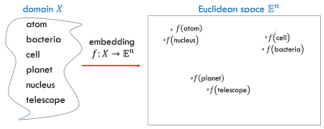
Quantification of Similarities: Embeddings into Euclidean Spaces and Taxonomies

Evgeny Dantsin and Alexander Wolpert
Roosevelt University, Chicago, IL, USA

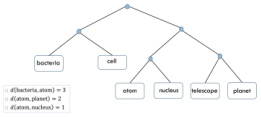
There is a domain X and an embedding

$$f: X \to \mathbb{E}^n$$

of this domain into a Euclidean space. The embedding f induces a metric on the domain.



Suppose new knowledge about the domain has emerged. Also suppose this knowledge is represented as a taxonomy on a subset $S \subseteq X$ of the domain.

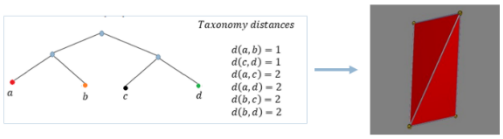


The metric induced by the taxonomy may be inconsistent with the "embedding metric".

Question. Can we adjust the embedding f so that the adjusted embedding is consistent with taxonomy?

Main result. We give a polynomial-time algorithm that takes the taxonomy as input and returns the required adjustment to the embedding f.

The algorithm is geometric in nature: it uses an isometric embedding of a finite ultrametric space into a Euclidean space.



The main idea of the algorithm