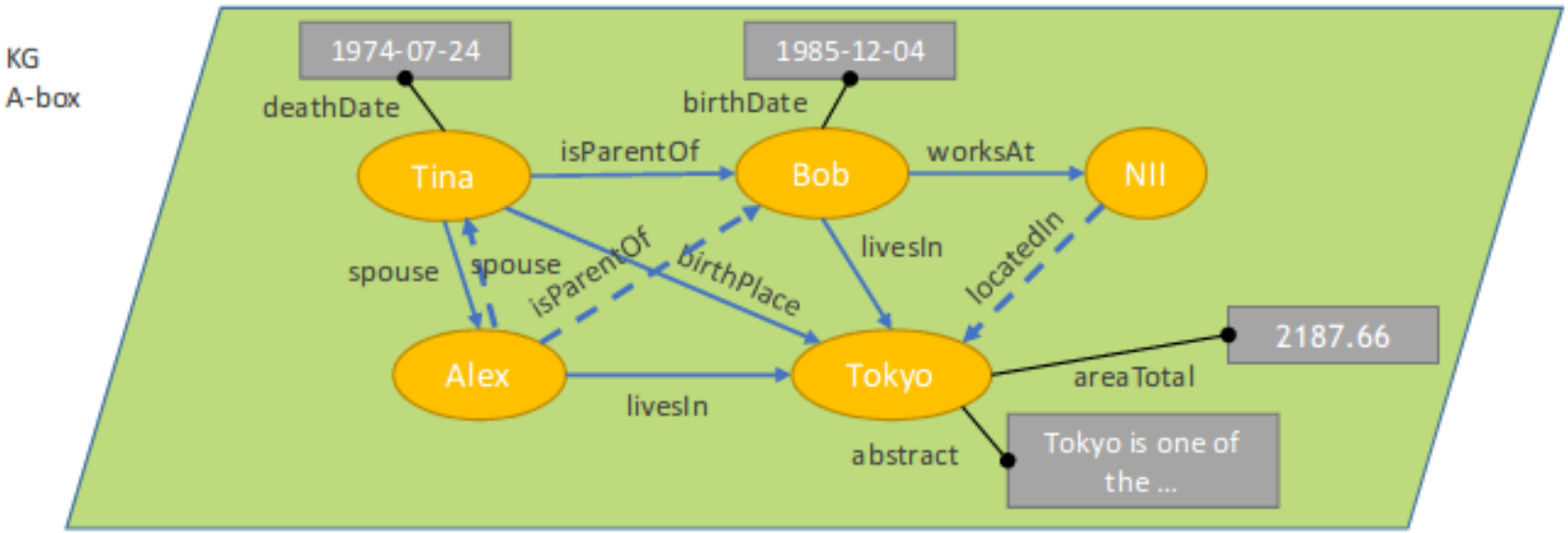
Guiding Graph Embeddings using Path-Ranking Methods for Error Detection in noisy Knowledge Graphs

Konstantinos Bougiatiotis, Romanos Fasoulis, Fotis Aisopos, Anastasios Nentidis, Georgios Paliouras

National Centre of Scientific Research (NCSR)
Demokritos

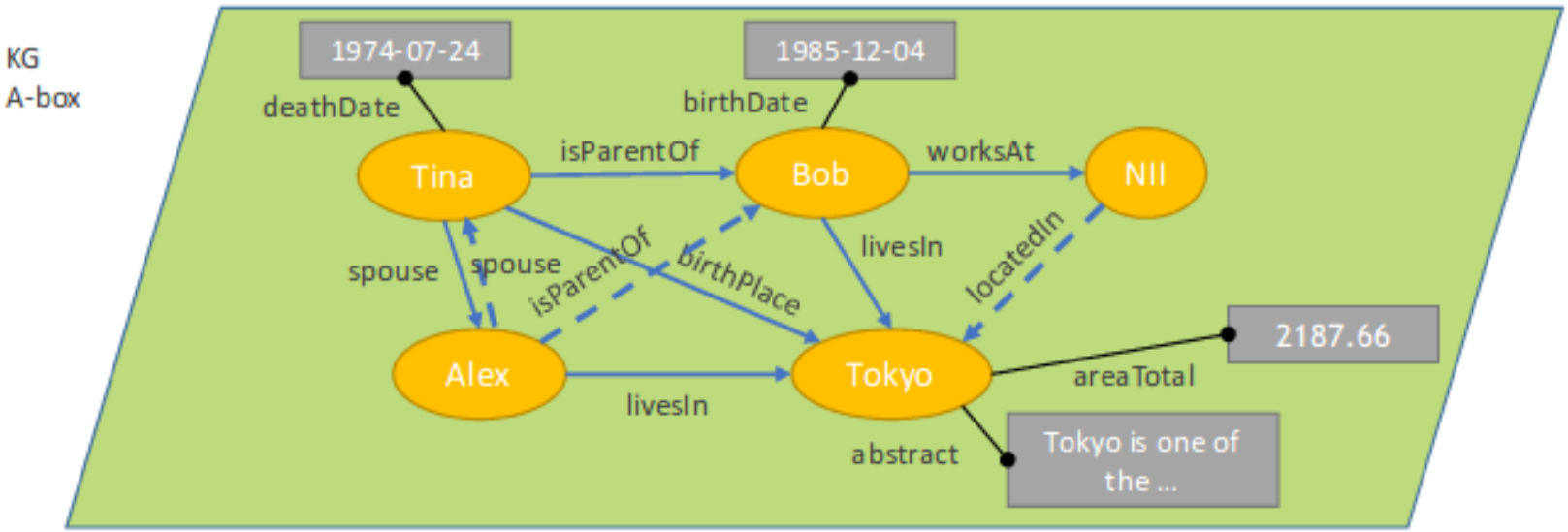
February 24, 2021

Knowledge Graphs



- **Knowledge Graph**: A heterogeneous graph with *context*
- Information is depicted in triples (s, r, o), where:
  1. s : subject
  2. o : object
  3. r : relation

The need for error detection



- KG construction mainly relied on manual annotation.
- Nowadays knowledge graphs are automatically constructed.
- Noise and errors are automatically introduced!

Table of contents

1. Introduction
2. Path Ranking
3. Graph Embeddings
4. Path Guided Embeddings
5. Results

Path Ranking Algorithm (PRA)

Original Implementation (first proposed in 2010 by Lao et al.)

- *Key Idea*: Explicitly use paths that connect two entities as features.
- *Path*: A sequence of relations $<r1, r2, ..., r_\lambda>$ that link two entities.
- 1. *Feature extraction*: Extract paths of **bounded** length.
  2. *Main computation*: Given an entity pair (s, o), and a path p, estimate probability $p(o|p, s)$ (mostly **random walks**).
  3. *Prediction*: Train individual classifier for each relation:

$$Score_r(s, o) = \sum_{p \in P} p(o|p, s)\theta_{pr}$$

Path Ranking Algorithm (PRA)

Original Implementation (first proposed in 2010 by Lao et al.)

- *Key Idea*: Explicitly use paths that connect two entities as features.
- *Path*: A sequence of relations $<r1, r2, ..., r_\lambda>$ that link two entities.
- 1. *Feature extraction*: Extract paths of **bounded** length.
  2. *Main computation*: Given an entity pair (s, o), and a path p, estimate probability $p(o|p, s)$ (mostly **random walks**).
  3. *Prediction*: Train individual classifier for each relation:

$$Score_r(s, o) = \sum_{p \in P} p(o|p, s)\theta_{pr}$$

PRA Extensions

- Incorporating a text corpus as additional evidence during inference + relaxing main computation (Gardner et al.).
- PaTyBRED: Paths and Types with Binary Relevance for Error Detection (Andre Melo, 2017).

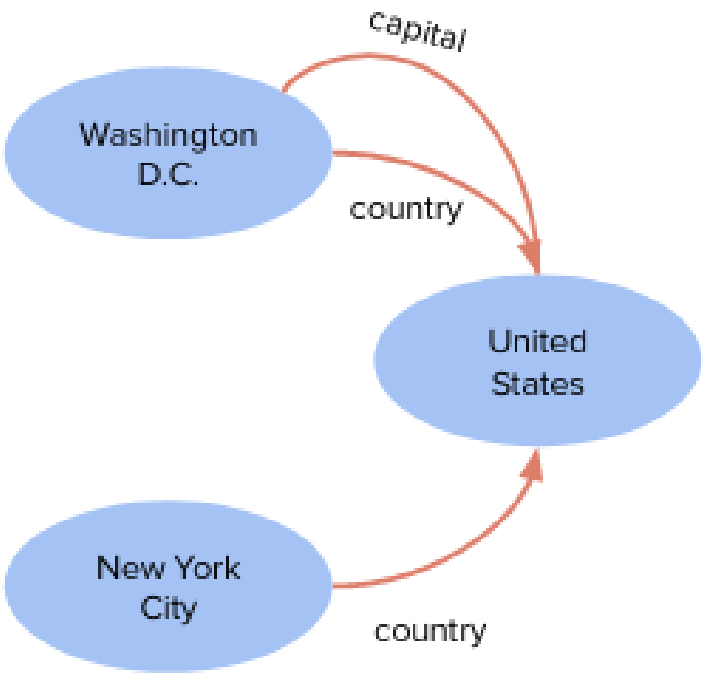
PaTyBRED

$$
\begin{array}{l}
 & \overbrace{(TREATS, IS\_A)}^{} \quad \overbrace{\text{feature paths}}^{} \\
(s, IS\_A, o) \\
\text{positives} \\
\\
\\
\\
\text{negatives} \\
\\
\\
\end{array}
\begin{bmatrix}
0 & 1 & 1 & \dots & 0 & 0 \\
1 & 0 & 1 & \dots & 0 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & 0 & \dots & 1 & 1 \\
\hdashline
0 & 0 & 0 & \dots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 1 & \dots & 0 & 0
\end{bmatrix}
\begin{array}{c}
\overbrace{\text{labels}}^{} \\
\\
\\
\\
\\
\\
\\
\end{array}
\begin{bmatrix}
1 \\
1 \\
\vdots \\
1 \\
\hdashline
0 \\
\vdots \\
0
\end{bmatrix}
=
\begin{array}{c}
\overbrace{\text{confidence}}^{} \\
\\
\\
\\
\\
\\
\\
\end{array}
\begin{bmatrix}
0.8 \\
1.0 \\
\vdots \\
0.6 \\
\hdashline
0.3 \\
\vdots \\
0.0
\end{bmatrix}
$$

Table of contents

1. Introduction
2. Path Ranking
3. Graph Embeddings
4. Path Guided Embeddings
5. Results

Graph Embeddings



Embedding

Washington D.C.	New York City	United States	country	capital
1	1	5	1	2
2	2	6	2	2
4	5	4	4	5
5	4	4	1	1
6	7	8	4	5
7	7	7	5	5

- Embedding methods generate representations of the elements of the knowledge graph that are embedded in a vector space.
- These representations encode latent properties of the graph.
- ...e.g. similar entities tend to be described with similar vectors.

Main KG Graph Embedding approaches

Translational approaches

- Often based on learning translations from the subject to the object.
- Have the advantages of having a concise definition and getting good performances while being scalable at the same time.
- *Most representative model*: **TransE**:

$$E(s, r, o) = ||s + r - o||_2$$

based on the idea that $s + r = o$.

Main KG Graph Embedding approaches

Translational approaches

- Often based on learning translations from the subject to the object.
- Have the advantages of having a concise definition and getting good performances while being scalable at the same time.
- *Most representative model*: **TransE**:

$$E(s, r, o) = ||s + r - o||_2$$

based on the idea that $s + r = o$.

BiLinear approaches

- High expressive power due to the use of a full rank matrix for each relationship.
- At the same time, the full rank matrix is prone to overfitting (constraints have been suggested)
- *Most representative model*: **RESCAL**: $E(s, r, o) = sW_r o$.

Table of contents

1. Introduction
2. Path Ranking
3. Graph Embeddings
4. Path Guided Embeddings
5. Results

Path Guided Embeddings

Utilize paths that connect an (s, o) tuple to train embeddings of triples:

PTransE

- Path : $p = r_1 + r_2 + ... + r_l$
- $Loss = \sum_{(s,r,o) \in S} L(s, r, o) + \frac{1}{Z} * \sum_{p \in P(s,o)} R(p|s, o) * L(p, r)$
- $L(s, r, o) = \sum_{(s',r,o') \in S'} max(0, \gamma + E(s, r, o) - E(s', r, o'))$
- $L(p, r) = \sum_{(s,r',o) \in S'} max(0, \gamma + ||p - r|| - ||p - r'||)$

Path Guided Embeddings

Utilize paths that connect an (s, o) tuple to train embeddings of triples:

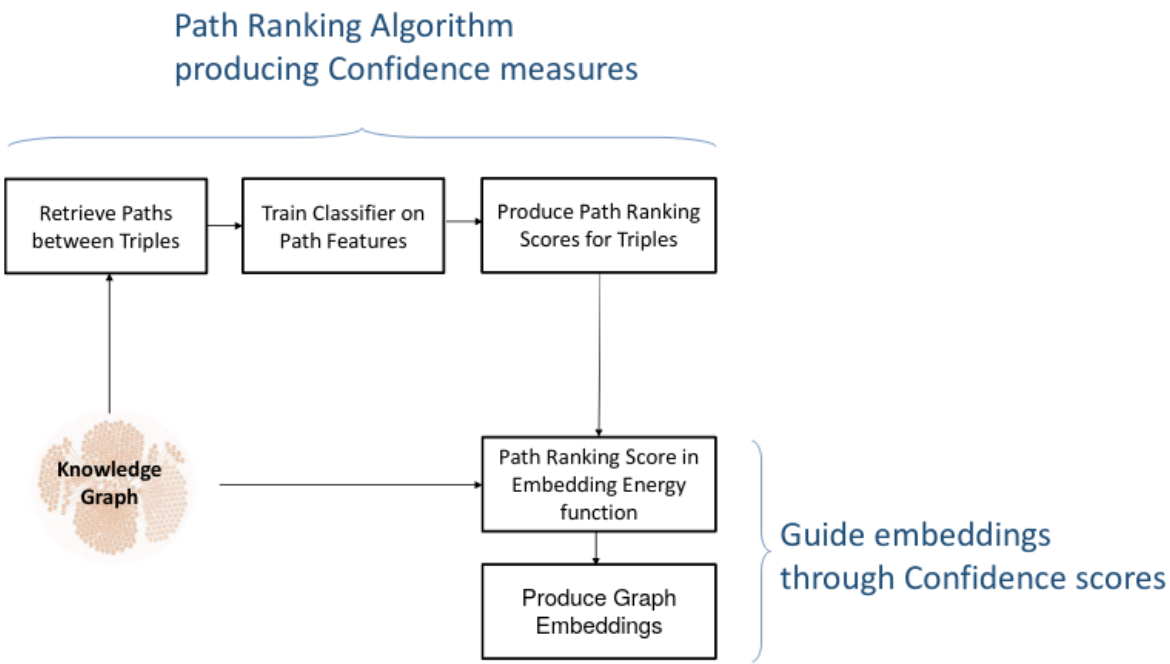
PTransE

- Path : $p = r_1 + r_2 + ... + r_l$
- $Loss = \sum_{(s,r,o) \in S} L(s, r, o) + \frac{1}{Z} * \sum_{p \in P(s,o)} R(p|s, o) * L(p, r)$
- $L(s, r, o) = \sum_{(s',r,o') \in S'} max(0, \gamma + E(s, r, o) - E(s', r, o'))$
- $L(p, r) = \sum_{(s,r',o) \in S'} max(0, \gamma + ||p - r|| - ||p - r'||)$

CKRL

- $Loss = \sum_{(s,r,o) \in S} L(s, r, o) * C(s, r, o)$
- $C(s, r, o) = \lambda_1 LT(s, r, o) + \lambda_2 PP(s, r, o) + \lambda_3 AP(s, r, o)$

PRGE



PRGE

- Instead of competing end-to-end, make something more modular.
- An embedding-based loss scaled with a Path Ranking Score:

$$L = \sum_{T \in S} \sum_{T \in S'} max(0, \gamma + E(s, r, o) - E(s', r, o')) \cdot P(s, r, o)^{\lambda}$$

Table of contents

1. Introduction
2. Path Ranking
3. Graph Embeddings
4. Path Guided Embeddings
5. Results

Experiments

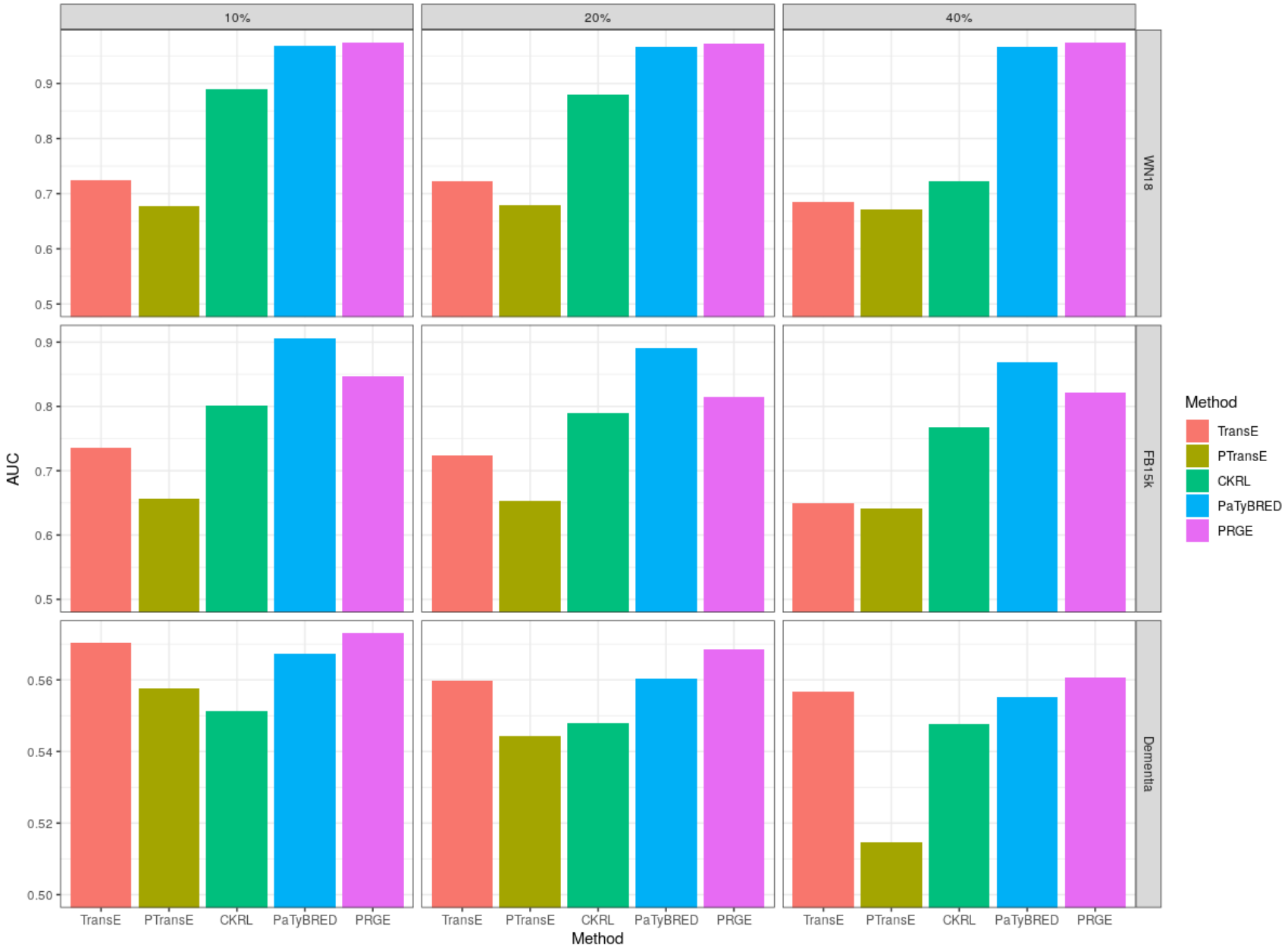
Datasets: Two benchmark datasets and one in-house dataset (Dementia from IASIS) were used for evaluation purposes:

Dataset	#Rel	#Ent	#Triples
WN18	18	40,943	141,442
FB15k	1345	14,951	483,142
Dementia	64	48,008	135,000

Evaluation Metrics:

- Ranking-based ones (Mean Rank, Mean Reciprocal Rank)
- Classification based ones (AUC)

Results


10%
20%
40%
WN18
FB15k
Dementia
AUC
0.9
0.8
0.7
0.6
0.5
0.56
0.54
0.52
0.50
TransE
PTransE
CKRL
PaTyBRED
PRGE
Method
Method
TransE
PTransE
CKRL
PaTyBRED
PRGE

Downstream Tasks

Advantages of the method:

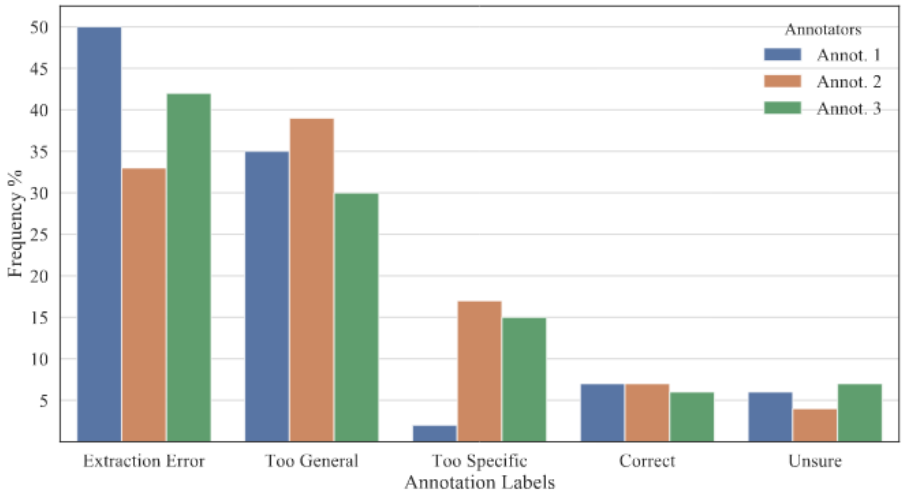
- *Modularity*: Any Embedding method can be used (Translational/Bilinear), and any PRA can be used.
- Other Triple confidence scores can be used from other sources (Text Similarity/Rule-based scores (Probabilistic soft Logic)).
- *Creation of Embeddings*: A result of this process is the creation of embeddings that can be used in other downstream tasks

TRIPLE CLASSIFICATION

Method	FB15K-10%	FB15K-20%	FB15K-40%
TransE	**0.717**	0.703	0.671
PTransE	0.686	0.678	0.67
CKRL	0.639	0.709	0.691
PRGE	0.715	**0.712**	**0.702**

Qualitative results

Applying PRGE to the IASIS Dementia dataset:



Erroneous triple (s - r - o)	Type	Supporting text
acetonitrile - AUGMENTS - 80%	ER	... The **acetonitrile** concentration was increased to 80% in 5 min and then held in **80%** acetonitrile for an additional 5 min ...
Denmark - PART_OF - Neurons	ER	... **Neurons** transfected with **DA**-GFP were found to have dendritic spines that had significantly lengthened necks compared to ...
Cells - PART_OF - Medial geniculate body	TG	... Dendritic spines of the polyhedral and elongated **cells** of the **medial geniculate bodies** were decreased in number ...
Entire thumb - PART_OF - Patients	TG	... The **patient** is asked to hold the ruler with his **thumb** and forefinger and to release the ruler while the investigator continues to ...
DDMS - PART_OF - Homo sapiens	TS	... HT-22 hippocampal cells and confirms observations using brain extracts from monkey, mouse, rat and **human DDM** ...
Brain - LOCATION_OF - Decreased plasmalogens	TS	... data suggest that long-term alterations in plasmalogen synthesis degradation result in **decreased brain plasmalogen** levels, a hallmark feature of AD ...

Conclusions/Future Work

Conclusions:

- The proposed PRGE method combines the path ranking score of a triple with the graph embedding framework, creating embeddings robust towards noise present in the graph
- The proposed PRGE method is agnostic of the underlying energy function and triple-scoring mechanism, as such, other path-ranking scores and energy function could work even better.
- Overall, we have proposed a generic framework to generate embeddings resilient to noise and we proved that they can also be used in multiple downstream tasks enhancing performance in the presence of noise.

Future Work:

- More experimentation and a thorough comparison with other state-of-the-art methods.
- An extended study with different combination of methodologies is needed (Different Path ranking algorithms/prior scores, different embedding approaches).