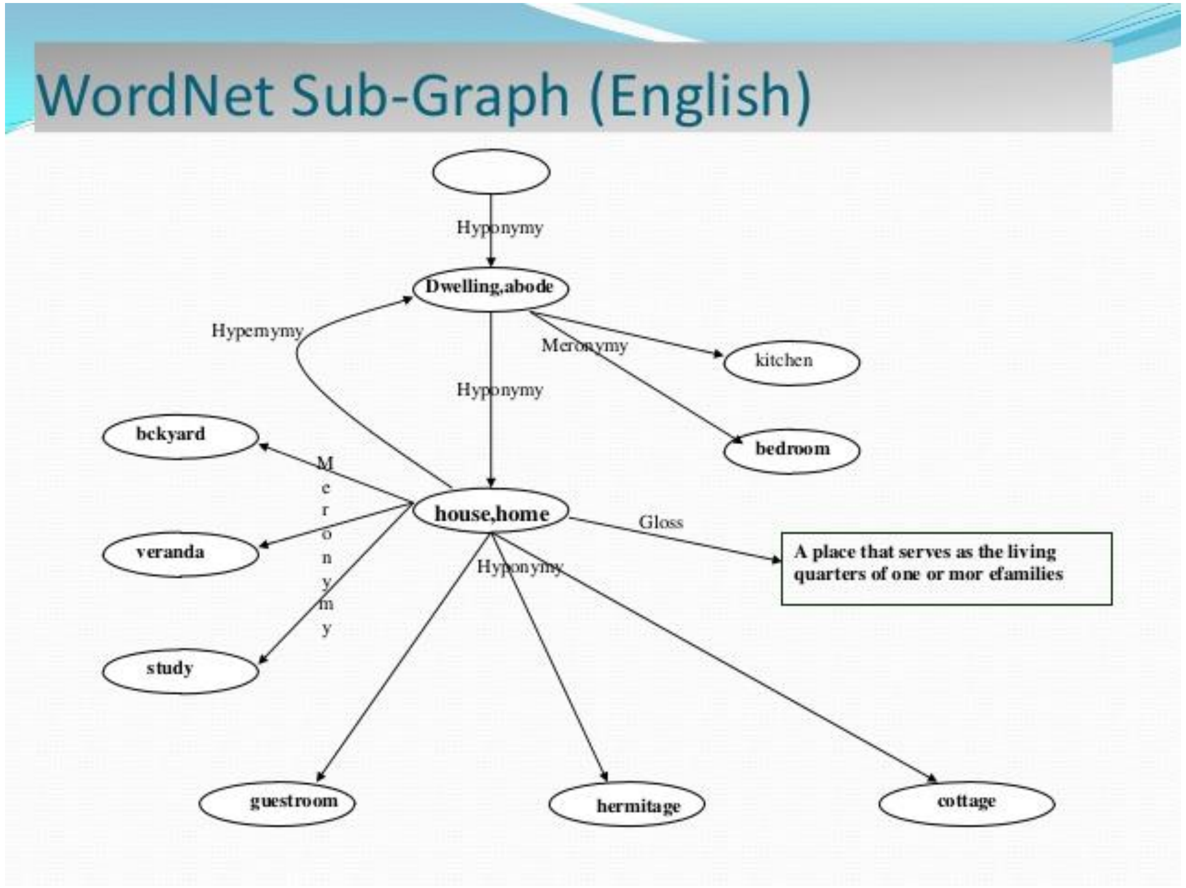
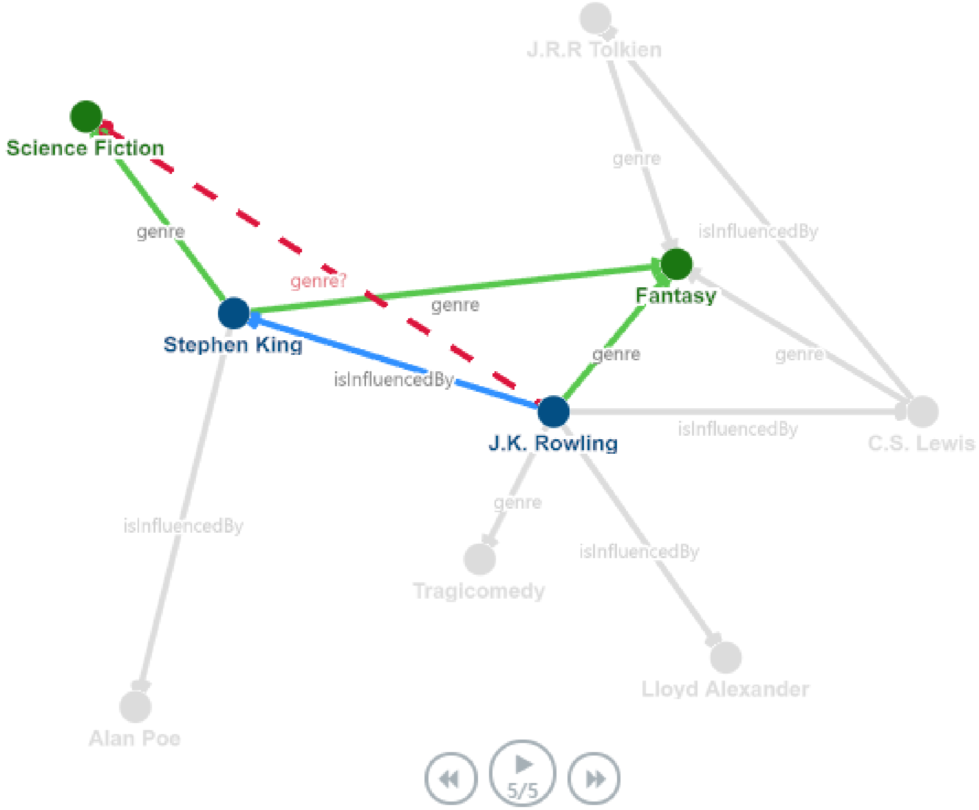
Poincaré Embeddings for Learning Hierarchical Representations

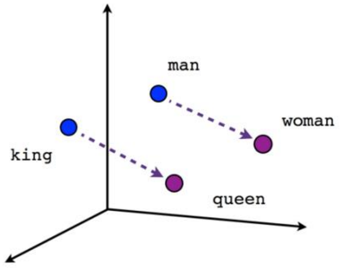
Neural Information Processing Systems 2017 (2019 NLP GROUP STUDY)

Data Mining & Information Systems Lab.
Department of Computer Science and Engineering,
College of Informatics, Korea University

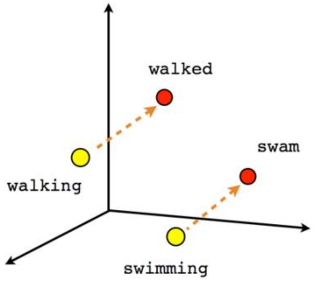
이잉걸



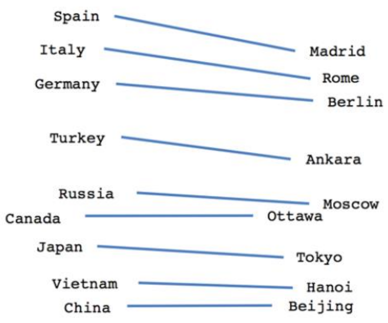




Male-Female

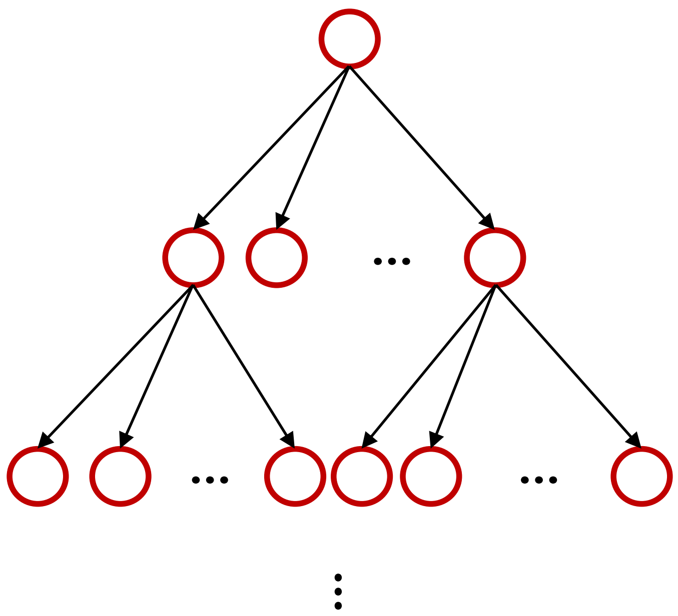


Verb tense



Country-Capital

Characteristics(?) of tree structured graph



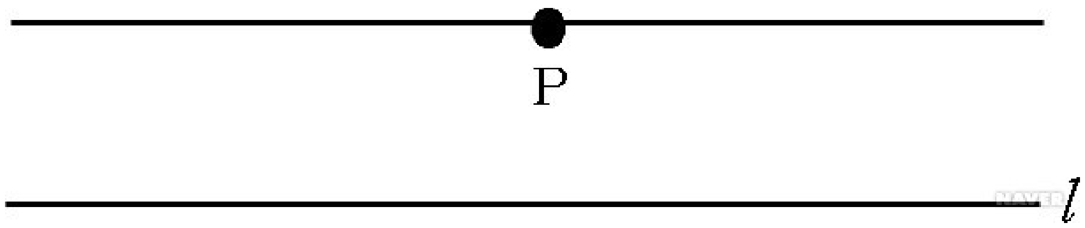
Suppose there are b branching factors on each nodes

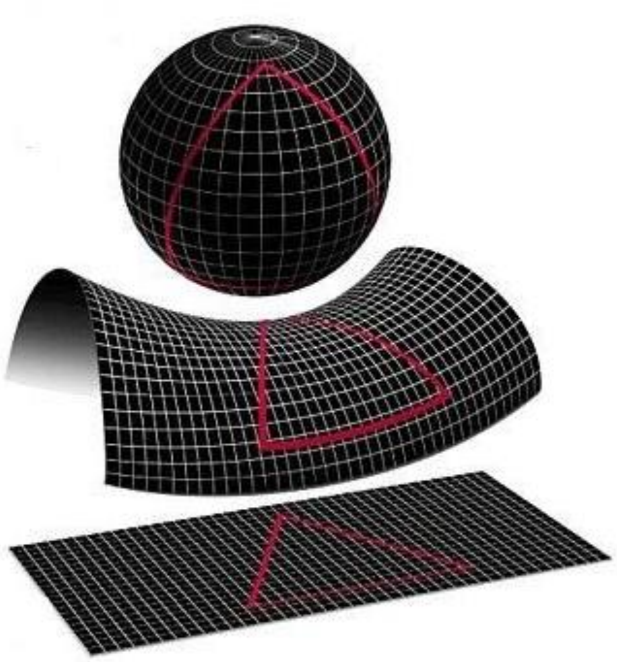
At level l, there are $(b+1)b^l$ nodes

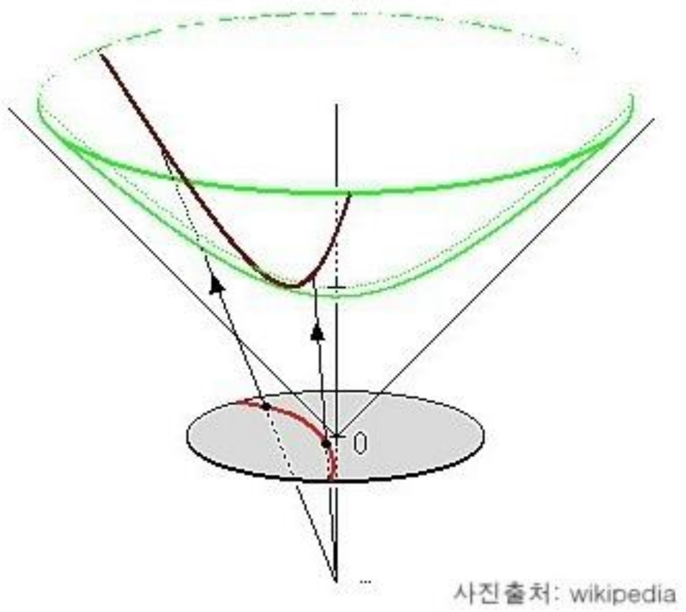
There are $\frac{\left((b+1)b^l-2\right)}{b-1}$ nodes on a level less or equal than l

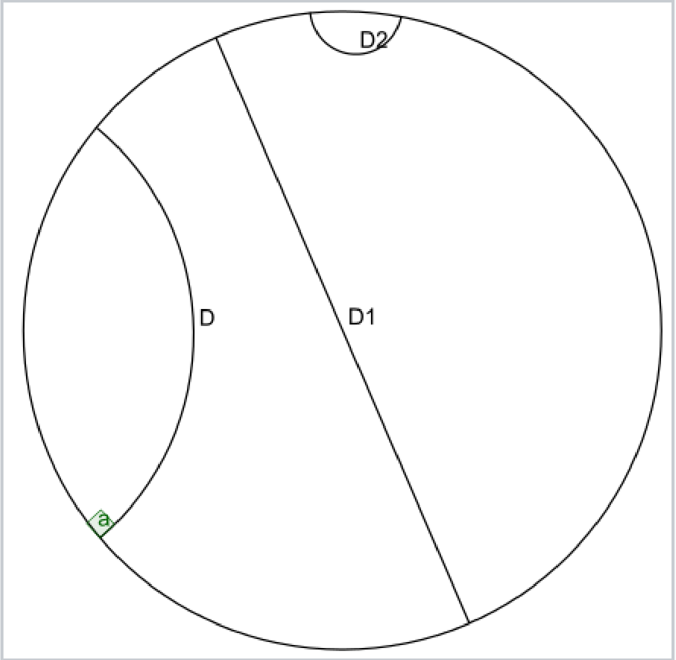
Note : the # of nodes are exponentially increasing
as level l (distance to the root of the tree) increases

선 밖의 한 점을 지나 그 직선에 평행한 직선은 단 하나만 존재한다.[1]

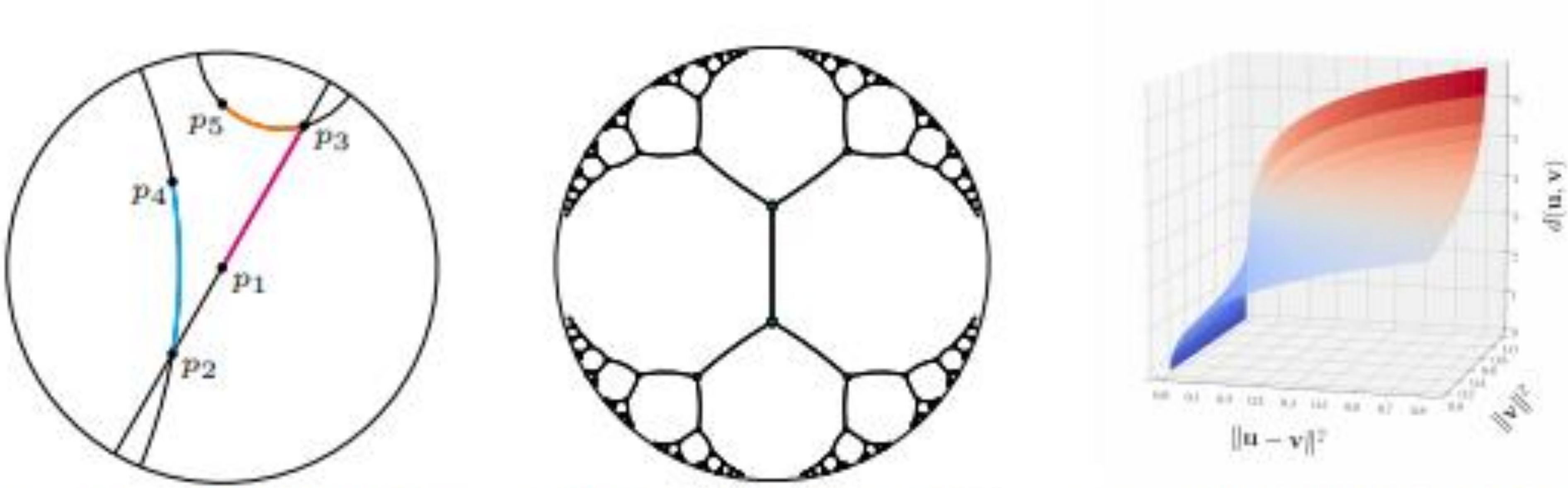








Poincaré disk with 3 ultraparallel (hyperbolic) straight lines



(a) Geodesics of the Poincaré disk (b) Embedding of a tree in $\mathcal{B}^2$ (c) Growth of Poincaré distance

Figure 1: (a) Due to the negative curvature of $\mathcal{B}$, the distance of points increases exponentially (relative to their Euclidean distance) the closer they are to the boundary. (c) Growth of the Poincaré distance $d(\boldsymbol{u}, \boldsymbol{v})$ relative to the Euclidean distance and the norm of $\boldsymbol{v}$ (for fixed $\|\boldsymbol{u}\| = 0.9$). (b) Embedding of a regular tree in $\mathcal{B}^2$ such that all connected nodes are spaced equally far apart (i.e., all black line segments have identical hyperbolic length).

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

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

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

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

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

$\alpha = 1 - \|\boldsymbol{\theta}\|^2$, $\beta = 1 - \|\boldsymbol{x}\|^2$ and let $\gamma = 1 + \frac{2}{\alpha\beta}\|\boldsymbol{\theta} - \boldsymbol{x}\|^2$

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

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

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

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

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

$\alpha = 1 - \|\boldsymbol{\theta}\|^2$, $\beta = 1 - \|\boldsymbol{x}\|^2$ and let $\gamma = 1 + \frac{2}{\alpha\beta}\|\boldsymbol{\theta} - \boldsymbol{x}\|^2$

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

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

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

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

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

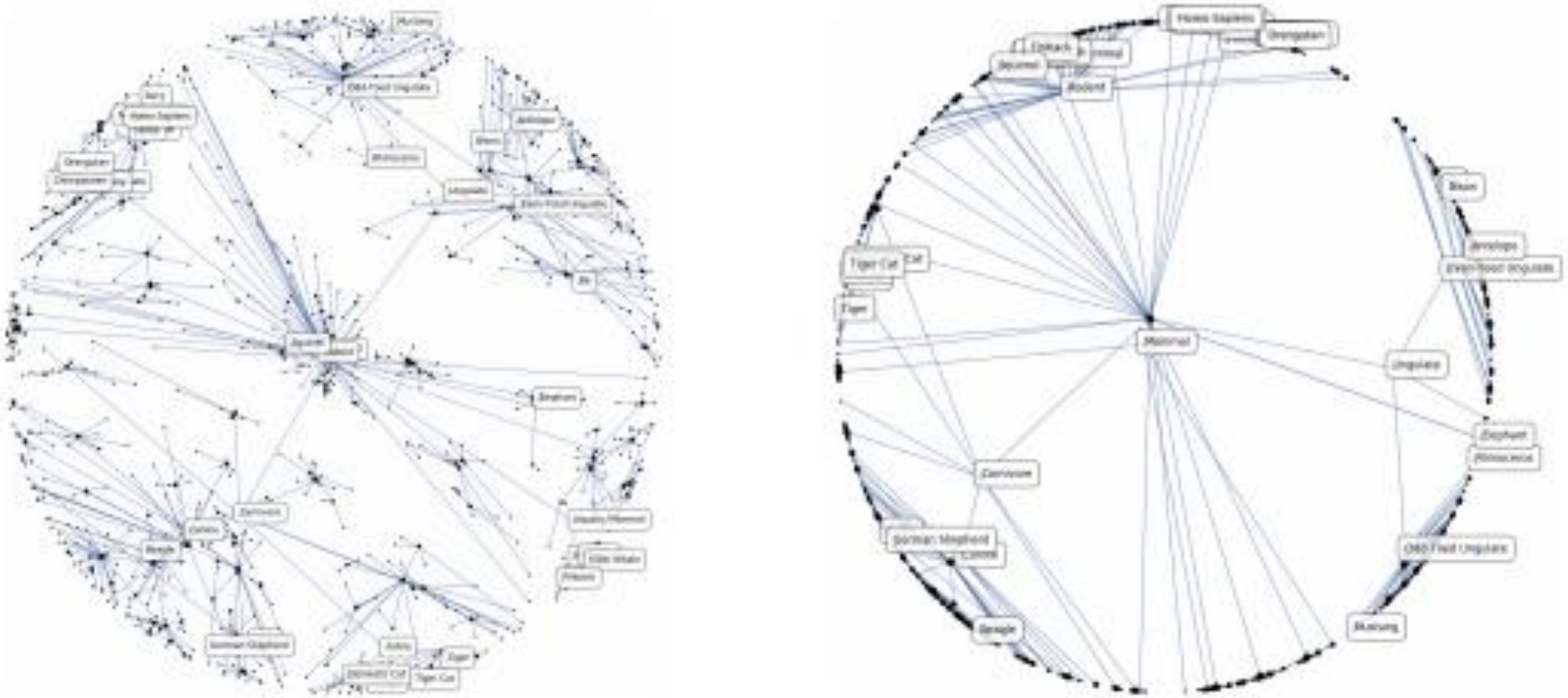
$$\mathfrak{R}_{\theta}(\boldsymbol{v}) = \boldsymbol{\theta} + \boldsymbol{v}$$

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

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

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 acheive 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



(a) Intermediate embedding after 20 epochs

(b) Embedding after convergence

Figure 2: Two-dimensional Poincaré embeddings of transitive closure of the WORDNET mammals subtree. Ground-truth `is-a` relations of the original WORDNET tree are indicated via blue edges. A Poincaré embedding with $d = 5$ achieves mean rank 1.26 and MAP 0.927 on this subtree.

Table 2: Mean average precision for Reconstruction and Link Prediction on network data.

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

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