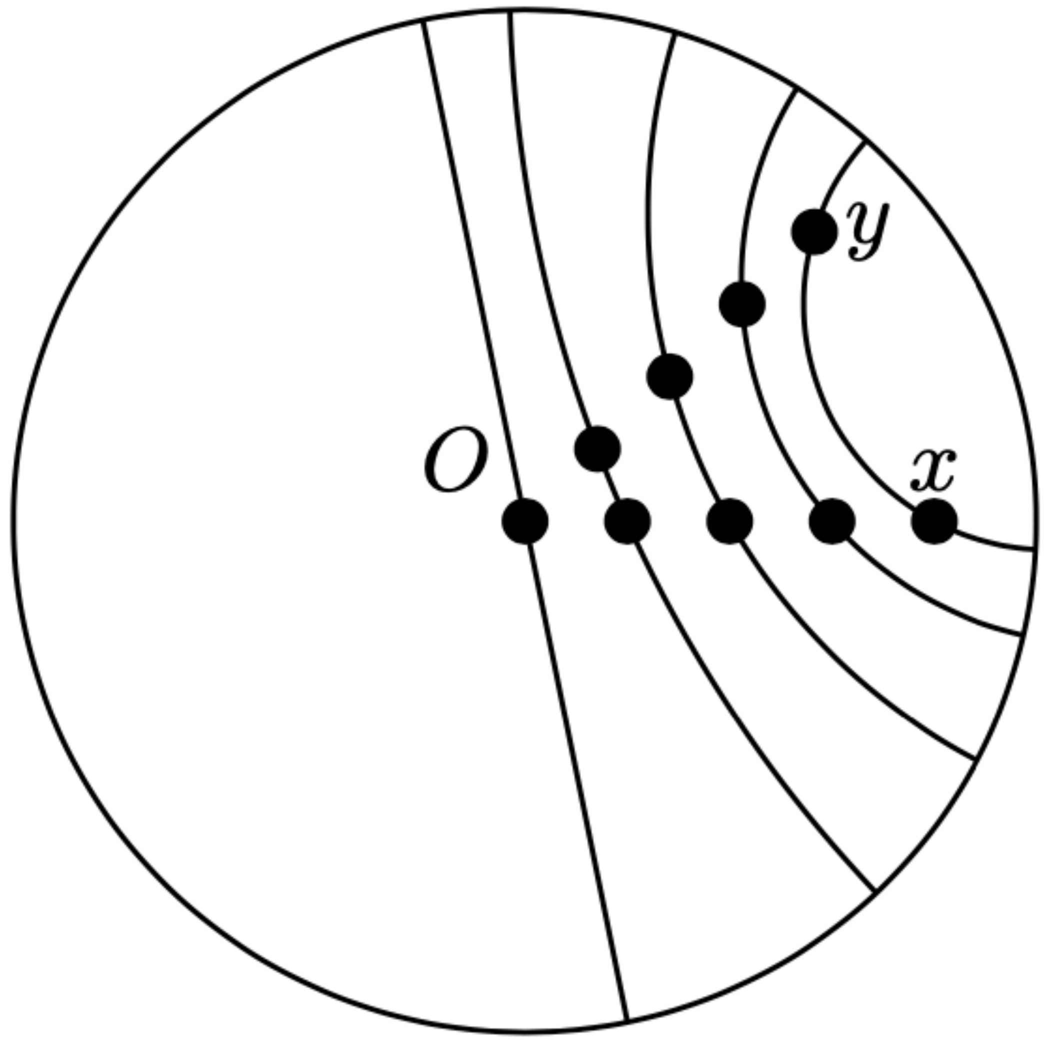
Representation Tradeoffs for Hyperbolic Embeddings

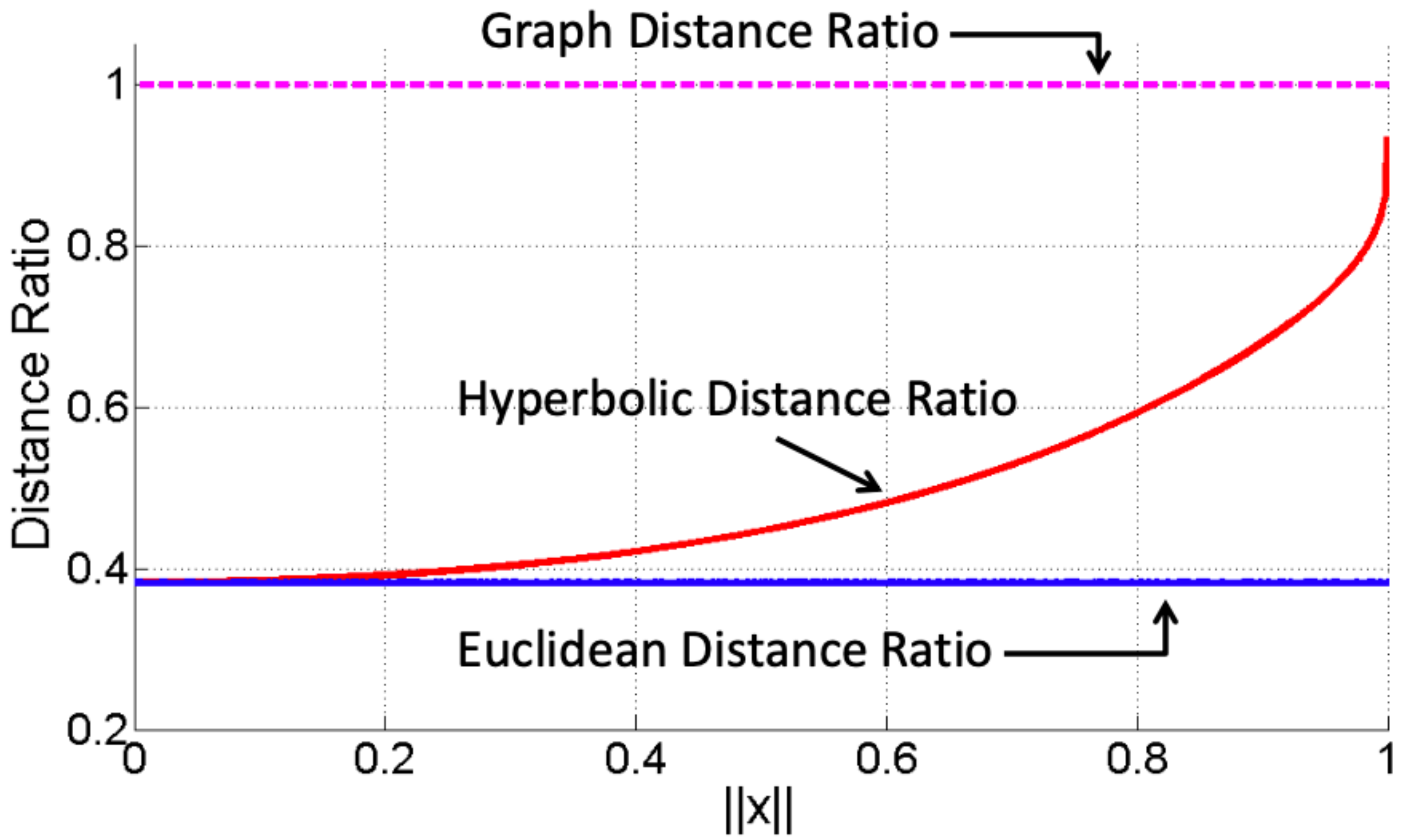
摘要

1. 提出一种组合嵌入的方法:
   1. 将 graph 嵌入到 a weighted tree
   2. 将 tree 嵌入到 hyperbolic disk
2. 理论上分析， 精度和维度之间的 trade-off
3. 十分充足的理论证明分析(30 页左右)

为什么双曲空间能够嵌入树结构

$\|x\| = \|y\| = t$





当 t —> 1

$$\frac{d_E(x, y)}{d_E(x,0) + d_E(0,y)}$$ **恒定**

$$\frac{d_H(x, y)}{d_H(x,0) + d_H(0,y)}$$ **越来越接近 1**

$d_H(x, y)$ **approaches** $d_H(x,0) + d_H(0,y)$

这类似于树的属性，其中两个兄弟节点之间的最短路径是通过它们的父节点的路径

Model

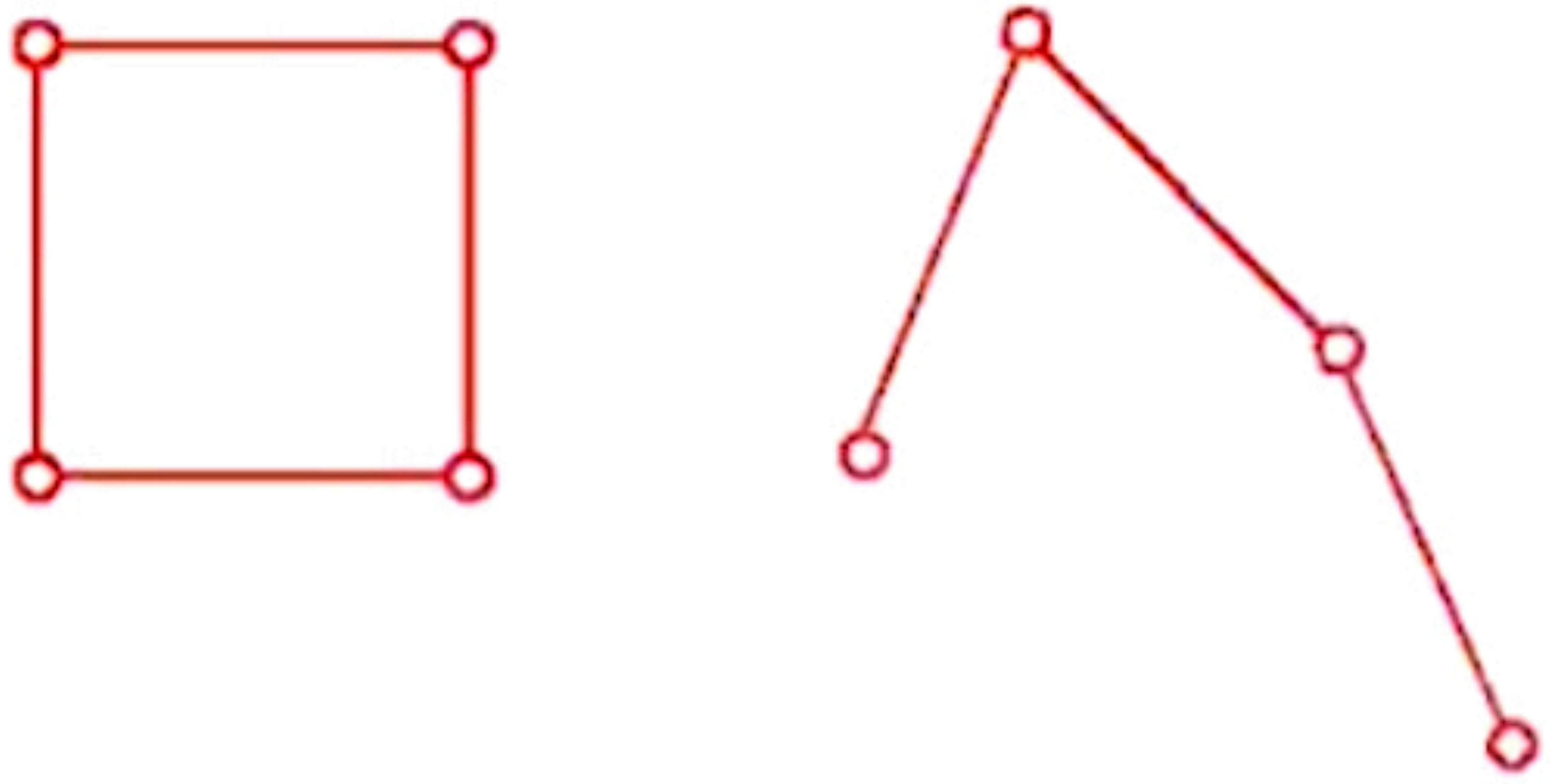
1. Embed the graph $G = (V, E)$ into a tree T,
2. Embed T into the Poincaré ball $\mathbb{H}_d$.

Sarkar's Construction

First, $f(a)$ and $f(b)$ are reflected across a geodesic (using circle inversion) so that $f(a)$ is mapped onto the origin 0 and $f(b)$ is mapped onto some point z. Next, we place the children nodes to vectors $y_1, \ldots, y_{d-1}$ equally spaced around a circle with radius $\frac{e^\tau - 1}{e^\tau + 1}$ (which is a circle of radius τ in the hyperbolic metric), and maximally separated from the reflected parent node embedding z. Lastly, we reflect all of the points back across the geodesic. Note that the isometric properties of reflections imply that all children are now at hyperbolic distance exactly τ from $f(a)$.

衡量网络的 tree-like

4-point condition



δ-hyperbolicity (δ-4PC)

有向图的双曲嵌入

Hyperbolic Entailment Cones for Learning Hierarchical Embedding

摘要