

# HYPERBOLIC METABOLITE-DISEASE ASSOCIATION PREDICTION

Domonkos Pogány<sup>1</sup>, Péter Antal<sup>1</sup>

European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning

9-11 October 2024  
Bruges, Belgium

## MOTIVATION

- In biomarker research, there is a growing demand for computational methods to identify new metabolite-disease associations (MDAs).
- However, current solutions overlook the geometry of the MDA space.

## OBJECTIVES

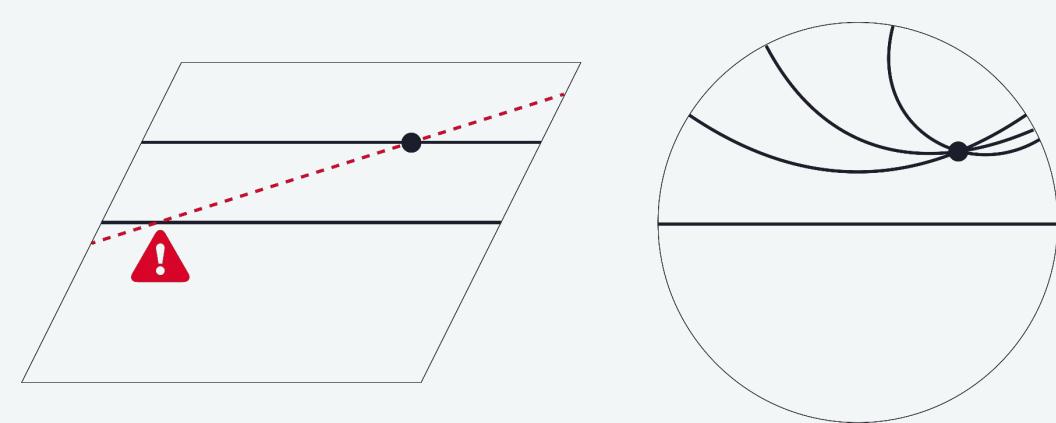
- Investigate the underlying geometry of the MDA network.
- Compare Euclidean ( $\mathcal{E}$ ) and hyperbolic ( $\mathcal{H}$ ) representation learning methods.
- Explain the differences between these approaches.

## CONCLUSIONS

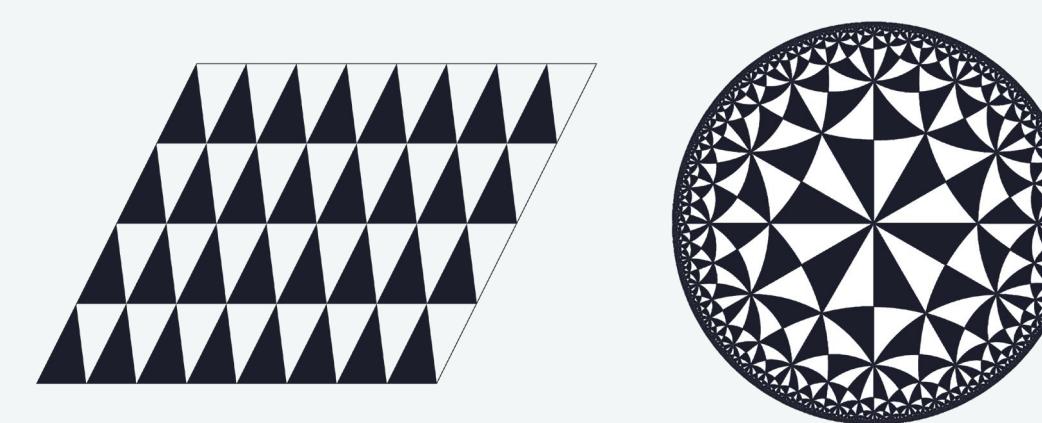
- I. The MDA network is scale-free, indicating hyperbolic geometry.
- II. Non-Euclidean classifiers require lower dimensionality.
- III. Hyperbolic embedding norm captures degree centrality.

## HYPERBOLIC GEOMETRY

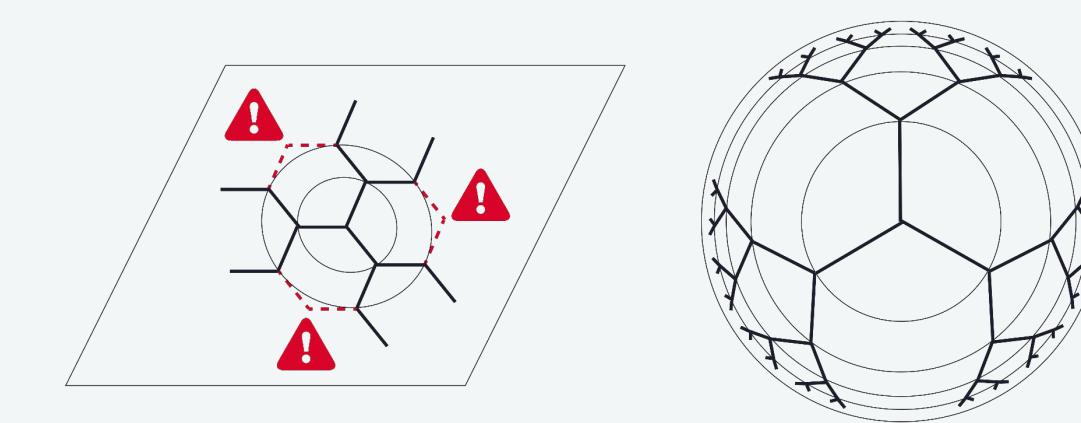
- In  $\mathcal{E}$ , there is exactly one; **in  $\mathcal{H}$ , there are at least two parallels** through any external point to a given line.



- In  $\mathcal{H}$ , the area grows exponentially;** in  $\mathcal{E}$ , it grows only polynomially with respect to the radius.

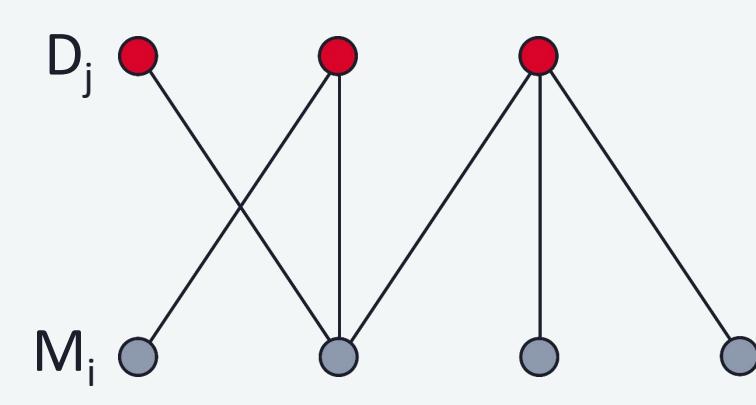


- $\mathcal{H}$  is, therefore, suitable for embedding trees / scale-free networks** with many leaves / small degree nodes.

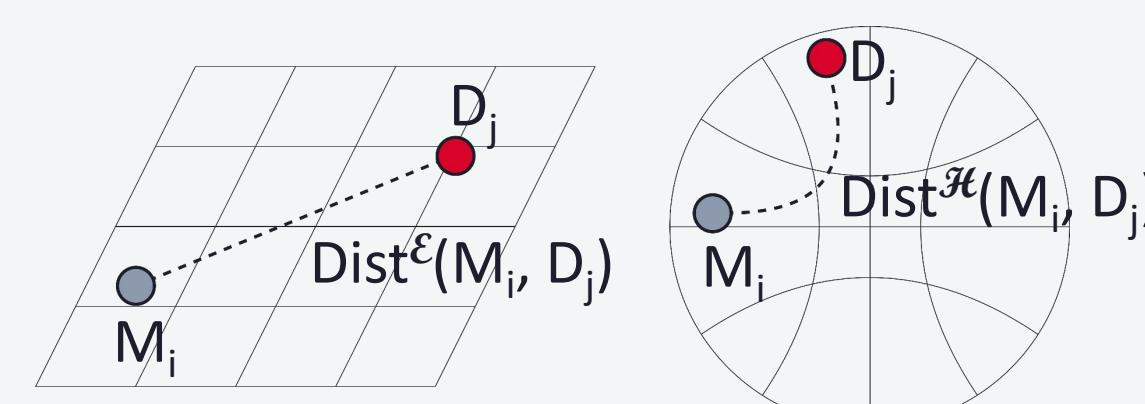


## METHODS

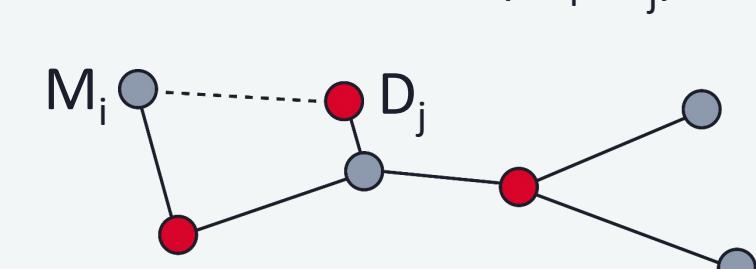
- Data: **MDA network** with 2.583 metabolites, 656 diseases, and 7.650 known associations.



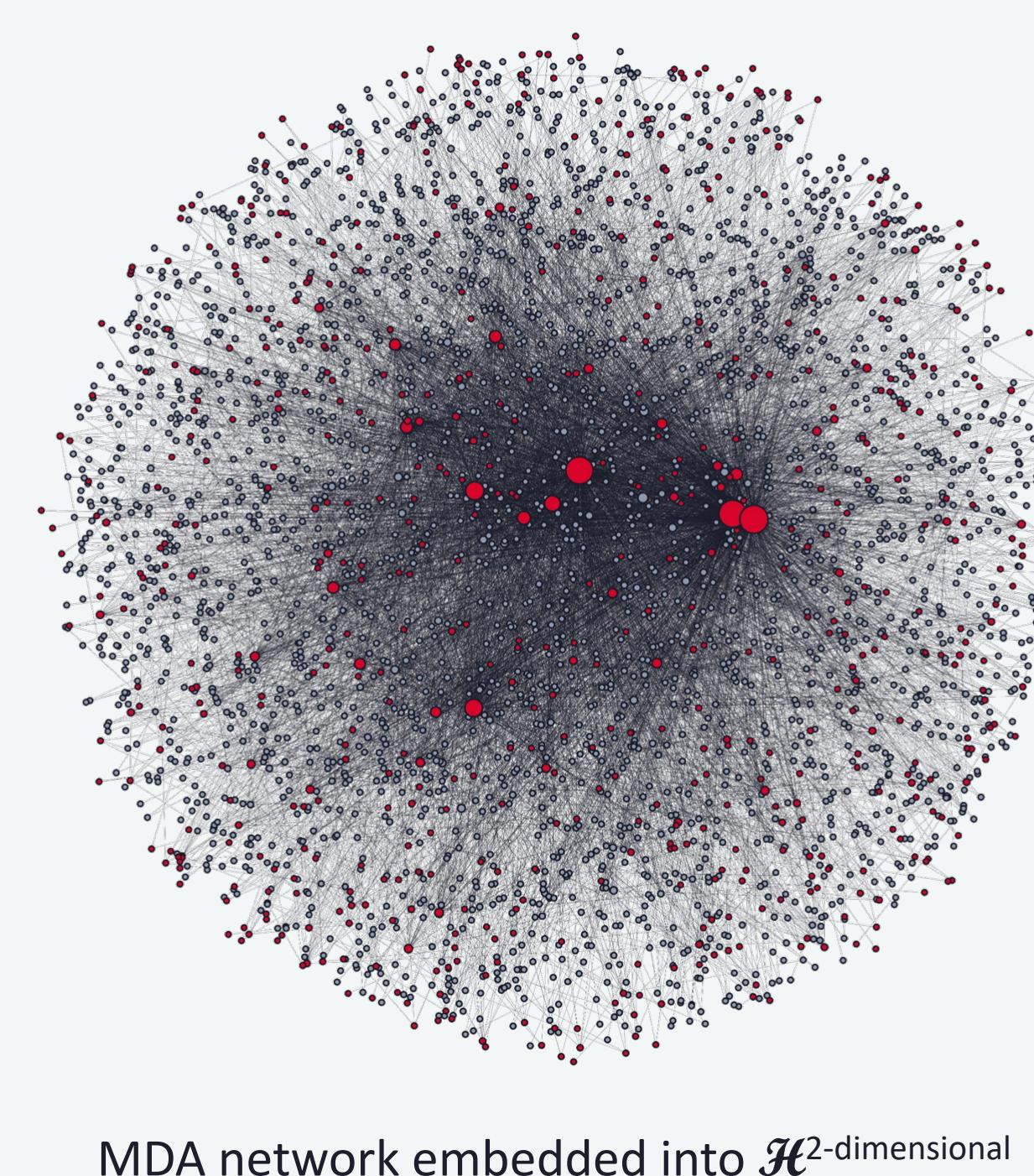
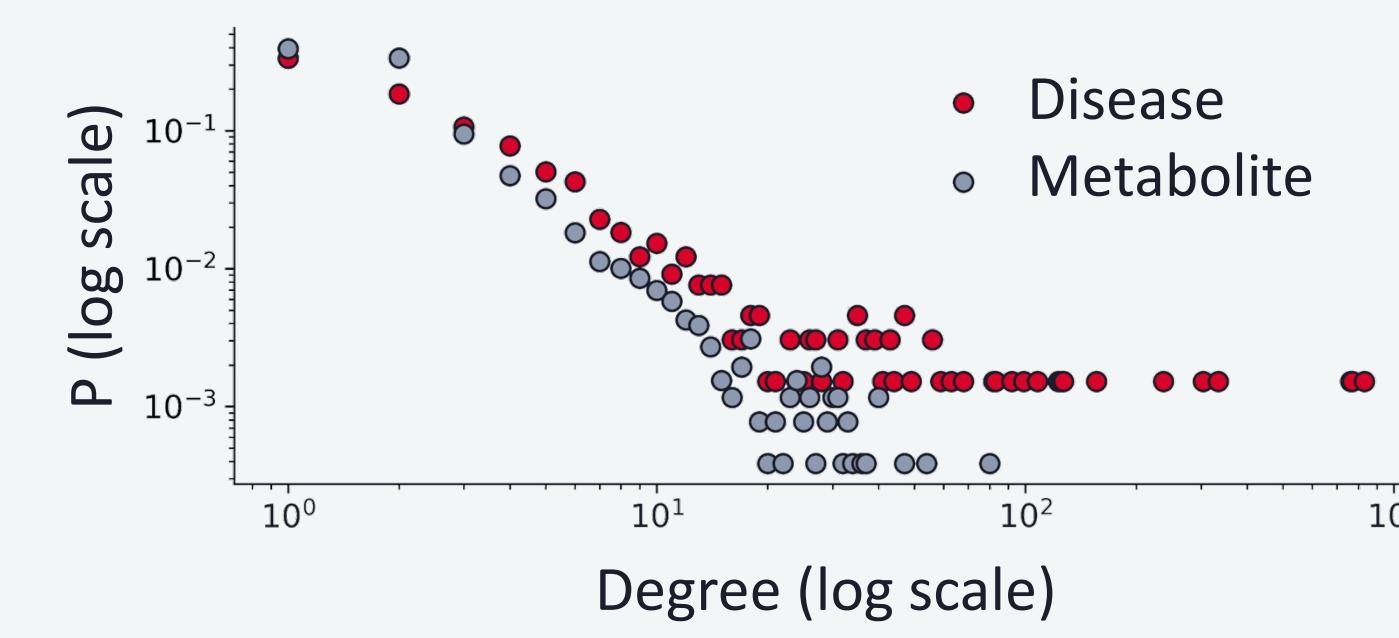
- Model: **Shallow node embeddings in  $\mathcal{E}$  and  $\mathcal{H}$** , placing associating nodes in close proximity.



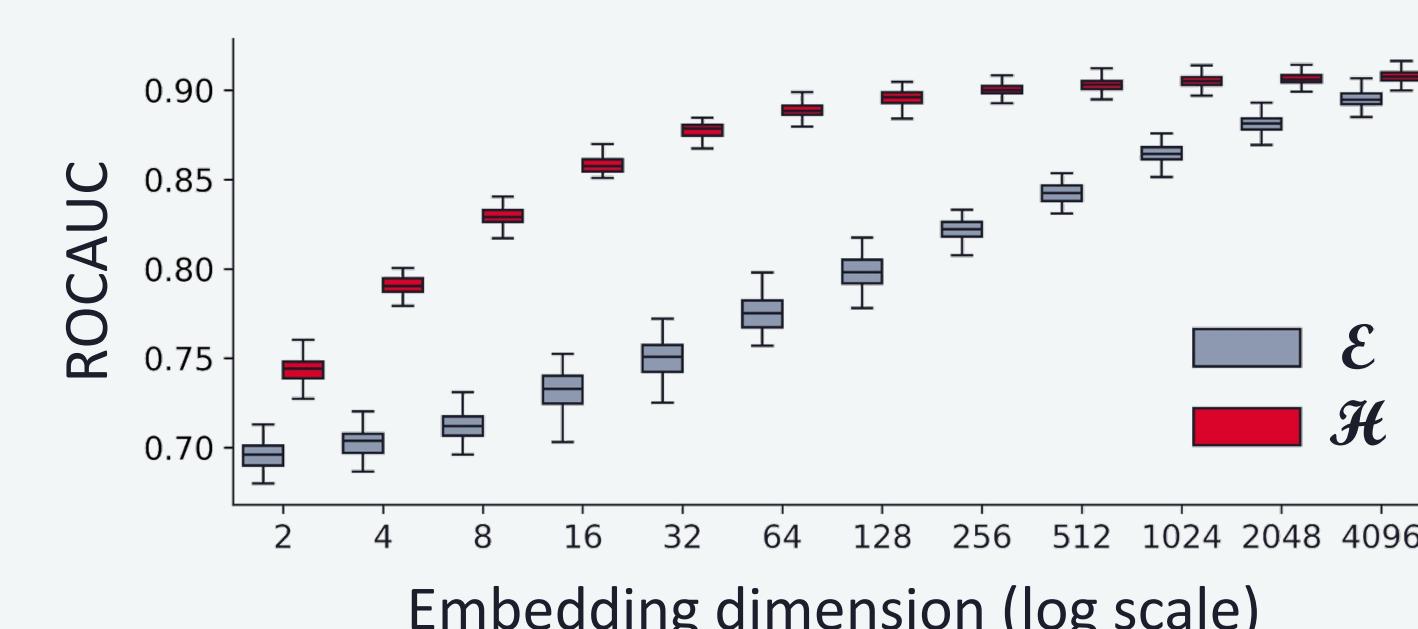
- Task: Binary classification formulation of **link prediction** based on manifold distances:  $P(\text{link}_{i,j}) = \frac{1}{1 + \text{Dist}(M_i, D_j)}$



- The MDA degree distribution follows a power law**, some hub nodes have a lot, while most have few connections.



- Embeddings in  $\mathcal{H}$  require orders of magnitude lower dimensionality** than in  $\mathcal{E}$ , as shown by the cross-validation results on classification performance.



- Embedding norm in  $\mathcal{H}$  encode degree centrality** by placing the few large-degree hub nodes near the origin and the low-degree nodes towards the exponentially growing periphery.

