RDF* Knowledge Graph Completion by Translation

Authors: Linda Kwan[1,†], Dr Pouya Ghiasnezhad Omran[2,†], Dr Armin Haller[3,†]

Our Contributions

- Proposed a novel translation method called ExtRet (Extended Reification) to carry out Knowledge Graph (KG) completion tasks on RDF* KGs.
- Applied translated KGs to different link predictors like StarE and AnyBURL to evaluate its ability to infer new knowledge.
- Demonstrated that ExtRet outperforms native translation methods with respect to KG completion.

Background & Motivation

- Knowledge Graphs (KGs) are important for modelling facts about real-world entities, but they are often incomplete. Thus, we require Link Prediction (LP) to infer facts which are not explicitly modelled in KGs.
- RDF (Resource Description Framework) is a general-purpose framework used to model information on the Web such as Wikipedia data.
- An RDF triple allows us to express facts in (subject, predicate, object) format. It describes a relation between two entities.
- Most simple facts can be expressed as RDF triples. However, it is difficult to model n-ary relations between entities using RDF triples.
- An RDF* triple allows us to express more complex statements. i.e., statements about other statements.
- In addition to the basic RDF structure, any RDF* triple can be a subject or an object of another RDF* triple. We can use qualifiers (predicate-entity pairs) to represent properties of an RDF* triple.
- RDF* triples without qualifiers are considered core facts.
- While there are a vast number of methods for carrying out LPs for RDF KGs like AnyBURL, there are a few LP methods that can handle RDF* directly like StarE.

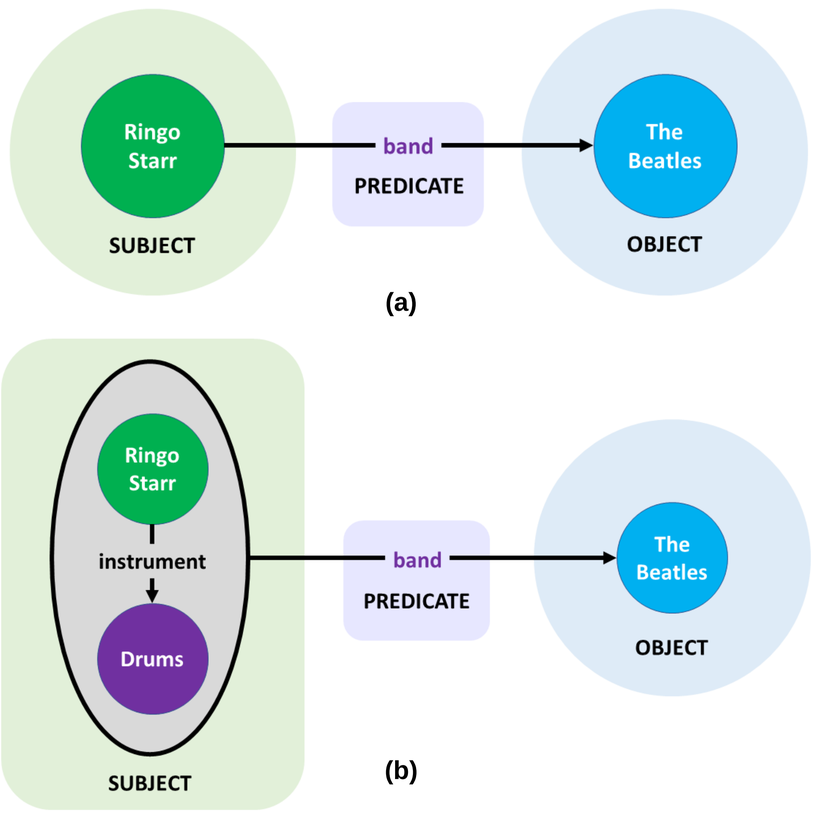

Fig. 1: (a) An example of an RDF triple. The statement modelled in this triple is *"Ringo Starr is in a band named The Beatles"*. (b) An example of an RDF-star triple. The statement modelled in this triple is *"Ringo Starr plays an instrument called the drums in a band named The Beatles"*.

Conclusion & Future Work

- ExtRet outperforms standard reification with respect to KG completion due to additional links between entities in core facts and entities in qualifiers which enable link predictors to recognise direct relations between those entities.
- Our proposed translation-based LP produces extra time complexity due to the number of RDF triples produced, which could be managed using more efficient link predictors such as AnyBURL.
- While training using AnyBURL is significantly faster than StarE, its performance overall is less optimal. Thus, we compare our algorithm against baselines like unqualification to measure the quality of LPs.
- For future work, we would consider using other benchmark datasets to compare LP quality of ExtRet against standard reification. Additionally, we could run our experiment on a machine with GPUs that is capable of handling larger datasets.

† College of Engineering & Computer Science, Australian National University

[1] Linda.Kwan@anu.edu.au [2] P.G.Omran@anu.edu.au [3] Armin.Haller@anu.edu.au

Translating RDF* Knowledge Graphs

- **Standard Reification**: A common method of translating RDF* KG to RDF KG. Replaces the core fact with an intermediate node, which is then connected to the subject, predicate and object of the core fact. However, link predictors fail to recognise relations between entities in core facts and outer entities.
- **Unqualification**: Extracts the core fact from every RDF* triple, then forms a separate KG from them. Used as a baseline to set a minimum performance standard. Not intended to represent the whole RDF* KG.
- **ExtRet (Extended Reification)**: Our proposed translation algorithm. Combines standard reification and unqualification, but includes direct relations between outer entities and entities in core facts. Aims to minimise structural information loss while extending its ability to make link predictions using core facts as queries.

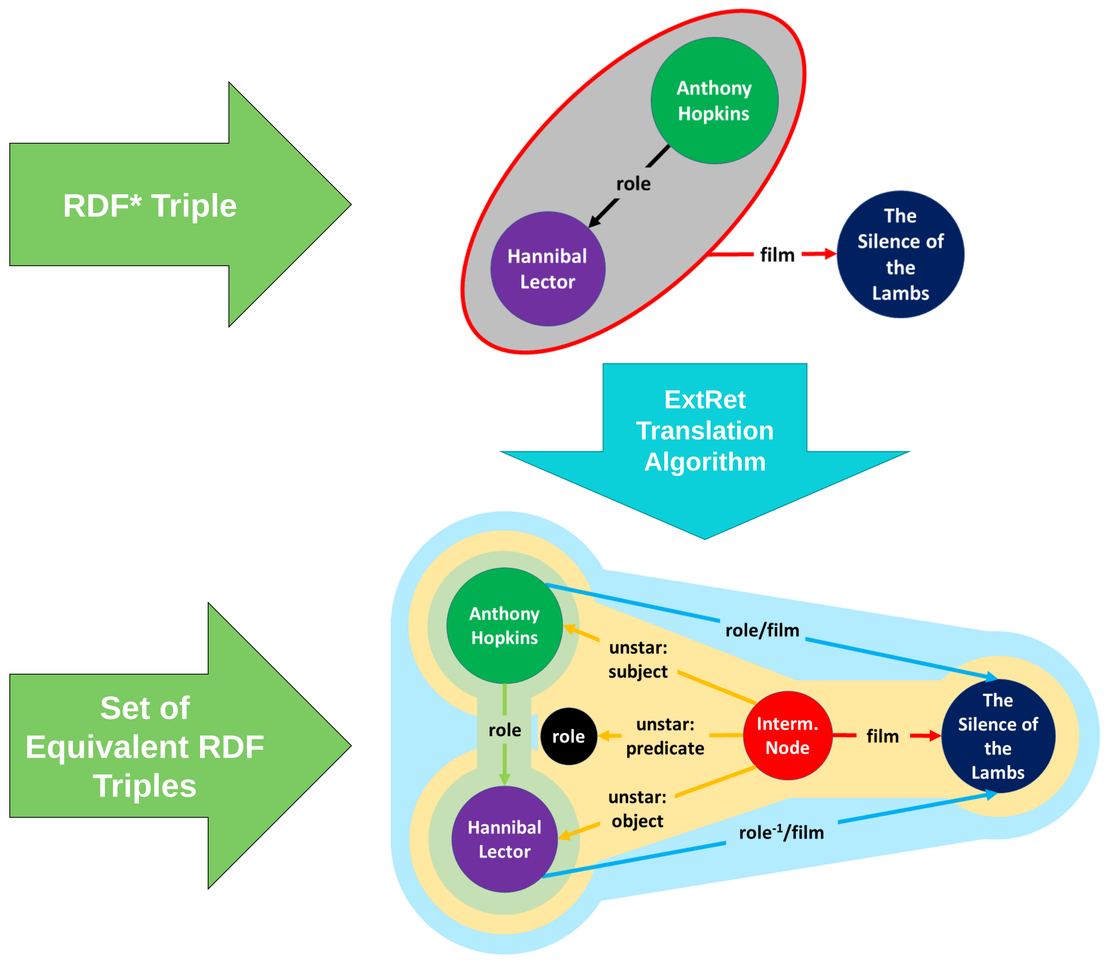

Fig. 2: A visualisation of our proposed translation method ExtRet (blue area). ExtRet combines standard reification (orange area) with unqualification (green area) in addition to direct links between entities.

Experimental Results

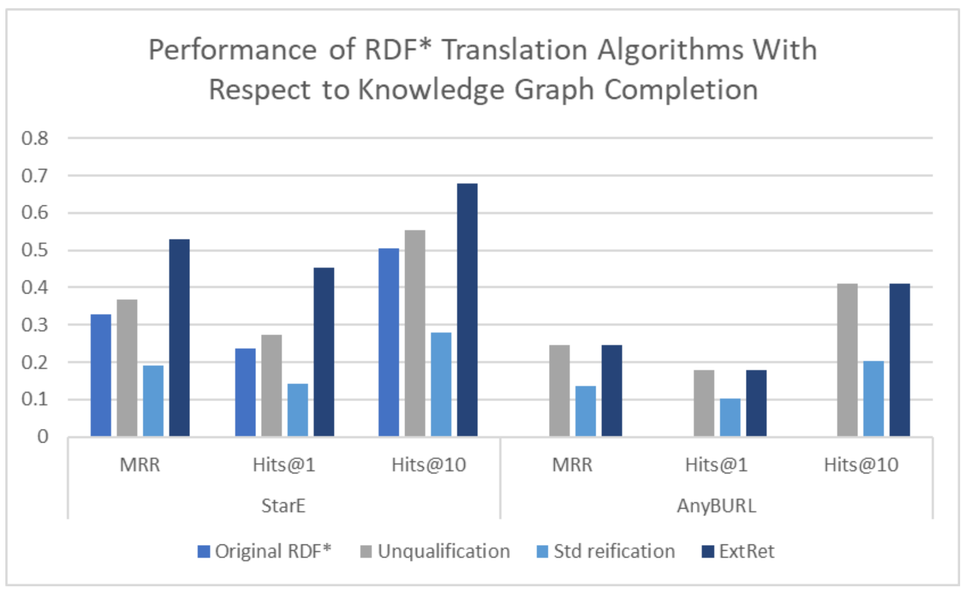

Fig. 3: Performance comparison of RDF* KG Translation Methods with respect to KG completion. A subset of JF17K dataset is used as the experimental benchmark. Only core facts are used as queries. Performances are evaluated using Mean Reciprocal Rank (MRR), Hits@1, and Hits@10.

	StarE	AnyBURL
Original RDF*	2h 30min	1min 40s
Unqualification	**1h 13min**	1min 40s
Std Reification	8h 33min	1min 40s
ExtRet	14h 27min	1min 40s

Table 1: Comparison of link prediction training times using KGs produced by different RDF* translation algorithms.

Australian National University

ANU College of Engineering & Computer Science