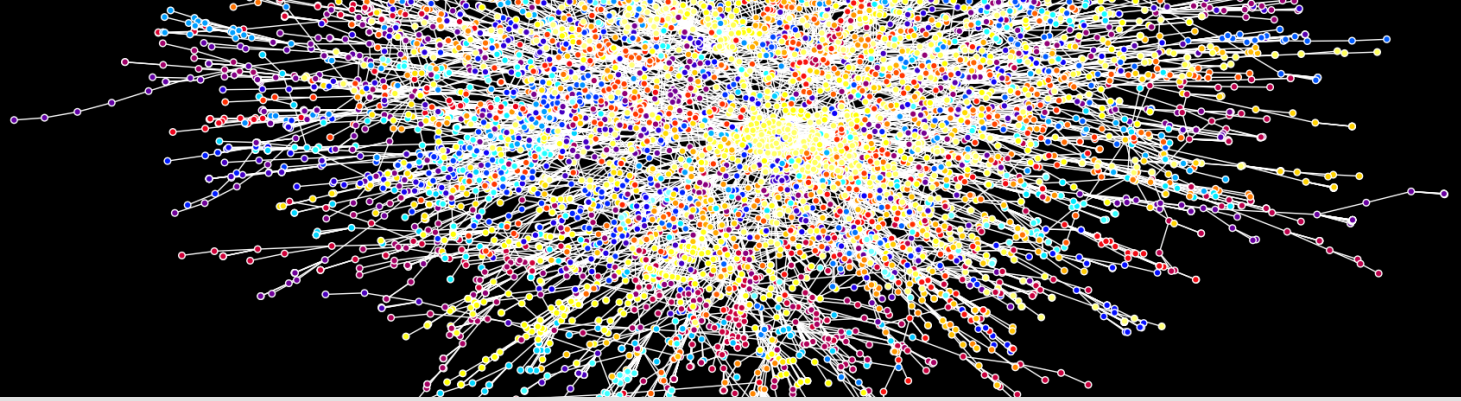


Targeted end-to-end knowledge graph decomposition

Blaž Škrlj, Jan Kralj and Nada Lavrač

Jožef Stefan Institute, Ljubljana, Slovenia
blaz.skrlj@ijs.si

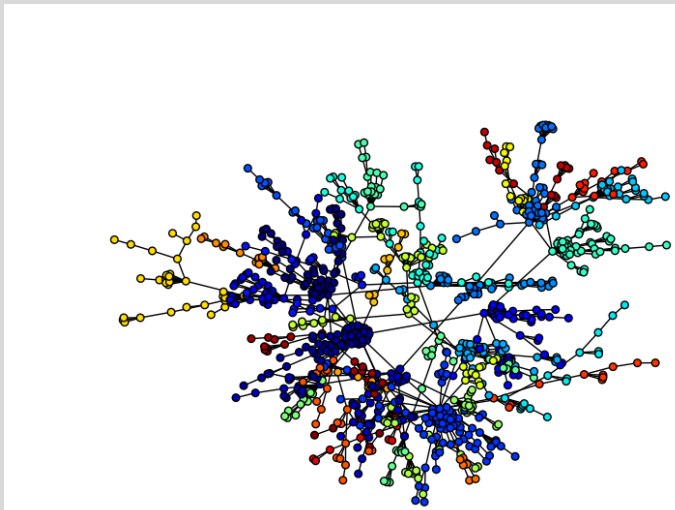
September 3, 2018

Jožef Stefan Institute

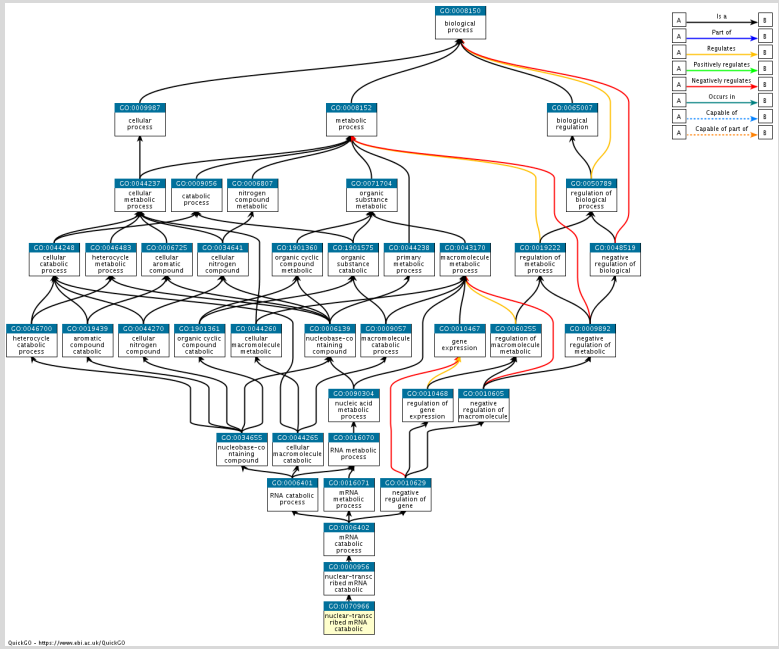
MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
SLOVENIJA
LJUBLJANA
2004

Introduction

Complex networks



Curated knowledge (e.g., Ontologies)

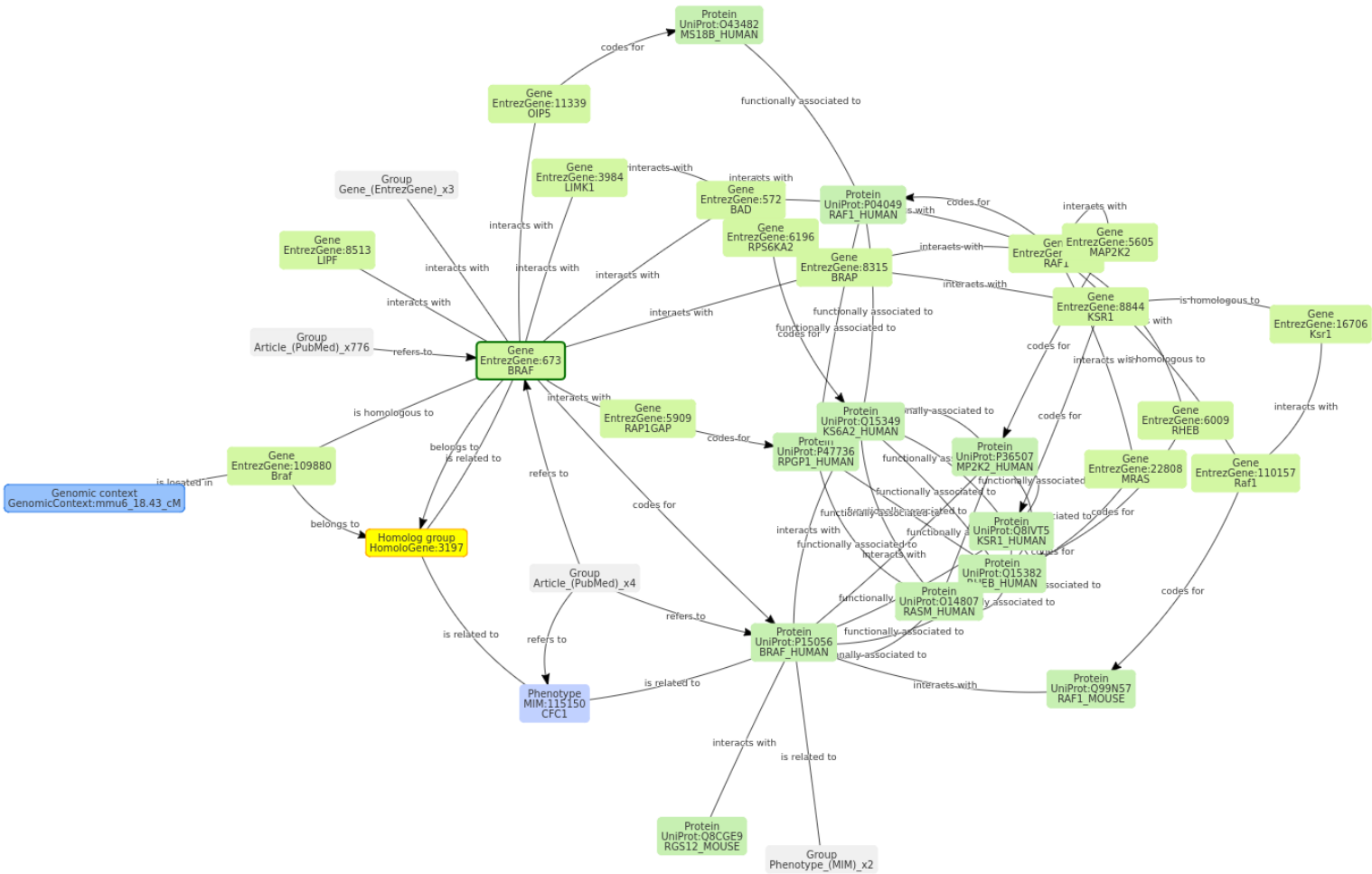


- Can we use the curated (background) knowledge to learn better from networks?

Jožef Stefan Institute

Knowledge graphs

- Complex networks + semantic relations (e.g., BioMine[1])



[1] Lauri Eronen and Hannu Toivonen. “Biomine: predicting links between biological entities using network models of heterogeneous databases”. In: *BMC bioinformatics* 13.1 (2012), p. 119.

Problem statement

Inputs

Given:

- A knowledge graph (with relation-labeled edges)
- A set of class-labeled target nodes

Outputs

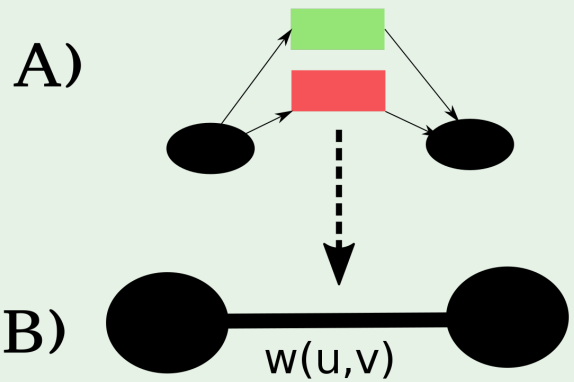
- An optimal decomposition of the knowledge graph with respect to target nodes and a *given task* (e.g., node classification)

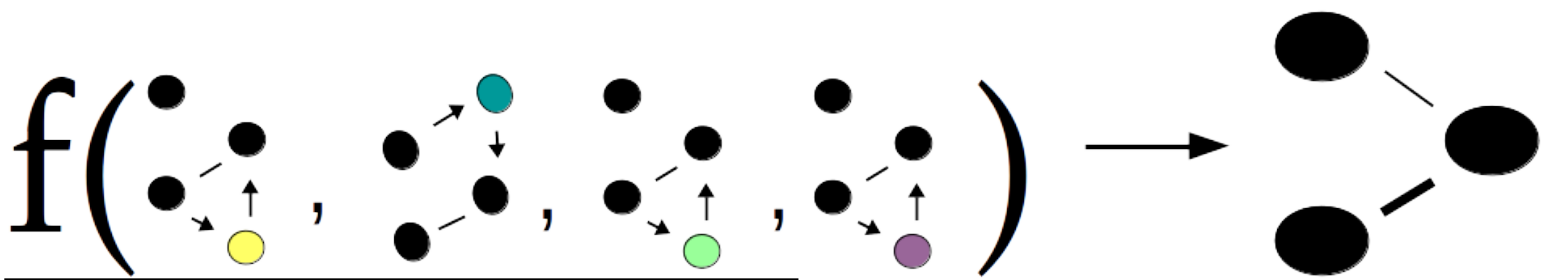
Open problem: **How to automatically exploit background knowledge (relation-labeled edges) during learning?**

Network decomposition—HINMINE[2] key idea

- Identify directed paths of length two between the target nodes of interest.
- Construct weighted edges between target nodes.

Edge construction.





[2] Jan Kralj, Marko Robnik-ikonja, and Nada Lavra. "HINMINE: Heterogeneous information network mining with information retrieval heuristics". In: *Journal of Intelligent Information Systems* (2017), pp. 1–33.

Edge weight computation

More formally, given a heuristic function f, a weight of an edge between the two nodes u and v is computed as

$$w(u,v) = \sum_{\substack{m \in M \\ (u,m) \in E \\ (m,v) \in E}} f(m);$$

where the $f(m)$ represents the weight function and m an intermediary node. Here, M represents the set of intermediary nodes and E the set of a knowledge graph's edges.

HINMINE and current state-of-the-art

Table 1: HINMINE term weighing schemes, tested for decomposition of knowledge graphs and their corresponding formulas in text mining.

Scheme	Formula
`tf`	$f(t,d)$
`if-idf`	$f(t,d) \cdot \log\left(\frac{\lvert D\rvert}{\lvert\{d' \in D : t \in d'\}\rvert}\right)$
`chi^2`	$f(t,d) \cdot \sum_{c\in C} \frac{(P(t\wedge c)P(\neg t\wedge\neg c) - P(t\wedge\neg c)P(\neg t\wedge c))^2}{P(t)P(\neg t)P(c)P(\neg c)}$
`ig`	$f(t,d) \cdot \sum_{c\in C, c'\in\{c,\neg c\} t'\in\{t,\neg t\}} \left(P(t',c') \cdot \log\frac{P(t'\wedge c')}{P(t')P(c')}\right)$
`gr`	$f(t,d) \cdot \sum_{c\in C} \frac{\sum_{c'\in\{c,\neg c\}}\sum_{t'\in\{t,\neg t\}}\left(P(t',c')\cdot\log\frac{P(t'\wedge c')}{P(t')P(c')}\right)}{-\sum_{c'\in\{c,\neg c\}} P(c)\cdot\log P(c)}$
`delta-idf`	$f(t,d) \cdot \sum_{c\in C}\left(\log\frac{\lvert c\rvert}{\lvert\{d'\in D : d'\in c\wedge t\in d'\}\rvert} - \log\frac{\lvert\neg c\rvert}{\lvert\{d'\in D : d'\notin c\wedge t\notin d'\}\rvert}\right)$
`rf`	$f(t,d) \cdot \sum_{c\in C}\log\left(2+\frac{\lvert\{d'\in D : d'\in c\wedge t\in d'\}\rvert}{\lvert\{d'\in D : d'\notin c\wedge t\notin d'\}\rvert}\right)$
`bm25`	$f(t,d)\cdot\log\left(\frac{\lvert D\rvert}{\lvert\{d'\in D : t\in d'\}\rvert}\right)\cdot\frac{k+1}{f(t,d)+k\cdot\left(1-b+b\cdot\frac{\lvert d\rvert}{\text{avgdl}}\right)}$

Towards end-to-end decomposition

HINMINE's heuristics are comparable to state-of-the-art methods, **BUT**

- A Heuristic's performance is dataset-dependent
- Paths, used for decomposition are manually selected (many possibilities)

In this paper we address the following questions:

- Can we automate the heuristic selection?
- Can decompositions be combined?
- Is domain expert knowledge really needed for path selection?

Decomposition as stochastic optimization

$$X_{opt} = \underset{(d,o,t)\in P(\mathfrak{D})\times\mathfrak{S}\times P(\mathfrak{T})}{\arg\min} \Big[\rho(\tau(d,o,t))\Big].$$

Where the:

- (d, o, t) corresponds to paths, operators and heuristics used
- τ corresponds to decomposition computation
- ρ represents a decomposition scoring function
- X_{opt} is the optimal decomposition

Combining decompositions

Set of heuristic combination operators. Let $\{h_1, h_2, \ldots, h_k\}$ be a set of matrices, obtained using different decomposition heuristics. We propose four different heuristic combination operators.

1. **Element-wise sum**. Let $\oplus$ denote elementwise matrix summation. Combined aggregated matrix is thus defined as $M = h_1 \oplus \cdots \oplus h_k$, a well defined expression as $\oplus$ represents a commutative and associative operation.
2. **Element-wise product**. Let $\otimes$ denote elementwise product. Combined aggregated matrix is thus defined as $M = h_1 \otimes \cdots \otimes h_k$.
3. **Normalized element-wise sum**. Let $\oplus$ denote elementwise summation, and $\max(A)$ denote the largest element of the matrix A. Combined aggregated matrix is thus defined as $M = \frac{1}{\max(h_1 \oplus \cdots \oplus h_k)}(h_1 \oplus \cdots \oplus h_k)$. As $\oplus$ represents a commutative operation, this operator can be generalized to arbitrary sets of heuristics without loss of generality.
4. **Normalized element-wise product**. Let $\otimes$ denote elementwise product, and $\max(A)$ denote the largest element of the matrix A. Combined aggregated matrix is thus defined as $M = \frac{1}{\max(h_1 \otimes \cdots \otimes h_k)}(h_1 \otimes \cdots \otimes h_k)$. This operator can also be generalized to arbitrary sets of heuristics.

Decomposition as stochastic optimization

Considering all possible paths + all possible heuristics + combinations of different decompositions results in **combinatorial explosion**.

- Obtaining the optimal decomposition can also be formulated as **differential evolution**:
  - A binary vector of size $|heuristics| + |triplets| + |combinationOP|$ is propagated through the parametric space
  - final solution represents a unique decomposition

Pseudocode of the approach

1. Select unique paths, heuristics and operators
2. evolve binary vector of solutions with respect to target task (e.g., classification)
3. Upon final number of iterations/convergence etc., use the vector to obtain dataset-specific decomposition

BUT, how are the node labels predicted (decompositions scored)?

Jožef Stefan Institute

P-PR and node prediction

Modern way: **Prediction via subnetwork embeddings**. We compute P-PR vectors for individual target nodes, hence obtaining $|k|^2$ feature matrices, where $|k| << |N|$.
These matrices are used to learn the labels.

P-PR embeddings

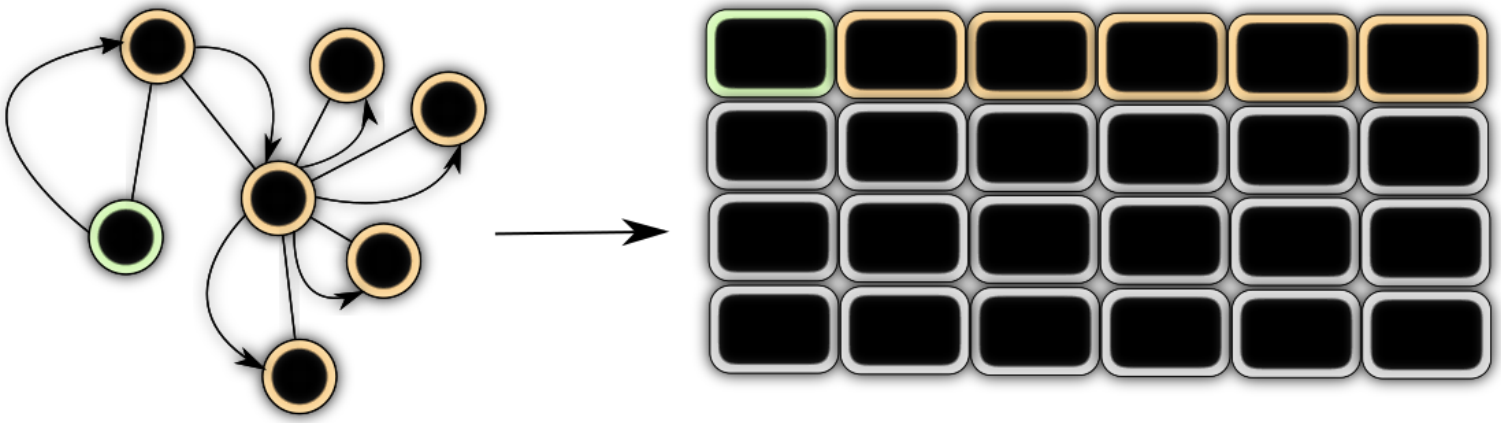


Figure 1: Personalized PageRank-based embedding. Repeated for each node, this iteration yields a $|k|^2$ matrix, directly usable for learning tasks.

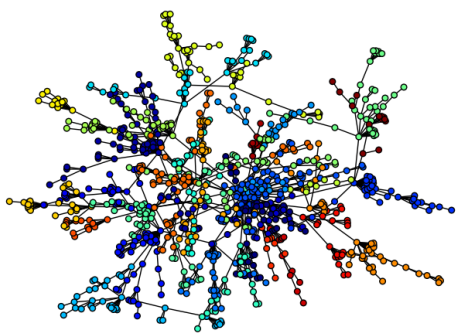
P-PR general use

Node classification

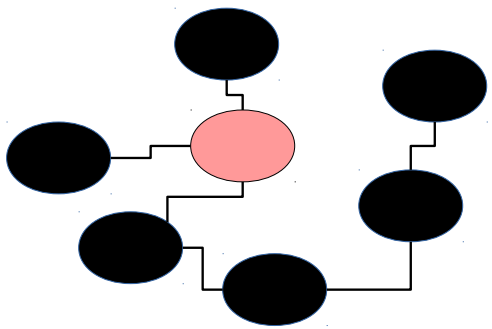
We try to classify individual nodes into target class (es). Relevant for e.g.,

- Protein function prediction
- Genre classification
- Recommendation etc.

Function prediction



Recommendation



Datasets

IMDB dataset—genre classification

The main classification task related to this dataset corresponds to classification of individual movie's genres, based on actors, directors and movies. Here, 300 nodes are labeled, whereas the whole network consists of 6,387 nodes and 14,714 edges. An example triplet yielding a valid decomposition for this dataset is: *Actor* $\xrightarrow{\text{actsIn}}$ *Movie* $\xrightarrow{\text{directedBy}}$ *Director*.

Protein function prediction

The classification goal for this dataset is thus protein function prediction[3]. The network consists of 2,204 nodes and 2,772 edges, 456 nodes are target (labeled) nodes.
Protein $\xrightarrow{\text{interactsWith}}$ *Protein* $\xrightarrow{\text{subsumes}}$ *Protein*.

[3] Sandra Orchard et al. "The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases". In: *Nucleic Acids Research* 42.Database issue (Jan. 2014), ISSN: 0305-1048. DOI: 10.1093/nar/gkt1115. URL: http://europepmc.org/articles/PMC3965093.

Results (1)

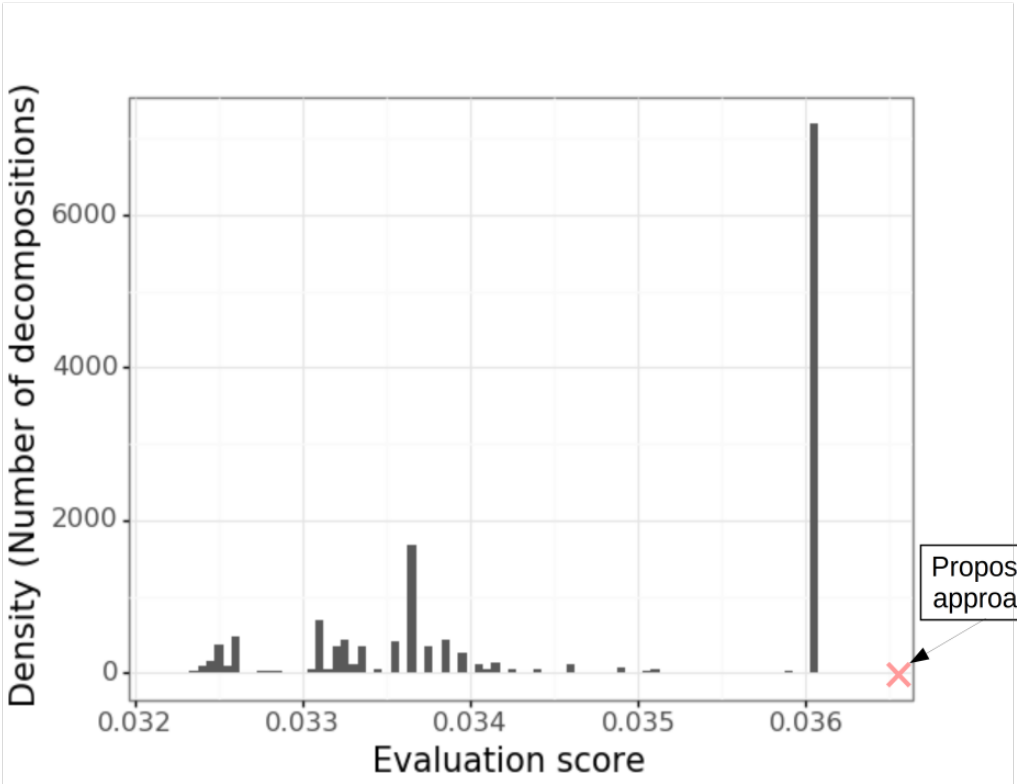


Figure 2: Global optimum found for the IMDB dataset.

Results (2)

The table of empirical results. The proposed approach was tested against random decomposition selection.

Dataset	Min F1	Max F1	Mean F1	Proposed approach	DE	Exhaustive search
IMDB	0.0315	**0.0372**	0.0346	**0.0372**	50min	$\approx 22h$
Epigenetics	0.0211	0.0296	0.0243	0.0284	6h	$> 1\mathrm{day}$

The result indicates significant speedups **(20x)** are possible even if no domain knowledge is present.

Example relations, relevant for classification

Epigenetics dataset (Target node = protein)

$$\textit{Protein} \xrightarrow{\text{contains}} \textit{Domain} \xrightarrow{\text{contains}} \textit{Protein}$$

$$\textit{Protein} \xrightarrow{\text{interactsWith}} \textit{Protein} \xrightarrow{\text{subsumes}} \textit{Protein}$$

$$\textit{Protein} \xrightarrow{\text{belongsTo}} \textit{Family} \xrightarrow{\text{belongsTo}} \textit{Protein}$$

$$\textit{Protein} \xrightarrow{\text{isRelatedTo}} \textit{Phenotype} \xrightarrow{\text{isRelatedTo}} \textit{Protein}$$

$$\textit{Protein} \xrightarrow{\text{interactsWith}} \textit{Protein} \xrightarrow{\text{interactsWith}} \textit{Protein}$$

IMDB (Target node = movie):

$$\textit{Movie} \xrightarrow{\text{features}} \textit{Person} \xrightarrow{\text{actsIn}} \textit{Movie},$$

$$\textit{Movie} \xrightarrow{\text{directedBy}} \textit{Person} \xrightarrow{\text{directed}} \textit{Movie},$$

$$\textit{Movie} \xrightarrow{\text{features}} \textit{Person} \xrightarrow{\text{directed}} \textit{Movie}.$$

Conclusions and further work

- One of the first end-to-end targeted decomposition approaches
- Used for classification task
- Relation relevance discovery
- Scalability (subnetworks in other domains)
- Extensibility (GA, ant colonies . . .)
- Generality of the approach (clustering?)
- Further use?

References I

Eronen, Lauri and Hannu Toivonen. “Biomine: predicting links between biological entities using network models of heterogeneous databases”. In: *BMC bioinformatics* 13.1 (2012), p. 119.

Kralj, Jan, Marko Robnik-ikonja, and Nada Lavra. “HINMINE: Heterogeneous information network mining with information retrieval heuristics”. In: *Journal of Intelligent Information Systems* (2017), pp. 1–33.

Orchard, Sandra et al. “The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases”. In: *Nucleic Acids Research* 42.Database issue (Jan. 2014), ISSN: 0305-1048. DOI: 10.1093/nar/gkt1115. URL: http://europepmc.org/articles/PMC3965093.

Jožef Stefan Institute