Towards the difference between RDF graphs

François Goasdoué, Hélène Jaudoin and Claire Vanoni
IRISA, Univ. Rennes, Lannion, France
{fg, helene.jaudoin,claire.vanoni}@irisa.fr

Context and main goal

Comparing RDF graphs, by identifying their commonalities and differences, is central to understand the data at hand. While identifying commonalities between RDF graphs has received considerable attention so far, identifying differences was only tackled in [1] with a main limitation (in our view): the difference between RDF graphs is not an RDF graph.

Some useful important applications:
- RDF graph exploration
- RDF graph evolution/maintenance

Goal: a principled definition of difference between RDF graphs that takes into account essentiel features of RDF: blank nodes, RDF Schema and RDF entailment

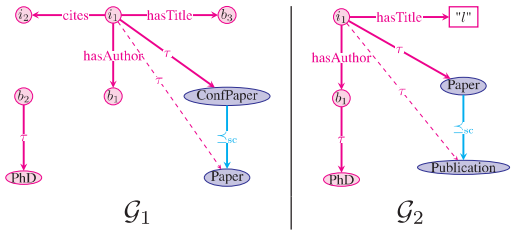
[1] Alina Petrova, Evgeny Sherkhonov, Bernardo Cuenca Grau and Ian Horrocks: *Entity Comparison in RDF graphs*. International Semantic Web Conference (ISWC), 2017.

RDF triples and entailment rules

RDF statement	Fact triple
Class assertion	$(s, \text{rdf:type}, o)$
Property assertion	(s, p, o) with $p \neq \text{rdf:type}$

RDFS statement	Ontological triple
Subclass	$(s, \preceq_{sc}, o)$
Subproperty	$(s, \preceq_{sp}, o)$
Domain typing	$(s, \hookleftarrow_d, o)$
Range typing	$(s, \hookrightarrow_r, o)$

Rule	Entailment rule
rdfs2	$(p, \hookleftarrow_d, o), (s_1, p, o_1) \rightarrow (s_1, \tau, o)$
rdfs3	$(p, \hookrightarrow_r, o), (s_1, p, o_1) \rightarrow (o_1, \tau, o)$
rdfs5	$(p_1, \preceq_{sp}, p_2), (p_2, \preceq_{sp}, p_3) \rightarrow (p_1, \preceq_{sp}, p_3)$
rdfs7	$(p_1, \preceq_{sp}, p_2), (s, p_1, o) \rightarrow (s, p_2, o)$
rdfs9	$(s, \preceq_{sc}, o), (s_1, \tau, s) \rightarrow (s_1, \tau, o)$
rdfs11	$(s, \preceq_{sc}, o), (o, \preceq_{sc}, o_1) \rightarrow (s, \preceq_{sc}, o_1)$
ext1	$(p, \hookleftarrow_d, o), (o, \preceq_{sc}, o_1) \rightarrow (p, \hookleftarrow_d, o_1)$
ext2	$(p, \hookrightarrow_r, o), (o, \preceq_{sc}, o_1) \rightarrow (p, \hookrightarrow_r, o_1)$
ext3	$(p, \preceq_{sp}, p_1), (p_1, \hookleftarrow_d, o) \rightarrow (p, \hookleftarrow_d, o)$



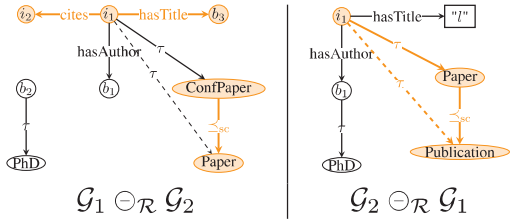
An RDF graph $\mathcal{G}$ can be seen as the pair $\mathcal{G} = \langle O, A \rangle$, where O is comprised of $\mathcal{G}$'s RDFS statements and A is comprised of $\mathcal{G}$'s RDF statements.

Intuitive difference between RDF graphs

Definition 1: Difference between RDF graphs

Let $\mathcal{G}_1 = \langle O_1, A_1 \rangle$ and $\mathcal{G}_2$ be two RDF graphs and $\mathcal{R}$ be a set of entailment rules. The *difference* between $\mathcal{G}_1$ and $\mathcal{G}_2$ w.r.t. $\mathcal{R}$, denoted by $\mathcal{G}_1 \ominus_{\mathcal{R}} \mathcal{G}_2$, is the RDF graph $\mathcal{G}_1$ minus every minimal subgraph a_1 of A_1 such that $\langle O_1, a_1 \rangle$ entails some subgraph of $\mathcal{G}_2$ w.r.t. $\mathcal{R}$.

In the figure on the right, the triples diplayed in orange correspond to the intuitive difference.



Negative result w.r.t. Definition 1 and research problem

Proposition 1: Combining n entailment relationships into one

Let $\mathcal{G}_1, \mathcal{G}'_1, \ldots, \mathcal{G}_n, \mathcal{G}'_n$ be RDF graphs such that $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are subgraphs of some RDF graph $\mathcal{G}$, $\mathcal{G}'_1, \ldots, \mathcal{G}'_n$ are subgraphs of some RDF graph $\mathcal{G}'$, and $\mathcal{G}_i \models_{\mathcal{R}} \mathcal{G}'_i$ for $1 \leq i \leq n$. $\bigcup_{i=1}^n \mathcal{G}_i \models_{\mathcal{R}} \biguplus_{i=1}^n \mathcal{G}'_i$ holds.

Let us consider the subgraphs a_1 and a'_1 of A_1 such that $a_1 = \{(i_1, hasAuthor, b_1)\}$ and $a'_1 = \{(b_2, \tau, PhD)\}$, and, a_2 and a'_2 of A_2 such that $a_2 = \{(i_1, hasAuthor, b_1)\}$ and $a'_2 = \{(b_1, \tau, PhD)\}$.
We have $a_1 \models_{\mathcal{R}} a_2$ and $a'_1 \models_{\mathcal{R}} a'_2$, hence as per **Proposition 1** $a_1 \bigcup a'_1 \models_{\mathcal{R}} a_2 \biguplus a'_2 = \{(i_1, hasAuthor, b_2), (b_3, \tau, PhD)\}$. However, $a_1 \bigcup a'_1 \not\models_{\mathcal{R}} a_2 \bigcup a'_2 = \{(i_1, hasAuthor, b_1), (b_1, \tau, PhD)\}$.

Proposition 1 explains the issue w.r.t. Definition 1

We have $\bigcup_{i=1}^n \mathcal{G}_i \subseteq \mathcal{G}$, while $\biguplus_{i=1}^n \mathcal{G}'_i \not\subseteq \mathcal{G}'$ and moreover $\biguplus_{i=1}^n \mathcal{G}'_i \not\equiv_{\mathcal{R}} \bigcup_{i=1}^n \mathcal{G}'_i$ in general!

Research problem

Under which extra conditions on $\mathcal{G}_1, \mathcal{G}'_1, \ldots, \mathcal{G}_n, \mathcal{G}'_n$ in Proposition 1, $\bigcup_{i=1}^n \mathcal{G}_i \models_{\mathcal{R}} \bigcup_{i=1}^n \mathcal{G}'_i$ holds?

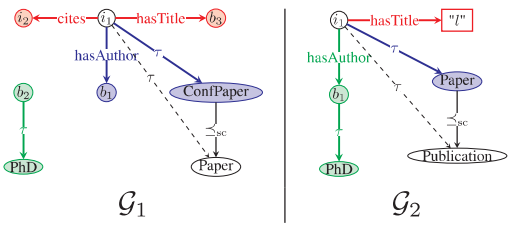
Solution: join-preserving graphs (join graphs in short)

Definition 2 : Join graph of an RDF graph

A *join graph* $\mathcal{G}'$ of an RDF graph $\mathcal{G}$ is a subgraph of $\mathcal{G}$ such that for each blank node b in $\mathcal{G}'$, any triple of $\mathcal{G}$ in which b occurs belongs to $\mathcal{G}'$.

Proposition 2: Combining n entailment relationships into one (revisited)

Let $\mathcal{G}_1, \mathcal{G}'_1, \ldots, \mathcal{G}_n, \mathcal{G}'_n$ be RDF graphs such that $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are subgraphs of some RDF graph $\mathcal{G}$, $\mathcal{G}'_1, \ldots, \mathcal{G}'_n$ are join graphs of some RDF graph $\mathcal{G}'$, and $\mathcal{G}_i \models_{\mathcal{R}} \mathcal{G}'_i$ for $1 \leq i \leq n$. $\bigcup_{i=1}^n \mathcal{G}_i \models_{\mathcal{R}} \bigcup_{i=1}^n \mathcal{G}'_i$ holds.



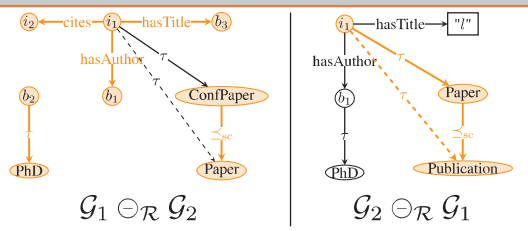
In different colors are examples of join graphs

Principled definition of Difference between RDF graphs

Definition 3 : Difference between RDF graphs (revised)

Let $\mathcal{G}_1 = \langle O_1, A_1 \rangle$ and $\mathcal{G}_2$ be two RDF graphs and $\mathcal{R}$ be a set of entailment rules. The *difference* between $\mathcal{G}_1$ and $\mathcal{G}_2$ w.r.t. $\mathcal{R}$, denoted by $\mathcal{G}_1 \ominus_{\mathcal{R}} \mathcal{G}_2$, is the RDF graph $\mathcal{G}_1$ minus every minimal subgraph a_1 of A_1 such that $\langle O_1, a_1 \rangle$ entails some join graph of $\mathcal{G}_2$ w.r.t. $\mathcal{R}$.

In the figure on the right, the triples diplayed in orange correspond to the difference.



MaDICS 2026, 2nd June - 3rd June 2026, Avignon, France