On embeddings as an alternative relational learning paradigm

Sebastijan Dumančić, Alberto García-Durán, Mathias Niepert
sebastijan.dumancic@cs.kuleuven.be, {alberto.duran,mathias.niepert}@neclab.eu
KU Leuven, Belgium; NEC Labs Europe, Germany

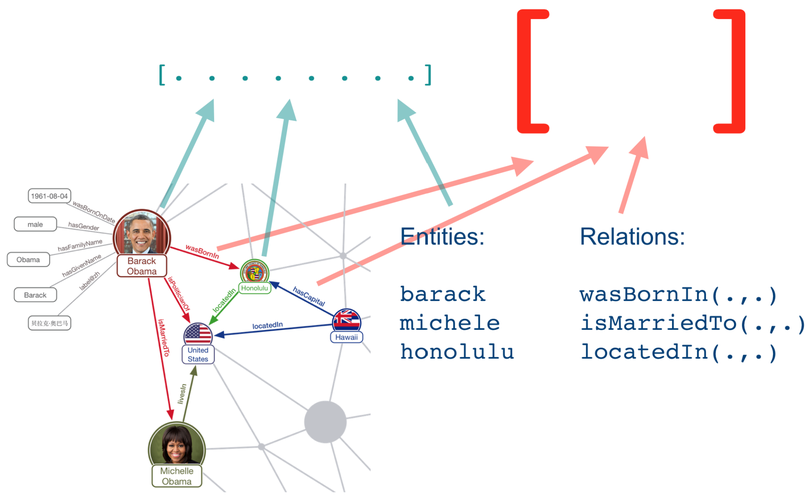
Paradigms of relational learning

Statistical relational learning
(Probabilistic) inductive logic programming
→ use predicate logic (and probability theory) to represent complex data

```
0.3::stress(X) :- person(X).
0.4::asthma(X) :- smokes(X).
smokes(X) :- stress(X).
smokes(X) :- friend(X,Y).

person(angelika).
person(joris).
person(jonas).
friend(joris,jonas).
friend(joris,angelika).
```

Knowledge graph embeddings
→ re-represent data as vectors



What are the relative strengths of the two paradigms?

Goal: compare the typical representatives of ILP (because of the better support for learning the logical theories) and KG embeddings on a series of classification and clustering tasks; also include relational *latent representation learning* approach CUR^{2}LED

Classification

ILP methods and **embeddings** often achieve comparable results

There is no significant difference between different embeddings methods

The amount of informative information in the immediate neighbourhood of target vertices seems to be indicative whether KGE of ILP performs better

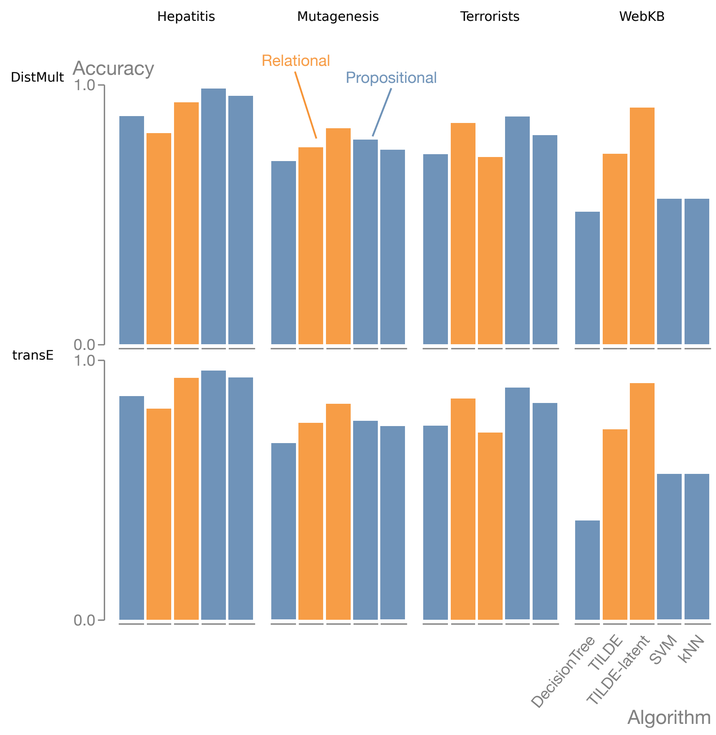
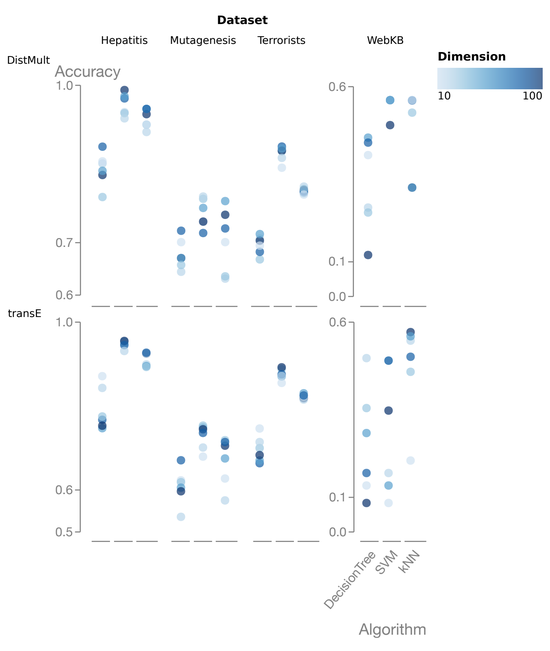


Table 1: Properties of extracted rules are summarized as a proportion of relation and attribute predicates

Dataset	Reasoning depth	Rule properties			Dataset properties		
		Relations	Attributes	Mix	Attributes	Relations	Neighbourhood
Hepatitis	1.4	10 %	38 %	52 %	100 %	66 %	70 %
Terrorists	1	0 %	95 %	5 %	35 %	100 %	0 %
Mutagenesis	0	0 %	100 %	0 %	100 %	0 %	0 %
WebKB	2	5 %	76 %	19 %	4 %	50 %	1 %

Hyper-parameters of the embeddings methods are very important

Performance is very sensitive to the choice of the dimension of the embedding vector

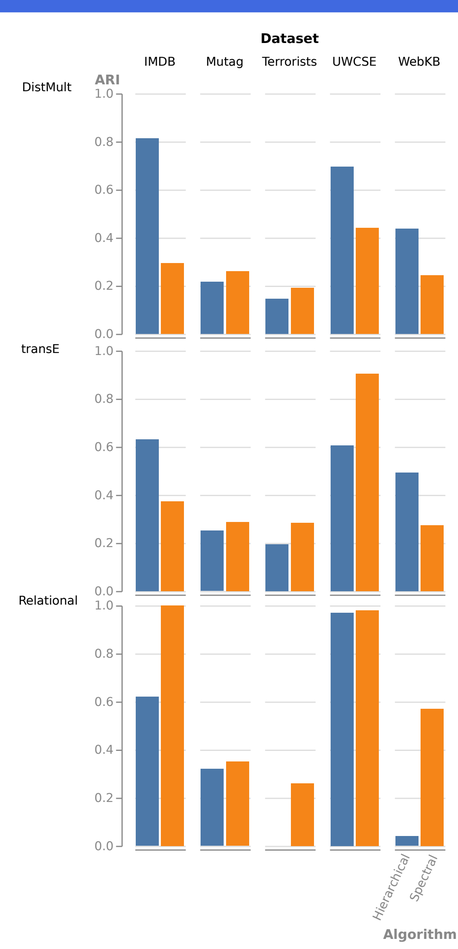


Clustering

Embedding methods show potential for relational clustering

Performance is similar to the state-of-the-art relational clustering method

The choice of the clustering algorithm seems to play an improtant rule



Unfortunately, the clustering performance is extremely sensitive to the dimension of the embedding vectors

How to make this choice without the labels?
No clear way to do that