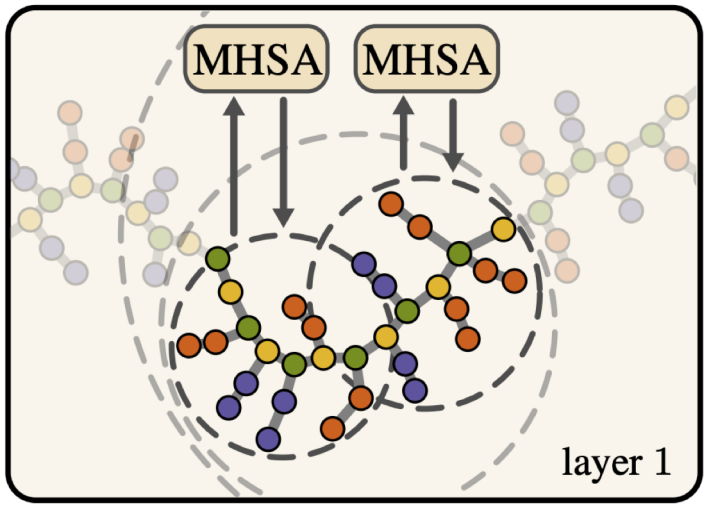
University of Amsterdam
Faculty of Science

Designing Hierarchies for Optimal Hyperbolic Embedding

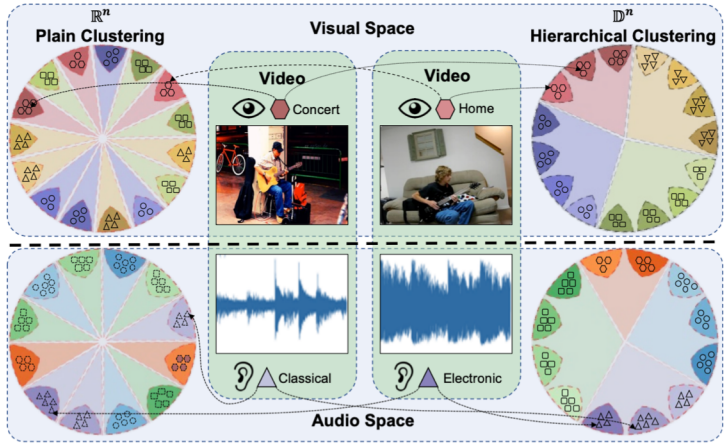
Melika Ayoughi, Max van Spengler, Pascal Mettes, and Paul Groth
m.ayoughi@uva.nl

June 3rd 2025

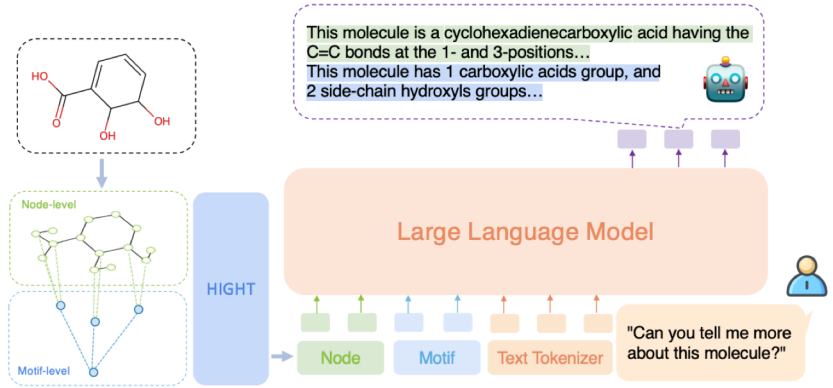
Embedded Hierarchies



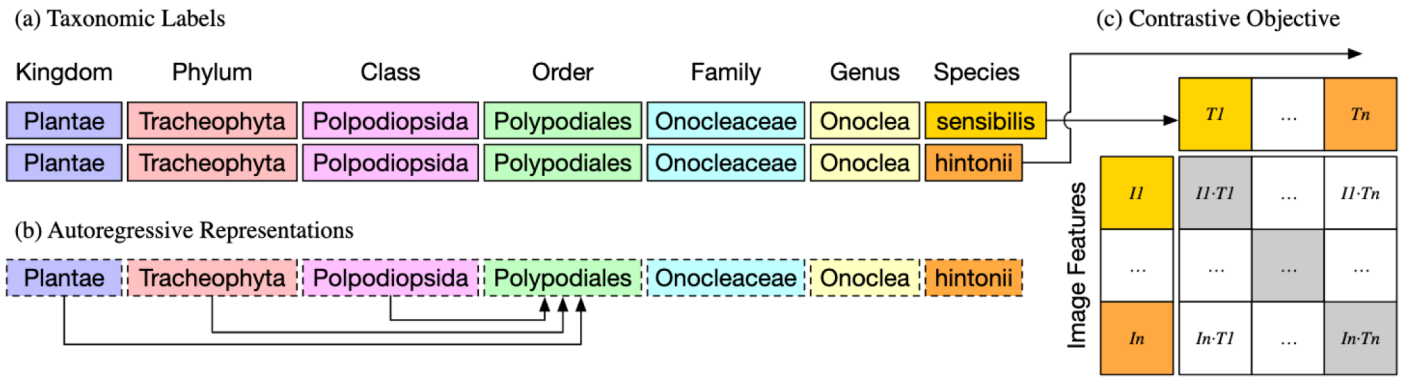
[Zhdanov et al. ICML 2025]



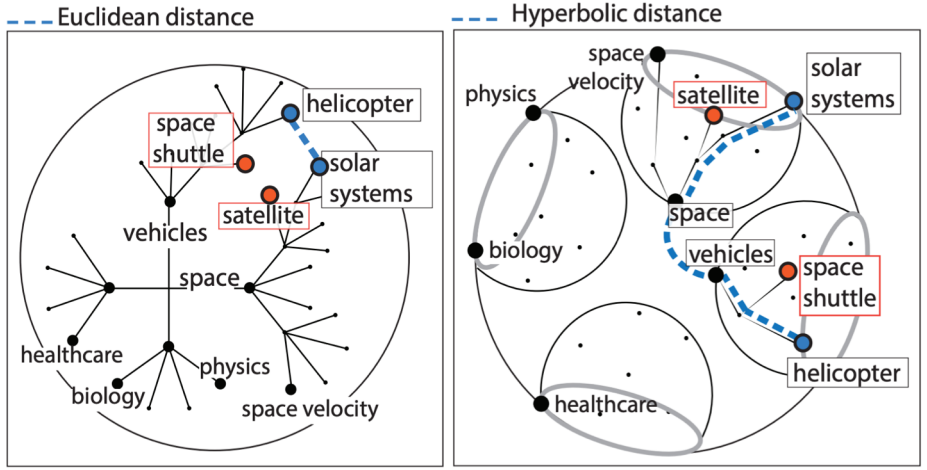
[Teng et al. ICCV 2023]



[Chen et al. ICML 2025]



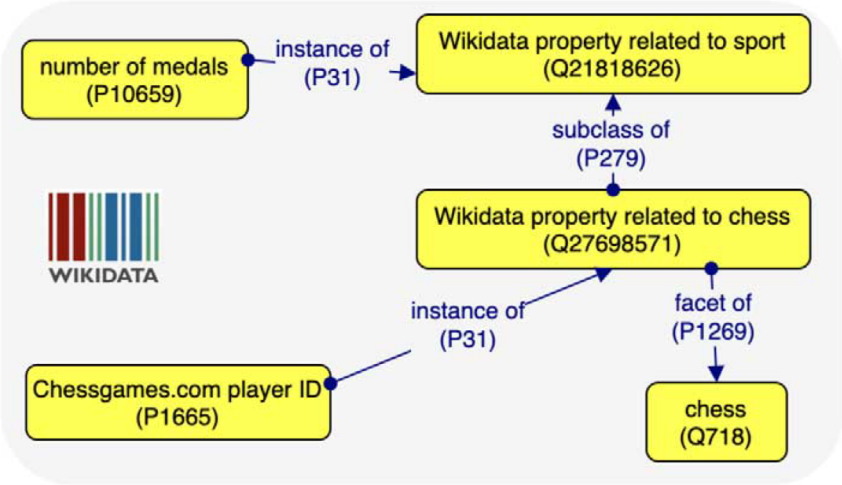
[Stevens et al. CVPR 2024]



[Shahid et al. ACL 2023]

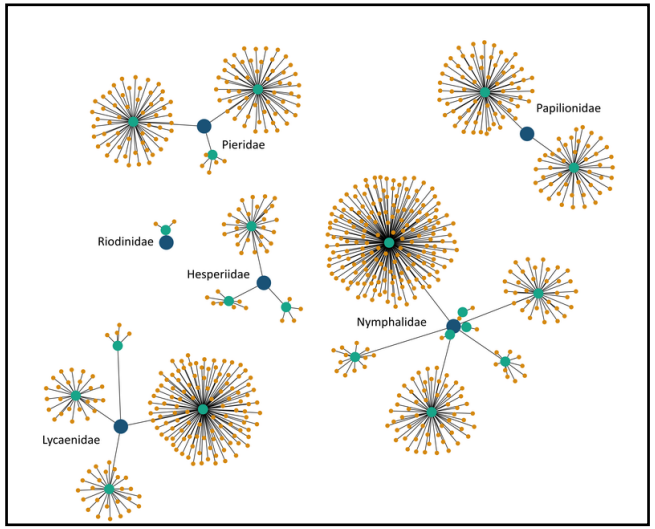
Knowledge Graphs and Ontologies

[schema.org]



[Carriero et al. Semantic Web 2024]

WordNet
A Lexical Database for English



[The ETHEC entomological hierarchy]

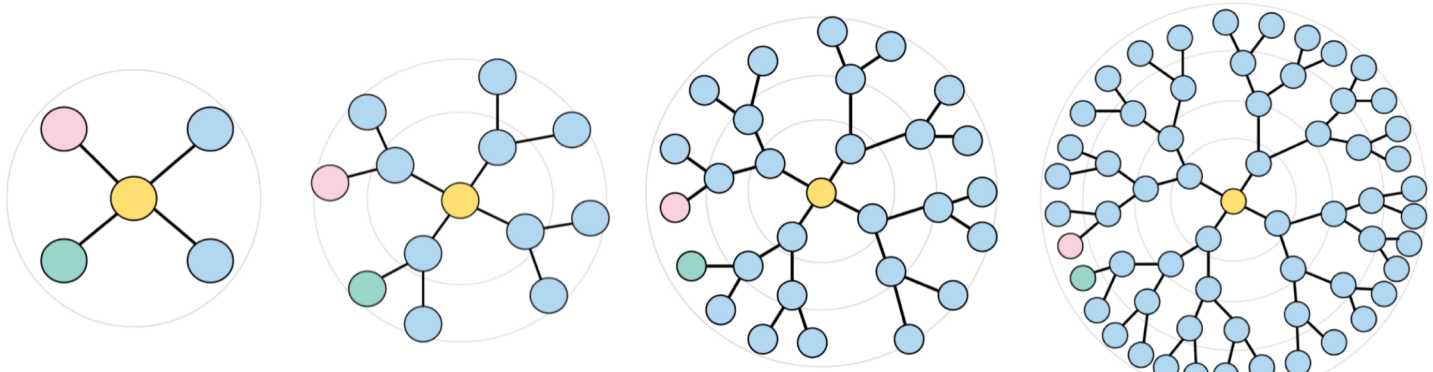
Most methods use the hierarchies **as is**, however, ontology engineers have **control** over the hierarchy.

How can hierarchies be designed to suite ML tasks?

[Tree of life by Gustav Klimt]

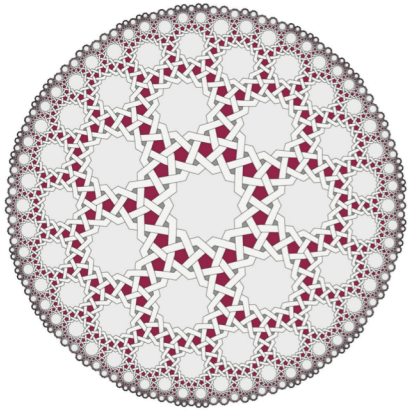
Optimal geometry for hierarchical representation

Euclidean



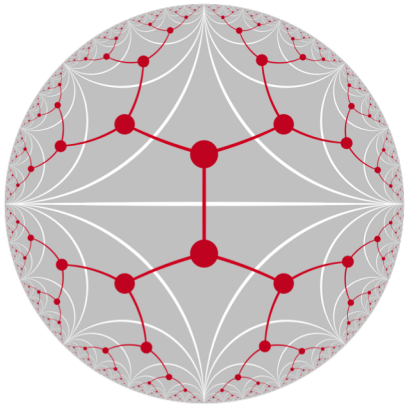
[Bachmann et al. ICML 2020]

Hyperbolic



[KAPLAN et al. TOG 2004]

[M.C.Escher's "Circle Limit III"]



"Hyperbolic space can be thought of as a **continuous** analogue to **discrete trees**"

Controlled Experiment Setup

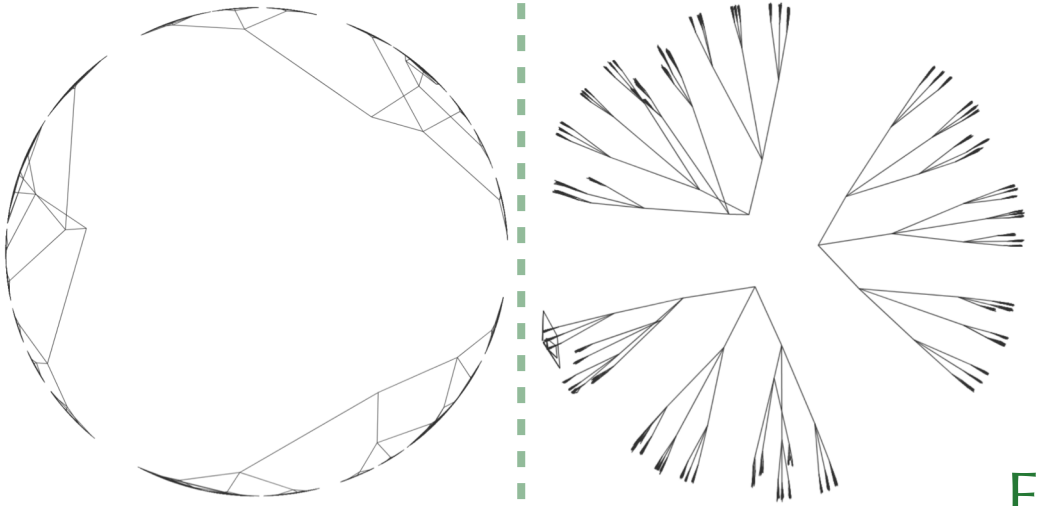
Hyperbolic Embedding Methods, Hierarchies, Evaluation Metrics

Hyperbolic Embedding Methods

Gradient-based

Poincaré

Entailment

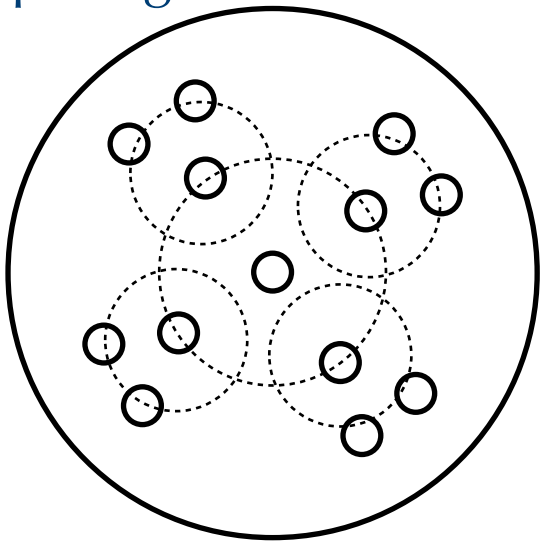


[Ganea et al. ICML 2018]

Construction-based

Precomputing

Hadamard



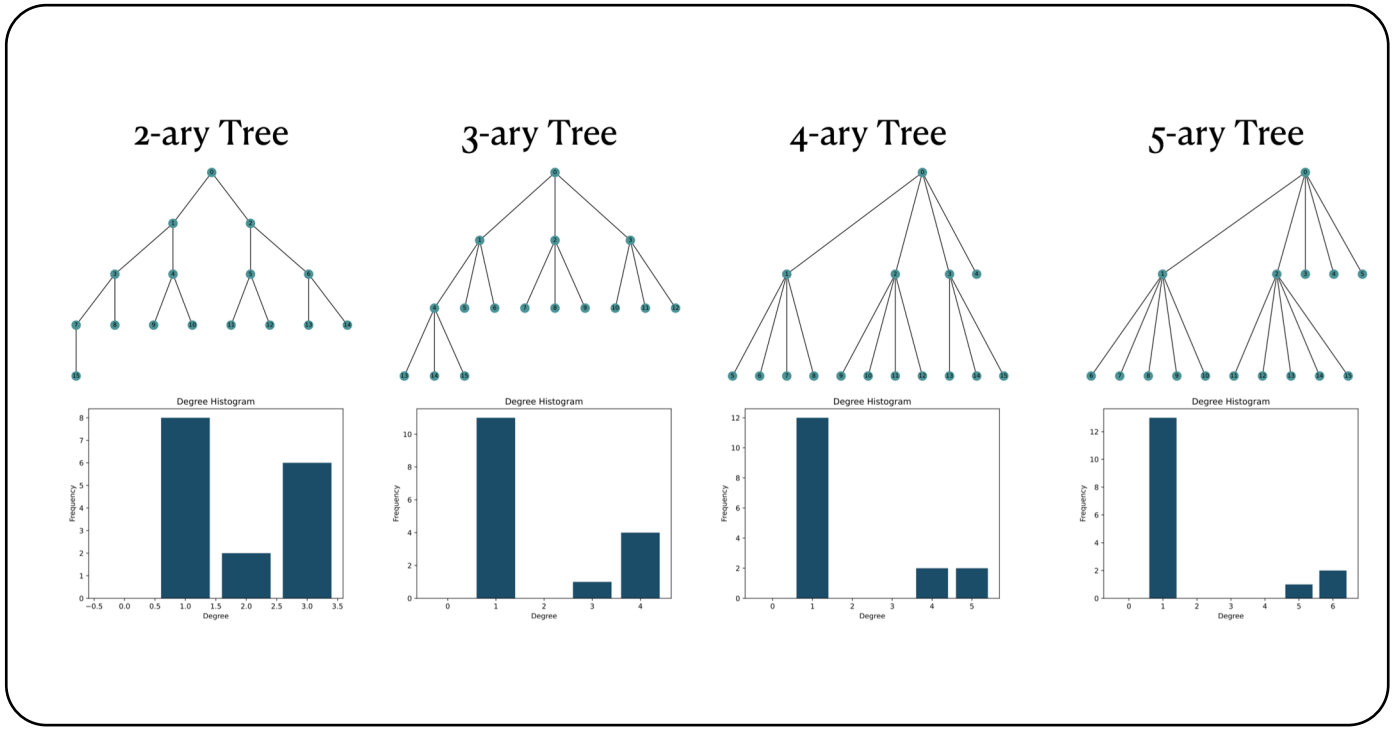
Enforce structure

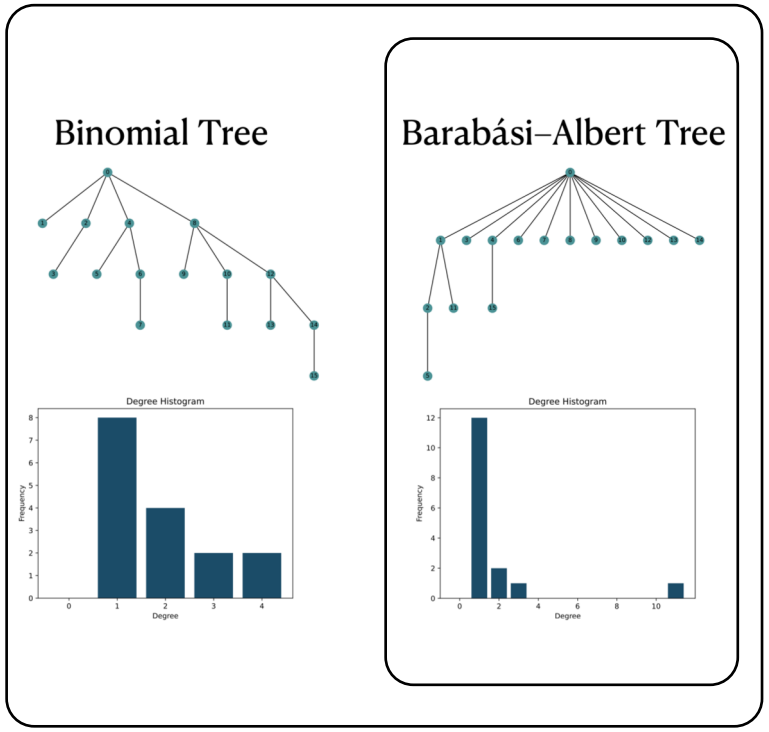
- ⊕ General-purpose
- ⊖ Quality not guarantied

- ⊕ High-quality embeddings
- ⊕ Fast
- ⊕ Preserve original hierarchy structure
- ⊖ Arbitrary-precision arithmetic

Diverse Hierarchies

Long-tailed

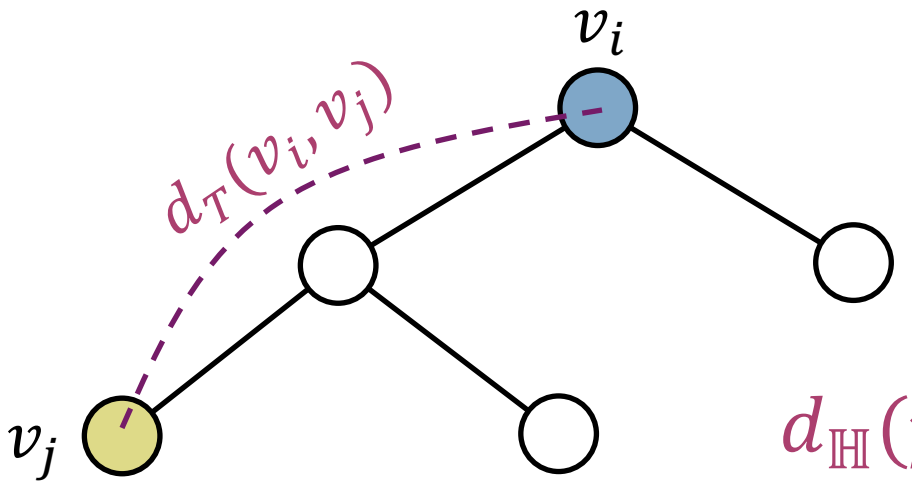




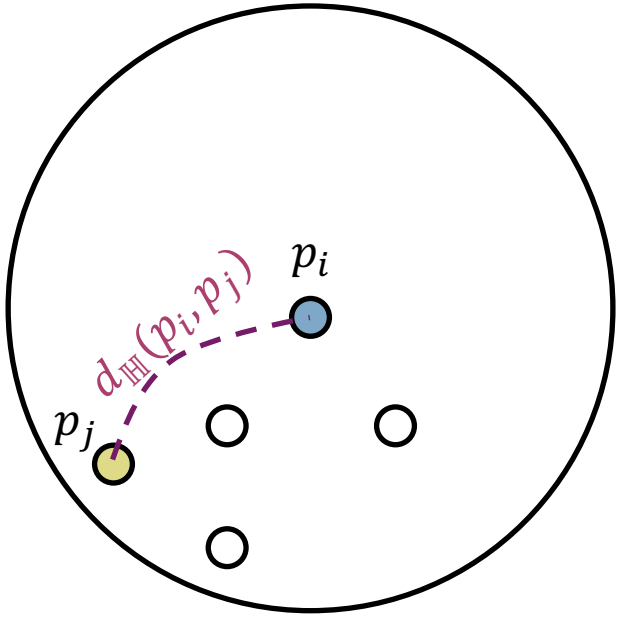
Balanced

Imbalanced

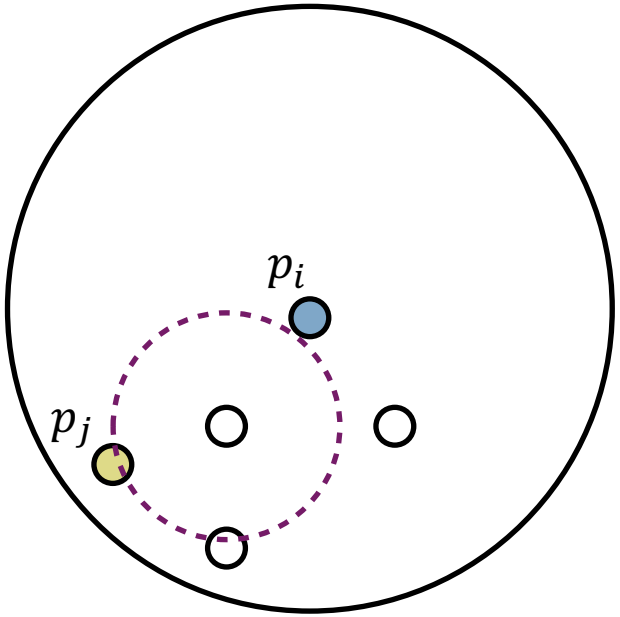
Evaluation Metrics



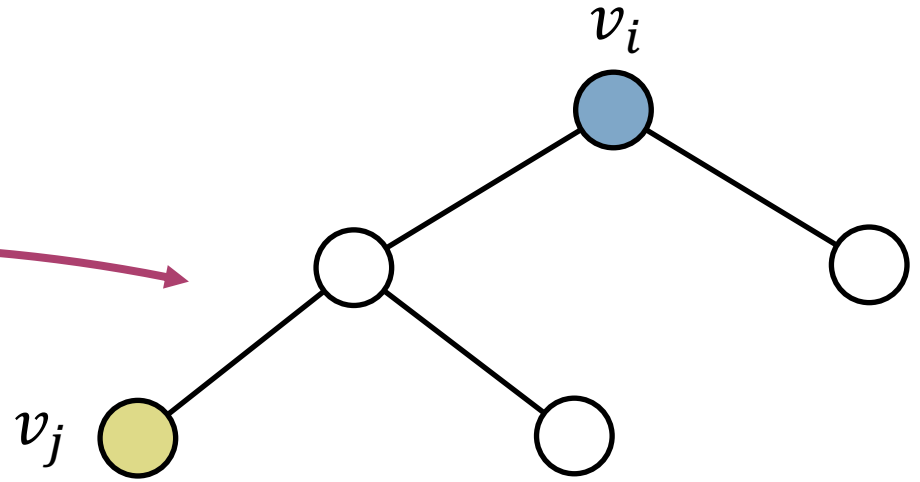
$$d_{\mathbb{H}}(p_i, p_j) \propto d_T(v_i, v_j)$$



$\downarrow D_{avg} \downarrow D_{wc}$
Distortion



Reconstruct

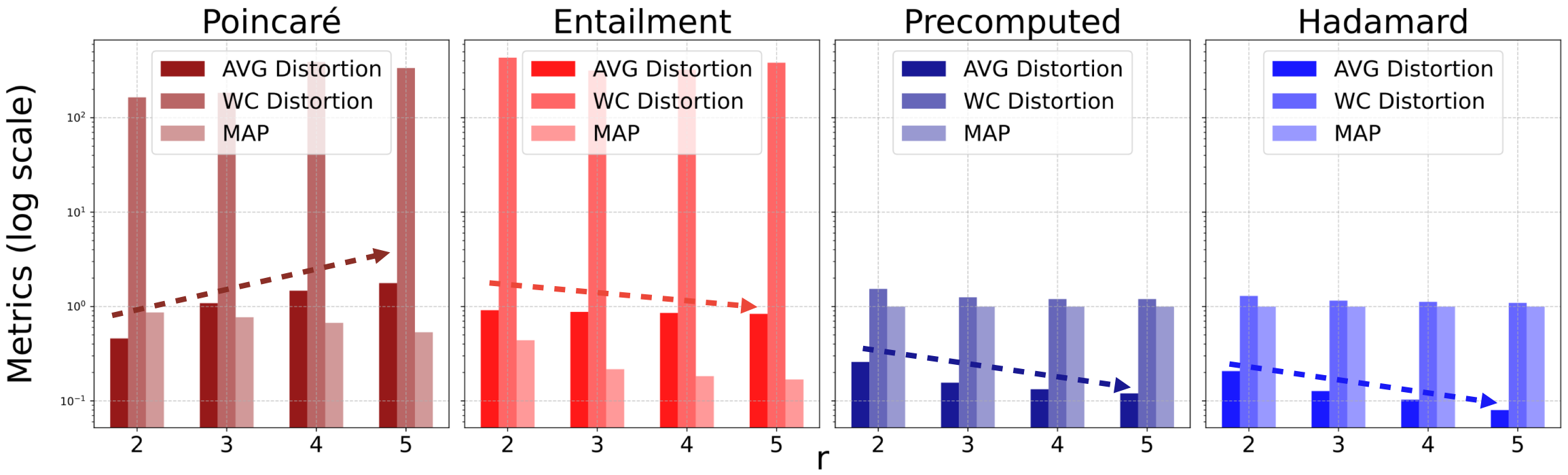


$\uparrow MAP$
Reconstruction

Experiments

Depth or Width?

Cheatsheet:
$\downarrow D_{avg} \downarrow D_{wc} \uparrow MAP$



I. All methods except for Poincaré: **wide and shallow** hierarchies have lower distortion.
II. **Construction-based** methods paired with **wide** hierarchies achieve **optimal** embeddings.

What is the impact of more nodes on embedding quality?

Cheatsheet:
$\downarrow D_{avg} \downarrow D_{wc} \uparrow MAP$

	Gradient-based						Construction-based					
	Poincaré			Entailment			Precomputed			Hadamard		
	256	512	1024	256	512	1024	256	512	1024	256	512	1024
Balanced												
2-ary	0.880	0.459	0.229	0.816	0.914	0.960	0.220	0.259	0.300	0.176	0.207	0.240
3-ary	1.439	1.085	0.752	0.742	0.878	0.940	0.124	0.156	0.160	0.102	0.127	0.130
4-ary	2.129	1.471	1.092	0.695	0.855	0.928	0.102	0.133	0.137	0.079	0.103	0.105
5-ary	2.472	1.770	1.385	0.657	0.837	0.919	0.115	0.120	0.156	0.078	0.080	0.103
Imbalanced												
Binomial	1.736	1.439	0.988	0.717	0.863	0.932	0.207	0.249	0.298	0.161	0.186	0.211
BA	3.444	2.791	2.206	0.595	0.802	0.903	0.108	0.140	0.178	-	-	-

A strong **increase in semantic complexity** has **minimal impact** on embedding quality.

Balanced or Imbalanced?

Cheatsheet:
$\downarrow D_{avg} \downarrow D_{wc} \uparrow MAP$

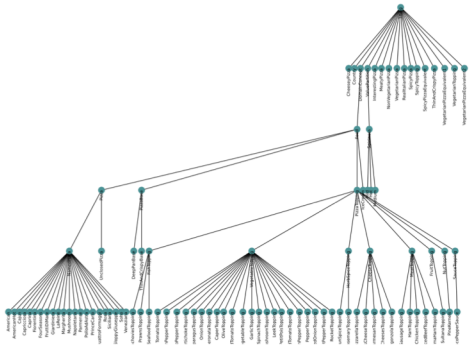
best and worst

	Gradient-based						Construction-based					
	Poincaré			Entailment			Precomputed			Hadamard		
	D_{avg}	D_{wc}	MAP	D_{avg}	D_{wc}	MAP	D_{avg}	D_{wc}	MAP	D_{avg}	D_{wc}	MAP
Balanced												
2-ary	0.459	164.777	0.866	0.914	434.177	0.439	0.259	1.539	1	0.207	1.297	1
3-ary	1.085	183.974	0.770	0.878	316.338	0.217	0.156	1.252	1	0.127	1.155	1
4-ary	1.471	390.397	0.671	0.855	323.967	0.183	0.133	1.201	1	0.103	1.121	1
5-ary	1.770	336.711	0.534	0.837	383.626	0.169	0.120	1.201	1	0.080	1.092	1
Imbalanced												
Binomial	1.439	69.530	0.171	0.863	224.731	0.304	0.249	1.542	1	0.186	1.257	1
BA	2.791	3607.95	0.020	0.802	731.914	0.231	0.140	1.329	1	-	-	-

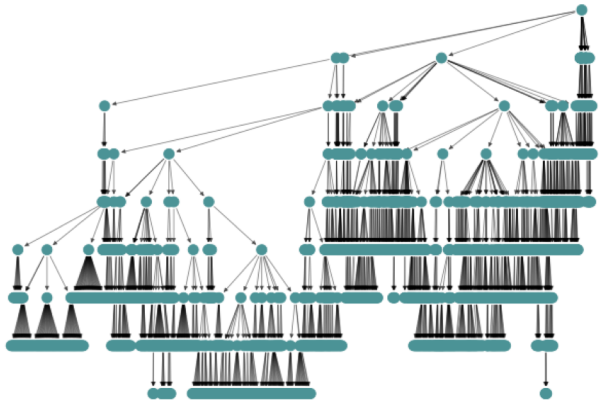
Better to have a wide imbalanced tree than a deep balanced one!

Case study: The Pizza and ImageNet ontologies

Pizza reorganized



ImageNet reorganized



Cheatsheet:
$\downarrow D_{avg} \downarrow D_{wc} \uparrow MAP$

	Gradient-based						Construction-based					
	Poincaré			Entailment			Precomputed			Hadamard		
	D_{avg}	D_{wc}	MAP	D_{avg}	D_{wc}	MAP	D_{avg}	D_{wc}	MAP	D_{avg}	D_{wc}	MAP
Pizza												
Original	3.321	7066.671	0.059	-	-	-	-	-	-	-	-	-
+ single inheritance	3.387	10509.346	0.051	0.499	511.594	0.195	0.234	1.538	1	0.126	1.180	1
+ reorganized	3.422	9343.566	0.045	0.452	1454.972	0.164	0.167	1.329	1	**0.089**	**1.118**	**1**
ImageNet												
Original	0.809	3983.563	0.087	-	-	-	-	-	-	-	-	-
+ single inheritance	0.722	2745.952	0.220	0.961	2364.827	0.293	0.725	885.622	0.725	0.297	1.647	1
+ reorganized	1.008	12715.625	0.156	0.955	4096.000	0.164	0.507	2.698	1	**0.171**	**1.232**	**1**

I. Poincaré: only method to handle multiple inheritance
II. Hierarchy reorganization leads to better distortion and MAP

Recommendations

- ✓ Design hierarchies for width
- ✓ Do not worry about balance
- ✓ Hyperbolic embeddings can handle additional node complexity
- ✓ Avoid multiple inheritance; o/w limited to Poincaré method

Wrapping up my PhD - Looking for What's Next!

Postdoc or research opportunities in industry

Learn more about my work and background

m.ayoughi@uva.nl