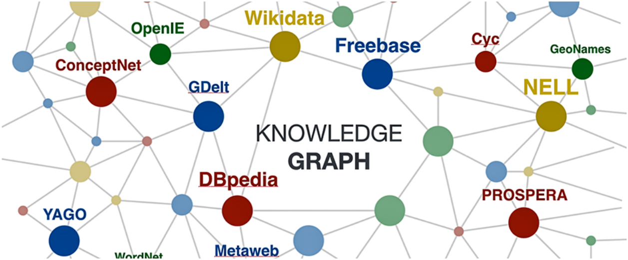
Link Prediction and Explanations on Knowledge Graphs: Issues and Perspectives

Claudia d'Amato

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

SEmantic web SeminAr MontpEllier Research Day
(SESAME Research Day 2025)

October 20, 2025 - Montpellier, France



[1]

Open KG
online with content freely accessible

- BabelNet
- DBpedia
- Freebase
- Wikidata
- YAGO
- ....

Enterprise KG
for commercial usage

- Google
- Amazon
- Facebook
- LinkedIn
- Microsoft
- ....

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

Applications

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

Research Fields

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

Knowledge Graph: Definition [Hogan *et al.*, 2021]

A graph of data intended to convey knowledge of the real world

- conforming to a graph-based data model
- nodes represent entities of interest
- edges represent different relations between these entities
- data graph potentially enhanced with schema

KGs: Main Features

- ontologies employed to define and reason about the semantics of nodes and edges
- RDF, RDFS, OWL representation languages largely adopted
- grounded on the Open World Assumption (OWA)
- very large data collections
- suffer of incompleteness and noise
  - since often result from a complex building process

Machine Learning
&
Knowledge Graphs

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

Issues and Why Semantics is Needed

Numeric-based methods mostly adopted

- highly scalable on KG volume
- opaque / black box
- no background knowledge and reasoning capabilities exploited
    - only factual information considered

$$\Downarrow$$

Knowledge within KG

- only partially considered
- and not always in a fully correct way (negatives)

ML as input to KG

(KG Refinement at Assertion Level: Link Prediction)

(Explanation of Link Predictions on KGs)

Tackling KG Semantics

ML as input to KG

(**KG Refinement: Link Prediction**)

(Explanation of Link Predictions on KGs)

Tackling KG Semantics

Incompleteness and **noise**

⇓

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

⇓

New scalable Machine Learning methods

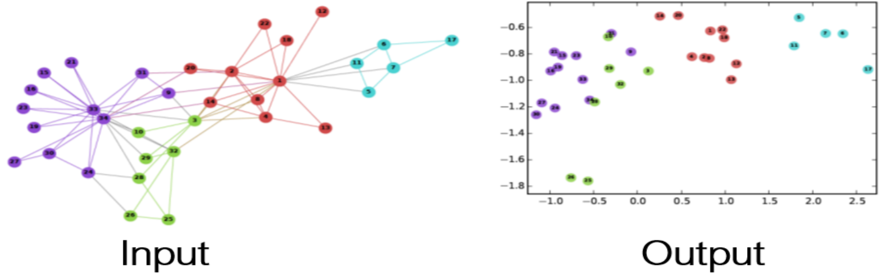
- grounded on numeric-based approaches
  - KG vector embedding models (KGE) largely investigated [Cai *et al.*, 2018]

ML/KGE for KGs: Issues

- CWA (or LCWA) mostly adopted vs. OWA
- schema level information and reasoning capabilities almost disregarded
- black box models ⇒ 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...

KGE methods differ in their main building blocks [Ji *et al.*, 2020]:

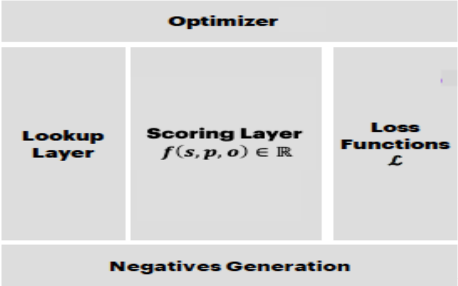
- **Representation Space**
  - where represeting relations and entities
  - point-wise, complex, discrete, Gaussian, manifold, etc.
- **Encoding Model**
  - for representing and learning relational interactions
  - linear, factorization, neural models, etc.
- **Scoring Function**
  - for measuring the plausibility of factual triples
  - based on distance, energy, semantic matching, other criteria
- **Auxiliary Information**
  - to be incorporated into the embedding methods
  - text, type, images

...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 Models:

Injecting Semantics

Enhancing KGE by Injecting Background Knowledge (BK) [d'Amato *et al.*, 2021c,b] [4] [5]

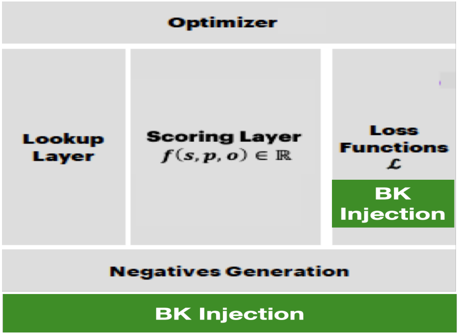
By two components:

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`.



[4] C. d'Amato, N. F. Quatraro, N. Fanizzi: Injecting Background Knowledge into Embedding Models for Predictive Tasks on Knowledge Graphs. ESWC 2021: 441-457 (2021)

[5] C. d'Amato, N. F. Quatraro, N. Fanizzi: Embedding Models for Knowledge Graphs Induced by Clusters of Relations and Background Knowledge. IJCLR 2021 Proceedings (2021)

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

- BoxEL KGE model the logical structure of ABox and TBox axioms in $\mathcal{EL}^{++}$ Description Logics [Xiong *et al.*, 2022]

An approach to learn embeddings exploiting BK [d'Amato *et al.*, 2021c,b] [6] [7]

TRANSOWL

TRANSROWL TRANSROWL^R

TransE
[Bordes *et al.*, 2013]

TransR
[Lin *et al.*, 2015]

Directly applicable to KGs

Could be applied to more complex KG embedding methods with additional formalization

[6] C. d'Amato, N. F. Quatraro, N. Fanizzi: Injecting Background Knowledge into Embedding Models for Predictive Tasks on Knowledge Graphs. ESWC 2021: 441-457 (2021)

[7] C. d'Amato, N. F. Quatraro, N. Fanizzi: Embedding Models for Knowledge Graphs Induced by Clusters of Relations and Background Knowledge. IJCLR 2021 Proceedings (2021)

TransOWL, TransROWL [d'Amato *et al.*, 2021c] [8]

- 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

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

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

[8] C. d'Amato, N. F. Quatraro, N. Fanizzi: Injecting Background Knowledge into Embedding Models for Predictive Tasks on Knowledge Graphs. ESWC 2021: 441-457 (2021)

Alternative Approach: TransROWL^R [d'Amato *et al.*, 2021c] [9]

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

$$
\begin{aligned}
L \;=\; & \sum_{\substack{\langle h,r,t\rangle \in \Delta \\ \langle h',r',t'\rangle \in \Delta'}} [\gamma + f'_r(h,t) - f'_r(h',t')]_+ \\
& + \lambda_1 \sum_{r \equiv q^- \in \mathcal{T}_{\textbf{inverseOf}}} \|r + q\| \; + \lambda_2 \sum_{r \equiv q^- \in \mathcal{T}_{\textbf{inverseOf}}} \|M_r - M_q\| \\
& + \lambda_3 \sum_{r \equiv p \in \mathcal{T}_{\textbf{equivProp}}} \|r - p\| \; + \lambda_4 \sum_{r \equiv p \in \mathcal{T}_{\textbf{equivProp}}} \|M_r - M_p\| \\
& + \lambda_5 \sum_{e' \equiv e'' \in \mathcal{T}_{\textbf{equivClass}}} \|e' - e''\| \; + \lambda_6 \sum_{s' \subseteq s'' \in \mathcal{T}_{\textbf{subClass}}} \|1 - \beta - (s' - s'')\|
\end{aligned}
$$

[9] C. d'Amato, N. F. Quatraro, N. Fanizzi: Injecting Background Knowledge into Embedding Models for Predictive Tasks on Knowledge Graphs. ESWC 2021: 441-457 (2021)

Lesson Learnt from Experiments

Goal: Assessing the benefit of exploiting BK

- Comparing[10] TransOWL, TransROWL, TransROWL^R over to the original models TransE and TransR as a baseline

KGs adopted:

KG	#Triples	#Entities	#Relationships
DBpedia15K	180000	12800	278
DBpedia100K	600000	100000	321
DBpediaYAGO	290000	88000	316
NELL[11]	150000	68000	272

- **Outcomes:**
  - TransROWL best performing method, in most of the cases
  - TransROWL slightly outperfroming TransROWL^R

Next Challenges

- assess the impact of more fine grained (probabilistic) solutions for generating negatives on KGE performance
- extend the framework to more complex KGE models

[10] All methods implemented as publicly available systems https://github.com/Keehl-Mihael/TransROWL-HRS

[11] equivalentClass and equivalentProperty missing; limited number of typeOf-triples; abundance of subClassOf-triples

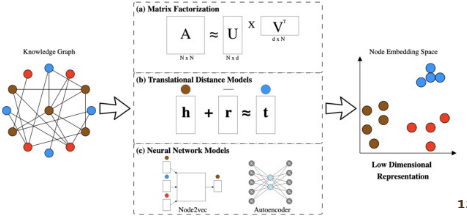
ML as input to KG

(KG Refinement: Link Prediction)

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

Tackling KG Semantics

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

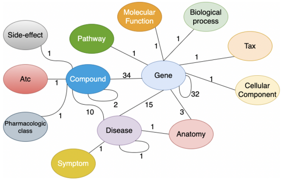


[12]

This may affects:

- the interpretability of the models
- the explainability of the results
- and possibly the trustworthiness of results

DRKG – Drug Repurposing Knowledge Graph

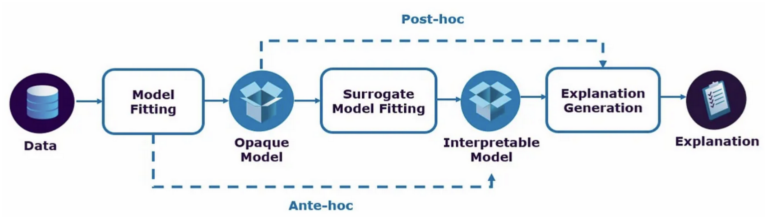


[13]

[12] 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

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

Computing Explanations: Approaches



Explainable Machine Learning Approaches [14]

Ante-hoc approach

- explain the model itself
- model-dependent

Post-hoc approach

- explain the result providing evidence from the data
- model-agnostic

[14] Picture from https://medium.com/@sparsha.stars/explainable-artificial-intelligence-technical-perspective-part-1-6eec91b4cc60

Post-hoc Explanations of Link Prediction (LP)

Post-hoc explanation methods are model agnostic. They find explanation(s) based on the output and the model input, independently on the KGE adopted

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⟩.

Post-hoc Explanations of LP: Current 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

- search for a path linking the subject s and object o of a predicted triple $\langle s, r, o\rangle$
  - **Max lenght 2** $\rightarrow$ six types of paths possible:
    Length 1: $P_1 = \{\langle s, r_q, o\rangle\}$, $P_2 = \{\langle o, r_q, s\rangle\}$
    Length 2: $P_3 = \{\langle e', r_q, s\rangle, \langle e', r', o\rangle\}$, $P_4 = \{\langle e', r_q, s\rangle, \langle o, r', e'\rangle\}$,
    $P_5 = \{\langle s, r_q, e'\rangle, \langle e', r', o\rangle\}$, $P_6 = \{\langle s, r_q, e'\rangle, \langle o, r', e'\rangle\}$,
    where r_q similar to r, r' any other relationship, e' any other entity;
- search driven by similarities between relation/entity embeddings via Euclidean distance
- structural comparisons with other paths in the KG to reinforce the reliability of the explanation found (referred to as support)

Exploiting Semantics for Providing Explanations to Link Predictions on KGs

SemanticCrossE

SemanticCrossE provide semantic-based explanations for LP on KGs

[d'Amato *et al.*, 2021a] [15]

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:

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

where $\mathbf{e}$ 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.

[15] C. d'Amato, P. Masella, N. Fanizzi: An Approach Based on Semantic Similarity to Explaining Link Predictions on Knowledge Graphs. In IEEE/WIC/ACM Internat. Conf. on Web Intelligence (WI-IAT 2021) pp. 170-177 (2021)

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] [16]
- Efficient computation via a preliminary clustering phase [d'Amato *et al.*, 2023] [17]

[16] C. d'Amato, P. Masella, N. Fanizzi: An Approach Based on Semantic Similarity to Explaining Link Predictions on Knowledge Graphs. In IEEE/WIC/ACM Internat. Conf. on Web Intelligence (WI-IAT 2021) pp. 170-177 (2021)

[17] C. d'Amato, F. Benedetti, N. Fanizzi. Efficient Explanation of Predictions on DL Knowledge Graphs through Enhanced Similarity Search. The 36th International Workshop on Description Logics. Vol. 3515, CEUR (2023)

Example (Computing the semantic Score)

Let us suppose $e := Bob \quad e' := Kathy$

$Cl(e) = \mathtt{Student}$ and $Cl(e') = \mathtt{Student} \sqcup \mathtt{Worker}$. Then:

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

Similarly, the semantic Score for relations can be computed by considering their domains and/or ranges, that are ultimately class expressions

Lesson Learnt from Experiments

Goal: Establishing the **impact of an added semantic component when computing explanations** of link prediction results

- Comparing ApproxSemanticCrossE and cosineCrossE to CrossE as baseline[18]
- CrossE adopted for the preliminary link prediction phase

KG	#Triples	#Entities	#Relationships
FB15k-237	310116	14541	237
WN18	151442	40943	18
DBpedia15K	183218	12862	279

Results:

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

Next Challenges

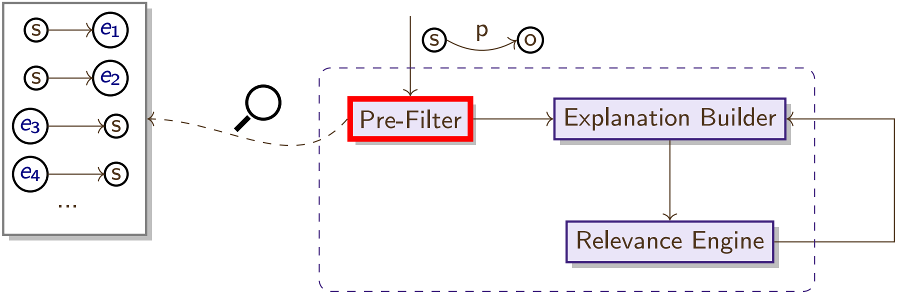
- Taking into account additional semantics in KGs (e.g. transitivity, symmetry etc.)

[18] Code and datasets publicly available https://github.com/pierulohacker/SemanticCrossE/tree/master/explanation

KELPIE++

KELPIE: High-Level Architecture...

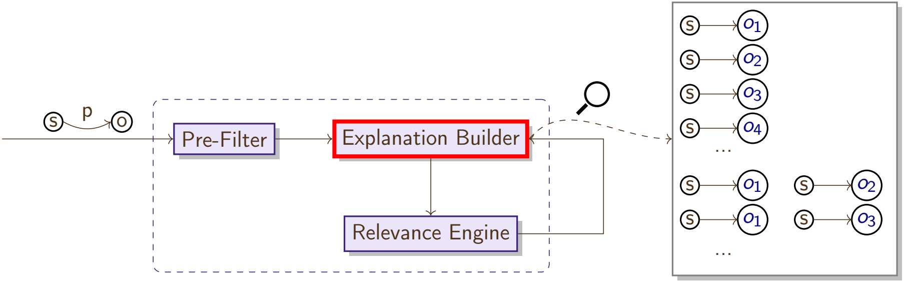
KELPIE [Rossi *et al.*, 2022]: generates necessary and sufficient (path) conditions and an articulated new evaluation protocol



Pre-Filter:

- extracts the triples featuring s (either as subject or object)
- selects the k ones most topologically related to the prediction - structural similarity
- ignores semantic relations between facts

...KELPIE: High-Level Architecture



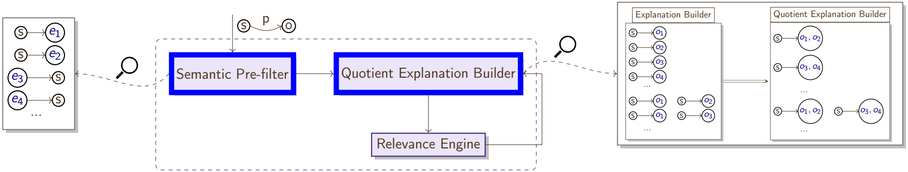
Explanation-Builder:

- combines the pre-filtered triples into candidate explanations
- leads the number of candidates to explode combinatorially

Relevance Engine: computes the relevance of candidate explanations

KELPIE++: Main features...

[Barile *et al.*, 2024b] [19]



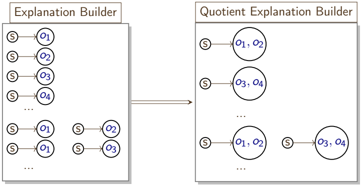
- Introduce Semantic Pre-filter by computing the Semantic Score
  - **Goal:** Improving the effectiveness/appropriateness of pre-filtered triples
- Introduce Quotient Explanation Builder of the pre-filtered triples
  - **Goal:** Improving the efficiency of the Explanation Builder

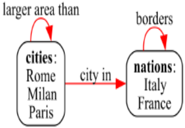
[19] R. Barile, C. d'Amato, N. Fanizzi. Explanation of link predictions on knowledge graphs via levelwise filtering and graph summarization. The Semantic Web - 21st International Conference, (ESWC 2024) Proceedings, Part I, volume 14664, pp. 26-30, LNCS Springer (2024)

...KELPIE++: Main features

[Barile *et al.*, 2024b] [20]

- Introduce Quotient Explanation Builder of the pre-filtered triples





  - three different quotient graphs formulated
    - Simulation, Bisimulation, Depth-1 Bisimulation
    - exploit semantic information on types
    - offer different levels of granularity
  - combines quotient triples into candidate explanations
- Provides explanations at different level of details

[20] R. Barile, C. d'Amato, N. Fanizzi. Explanation of link predictions on knowledge graphs via levelwise filtering and graph summarization. The Semantic Web - 21st International Conference, (ESWC 2024) Proceedings, Part I, volume 14664, pp. 26-30, LNCS Springer (2024)

KELPIE++: explanations at different level of details

prediction: ⟨*Gabourey Sidibe*, *nationality*, *United States*⟩

dataset: YAGO4-20, **model**: ComplEx

KELPIE++ - Simulation quotient	
quotient triples	**entities**
⟨*Movie*, *actor*, *Gabourey Sidibe*⟩	*Casting By, Top Five, Seven Psychopaths, White Bird in a Blizzard, Tower Heist, Precious*
⟨*TVSeries*, *actor*, *Gabourey Sidibe*⟩	*Empire (2015 TV Series)*

KELPIE
⟨*Empire (2015 TV Series)*, *actor*, *Gabourey Sidibe*⟩ ⟨*Top Five*, *actor*, *Gabourey Sidibe*⟩ ⟨*White Bird in a Blizzard*, *actor*, *Gabourey Sidibe*⟩

Lesson Learnt from Experiments

Goal: Establishing the **impact of the Semantic Pre-Filter and Quotient Explanation Builder components when computing explanations** of link prediction results

- Comparing KELPIE++[21] to KELPIE as baseline
- TransE, ConvE, ComplEx adopted for the preliminary link prediction phase

KG	#Triples	#Entities	#Relationships
DBpedia50K	34289	24620	351
DBpedia100K	636802	98776	464
YAGO4-20	624580	96910	70

Results:

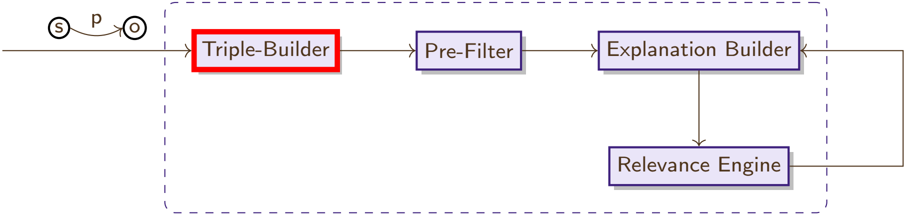
- KELPIE++ outperformed KELPIE in terms of
  - effectiveness (Δ HITS@1, Δ MRR)
  - efficiency (number of calls to the relevance engine)
  - particularly with **simulation** quotient graph

[21] Code and datasets publicly available https://github.com/rbarile17/kelpiePP

IMAGINE

IMAGINE: High-Level Architecture

The first solution generating additive counterfactual explanations [Barile *et al.*, 2024a] [22]
new facts to be used for providing explanations
motivated by KG incompleteness



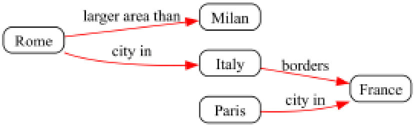
Triple-Builder: generates a set of additional triples featuring s

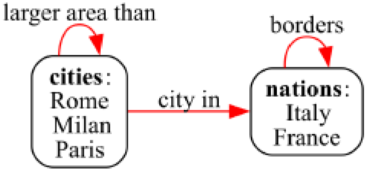
- compute the quotient graph $\mathcal{Q}_{train}$
- identifies the quotient node S in $\mathcal{Q}_{train}$ containing the prediction's subject s
- given the quotient triple $\langle S, r, O\rangle$, output the set $\mathcal{A}^s$ given by the triples $\langle s, r, o\rangle$ for all o in O (viceversa if s is in O)

[22] R. Barile, C. d'Amato, N. Fanizzi. Additive Counterfactuals for Explaining Link Predictions on Knowledge Graphs. Knowledge Engineering and Knowledge Management - 24th International Conference, (EKAW) 2024 Proceedings, Volume 15370, pp. 346–363, LNCS Springer (2024)

IMAGINE: Triple Builder

Predicted triple: $\langle$*Milan*, *located in*, *Italy*$\rangle$





For the quotient triple $\langle\{$*Milan, Rome, Paris*$\}$, *city in*, $\{$*Italy, France*$\}\rangle$

Generated additional triples: $\{\langle$*Milan*, *city in*, *Italy*$\rangle$, $\langle$*Milan*, *city in*, *France*$\rangle\}$

Lesson Learnt from Experiments

Goal: Establishing the **effectiveness of an additive counterfactual based approach when computing explanations** of link prediction results

- Evaluate improvements in solving link prediction tasks
- Comparing IMAGINE[23] to an adapted version KELPIE++ and KELPIE as baseline
- TransE, ConvE, ComplEx adopted for the preliminary link prediction phase

KG	#Triples	#Entities	#Relationships
DBpedia50K	34289	24620	351
DBpedia100K	636802	98776	464
YAGO4-20	624580	96910	70

Results:

- IMAGINE outperformed KELPIE++ and KELPIE in almost all cases
- IMAGINE suitable for complementing subtractive counterfactual based approaches like KELPIE++ and KELPIE

Next Challenge

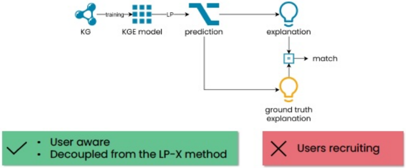
- Coming up with a unified solution integrating subtractive and additive counterfactual approaches the prediction

[23] Code and datasets publicly available https://github.com/rbarile17/imagine

Evaluating Post-Hoc Explanations of Link Prediction on KGs

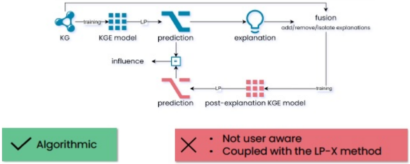
Evaluating Explanations: State of the Art and Issues

Approach 1: compare the computed explanations with ground truth



- Costly
  - needing recruiting users
- Hardly repeatable on new KG
  - ground truth missing

Approach 2: assess the impact-influence of explanations on solving the same LP task



- Different protocols adopted
  - Recall/Support vs. ΔHITS@1, ΔMRR
- Hard to compare different explanation solutions

Evaluation Explanation Protocol: Requirements

- **User-aware:** users can assess the utility of explanations
- **Algorithmic:** the evaluation of explanations can be automatized
  - user studies are costly
- **General:** explanations coming from different solutions can be compared

LP-DIXIT [Barile *et al.*, 2025] [24]...

LP-DIXIT - the only protocol for evaluating explanations that is:

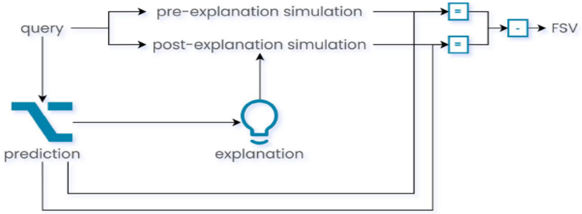
- user-aware
- fully algorithmic
- decoupled from the explanation method
- allows comparative analysis of different explanation methods - previously not possible

[24] R. Barile, C. d'Amato, N. Fanizzi. LP-DIXIT: Evaluating Explanations for Link Predictions on Knowledge Graphs using Large Language Models. Proceedings of the ACM on Web Conference 2025, (WWW 2025), pp. 4034–4042, ACM (2025)

...LP-DIXIT... [Barile *et al.*, 2025] [25]

LP-DIXIT grounded on forward simulatability cognitive theory [Hoffman *et al.*, 2023]

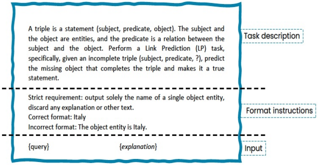
- a prediction is understandable if it is simulatable
- a prediction is simulatable if a verifier can guess its output given the same input
  - Forward Simulatability Variation (FSV) computed



Large Language Models (LLMs) adopted for mimic the user(s) as verifier(s)

[25] R. Barile, C. d'Amato, N. Fanizzi. LP-DIXIT: Evaluating Explanations for Link Predictions on Knowledge Graphs using Large Language Models. Proceedings of the ACM on Web Conference 2025, (WWW 2025), pp. 4034–4042, ACM (2025)

...LP-DIXIT [Barile *et al.*, 2025] [27]



LP-DIXIT$_Z$: zero-shot without output constraint

task description	format instructions	input

LP-DIXIT$_O$: zero-shot with output constraint

task description	format instructions	candidate entities	input

- avoids that the LLMs output is not an entity in the KG.
- *candidate entities*: top-k ranked according to the LP method.

LP-DIXIT$_F$: few-shot without output constraint

task description	format instructions	demonstrations	input

- LLMs are not trained for LP on KGs
- demonstrations: triples with the same predicate of Q

LP-DIXIT$_F$: few-shot without output constraint

task description	format instructions	candidate entities	demonstrations	input

LP-DIXIT validated by:

- assessing its agreement with ground truth user judgements
- performing comparative study of different explanation methods - previously not possible
- delivering GRainsaCK: open source library for benchmarking explanations [26]

Next Challenges

- Enhance LP-DIXIT with a more fine-grained score (FSV)
- Conduct a dedicated user study in comparison to LP-DIXIT with users

[26] Code, datasets and documentation publicly available at https://github.com/rbarile17/grainsack

[27] R. Barile, C. d'Amato, N. Fanizzi. LP-DIXIT: Evaluating Explanations for Link Predictions on Knowledge Graphs using Large Language Models. Proceedings of the ACM on Web Conference 2025, (WWW 2025), pp. 4034–4042, ACM (2025)

A Broader Perspective on the Evaluation of Explanations

Assessment of Explanations: Multiple Dimensions

In XAI, explanations assessed across various dimensions [Schwalbe and Finzel, 2024], e.g.

- explanation impact on predictive task performance
- stability of the explanation wrt to changes in the underlying data
- usefulness of the explanation to the user
- overall explanation clarity / understandability
- ...

$\Downarrow$ $\Downarrow$

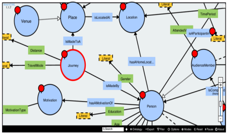
Method Perspective

Difficult identifying their trade-off and benchmarking

User Perspective

Difficult to find the most effective solutions for certain dimension(s)

Idea: Ontologies for Systematizing and Automatizing Evaluation of Explanations



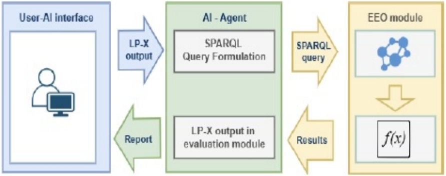
Shared conceptualization needed for:

- systematize and modeling
  - explanation approaches and methods
  - evaluation dimensions and corresponding metrics
- easily expand to novel methods and evaluation dimensions
- enabling automation of evaluation of explanations

Architectural View [Balbi *et al.*, 2025] [28]

Given a (LP) problem and explanations as input, allows:

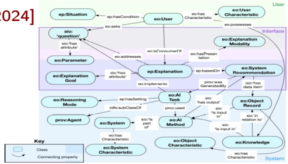
- querying for different explanation evaluation dimension(s) (for an approach)
- collecting evaluation methods and metrics for a dimension
- automating the execution of explanation evaluation protocols
- querying for methods supporting certain evaluation dimension(s)



[28] L. Balbi, F. Bindt, K. Breitenfelder, R. Campi, J. De Smet, C. d'Amato. A Semantic Framework for Evaluating Post-hoc Explanations in Link Prediction. Proceedings of the 1st Workshop on Explainable AI, Knowledge Representation, and Knowledge Graphs 2025 (XAI-KRKG@ECAI2025), CEUR (To appear)

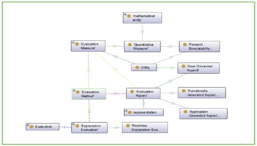
Extended Explanation Ontology (EEO) [Balbi *et al.*, 2025] [30]

EEO [29] extends the Explanation Ontology (EO) [Chari *et al.*, 2024]



- describe XAI methods in terms of
  - inputs, outputs, and underlying data
- lacks any formalization of their evaluation

EEO add



- classes for `Explanation Evaluation`, `Measure` and `Quantitative Measure`
- subclasses, object properties for `dimensions, measures and metrics` identified at SOTA [Schwalbe and Finzel, 2024]

EEO queried via SPARQL (for answering questions, e.g.)

- which measure(s) are available for an evaluation dimension?
- which method(s) should be used to evaluate explanations on a given dimension?

Proof-of-concept: instantiation of EEO with LP-DIXIT and querying

[29] EEO publicly available at https://doi.org/10.5281/zenodo.15658539

[30] L. Balbi, F. Bindt, K. Breitenfelder, R. Campi, J. De Smet, C. d'Amato. A Semantic Framework for Evaluating Post-hoc Explanations in Link Prediction. Proceedings of the 1st Workshop on Explainable AI, Knowledge Representation, and Knowledge Graphs 2025 (XAI-KRKG@ECAI2025), CEUR (To appear)

Conclusions

Conclusions:

- Improved ML and explanation solutions for KGs can be formalized taking into account semantics and reasoning capabilities
  - Framework for injecting semantics into KGE models
  - Solutions for injecting semantics when computing post-hoc explanations to link predictions
- A unified and standardized explanation evaluation protocol targeting multiple dimension is needed

Next Research Challenges:

- Extend the framework for injecting semantics to complex KGE models
- Empower semantic explanation with additional schema axioms
- Computing semantic explanation via an integrated additive and subtractive couterfactual approach
- Fully operationalize the semantic framework for evaluating explanations

Thank you

Roberto Barile

Nicola Flavio Quatraro

Pierpaolo Masella

Nicola Fanizzi

Laura Balbi

Felix Bindt

Katja Breitenfelder

Riccardo Campi

Jitse De Smet

References

Balbi, L., Bindt, F., Breitenfelder, K., Campi, R., Smet, J. D., and d'Amato, C., editors (2025). *A Semantic Framework for Evaluating Post-hoc Explanations in Link Prediction*, CEUR Workshop Proceedings. CEUR-WS.org. to appear.

Barile, R., d'Amato, C., and Fanizzi, N. (2024a). Additive counterfactuals for explaining link predictions on knowledge graphs. In *he 24th International Conference on Knowledge Engineering and Knowledge Management, EKAW 2024, Amsterdam, Netherlands, November 26-28, 2024, Proceedings*, Lecture Notes in Computer Science. Springer.

Barile, R., d'Amato, C., and Fanizzi, N. (2024b). 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.

Barile, R., d'Amato, C., and Fanizzi, N. (2025). LP-DIXIT: evaluating explanations for link predictions on knowledge graphs using large language models. In G. Long, M. Blumestein, Y. Chang, L. Lewin-Eytan, Z. H. Huang, and E. Yom-Tov, editors, *Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025- 2 May 2025*, pages 4034–4042. ACM.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *NIPS 2013*. Curran Assoc., Inc.

References

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.

Chari, S., Seneviratne, O., Ghalwash, M. F., Shirai, S., Gruen, D. M., Meyer, P., Chakraborty, P., and McGuinness, D. L. (2024). Explanation ontology: A general-purpose, semantic representation for supporting user-centered explanations. *Semantic Web*, **15**(4), 959–989.

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

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.

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.

Hoffman, R. R., Mueller, S. T., Klein, G., and Litman, J. (2023). Measures for explainable AI: explanation goodness, user satisfaction, mental models, curiosity, trust, and human-ai performance. *Frontiers Comput. Sci.*, **5**.

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., and Zimmermann, A. (2021). *Knowledge Graphs*. Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool Publishers.

References

Ji, S., Pan, S., Cambria, E., Marttinen, P., and Yu, P. S. (2020). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, **33**, 494–514.

Lin, Y., Liu, Z., Sun, M., Liu, Y., and Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In *AAAI 2015 Proceedings*, pages 2181–2187. AAAI Press.

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.

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.

References

Schwalbe, G. and Finzel, B. (2024). A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts. *Data Mining and Knowledge Discovery*, **38**(5), 3043–3101.

Xiong, B., Potyka, N., Tran, T.-K., Nayyeri, M., and Staab, S. (2022). Faithful embeddings for $\mathcal{EL}^{++}$ knowledge bases. In U. Sattler, A. Hogan, M. Keet, V. Presutti, J. P. A. Almeida, H. Takeda, P. Monnin, G. Pirrò, and C. d'Amato, editors, *The Semantic Web – ISWC 2022*, pages 22–38. Springer International Publishing.

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.