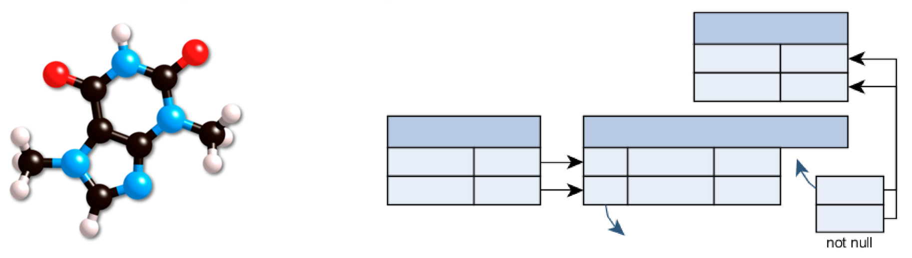
Clustering-Based Unsupervised Relational Representation Learning with an Explicit Distributed Representation

{Sebastijan.Dumancic and Hendrik.Blockeel }@cs.kuleuven.be

Problem

Deep learning revolutionized machine learning by automatically learning multiple layers of **abstract, re-usable** and **compositional** features for a given task. Despite huge interest, it still heavily focuses on *sensory* data such as images and speech – here we focus on **learning latent features of rich relational data formats** such as *graphs* and **relational databases**!

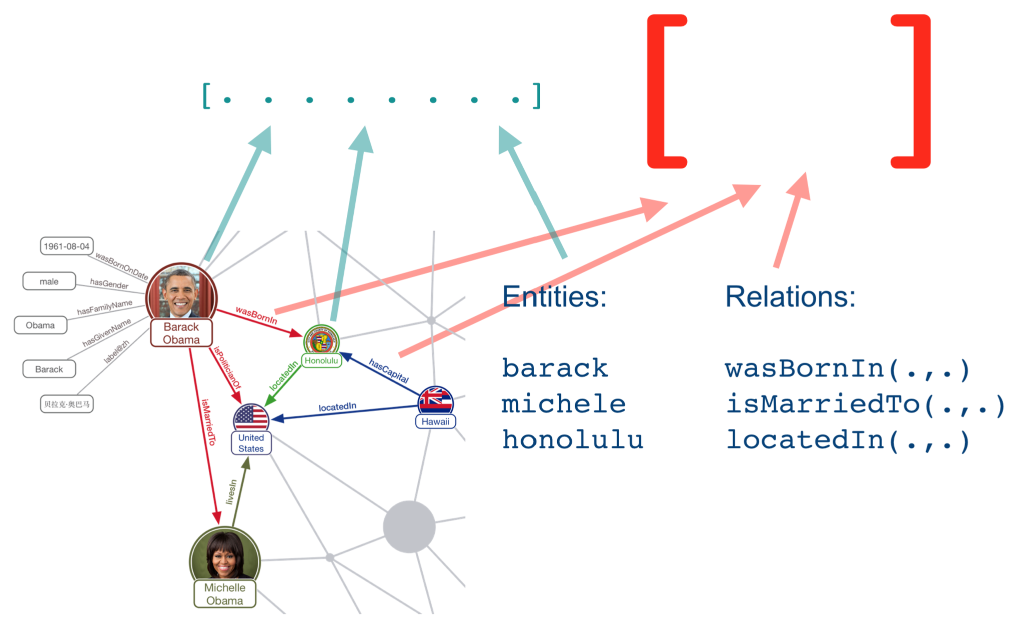
Relational data contains both instances and relationships amongst them



Current state of affairs

Vectors spaces in knowledge graphs: *Replace symbols with vectors, and logic with algebra*

Learning representation = learning vector representation on entities, and matrices/functions for relations



A good vector representation is the one that, given a **true** fact `wasBornIn(barack,honolulu)`, results in a high value of the vector-matrix multiplication of the corresponding entities

and a *low* value for the *false* example., e.g. `wasBornIn(barack,nairobi)`

Problems: uninterpretable latent features, need big amounts of data, difficult to handle unseen entities, does not integrate in existing relational learners

Goal

Develop an *relational* representation learning method that is:

- **relational** – considers both instances and their relationships
- **unsupervised** – no labels provided
- **interpretable** – latent features defined in logic
- integrates with existing relational learners

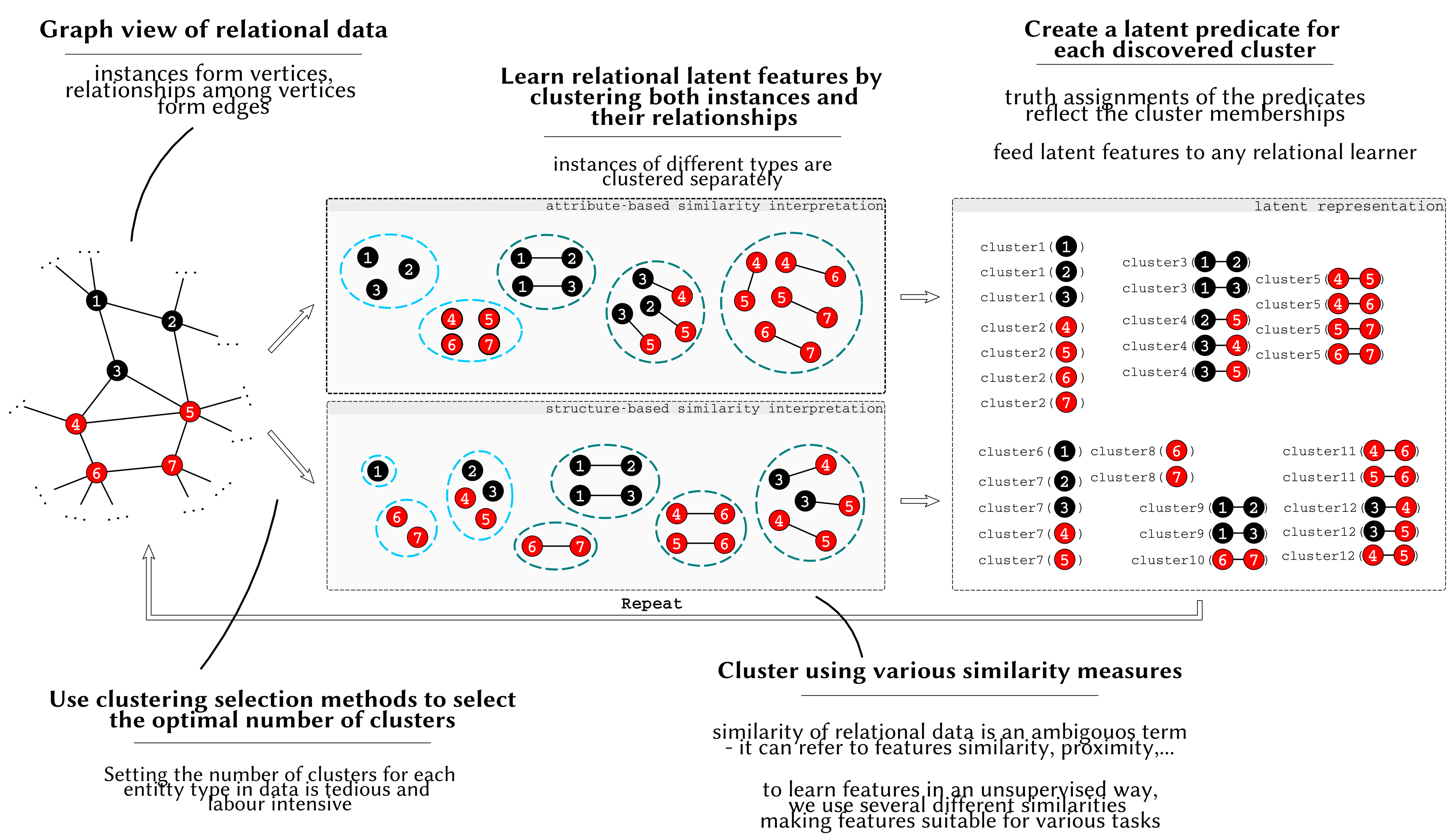
Statistical relational models

```
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).
```

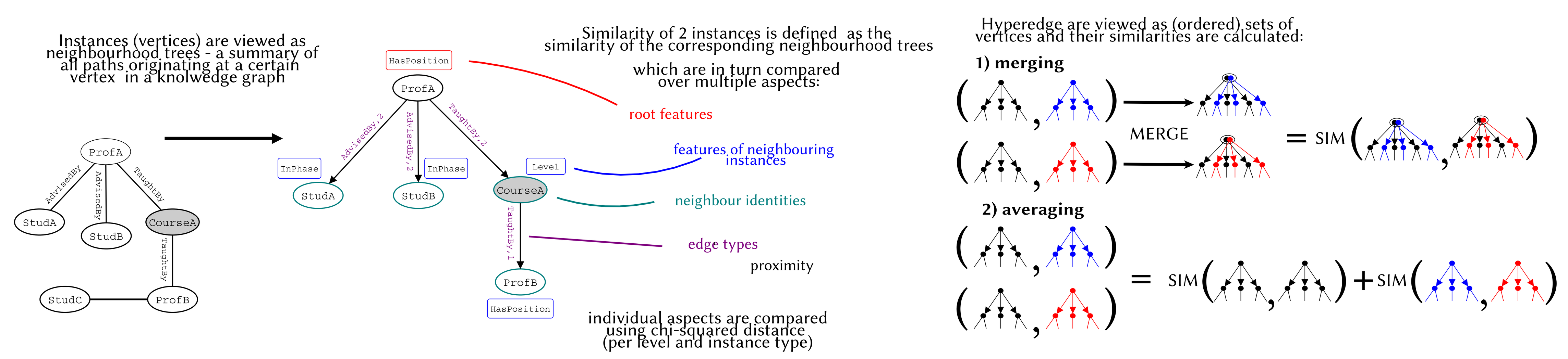
Learning relational latent features with clustering

The proposed pipeline is inspired by the *k-means feature learning pipeline* [Coates and NG, 2001]



Similarity of relational data

Similarity of instances and their relationships is assessed with ReCeNT [Dumancic & Blockeel, MLJ 2017]



Experiments

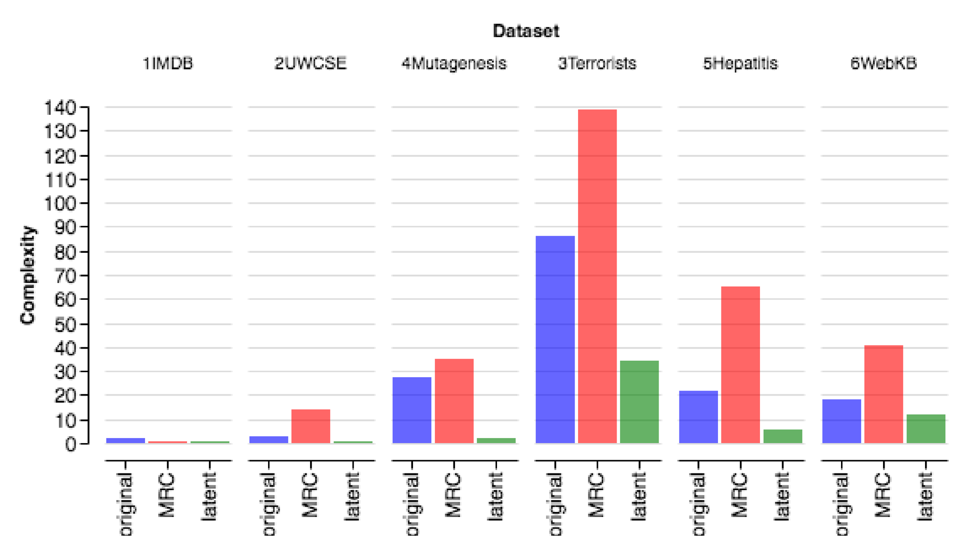
Question: *Does learning from latent spaces benefit relational learners compared to learning in the original space?*

- lower model complexity
- improved performance

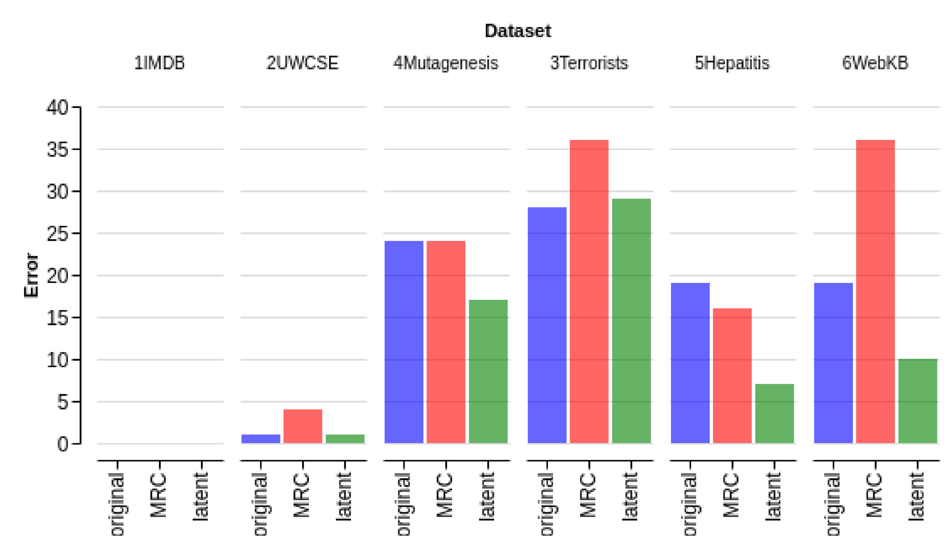
Setup

- learn features on training set; map test data to learned clusters (cross validation)
- learn relational decision tree `TILDE` on original and latent representations

Results



`TILDE` models learned on latent representations are less complex in terms of the number of nodes in a tree



`TILDE` models learned on latent representations often perform better (MRC = related approach)

References

Coates, A., Lee, H. & Ng, A.: *An analysis of single-layer networks in unsupervised feature learning*, AISTATS 2011

Dumancic, S., & Blockeel, H.: *An expressive dissimilarity measure for relational clustering using neighbourhood trees*, MLJ, to appear

Blockeel, H., & De Raedt, L.: *Top-down induction of first-order logical decision trees*, Artif. Intell. 1998

Kok, S., & Domingos, P.: *Statistical predicate invention*, ICML 2007

CODE:

dtai.cs.kuleuven.be/software/curled