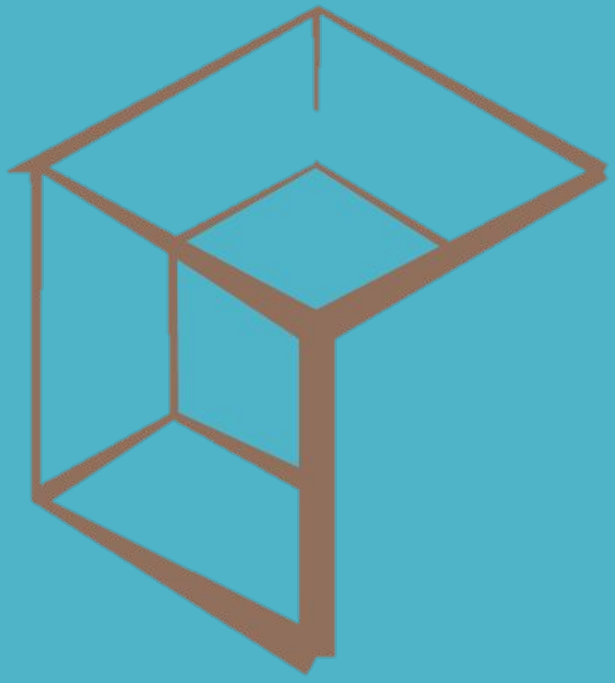
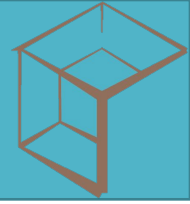
Graph? Yes! Which one? Help!



Lassila et al., Amazon Neptune

Nikolaos Karalis
KnowGraphs Colloquium
January 21, 2022

Graph Interoperability



Motivation

- Multiple graph models for the representation of KGs
  - Resource Description Framework (RDF)
  - Labeled Property Graphs (LPGs)

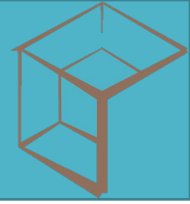
- Model specific query languages
  - SPARQL for RDF
  - Gremlin and openCypher for LPGs

Goal

- A data model that models both RDF graphs and LPGs

- Cross-use of query languages (e.g., SPARQL over LPGs)

One Graph (1G)



- Data model for the representation of RDF graphs and LPGs

- Quad-based representation
  - *src* —*label*→ *value: sid*
  - *src*: source vertex
  - *value*: target vertex or vertex property value
  - *label*: edge label or vertex property key
  - *sid*: unique identifier for each statement

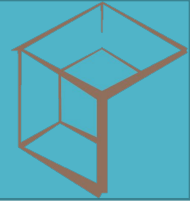
- Statement identifiers can be used for edge properties
  Example:

$$Bob \text{ —} knows \rightarrow Alice: sid_1 \text{ (ground statement)}$$

$$sid_1 \text{ —} since \rightarrow 2020: sid_2 \text{ (assertion)}$$

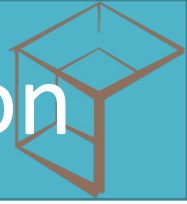
Challenges of Graph Interoperability

Challenges (Overview)



1. Edge properties, multiple edge instances, and reification
2. Triples vs. graph abstraction
3. Datatype alignment
4. Graph partitioning
5. Graph merging and external identifiers
6. Lack of formal foundation
7. Update query semantics

Edge Properties, Multiple Edge Instances, and Reification



- RDF has no built-in support for edge properties
  - Reification
  - RDF-star
- LPGs support multiple instances of an edge
  Example:

$Bob \text{ —} knows \rightarrow Alice: sid_1$

$sid_1 \text{ —} since \rightarrow 2020: sid_3$

$sid_1 \text{ —} statedBy \rightarrow NYT : sid_4$

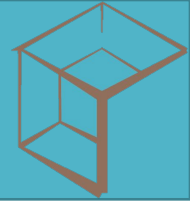
$Bob \text{ —} knows \rightarrow Alice: sid_2$

$sid_1 \text{ —} since \rightarrow 2021: sid_5$

$sid_1 \text{ —} statedBy \rightarrow Guardian : sid_6$

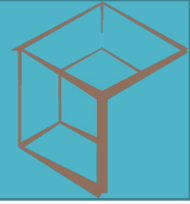
- In RDF(-star) each statement is unique
- How to treat multiple instances of an edge and their properties in RDF(-star)?

Triples vs. Graph Abstraction



- RDF graphs ⟶ set of triples
- LPGs ⟶ disjoint sets of vertices and edges
- The mapping of 1G statements to RDF statements is straightforward
- **Key challenge:** The definition of graph elements over 1G
  - How to distinguish edges from vertex properties
  - How to iterate over the vertices of a graph

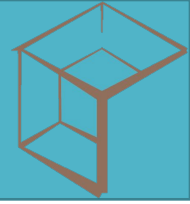
Datatype Alignment



- RDF has a formal type system
  - Builds upon the XML Schema Definition (XSD)
  - Extensibility and flexibility
  - Composite types (e.g., lists) are modelled via RDF containers
- No formal definition for the type system of LPGs
  - Built-in composite types
  - Semantics are delegated to the implementation language
- **Challenge:** Definition of a meta type system
- **Possible solution:** A meta type system based on user-defined RDF literals
  Example:

  "[1, 2, 3]"^^:OneGraphList

Graph Partitioning



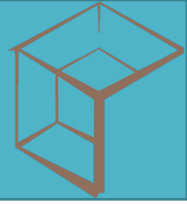
- RDF supports *named graphs*
- RDF uses quads to represent statements involving named graphs: *(s, p, o, g)*
- 1G introduces a new edge label: *inGraph*
  Example:

$$Bob \text{ —} knows \rightarrow Alice: sid_1$$

$$sid_1 \text{ —} inGraph \rightarrow g: sid_2$$

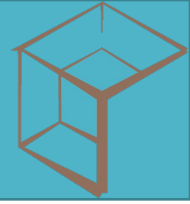
- The evaluation of SPARQL queries needs to consider the new edge label

Graph Merging and External Identifiers



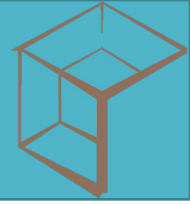
- RDF provides a specification for graph merging
- The merging of RDF graphs makes use of the global identifiers (IRIs)
- There is no standard procedure for the merging of LPGs
  - How to deal with edge properties
  - How to deal with multiple edge instances
- **Interesting use case:** Merging of an RDF graph and a LPG

Lack of Formal Foundation

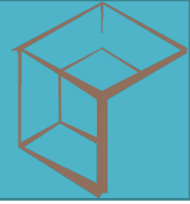


- Query languages for LPGs lack formal semantics / a query algebra
- How to assess that queries of different languages are semantically equal?

Update Query Semantics



- Ambiguity when updating over a simplified view of 1G
- How should a statement be removed?
  - RDF does not allow multiple instances of a statement
  - Should a SPARQL query over LPG remove all instances of an edge and its associated edge properties?
- How to insert multiple edge instances using SPARQL?



Questions?

Original Slides: https://www.lassila.org/publications/2021/scg2021-lassila+etal-preso.pdf

Talk by Ora Lassila: https://www.youtube.com/watch?v=f9wautaqWUs