Poincaré Embeddings for Learning Hierarchical Representations

Урьев Максим

Декабрь 11, 2017

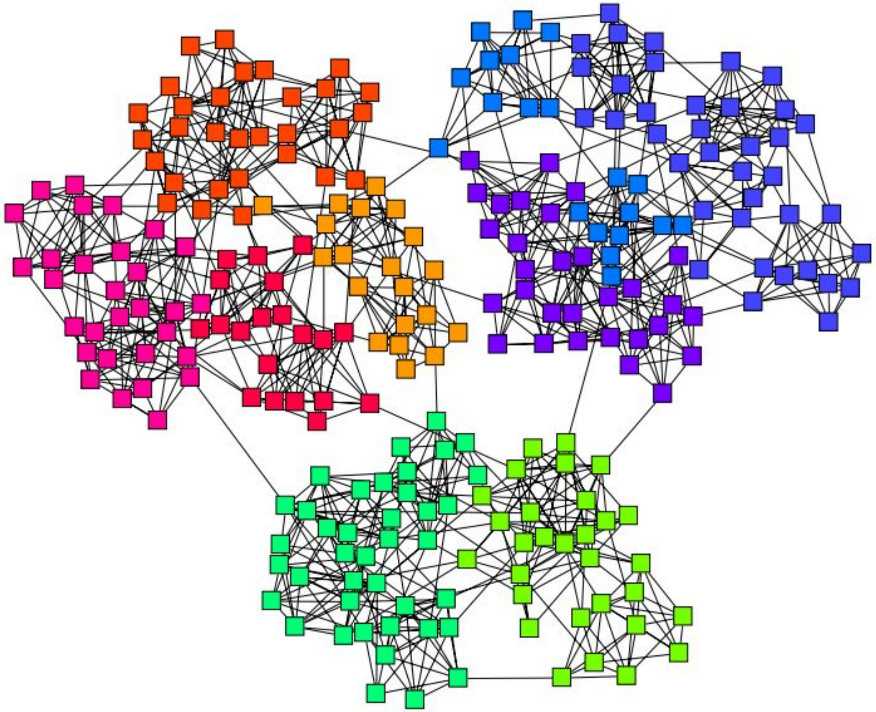
Statement

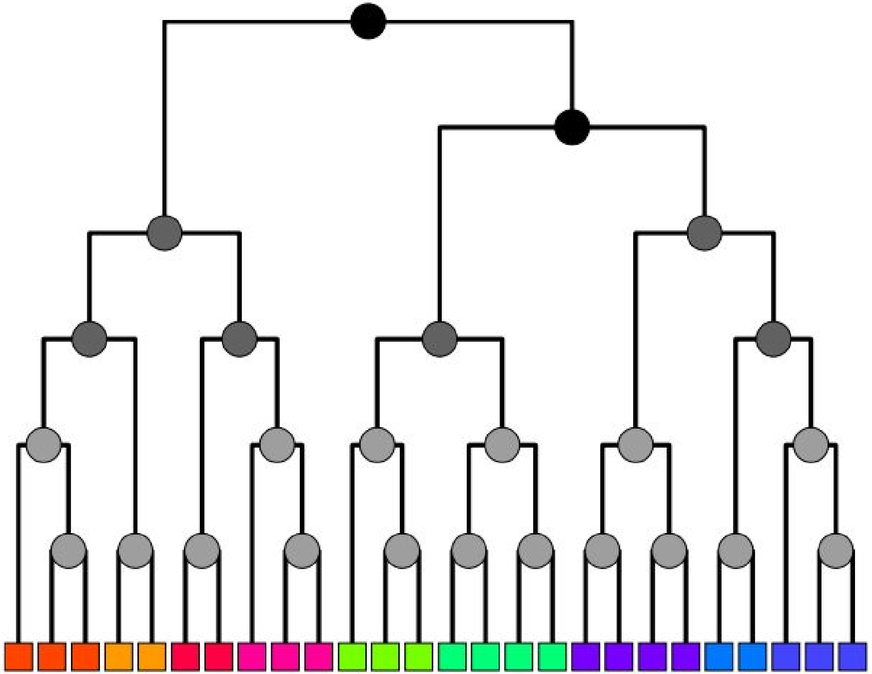
We are then interested in finding embeddings $\Theta = \{\boldsymbol{\theta}_i\}_{i=1}^n$ for a set of symbols $\mathcal{S} = \{x_i\}_{i=1}^n$

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

- Reconstruction
- Generalization
- Gradation

Hierarchical Structure





Order Embeddings

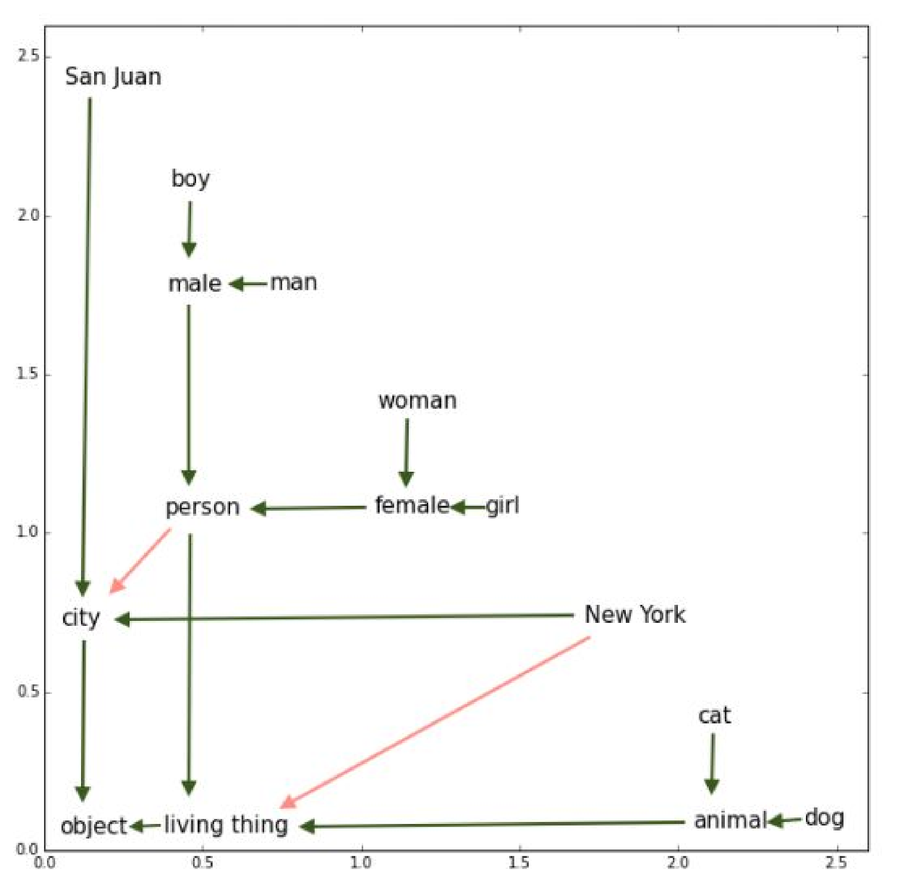
A function $f : (X, \preceq_X) \rightarrow (Y, \preceq_Y)$ *is an order-embedding if for all* $u, v \in X$,

$$u \preceq_X v \text{ if and only if } f(u) \preceq_Y f(v)$$

$$E(x, y) = ||\max(0, y - x)||^2$$

$$\sum_{(u,v) \in P} E(f(u), f(v)) + \sum_{(u',v') \in N} \max\{0, \alpha - E(f(u'), f(v'))\}$$

Order Embeddings



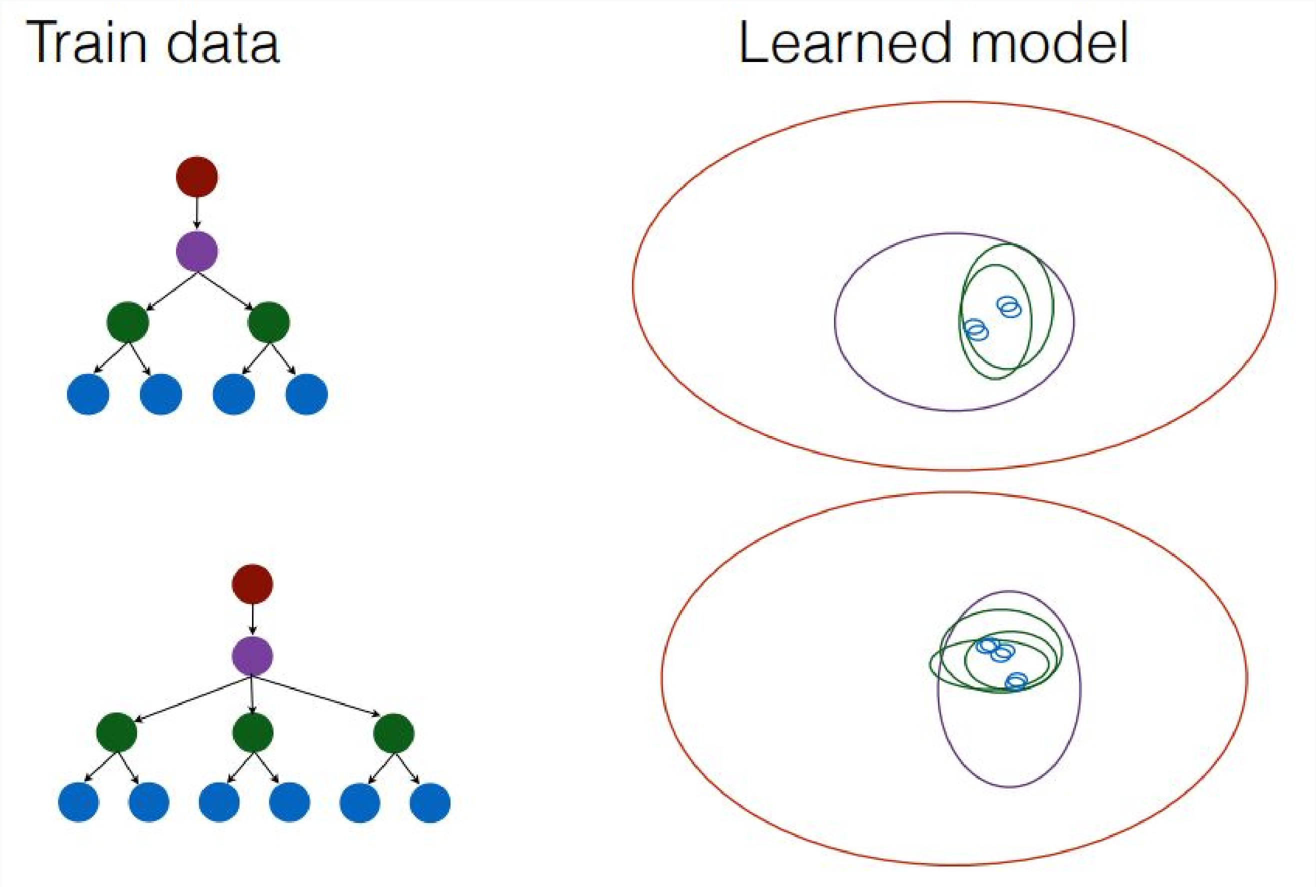
Gaussian Embeddings

$$E(P_i, P_j) = \int_{x \in \mathbb{R}^n} \mathcal{N}(x; \mu_i, \Sigma_i) \mathcal{N}(x; \mu_j, \Sigma_j) dx = \mathcal{N}(0; \mu_i - \mu_j, \Sigma_i + \Sigma_j)$$

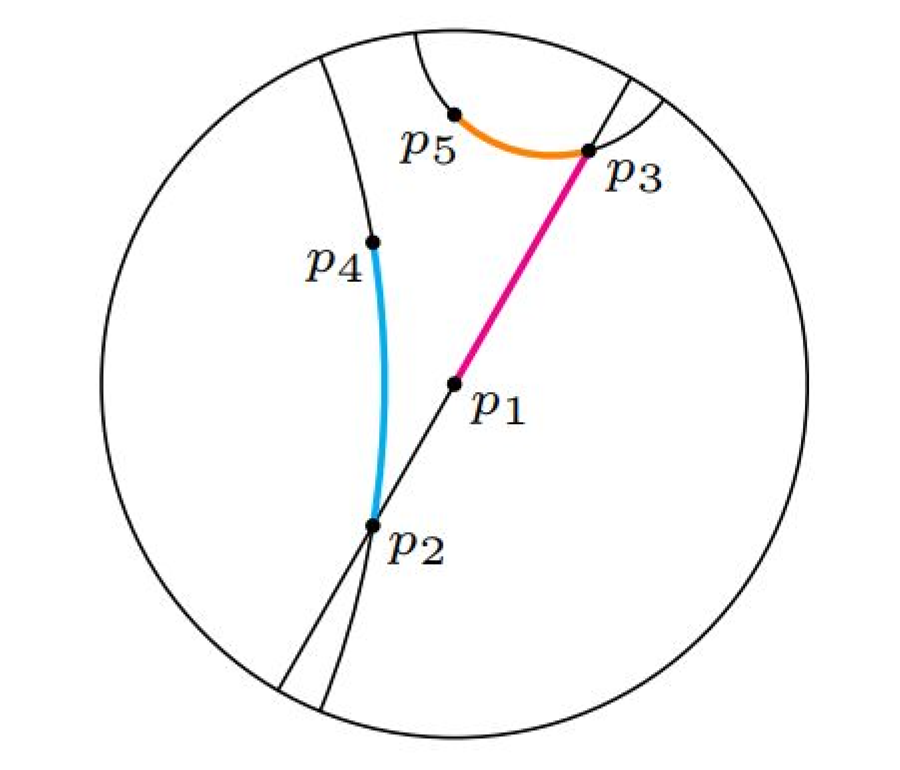
$$-E(P_i, P_j) = D_{KL}(\mathcal{N}_j || \mathcal{N}_i) = \int_{x \in \mathbb{R}^n} \mathcal{N}(x; \mu_i, \Sigma_i) \log \frac{\mathcal{N}(x; \mu_j, \Sigma_j)}{\mathcal{N}(x; \mu_i, \Sigma_i)} dx$$

$$L_m(w, c_p, c_n) = \max(0, m - E(w, c_p) + E(w, c_n))$$

Gaussian Embeddings



Poincar´e unit disk model

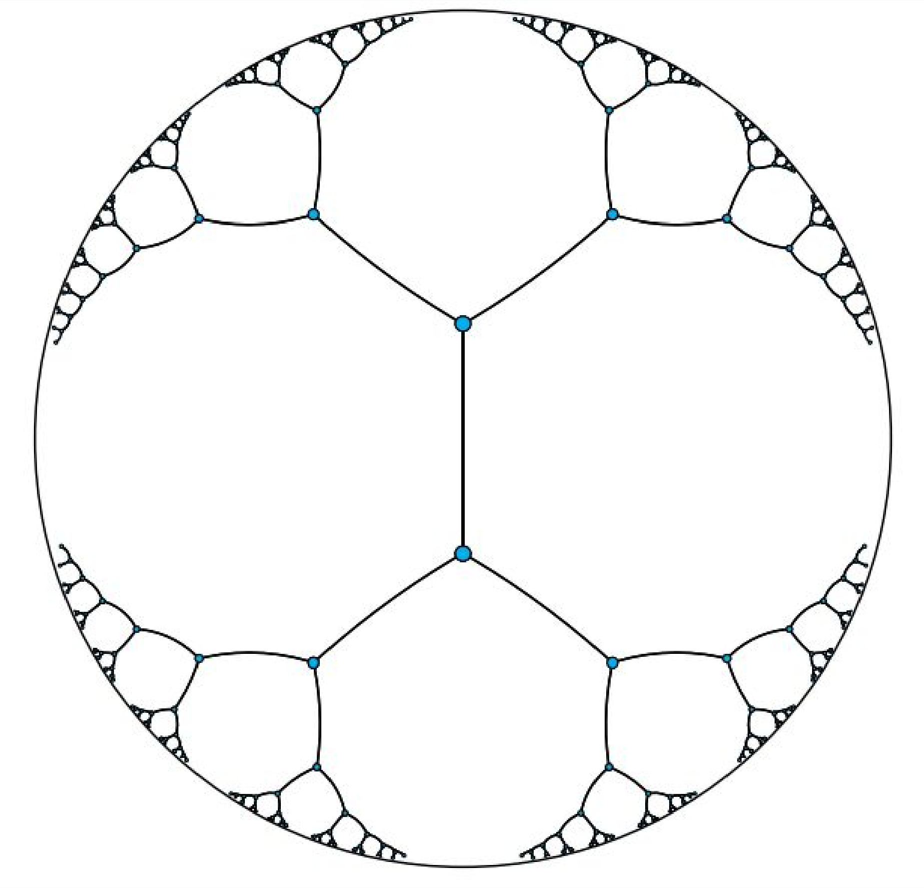


Poincar´e unit disk model

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

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

Embeddings and Hyperbolic Geometry

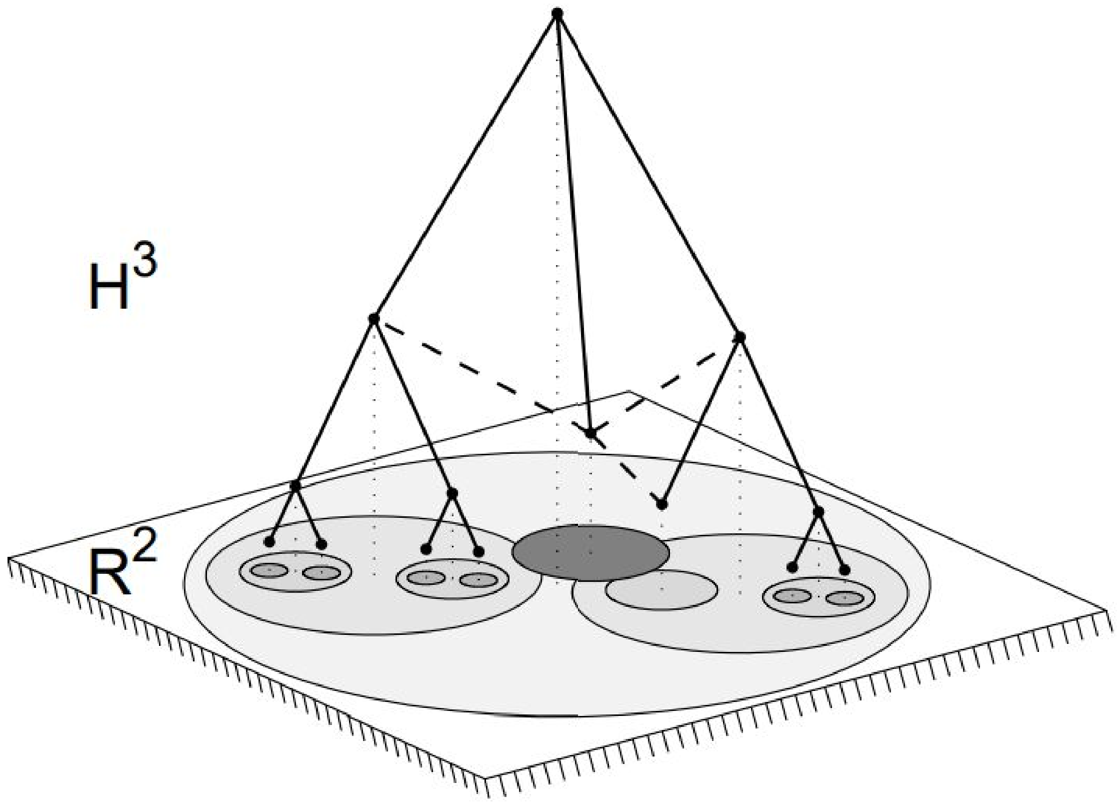


Embeddings and Hyperbolic Geometry

Definition 1: Let $0 \leq \delta < \infty$. (X, ℓ) is called *4-point δ-hyperbolic* if and only if for all $x, y, u, v \in X$, ordered such that $\ell(x,y) + \ell(u,v) \geq \ell(x,u) + \ell(y,v) \geq \ell(x,v) + \ell(y,u)$, the following condition holds:

$$(\ell(x,y) + \ell(u,v)) - (\ell(x,u) + \ell(y,v)) \leq 2\delta. \qquad (1)$$

Embeddings and Hyperbolic Geometry



Optimization

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

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

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

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

The algorithm is straightforward to parallelize via methods such as Hogwild.

Training Details

- Good initial angular layout can be helpful to find good embeddings
- Initialize all embeddings randomly from the uniform distribution

  $U(-0.001, 0.001)$.
- Train during an initial phase with a reduced learning rate

Evaluation: WordNet

Euclidean: $$d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|^2$$

Translational: $$d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v} + \boldsymbol{r}\|^2$$

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

Evaluation: WordNet

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

Evaluation: Network Embeddings

$$P((u,v) = 1 \mid \Theta) = \frac{1}{e^{(d(\boldsymbol{u},\boldsymbol{v})-r)/t} + 1}$$

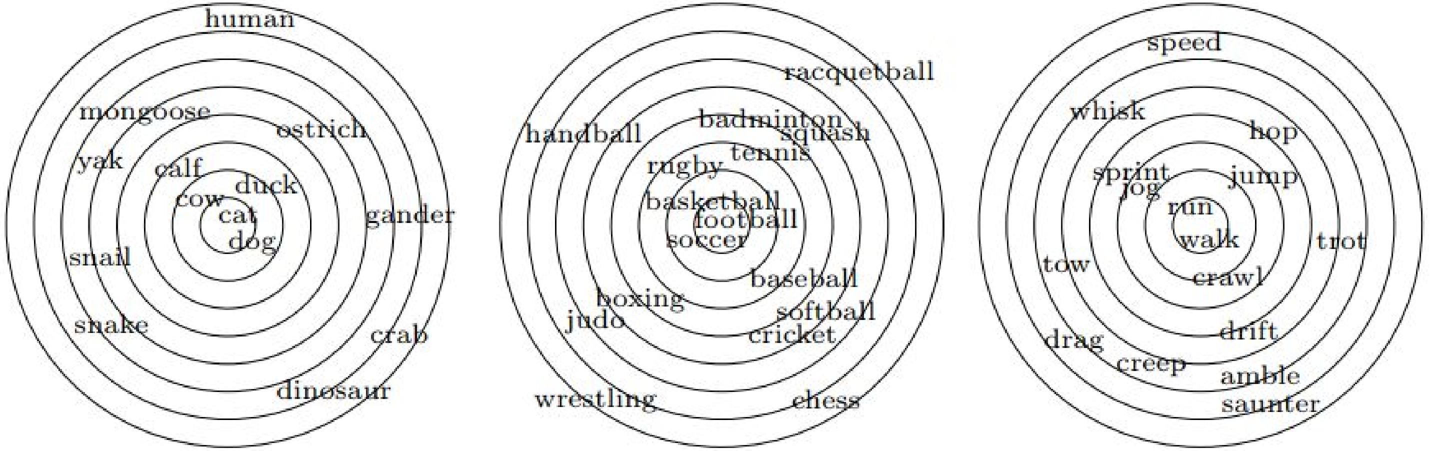
Cross-entropy loss to learn the embeddings and sample negatives

Evaluation: Network Embeddings

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 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

Evaluation: Lexical Entailment



(a) ANIMAL (b) SPORT (c) TO MOVE

$$\text{score}(\texttt{is-a}(u,v)) = -(1+\alpha(\|\boldsymbol{v}\|-\|\boldsymbol{u}\|))d(\boldsymbol{u},\boldsymbol{v})$$

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

Future Work

- Expand model to multi-relational data
- Full Riemannian optimization approach

References

- Maximillian Nickel, Douwe Kiela, **Poincaré Embeddings for Learning Hierarchical Representations.** https://arxiv.org/pdf/1705.08039.pdf
- Aaron B Adcock, Blair D Sullivan, and Michael W Mahoney. Tree-like structure in large social and information networks. *In Data Mining (ICDM), 2013 IEEE 13th International Conference on, pages 1–10. IEEE*, 2013. https://www.stat.berkeley.edu/~mmahoney/pubs/treelike-icdm13.pdf
- Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná. Hyperbolic geometry of complex networks. *Physical Review E, 82(3):036106*, 2010. https://arxiv.org/pdf/1006.5169.pdf
- Ivan Vulic, Daniela Gerz, Douwe Kiela, Felix Hill, and Anna Korhonen. Hyperlex: A large-scale evaluation of graded lexical entailment. https://arxiv.org/pdf/1608.02117.pdf