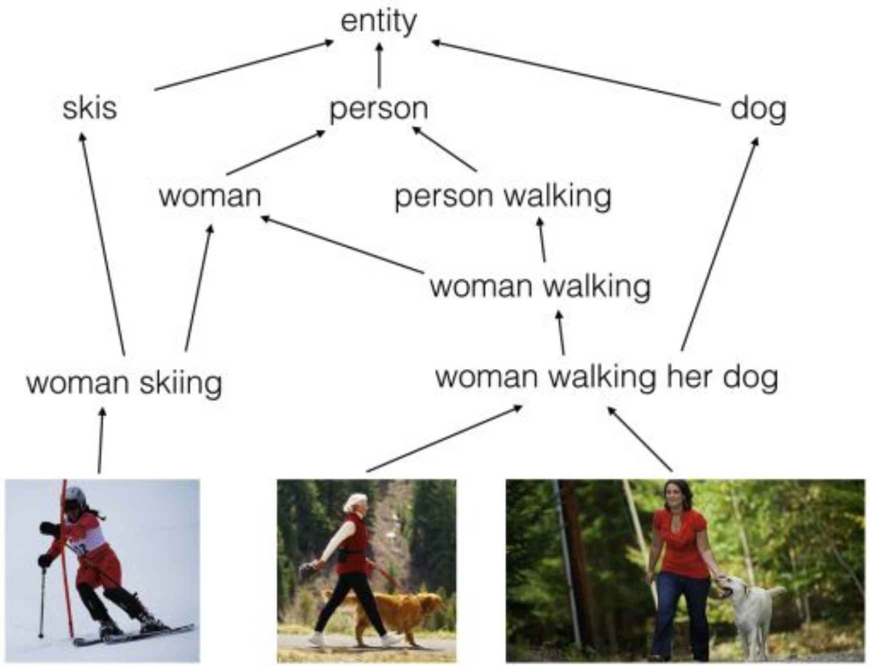
Hierarchical Image-Text Representations

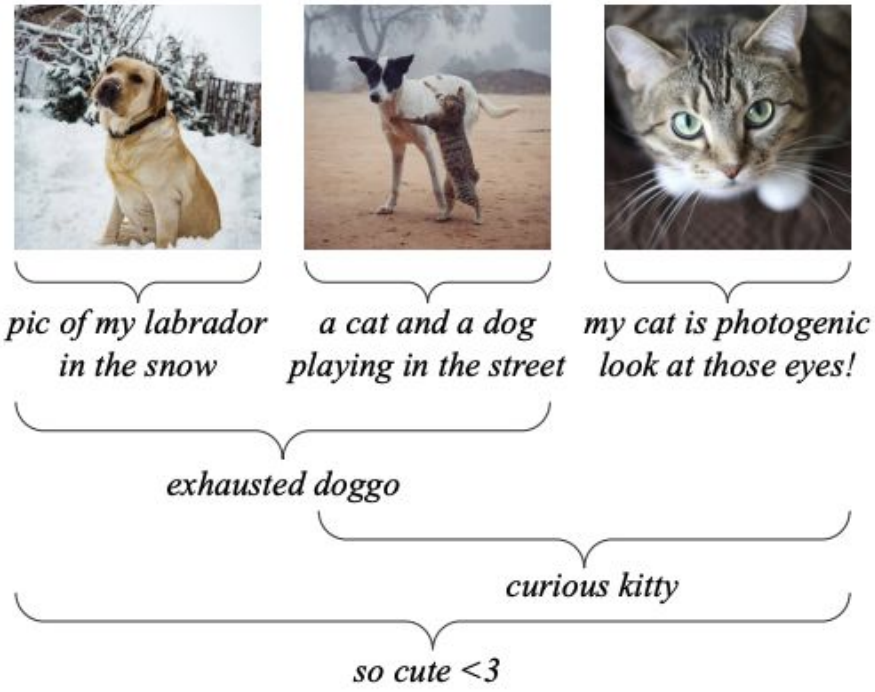
Aksenov Yaroslav

Partial order

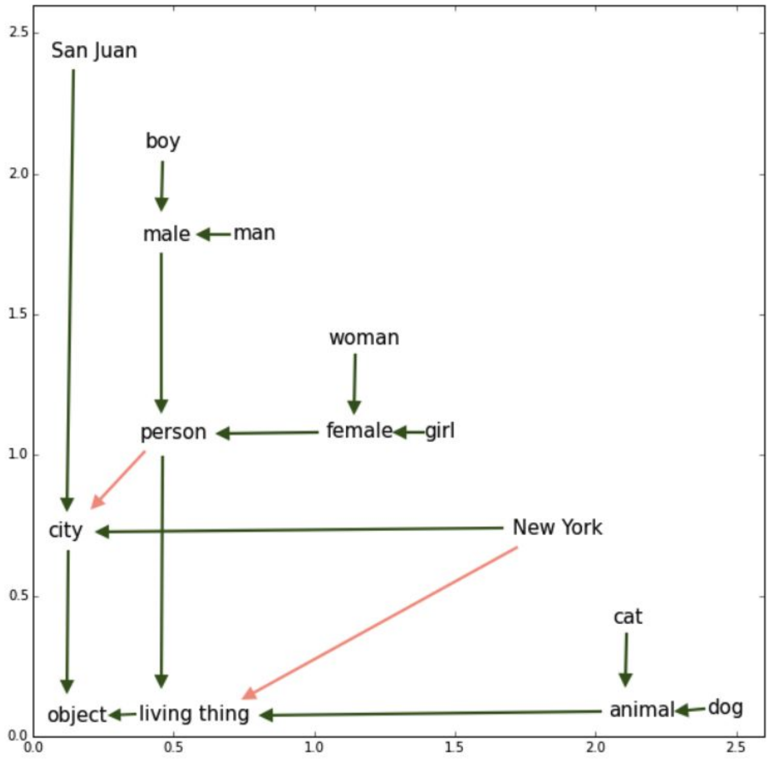
Definition 1. *A function* $f : (X, \preceq_X) \rightarrow (Y, \preceq_Y)$ *is an order-embedding if for all* $u, v \in X$,

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





Order-Embeddings of Images and Language (Vendrov at al. 2016)



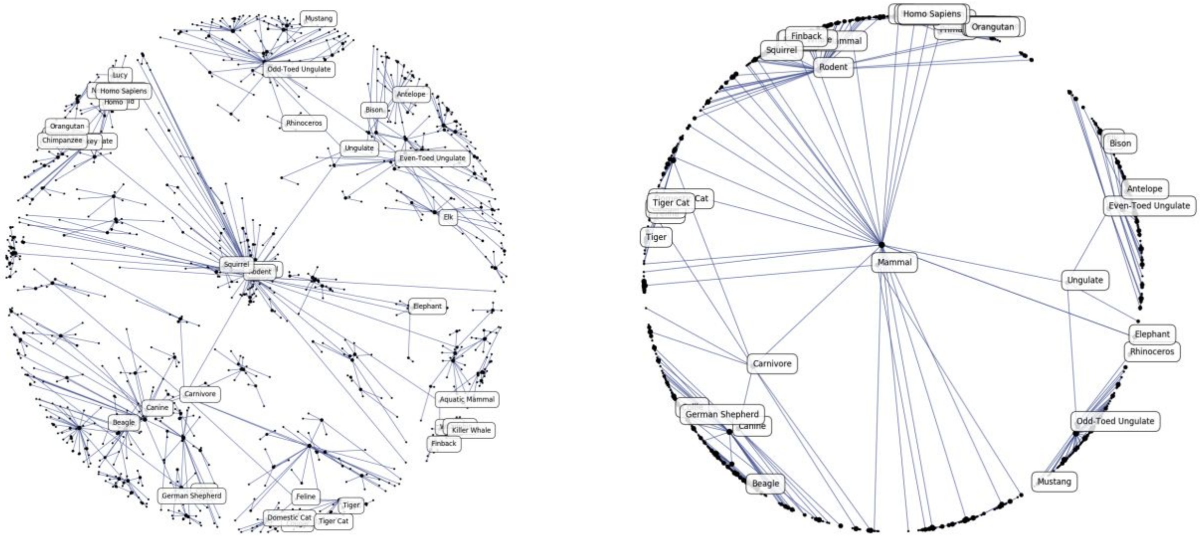
$$x \preceq y \text{ if and only if } \bigwedge_{i=1}^{N} x_i \geq y_i$$

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

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

WordNet – pairs of (common, specific)

Poincaré Embeddings for Learning Hierarchical Representations (Nickel, Kiela 2017)



(a) Intermediate embedding after 20 epochs

(b) Embedding after convergence

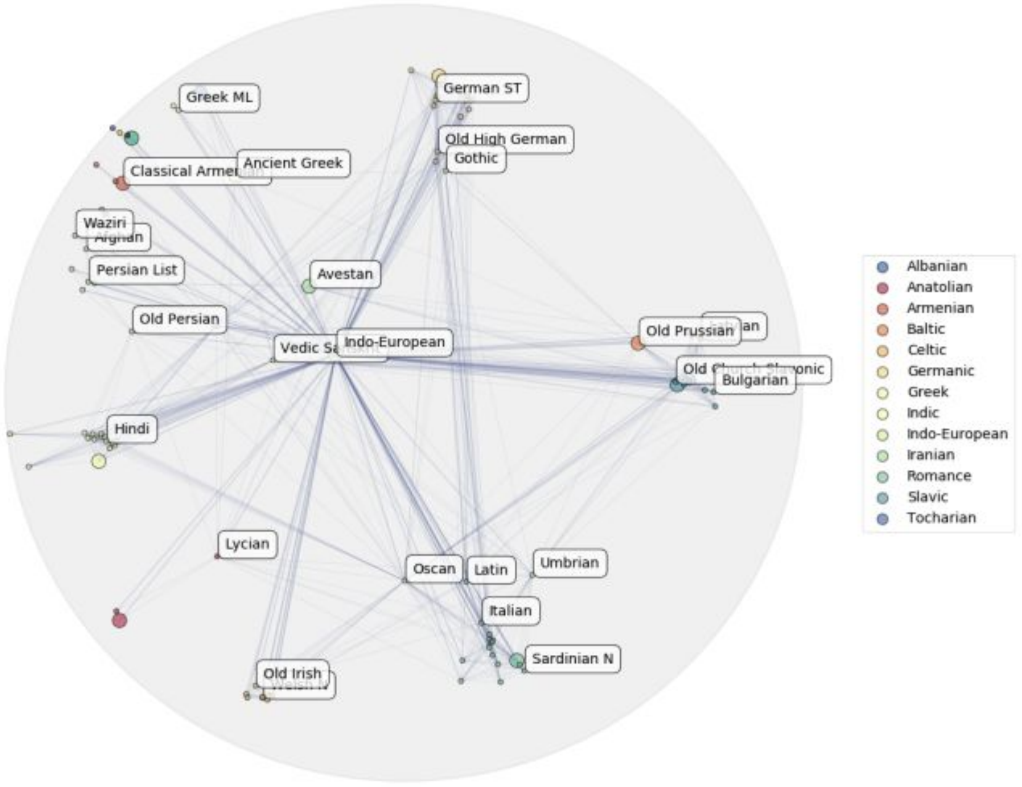
$$d(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{arcosh}\left(1 + 2\frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{(1 - \|\boldsymbol{u}\|^2)(1 - \|\boldsymbol{v}\|^2)}\right)$$

Poincaré Embeddings for Learning Hierarchical Representations (Nickel, Kiela 2017)

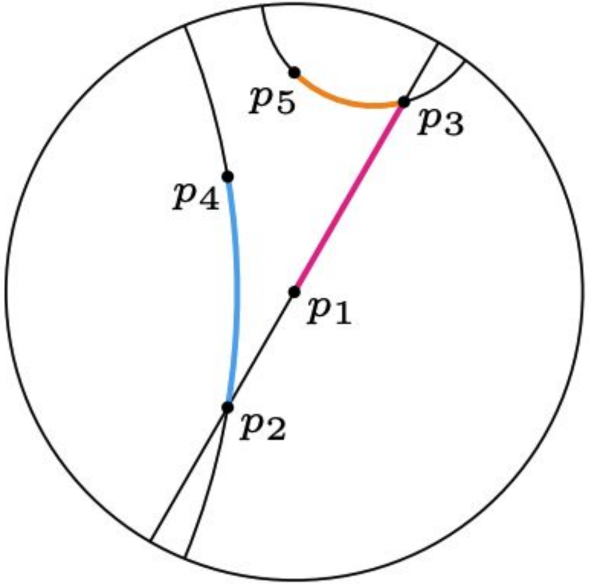
$d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v} + \boldsymbol{r}\|^2$ – translational distance

			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

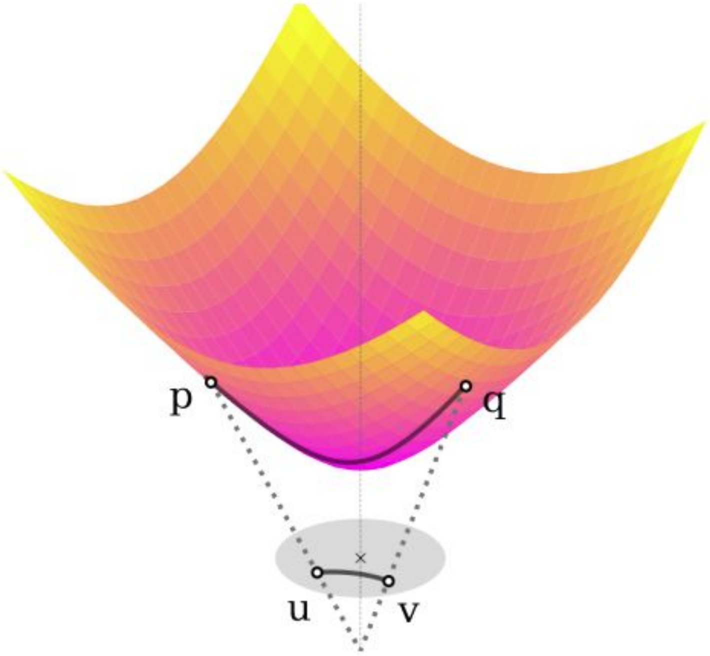
Poincaré Embeddings for Learning Hierarchical Representations (Nickel, Kiela 2017)



Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry (Nickel, Kiela 2018)



(a) Geodesics in the Poincaré disk.



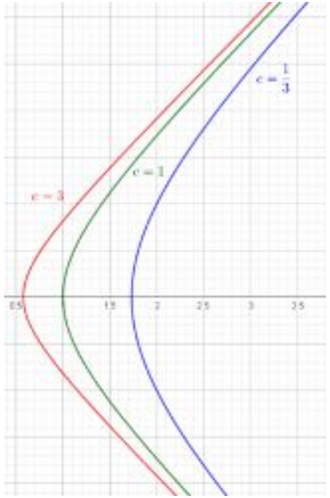
(b) Lorentz model of hyperbolic geometry.

Lorenz model of hyperbolic space

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}} = \langle \mathbf{x}_{space}, \mathbf{y}_{space} \rangle - x_{time}\, y_{time}$$

$$\mathcal{L}^n = \{ \mathbf{x} \in \mathbb{R}^{n+1} : \langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{L}} = -{}^1\!/c \;,\; c > 0 \}$$

$$x_{time} = \sqrt{{}^1\!/c + \|\mathbf{x}_{space}\|^2}$$



$$\mathcal{T}_{\mathbf{z}}\mathcal{L}^n = \{ \mathbf{v} \in \mathbb{R}^{n+1} : \langle \mathbf{z}, \mathbf{v} \rangle_{\mathcal{L}} = 0 \}$$

$$\|\mathbf{x}\|_{\mathcal{L}} = \sqrt{|\langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{L}}|}$$

$$\mathbf{x}_{space} = \frac{sinh(\sqrt{c}\, \|\mathbf{v}_{space}\|)}{\sqrt{c}\, \|\mathbf{v}_{space}\|} \mathbf{v}_{space}$$

Вопросы