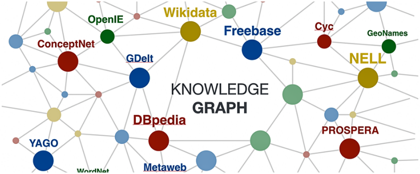
A perspective on Neuro-symbolic Approaches for Knowledge Graph Refinement and Explanation

Claudia d'Amato

Computer Science Department
University of Bari "Aldo Moro", Bari, Italy

Network NeSy AI Med - Zurich Meeting

July 4-5, 2024



[1]

Applications

- e-Commerce
- Semantic Search
- Fact Checking
- Personalization
- Recommendation
- Medical decision support system
- Question Answering
- Machine Translation
- ...

Research Fields

- Information Extraction
- Natural Language Processing
- Machine Learnig (ML)
- Knowledge Representation
- Web
- Robotics
- ...

[1] picture from https://www.csee.umbc.edu/courses/graduate/691/fall19/07/

ML and KGs

Two perspectives:

KG as input to ML

Goal: improving the performance in many learning tasks, e.g.

- Question Answering (QA)
- image classification
- instance disambiguation
- text summarization
- ....

ML as input to KG

Goal: improving the KG itself

- creating new facts
- creating generalizations
- prototyping
- improving the size, coverage, depth and accuracy of KGs $\rightarrow$ reducing their production costs

ML as input to KG

(KG Refinement: Link Prediction)

(Explanation of Link Predictions on KGs)

(KG Refinement at Schema Level)

ML as input to KG

(**KG Refinement: Link Prediction**)

(Explanation of Link Predictions on KGs)

(KG Refinement at Schema Level)

Incompleteness and **noise**

$\Downarrow$

Knowledge Graph Refinement

- Link Prediction: predicts missing links between entities
  - regarded as a learning to rank problem
- Triple Classification: assesses correctness of a statement wrt a KG
  - regarded as a binary classification problem

Very Large Data Collections

$\Downarrow$

New scalable Machine Learning methods

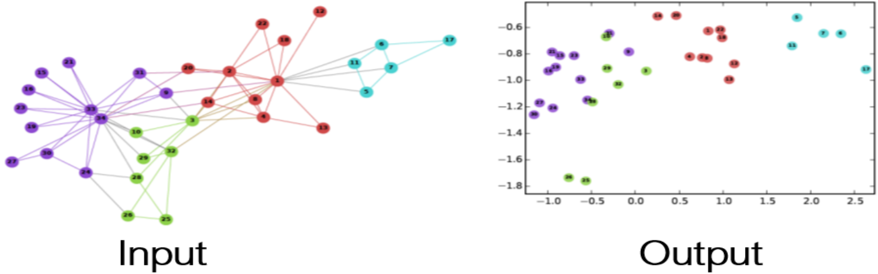
- grounded on numeric-based approaches
  - KG vector embedding models (KGE) largely investigated [Cai *et al.*, 2018]

ML/KGE for KGs: Issues

- CWA (or LCWA) mostly adopted vs. OWA
- schema level information and reasoning capabilities almost disregarded
- no interpretable models $\Rightarrow$ hard to motivate results

KG Embedding Models...

KGE models convert data graph into an optimal low-dimensional space [Cai *et al.*, 2018]



[2]

Graph structural information and properties preserved as much as possible

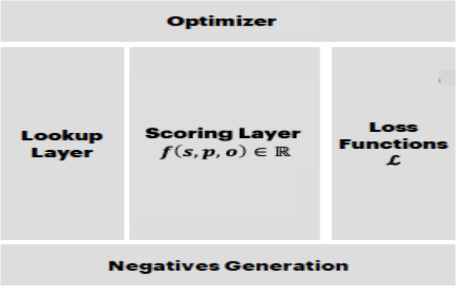
[2] Picture from https://laptrinhx.com/node2vec-graph-embedding-method-2620064815/

...KG Embedding Models

Goal

Learning embeddings s.t.

- score of a valid (positive) triple is higher than
- the score of an invalid (negative) triple



[3]

Negative examples generated by **random corruption of triples**

- false negatives may be generated
- only triple directly observable are considered

[3] Picture from "ECAI-20 Tutorial: Knowledge Graph Embeddings: From Theory to Practice"

KG Refinement by KG Embedding Methods:

Injecting Semantics

Enhancing KGE by Injecting Background Knowledge

[d'Amato *et al.*, 2021c,b]

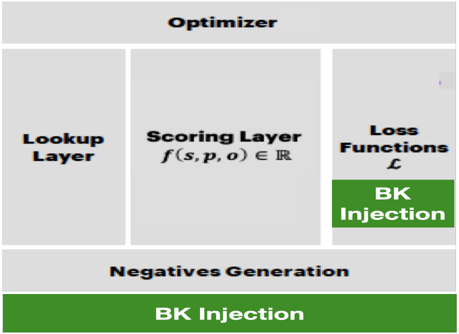
By two components:

Reasoning: used for generating negative triples

Axioms: `domain`, `range`, `disjointWith`, `functionalProperty`;

BK Injection: defines constraints on functions, corresponding to the considered axioms, guiding the way embedding are learned

Axioms: `equivClass`, `equivProperty`, `inverseOf` and `subClassOf`.



A First Approach: TransOWL, TransROWL

[d'Amato *et al.*, 2021c]

- Derive further triples to be considered for training via schema axioms
  - `equivClass`, `equivProperty`, `inverseOf` and `subClassOf`
- More complex loss function
  - adding a number of terms consistently with the constraints

$$
\begin{aligned}
L \; = \; & \overbrace{\sum_{\substack{\langle h,r,t\rangle\in\Delta \\ \langle h',r,t'\rangle\in\Delta'}} [\gamma + f_r(h,t) - f_r(h',t')]_+}^{\textbf{TransE}\ \textit{loss function}} + \sum_{\substack{\langle t,q,h\rangle\in\Delta_{\textbf{inverseOf}} \\ \langle t',q,h'\rangle\in\Delta'_{\textbf{inverseOf}}}} [\gamma + f_q(t,h) - f_q(t',h')]_+ \\
& + \sum_{\substack{\langle h,s,t\rangle\in\Delta_{\textbf{equivProperty}} \\ \langle h',s,t'\rangle\in\Delta'_{\textbf{equivProperty}}}} [\gamma + f_s(h,t) - f_s(h',t')]_+ + \sum_{\substack{\langle h,\textbf{typeOf},l\rangle\in\Delta\cup\in\Delta_{\textbf{equivClass}} \\ \langle h',\textbf{typeOf},l'\rangle\in\Delta'\cup\Delta'_{\textbf{equivClass}}}} [\gamma + f_{\textbf{typeOf}}(h,l) - f_{\textbf{typeOf}}(h',l')]_+ \\
& + \sum_{\substack{\langle h,\textbf{subClassOf},p\rangle\in\Delta_{\textbf{subClass}} \\ \langle h',\textbf{subClassOf},p'\rangle\in\Delta'_{\textbf{subClass}}}} [(\gamma - \beta) + f(h,p) - f(h',p')]_+
\end{aligned}
$$

where $q \equiv r^-$, $s \equiv r$ (properties), $l \equiv t$ and $t \sqsubseteq p$ (classes) and $f(h,p) = \|\mathbf{e}_h - \mathbf{e}_p\|$

An Alternative Approach: TransROWL^R

[d'Amato *et al.*, 2021c]

Adopting an axiom-based regularization of the loss function
as for TransE^R [Minervini *et al.*, 2017b]

- by adding specific constraints to the loss function rather than
- explicitly derive additional triples during training

Loss function

$$
\begin{aligned}
L \quad = & \sum_{\substack{\langle h,r,t\rangle \in \Delta \\ \langle h',r',t'\rangle \in \Delta'}} [\gamma + f'_r(h,t) - f'_r(h',t')]_+ \\
& +\lambda_{\mathbf{1}} \sum_{r \equiv q^- \in \mathcal{T}_{\mathbf{inverseOf}}} \|r + q\| \; + \lambda_{\mathbf{2}} \sum_{r \equiv q^- \in \mathcal{T}_{\mathbf{inverseOf}}} \|M_r - M_q\| \\
& +\lambda_{\mathbf{3}} \sum_{r \equiv p \in \mathcal{T}_{\mathbf{equivProp}}} \|r - p\| \; + \lambda_{\mathbf{4}} \sum_{r \equiv p \in \mathcal{T}_{\mathbf{equivProp}}} \|M_r - M_p\| \\
& +\lambda_{\mathbf{5}} \sum_{e' \equiv e'' \in \mathcal{T}_{\mathbf{equivClass}}} \|e' - e''\| \; + \lambda_{\mathbf{6}} \sum_{s' \subseteq s'' \in \mathcal{T}_{\mathbf{subClass}}} \|\mathbf{1} - \beta - (s' - s'')\|
\end{aligned}
$$

Main Outcome: TransROWL slightly outperforming TransROWL^R

Other KG Embedding Methods Leveraging BK

- Jointly embedding KGs and logical rules [Guo *et al.*, 2016]
  - triples represented as atomic formulae
  - rules represented as complex formulae modeled by t-norm fuzzy logics
- Adversarial training exploiting Datalog clauses encoding assumptions to regularize neural link predictors [Minervini *et al.*, 2017a]

A specific form of BK required, not directly applicable to KGs

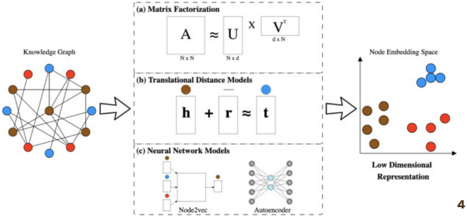
ML as input to KG

(KG Refinement: Link Prediction)

(**Explanation of Link Predictions on KGs**)

(KG Refinement at Schema Level)

Numeric-based methods consist of series of numbers without any obvious human interpretation

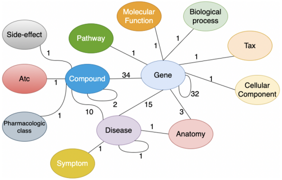


[4]

This may affects:

- the interpretability of the results
- the explainability
- and thus also somehow the trustworthiness of results

DRKG – Drug Repurposing Knowledge Graph



[5]

[4] Picture from D. N. Nicholson et al. Constructing knowledge graphs and their biomedical applications, Computational and Structural Biotechnology Journal, Vol. 18, pp. 1414–1428, (2020) ISSN 2001-0370

[5] Picture from https://github.com/topics/knowledge-graph-embeddings

A-Posteriori Explanations of Link Predictions

A-posteriori methods: find suitable explanation(s) based on the observed output and the model input, independently on the KGE adopted - Developed solutions

KELPIE [Rossi *et al.*, 2022]: generates necessary and sufficient (path) conditions and an articulated new evaluation protocol

CrossE [Zhang *et al.*, 2019]: embedding model for link predictions providing explanations

- the search for a path linking the subject h and object t of a predicted triple $\langle h, r, t \rangle$
  - **Max lenght 2** $\rightarrow$ six types of paths possible:
    Length 1: $P_1 = \{\langle h, r_s, t \rangle\}$, $P_2 = \{\langle t, r_s, h \rangle\}$
    Length 2: $P_3 = \{\langle e', r_s, h \rangle, \langle e', r', t \rangle\}$, $P_4 = \{\langle e', r_s, h \rangle, \langle t, r', e' \rangle\}$,
    $P_5 = \{\langle h, r_s, e' \rangle, \langle e', r', t \rangle\}$, $P_6 = \{\langle h, r_s, e' \rangle, \langle t, r', e' \rangle\}$,
    where r_s similar to r, r' any other relationship, e' any other entity;
- search driven by similarities between relation/entity embeddings by Euclidean distance
- structural comparisons with other paths in the KG to reinforce the reliability of the explanation found (referred to as support)

Explaining Link Predictions with CrossE

Given the predicted triple: ⟨`NickMason`, `recordLabel`, `CapitolRecords`⟩

why is it provided?

User is able to understand motivations, and trust (or not) the prediction

Example of exmplanation

⟨`NickMason`, `associatedBand`, `PinkFloyd`⟩,
⟨`PinkFloyd`, `recordLabel`, `CapitolRecords`⟩

Ideally supported by analogous situations to be found in the KG e.g.

⟨`RingoStarr`, `recordLabel`, `Parlophone`⟩

for which the computed explanation is:
⟨`RingoStarr`, `associatedBand`, `TheBeatles`⟩,
⟨`TheBeatles`, `recordLabel`, `Parlophone`⟩.

Exploiting Semantics for Providing Explanations to Link Predictions on KGs

Idea: SemanticCrossE provide semantic-based explanations for link predictions on KGs

[d'Amato *et al.*, 2021a]

Extends CrossE by adopting a Semantic Cosine similarity that leads the explanation process

- exploits the underlying KG semantics → Domain, Range and Classes considered
- increases the cosine similarity of two (entities / relationships) vector embeddings on the ground of available additional semantic information which is captured by a semantic Score function

Definition (semantic Cosine)

Given KG $\mathcal{K}(\mathcal{E}, \mathbb{R})$, the semantic Cosine measure for two entities $e, e' \in \mathcal{E}$ is defined by:

$$\mathrm{semCos}_{\alpha,\beta}(e, e') = \alpha \cdot \mathrm{sScore}(e, e') + \beta \cdot \mathrm{sim}_{\mathsf{cos}}(\mathsf{e}, \mathsf{e}')$$

where e the respective entity embedding vector; $\alpha, \beta \in [0, 1]$ chosen s.t. $\alpha + \beta = 1$.

In the case of relations $r, r' \in \mathbb{R}$ the measure is defined analogously.

Definition (semantic Score)

Given $\mathcal{C}$ set of the classes occurring in $\mathcal{K}(\mathcal{E}, \mathbb{R})$, and the functions $Cl \colon \mathcal{E} \to \mathcal{C}$, $Do \colon \mathbb{R} \to \mathcal{C}$, and $Ra \colon \mathbb{R} \to \mathcal{C}$ that return, resp., the conjunction of the classes an entity belongs to, and the domain and range of a relation, the semantic Score function for pairs of entities $e, e' \in \mathcal{E}$ is defined by:

$$\mathrm{sScore}(e, e') = \frac{|\mathrm{ret}[Cl(e) \sqcap Cl(e')]|}{|\mathrm{ret}[Cl(e) \sqcup Cl(e')]|}$$

where $\mathrm{ret}_{\mathcal{K}}(C)$ returns the entities that can be proven to belong to a given class C

Analogously, given any two relationships $r, r' \in \mathbb{R}$, it is defined:

$$\mathrm{sScore}(r, r') = \frac{|\mathrm{ret}[Do(r) \sqcap Do(r')]|}{|\mathrm{ret}[Do(r) \sqcup Do(r')]|} + \frac{|\mathrm{ret}[Ra(r) \sqcap Ra(r')]|}{|\mathrm{ret}[Ra(r) \sqcup Ra(r')]|}$$

- Approximated form of semantic Cosine measure (specifically of the semantic Score function) employed [d'Amato *et al.*, 2021a]
- Efficient computation obtained by a preliminary clustering phase [d'Amato *et al.*, 2023]

Main Outcome:

- ApproxSemanticCrossE showed improved results both in terms of recall and support
- ApproxSemanticCrossE not affected by noisy (irrelevant) explanations as for CrossE and partially cosineCrossE → qualitative evaluation conducted

A more efficient solution developed with **KELPIE++** [Barile *et al.*, 2024]

ML as input to KG

(KG Refinement: Link Prediction)

(Explanation of Link Predictions on KGs)

(**KG Refinement at Schema Level**)

Learning Disjointness Axioms

Missing Disjointness Axioms: Issues

Disjointness axioms often missing [Wang *et al.*, 2006]

Problems:

- introduction of noise

$\mathcal{K} = \{JournalPaper \sqsubseteq Paper,\ ConferencePaper \sqsubseteq Paper,\ ConferencePaper(a), Author(a)\ \}$
$\mathcal{K}$ is Consistent !!!
Cause Axiom: $Author \equiv \neg ConferencePaper$ **missing**

- counterintuitive inferences

$\mathcal{K} = \{JournalPaper \sqsubseteq Paper,\ ConferencePaper \sqsubseteq Paper,\ ConferencePaper(a)\ \}$

$\mathcal{K} \models JournalPaper(a)?$
Answer: Unknown
Cause Axiom: $JournalPaper \equiv \neg ConferencePaper$ **missing**

- **hard collecting negative examples when adopting numeric approaches**

Observation: extensions of disjoint concepts do not overlap

Question: would it be possible to automatically capture disjointness axioms by analyzing the data configuration/distribution?

Idea: Exploiting **(Conceptual) clustering methods** for mining disjointness axioms

[Rizzo *et al.*, 2021]

Definition (Problem Definition)

Given

- an ontological knowledge base $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$
- a set of individuals (aka entities) $\mathsf{I} \subseteq \mathrm{Ind}(\mathcal{A})$

Find

- n pairwise disjoint clusters $\{\mathsf{C}_1, \ldots, \mathsf{C}_n\}$
- for each $i = 1, \ldots, n$, a concept description D_i that describes C_i, such that:
  - $\forall a \in \mathsf{C}_i : \ \mathcal{K} \models D_i(a)$
  - $\forall b \in \mathsf{C}_j, j \neq i : \ \mathcal{K} \models \neg D_i(b)$.
- Hence $\forall D_i, D_j, i \neq j : \ \mathcal{K} \models D_j \sqsubseteq \neg D_i$.

Learning Disjointness Axioms: Developed Methods

Statistical-based approach

- NAR - exploiting negative association rules [Fleischhacker and Völker, 2011]
- PCC - exploiting Pearson's correlation coeff. [Völker *et al.*, 2015]

do not exploit any background knowledge and reasoning capabilities

Disjointness axioms learning/discovery can be hardly performed without symbol-based methods

Terminological Cluster Tree

Defined a method [6] for eliciting disjointness axioms [Rizzo *et al.*, 2021]

- solving a clustering problem via learning Terminological Cluster Trees
- providing a concept description for each cluster

Definition (Terminological cluster tree (TCT))

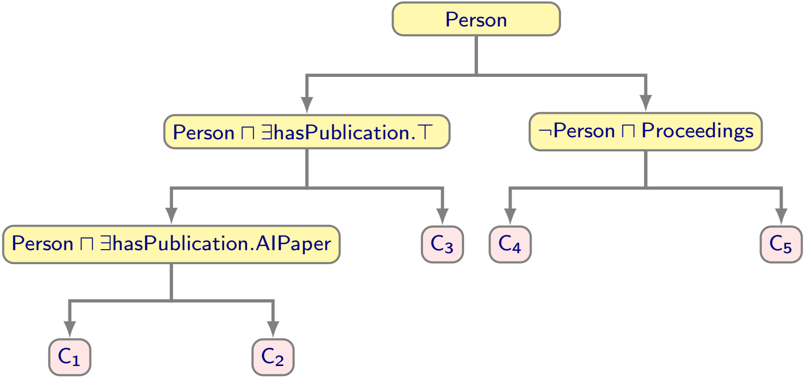
A binary logical tree where

- a leaf node stands for a cluster of individuals C
- each inner node contains a description D (over the signature of $\mathcal{K}$)
- each departing edge corresponds to positive (left) and negative (right) examples of D

[6] Implemented system publicly available at https://github.com/Giuseppe-Rizzo/TCTnew

Example of TCT

Given $\mathsf{I} \subseteq \mathrm{Ind}(\mathcal{A})$, an example of TCT describing the AI research community



Collecting Disjointness Axioms

Given a TCT T:

Step I:

- Traverse the T to collect the concept descriptions describing the clusters at the leaves
- A set of concepts CS is obtained

Step II:

- A set of candidate axioms A is generated from CS:
  - an axiom $D \sqsubseteq \neg E$ ($D, E \in$ CS) is generated if
    - $D \not\equiv E$ (or $D \not\sqsubseteq E$ or viceversa - reasoner needed)
    - $E \sqsubseteq \neg D$ has not been generated

Conclusions

Conclusions:

- Injecting semantics and exploiting reasoning capabilities may improve the effectiveness of ML solutions for KG
    - Framework for injecting BK into KGE models
    - Solution for injecting semantics when computing a-posteriori explanations to link predictions
- Supplementing schema level information (disjointness axioms) needed

Next Research Challenges:

- Extend the framework for injecting BK to more complex KGE models
- Learning disjointness axioms from KGE
- Develop a standardized evaluation protocol for evaluating explanations
- New solutions required for enhancing LLMs with KG semantics

Thank you

References

Barile, R., d'Amato, C., and Fanizzi, N. (2024). Explanation of link predictions on knowledge graphs via levelwise filtering and graph summarization. In A. Meroño-Peñuela, A. Dimou, R. Troncy, O. Hartig, M. Acosta, M. Alam, H. Paulheim, and P. Lisena, editors, *The Semantic Web - 21st International Conference, ESWC 2024, Hersonissos, Crete, Greece, May 26-30, 2024, Proceedings, Part I*, volume 14664 of *Lecture Notes in Computer Science*, pages 180–198. Springer.

Cai, H., Zheng, V. W., and Chang, K. C.-C. (2018). A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, **30**(9), 1616–1637.

d'Amato, C., Masella, P., and Fanizzi, N. (2021a). An approach based on semantic similarity to explaining link predictions on knowledge graphs. In J. He, R. Unland, E. S. Jr., X. Tao, H. Purohit, W. van den Heuvel, J. Yearwood, and J. Cao, editors, *WI-IAT '21: IEEE/WIC/ACM International Conference on Web Intelligence, Melbourne VIC Australia, December 14 - 17, 2021*, pages 170–177. ACM.

d'Amato, C., Quatraro, N. F., and Fanizzi, N. (2021b). Embedding models for knowledge graphs induced by clusters of relations and background knowledge. In N. Katzouris and A. Artikis, editors, *Inductive Logic Programming - 30th International Conference, ILP 2021, Virtual Event, October 25-27, 2021, Proceedings*, volume 13191 of *Lecture Notes in Computer Science*, pages 1–16. Springer.

References [...cont.]

d'Amato, C., Quatraro, N. F., and Fanizzi, N. (2021c). Injecting background knowledge into embedding models for predictive tasks on knowledge graphs. In R. Verborgh, K. Hose, H. Paulheim, P. Champin, M. Maleshkova, Ó. Corcho, P. Ristoski, and M. Alam, editors, *The Semantic Web - 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings*, volume 12731 of *Lecture Notes in Computer Science*, pages 441–457. Springer.

d'Amato, C., Benedetti, F., and Fanizzi, N. (2023). Efficient explanation of predictions on dl knowledge graphs through an enhanced similarity search. In O. Kutz, A. Ozaki, and C. Lutz, editors, *DL 2023: 36th International Workshop on Description Logics*. CEUR.

Fleischhacker, D. and Völker, J. (2011). Inductive learning of disjointness axioms. In R. Meersman and et. al., editors, *On the Move to Meaningful Internet Systems: OTM 2011 - Confederated International Conferences: CoopIS, DOA-SVI, and ODBASE 2011, Proceedings, Part II*, volume 7045 of *Lecture Notes in Computer Science*, pages 680–697. Springer.

Guo, S., Wang, Q., Wang, L., Wang, B., and Guo, L. (2016). Jointly embedding knowledge graphs and logical rules. In J. Su, X. Carreras, and K. Duh, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 192–202. The Association for Computational Linguistics.

References [. . . cont.]

Minervini, P., Demeester, T., Rocktäschel, T., and Riedel, S. (2017a). Adversarial sets for regularising neural link predictors. In G. Elidan, K. Kersting, and A. Ihler, editors, *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence, UAI 2017, Sydney, Australia, August 11-15, 2017*. AUAI Press.

Minervini, P., Costabello, L., Muñoz, E., Novácek, V., and Vandenbussche, P. (2017b). Regularizing knowledge graph embeddings via equivalence and inversion axioms. In M. Ceci, J. Hollmén, L. Todorovski, C. Vens, and S. Dzeroski, editors, *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18-22, 2017, Proceedings, Part I*, volume 10534 of *Lecture Notes in Computer Science*, pages 668–683. Springer.

Rizzo, G., d'Amato, C., and Fanizzi, N. (2021). An unsupervised approach to disjointness learning based on terminological cluster trees. *Semantic Web*, **12**(3), 423–447.

Rossi, A., Firmani, D., Merialdo, P., and Teofili, T. (2022). Explaining link prediction systems based on knowledge graph embeddings. In Z. G. Ives, A. Bonifati, and A. E. Abbadi, editors, *SIGMOD '22: International Conference on Management of Data, Philadelphia, PA, USA, June 12 - 17, 2022*, pages 2062–2075. ACM.

Völker, J., Fleischhacker, D., and Stuckenschmidt, H. (2015). Automatic acquisition of class disjointness. *Journal of Web Semantics*, **35**, 124–139.

References [...cont.]

Wang, T. D., Parsia, B., and Hendler, J. A. (2006). A survey of the web ontology landscape. In I. F. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo, editors, *The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings*, volume 4273 of *Lecture Notes in Computer Science*, pages 682–694. Springer.

Zhang, W., Paudel, B., Zhang, W., Bernstein, A., and Chen, H. (2019). Interaction embeddings for prediction and explanation in knowledge graphs. In *Proceedings of WSDM 2019*, pages 96–104. ACM.

Inducing a TCT

Given the set of individuals I and $\top$ concept

Divide-and-conquere approach adopted

- **Base Case:** test the stopCondition
  - the cohesion of the cluster I exceeds a threshold ν
    - distance between medoids below a threshold ν
- **Recursive Step** (stopCondition does not hold):
  - a set S of refinements of the current (parent) description C generated
  - the bestConcept $E^* \in$ S is selected and installed as current node
    - the one showing the best cluster separation $\Leftrightarrow$ with max distance between the medoids of its positive P and negative N individuals
  - I is split in:
    - $I_{left} \subseteq$ I $\leftrightarrow$ individuals with the smallest distance wrt the medoid of P
    - $I_{right} \subseteq$ I $\leftrightarrow$ individuals with the smallest distance wrt the medoid of N
    - reasoner employed for collecting P and N

Note: Number of clusters not required - obtained from data distribution

Distance measure between individuals adopted for TCT

Distance Function (adapted from [d'Amato et al.@ESWC2008]):

$$d_n^{\mathcal{C}} : \mathrm{Ind}(\mathcal{A}) \times \mathrm{Ind}(\mathcal{A}) \to [0, 1]$$

$$d_n^{\mathcal{C}}(a, b) = \left[\sum_{i=1}^{m} w_i \left[1 - \pi_i(a)\pi_i(b)\right]^n \right]^{1/n}$$

Context: a set of atomic concepts $\mathcal{C} = \{B_1, B_2, \ldots, B_m\}$

Projection Function:

$$\forall\, a \in \mathrm{Ind}(\mathcal{A}) \qquad \pi_i(a) = \begin{cases} 1 & \text{if } \mathcal{K} \models B_i(a) \\ 0 & \text{if } \mathcal{K} \models \neg B_i(a) \\ 0.5 & \text{otherwise} \end{cases}$$