KU LEUVEN

# Demystifying Relational Latent Representations

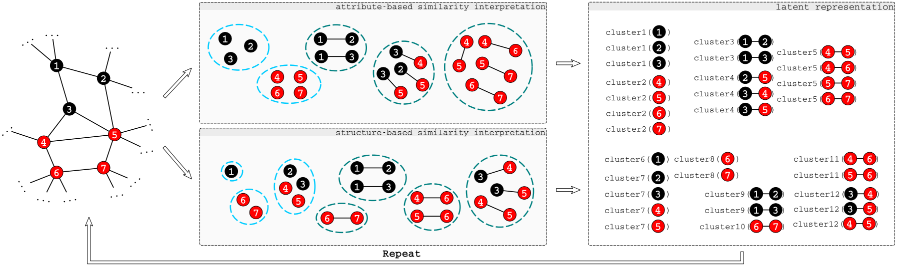
Sebastijan Dumančić, Hendrik Blockeel

DTAI, KU Leuven

September 6, ILP 2017

1. Introduction
2. Understanding latent features
3. Properties of latent spaces

Learning versatile relational latent features with clustering and variety of similarities (CUR$^2$LED)




[Dumančić and Blockeel, IJCAI 2017]

Benefits:

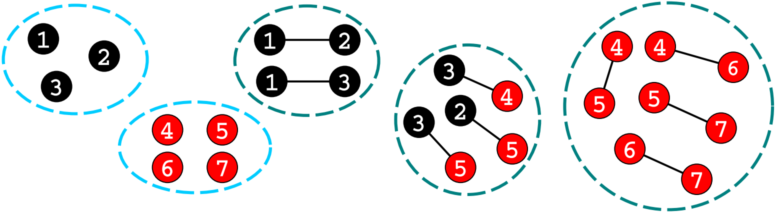
- better performance
- simpler models

[with some overhead]

Questions to be answered

1. *Can we interpret latent features?*
   - (approximate) definition of latent features/relations
2. *What makes them effective?*
   - distinctive properties?

latent features = **clusters of vertices (instances) and edges (relationships)**

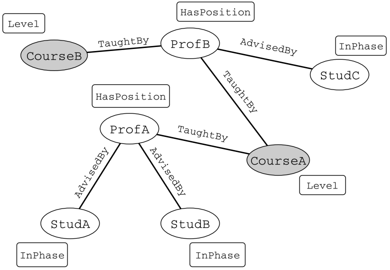



**key idea:** cluster prototype represents the meaning of a feature

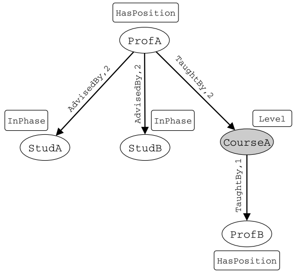
# 2 – Understanding features - technicalities (1)

CUR$^2$LED uses ReCeNT as a similarity measure for relational data

$\Rightarrow$ views instances as **neighbourhood trees**




Data



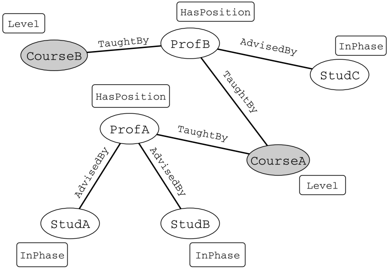

Neighbourhood tree

[Dumančić and Blockeel, MLJ 2017]

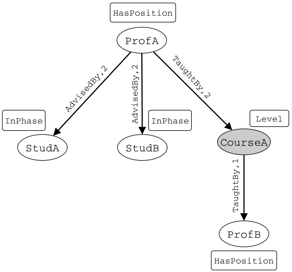
KU LEUVEN

CUR$^2$LED uses ReCeNT as a similarity measure for relational data

⇒ views instances as **neighbourhood trees**




Data



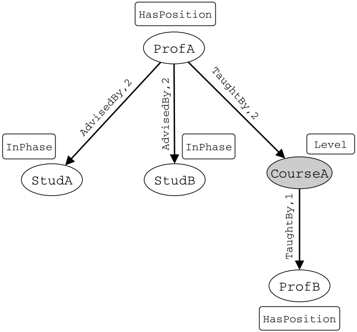

Neighbourhood tree

**key idea:** *mean tree* represents the meaning of a feature

CUR$^2$LED requires a (set of) **similarity interpretation(s)**

→ a specification what similarity reflects

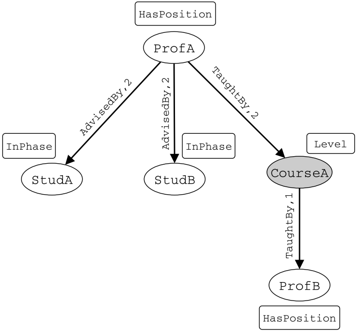
- attribute sim
- neighbourhood attributes sim
- neighbourhood identity
- edge labels
- connectedness




**key idea:** find a mean tree given the similarity interpretation

CUR²LED compares neighbourhood trees by comparing distributions of elements within them

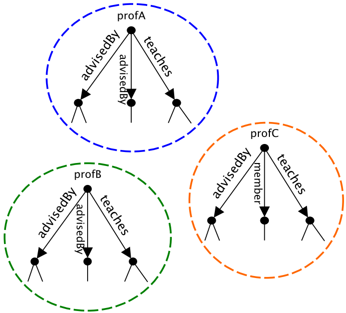
- elements selected by the similarity interpretation
- attributes values, edge types, identities ...




**key idea:** <u>mean tree</u> ≈ elements that appear in all NTs (in a cluster) with similar frequency

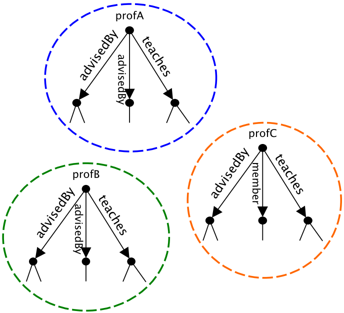
Given a set of neighbourhood tree and a similarity interpretation ...

1.

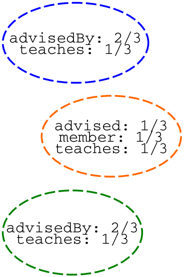


Calculate the relative frequencies of elements within a tree

1.



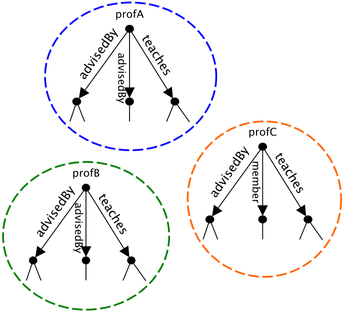

2.



# 2 – Understanding features - outline

Summarize the relative frequencies of unique elements across trees

1.




2.

```
advisedBy: 2/3
teaches: 1/3
```

```
advised: 1/3
member: 1/3
teaches: 1/3
```

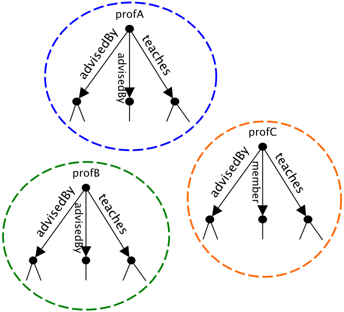
```
advisedBy: 2/3
teaches: 1/3
```

3.

```
advisedBy -> mean: 0.55
              std: 0.15
teaches -> mean: 0.33
            std: 0.0
member -> mean: 0.11
           std: 0.15
```

Select elements with low standard deviation

1.




2.

advisedBy: 2/3
teaches: 1/3

advised: 1/3
member: 1/3
teaches: 1/3

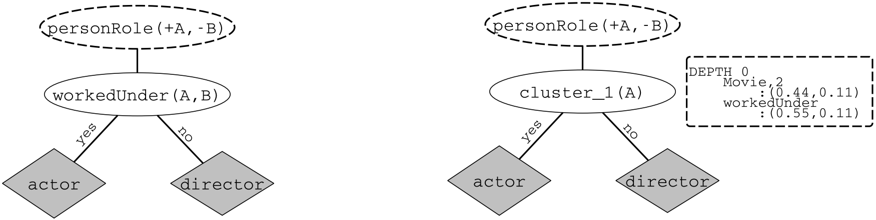
advisedBy: 2/3
teaches: 1/3

3.

```
advisedBy -> mean: 0.55
             std: 0.15
teaches -> mean: 0.33
           std: 0.0
member -> mean: 0.11
          std: 0.15
```

**($\theta$-confidence)** An element with mean value $\mu$ and standard deviation $\sigma$ in a cluster, is said to be $\theta$-confident if $\sigma < \theta \cdot \mu$.

## Use case: **IMDB**



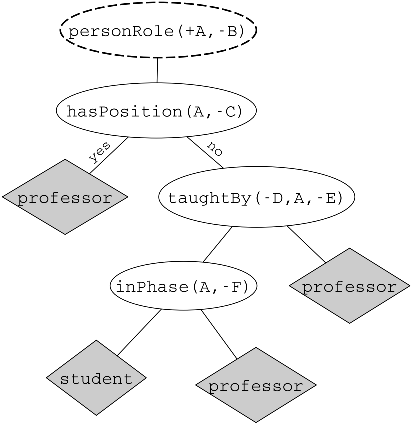

a) b)



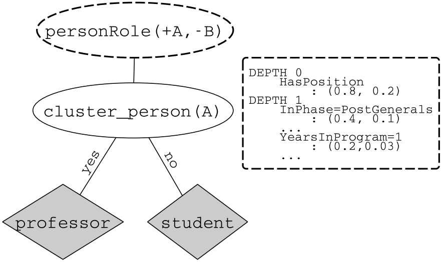
KU LEUVEN

# 2 – Understanding features - results

## Use case: **UWCSE**




a)

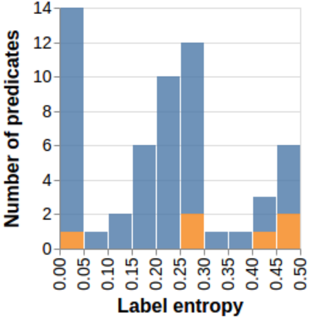



b)

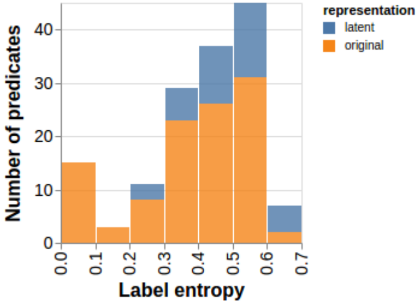
1. Introduction
2. Understanding latent features
3. Properties of latent spaces

Properties of latent spaces:

- **label entropy**
  - distribution of labels within true instantiations of predicates
  - proxy to a quantification of learning difficulty
- **sparsity**
  - modelling *local* vs. *global*
  - concept spread across a small number of local regions is easier to capture
- **redundancy**
  - $\text{CUR}^2\text{LED}$ creates many features - are all of them necessary?

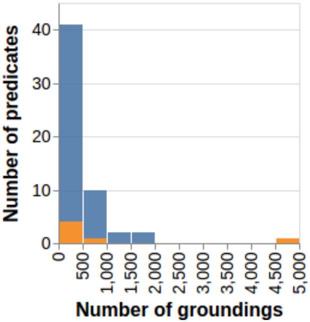



improved performance (3)

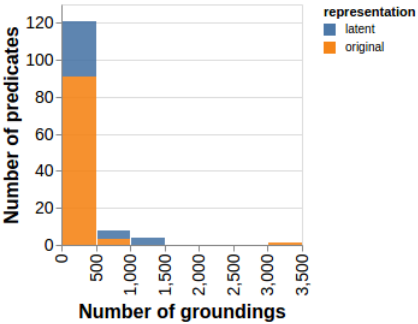



no improvement (1)

when performance increases, latent representation has many predicates of low label entropy



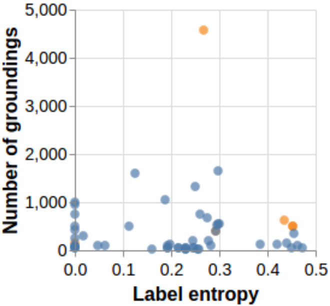

improved performance (3)



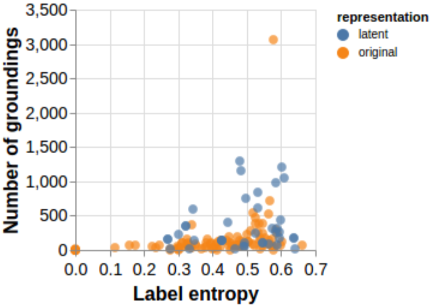

no improvement (1)

when performance increases, latent representation is sparser than the original one

Trivial explanation: many predicates with a very small number of true instantiations (not helpful)




improved performance (3)

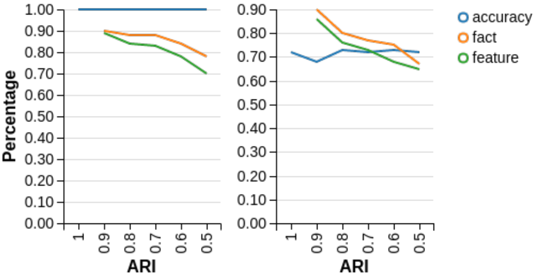



no improvement (1)

... not what's happening here: latent predicates have a comparable number of true instantiations latent features successfully identify local regions in the instance space that match well with the provided labels

CUR$^2$LED creates a lot of features

- similarity interpretations considered independently
- but many instances might be identical in several similarity interpretations
- every time a new clustering is obtained, check how much it overlaps with an existing clusterings using the adjusted Rand index






KU LEUVEN

Thank you!

Questions?