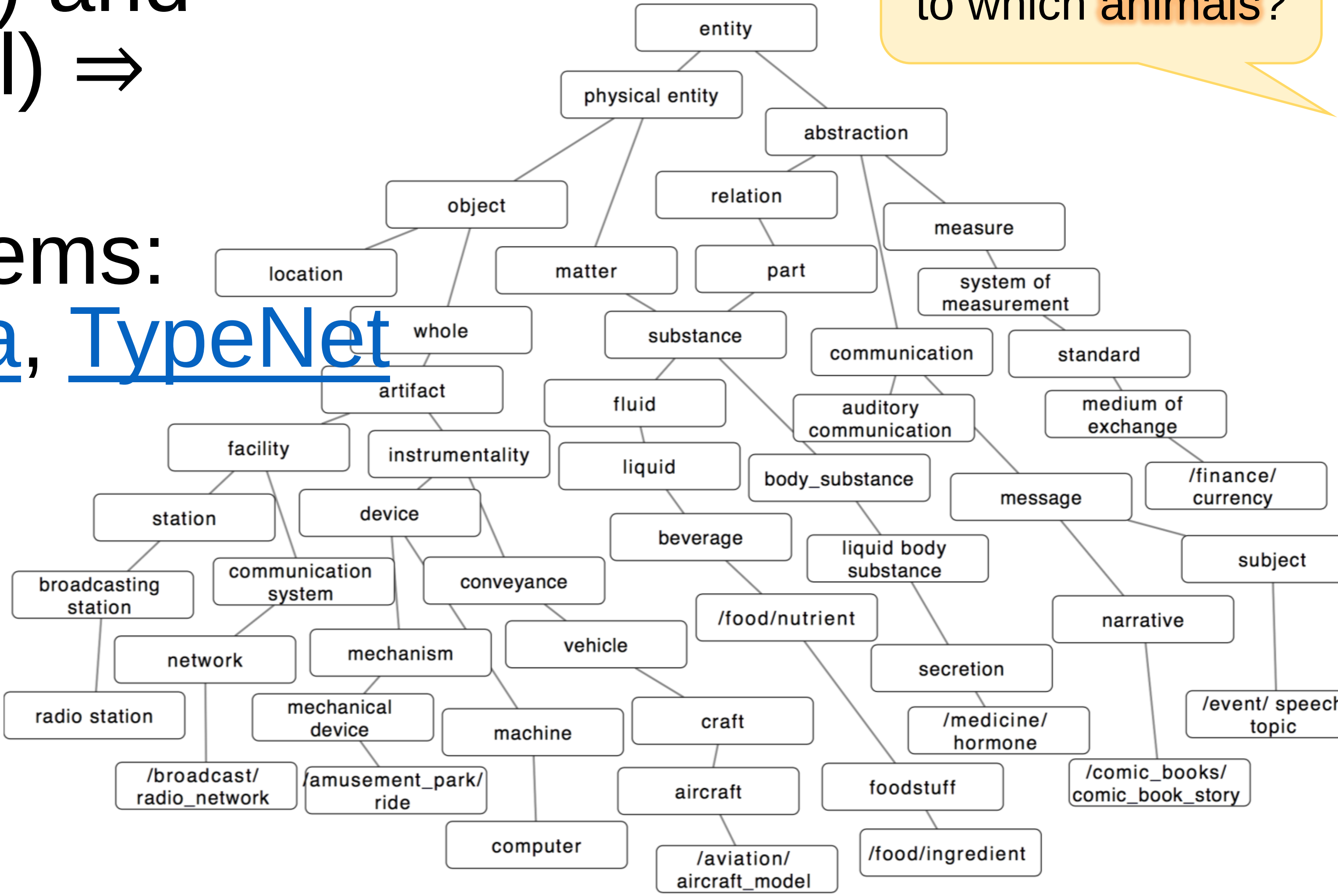
Knowledge and Retrieval

Modeling Hierarchies

Hierarchies in KGs

- Anti/symmetry … how about transitivity?
- (camel, is-a, mammal) and (mammal, is-a, animal) ⇒ (camel, is-a, animal)
- Popular KG type systems: [FIGER](), [GFT](), [DbPedia](), [TypeNet]()
- None of the KG embedding methods discussed thus far handle hierarchies

Deserts are home to which animals?



Two views of embedding hierarchies

- Embed a hierarchy (DAG) in a space with a distance such that distances on DAG are approximately preserved by distance in the embedding space
    - Euclidean → Poincaré balls, hyperbolic embeddings
- Encode DAG nodes so that ancestor-descendant queries can be answered efficiently
    - Gaussian, order and box embeddings
- Should work with incomplete supervision
- Should play well with other embeddings

Low-distortion graph embeddings

- Each node v in graph $G = (V, E)$ embedded to $x(v) \in \mathbb{R}^D$
- Graph distance between $u, v \in V$ is $d_G(u, v)$
- Distortion of embedding g is given by

$$\operatorname{distor}(x, G) = \frac{\max_{u,v} \frac{\|x(u) - x(v)\|}{d_G(u,v)}}{\min_{u,v} \frac{\|x(u) - x(v)\|}{d_G(u,v)}}$$

Maximum stretch

Minimum stretch

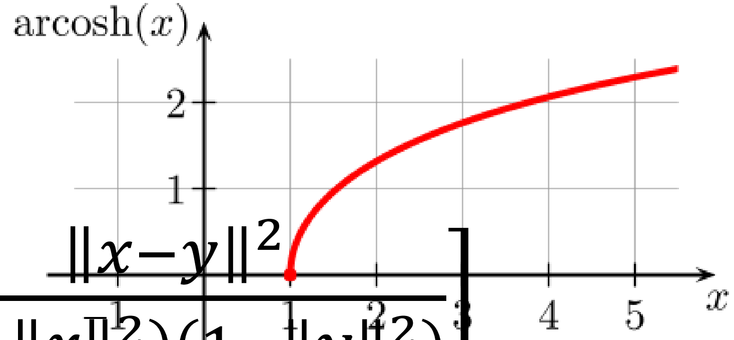
- Distortion of G is $\inf_x \operatorname{distor}(g, G)$

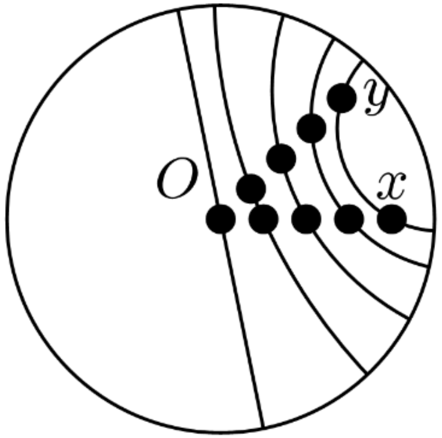
Euclidean distortion facts

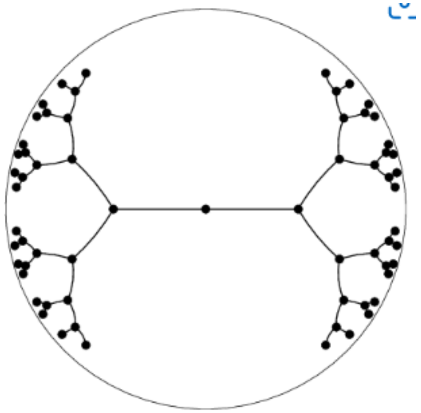
- With large enough D, any graph with n nodes can be embedded in $\mathbb{R}^D$ with $O(\log n)$ distortion
- Any connected planar graph can be embedded in $\mathbb{R}^2$ with $O(n)$ distortion; trees with $O(\sqrt{n})$
- With large enough D, any tree can be embedded in $\mathbb{R}^D$ with $O(\log \log n)$ distortion
- Binary trees can be embedded in a line with $O\left(\frac{n}{\log n}\right)$ distortion
- Binary trees can be embedded in $\mathbb{R}^D$ with $O\left(\frac{n^{1/D}}{\log n}\right)$ distortion
- Distortion of $(1+\epsilon)$ is possible in hyperbolic space

Poincaré disk, hyperbolic space

- $\operatorname{acosh}(a) = \ln\left(a + \sqrt{a^2 + 1}\right)$
- Points x, y strictly inside unit circle
- Hyperbolic distance $d_H(x, y) = \operatorname{acosh}\left[1 + 2\frac{\|x-y\|^2}{(1-\|x\|^2)(1-\|y\|^2)}\right]$
- As x, y approach perimeter, $\boxed{d_H \to \infty}$
- Natural tree embedding
- Precision bits not free!

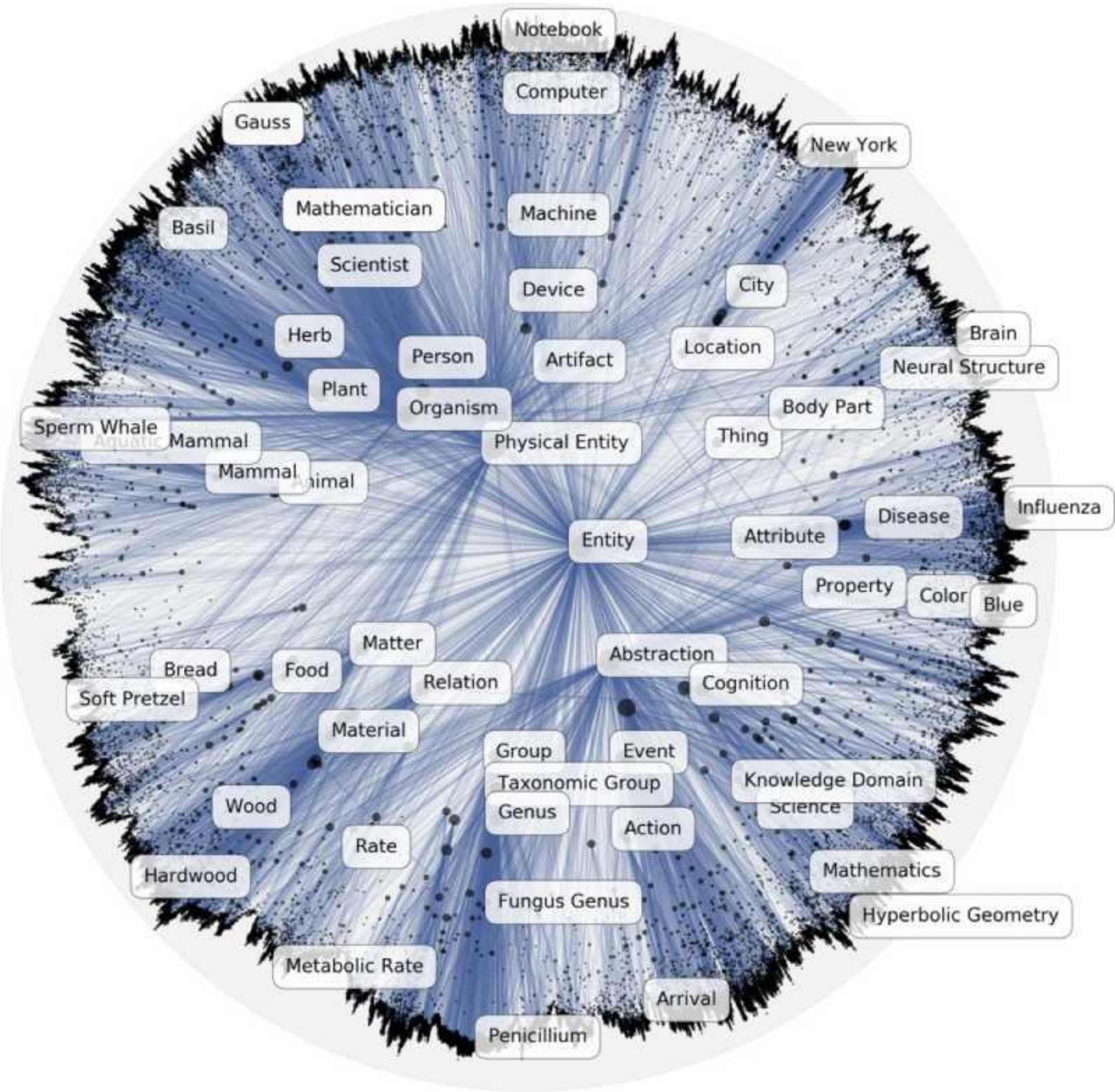






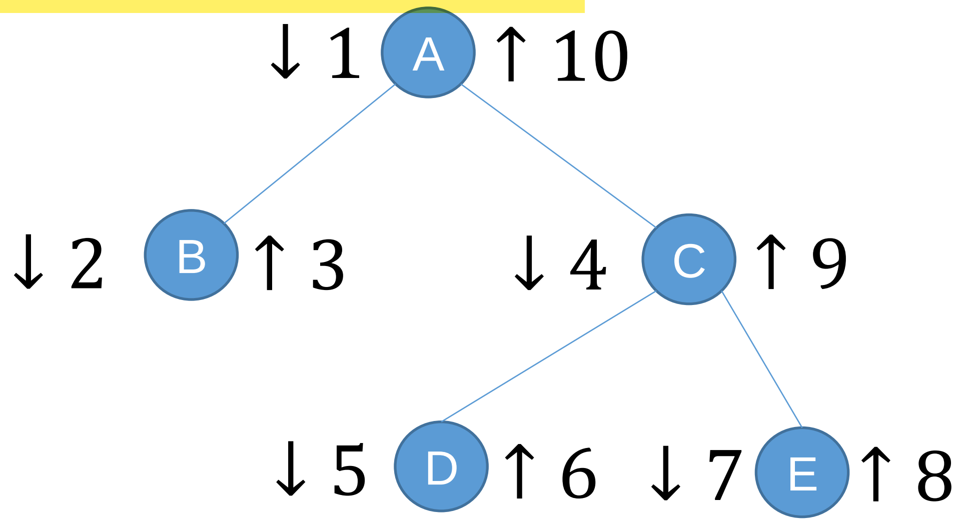
WordNet nouns in Poincaré disk

- As expected, generic synsets near disk center, specific near periphery
- What do applications (e.g., QA) need?
  - “scientists who played musical instruments”
  - “mammals living in the desert”
- word2vec ↔ Poincaré?



Toward order embeddings

- Computing Euclidean or hyperbolic distance between items not the only choice
- Denote partial order by "$x \prec y$" meaning x is a descendant of y
- In case of a tree, associate with each item x the in-order traversal interval $I(x)$
  - $I(B) = [2,3], I(A) = [1,10]$, $I(D) = [5,6]$
  - $I(x) \subset I(y) \Leftrightarrow x \prec y$

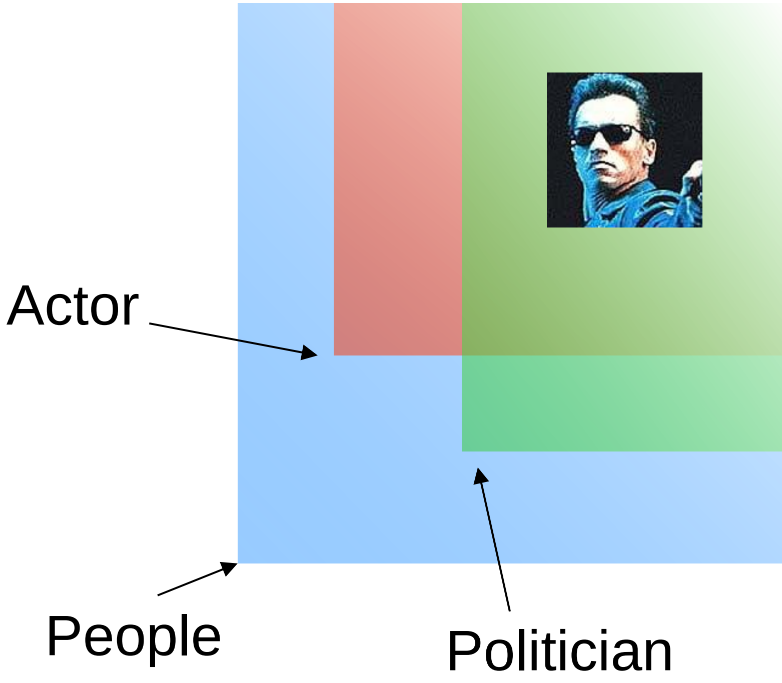


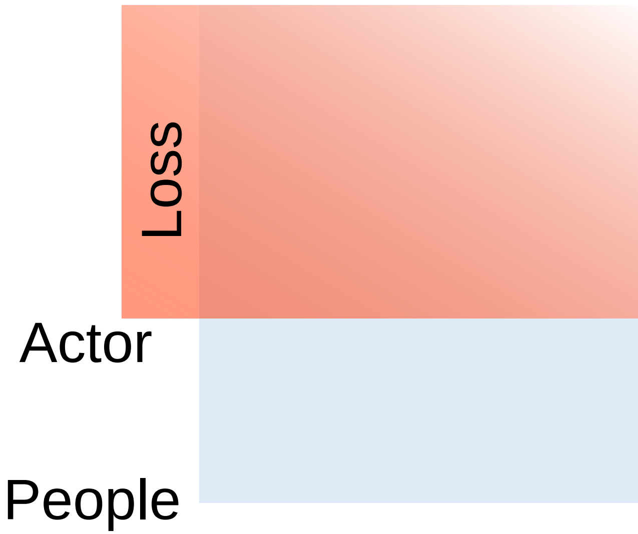
Nodes of a tree can be embedded in one dimension to answer ancestor-descendant queries in constant time.

Apex of axis-aligned open cones

- Item x represented by $\boldsymbol{u}_x \in \mathbb{R}^D$
- $\boldsymbol{u}_x$ is the apex of an open cone
- $x \prec y \Leftrightarrow \boldsymbol{u}_x \geq \boldsymbol{u}_y$, elementwise
- Design training loss function
  - Notation: $\mathrm{ReLU}(a) = [\mathrm{a}]_+ = \max\{0, a\}$
  - If $x \prec y$, $\left\|\mathrm{ReLU}\left(\boldsymbol{u}_y - \boldsymbol{u}_x\right)\right\|$, i.e. all D dims must satisfy constraint
  - If $x \nprec y$, $\mathrm{ReLU}\left[\alpha - \left\|\mathrm{ReLU}\left(\boldsymbol{u}_y - \boldsymbol{u}_x\right)\right\|\right]$
- Does not recognize asymmetry in losses

Margin





Addressing loss asymmetry

- If $x \prec y$, then for all dim d, want $u_x[d] \geq u_y[d]$
  - E.g., $\ell_+(x,y) = \max_{d \in [D]} \mathrm{ReLU}\big(u_y[d] - u_x[d]\big)$
- If $x \nprec y$, then for some d, want $u_x[d] < u_y[d]$
  - $\ell_-(x,y) = \min_{d \in [D]} \mathrm{ReLU}\Big(\alpha - \mathrm{ReLU}\big(u_y[d] - u_x[d]\big)\Big)$
- All open cones and their intersections have same measure of volume (unlike in-order intervals)
  - … even though Politicians $\subsetneq$ People
- Hard to model negative correlation
  - X is-a fruit, or X is-not-a scientist

(Hyper)rectangle/box embeddings

- Each type/item x characterized by interval $I_x[d] = \left[b_{x,d}, h_{x,d}\right]$ for each dimension d
- Want $I_x[d] \subseteq I_y[d]\ \forall d$, iff $x \prec y$
- Learning to lay out Venn diagrams