COMP318

Ontologies and Semantic Web

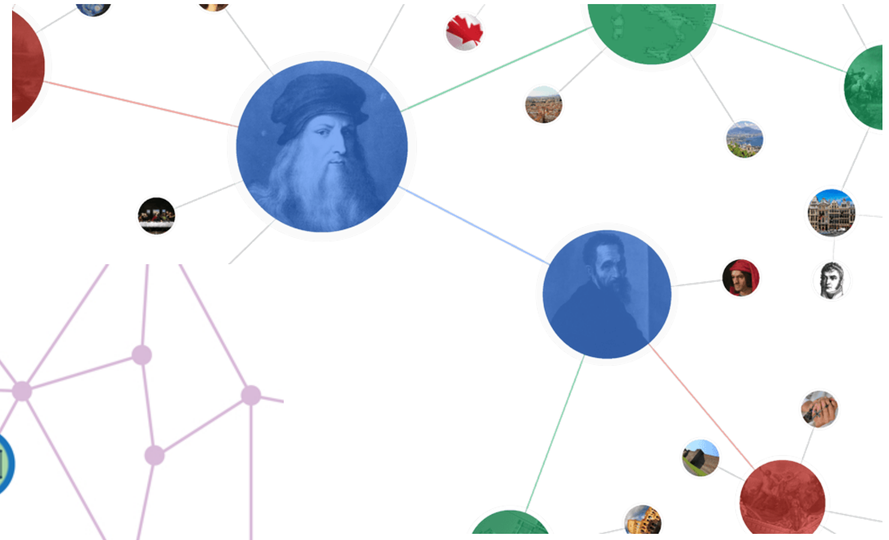
Knowledge Graphs - Part 1

Dr Valentina Tamma

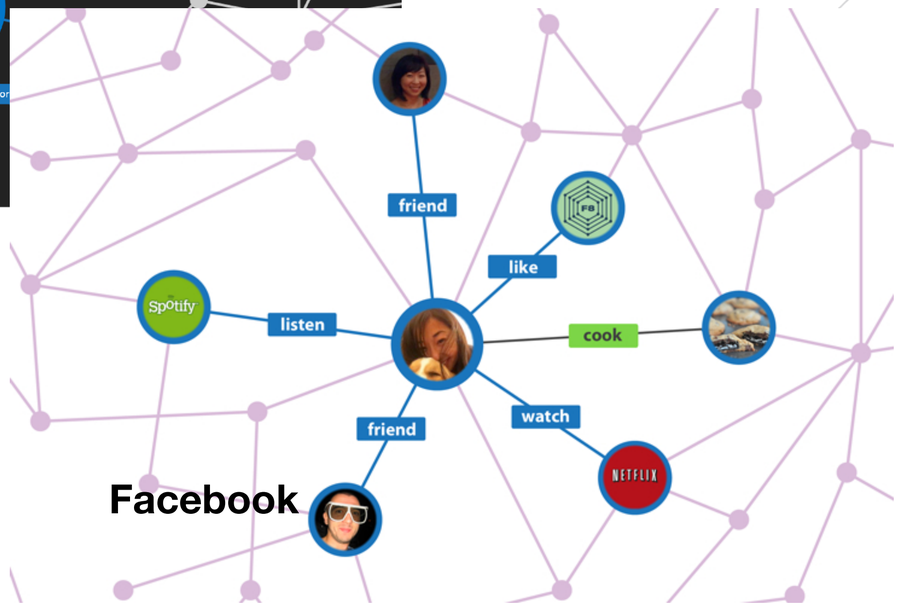
V.Tamma@liverpool.ac.uk

Graphs are everywhere

Microsoft



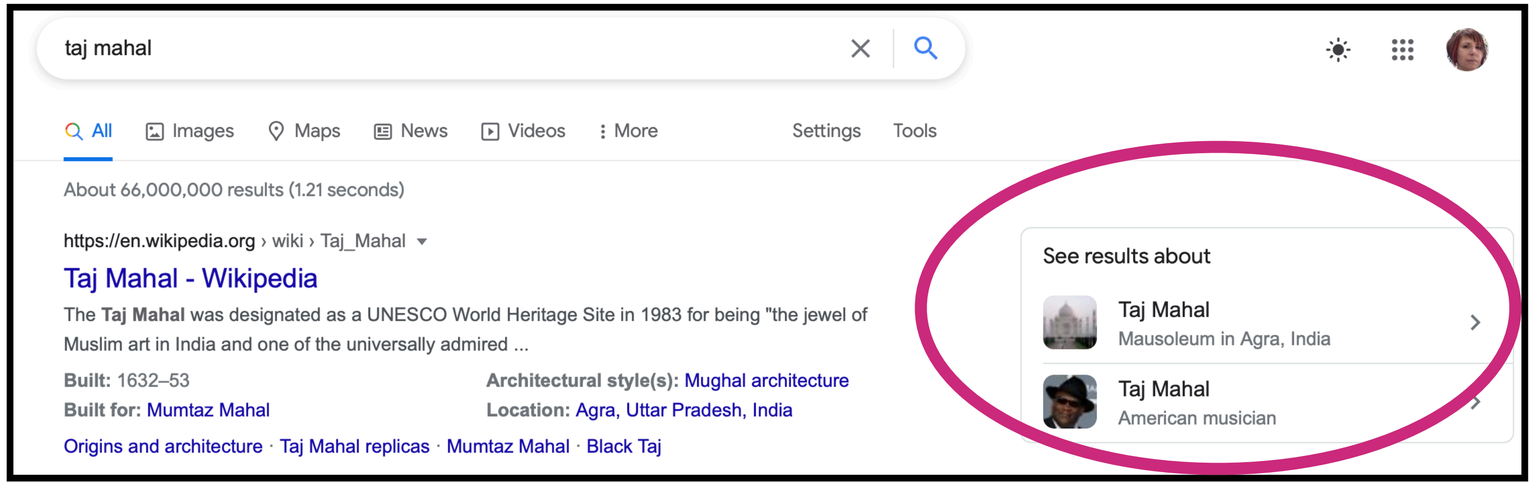
Google



Facebook

Graphs are the key to smart data

- It's all about **things**
  - URIs
- Not **strings**



- **KGs** related to the notions of:
  - ontologies,
  - knowledge bases, and
  - knowledge-based systems

What is a knowledge graph

- The definition of a “knowledge graph” remains contentious, with several (sometimes conflicting definitions emerging):
- A knowledge graph is a **graph of data** intended to accumulate and **convey knowledge** of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities.

A. Hogan et al: Knowledge graphs. Report on Dagstuhl Knowledge Graph Symposium, 2018

What is a knowledge graph

- **A knowledge graph as a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities.**

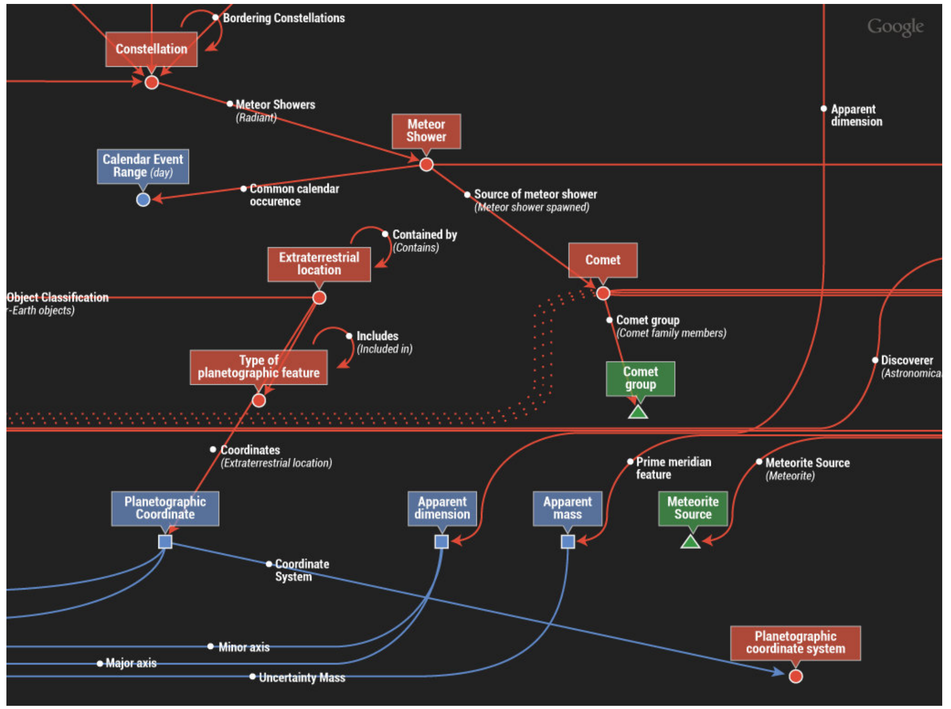
- Data vs Information vs Knowledge:
  - Data: uninterpreted symbols, e.g. strings of characters, integers, etc.
  - Information: data equipped with some meaning that allows interpretation by a human being or an autonomous systems
  - Knowledge: assimilated information and data that can be used to carry out tasks and create more information. Hence,
    - Knowledge is actionable because it can support the achievement of some task or some decision making process;
    - Knowledge is generative because it can be used to create more information, for instance through the use of rules that indicate the action to perform when some conditions are satisfied.

What is a knowledge graph

- By knowledge, we refer to something that is known or is inferred:
  - Knowledge may be accumulated from external sources, or extracted from the knowledge graph itself.
  - Knowledge may be composed of simple statements:
    - “London is the capital of the United Kingdom”
  - or quantified statements:
    - “all capitals are cities”

What is a knowledge graph

- Given a set of entities E, and relations R, a KG is a directed multi-relational graph G that contains triples of type (s, p, o)
  - $G \subseteq E \times R \times E : (s, p, o) \in G$
  - describes entities and relations
  - defines a schema interrelating arbitrary entities on various topical domains

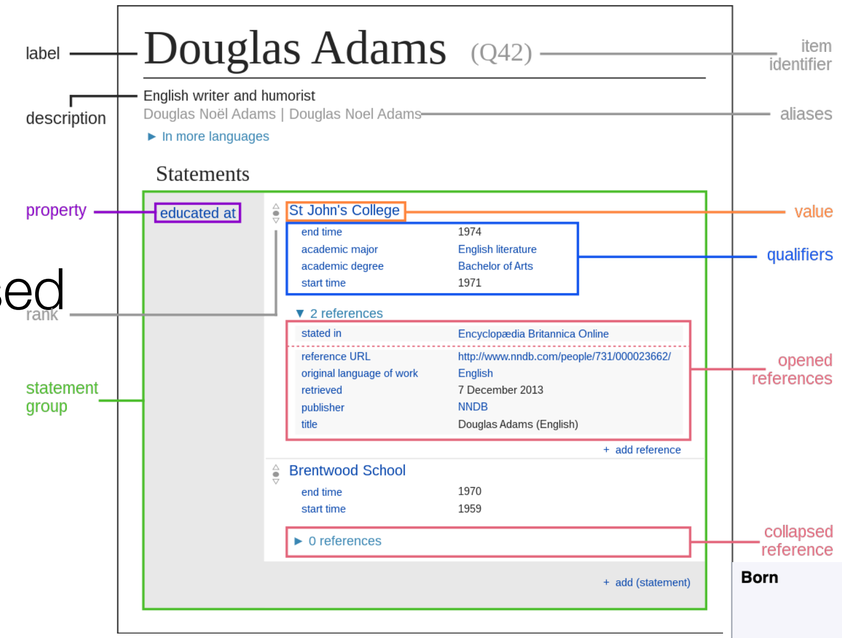


Domain Generic KGs

- Some knowledge graphs are generic and cover multiple domains
- Typically built automatically by extracting information, e.g. semi-structured content from Wikipedia, such as infoboxes.
  - DBPedia
  - Yago
- Oher generic KGs are manually curated for a specific purpose:
  - Wikidata is a KG built collaboratively and manually built as a background resource for Wikipedia

Wikipedia and Wikidata

- Wikidata acts as a centralised hub for Wikipedia's inter-language links:
  - Wikidata's language-independent identifiers are used by Wikipedia local versions, instead of localised versions of items and properties
  - Wikipedia identifiers store all the links connecting different language versions of Wikipedia articles
    - A Wikidata item exists for each Wikipedia article, but not the other way round, i.e. there may be Wikidata items with no corresponding Wikipedia article.
  - Wikidata multilinguality prevents contradictory information on the same topic due to its different language versions
    - an issue known to affect Wikipedia,



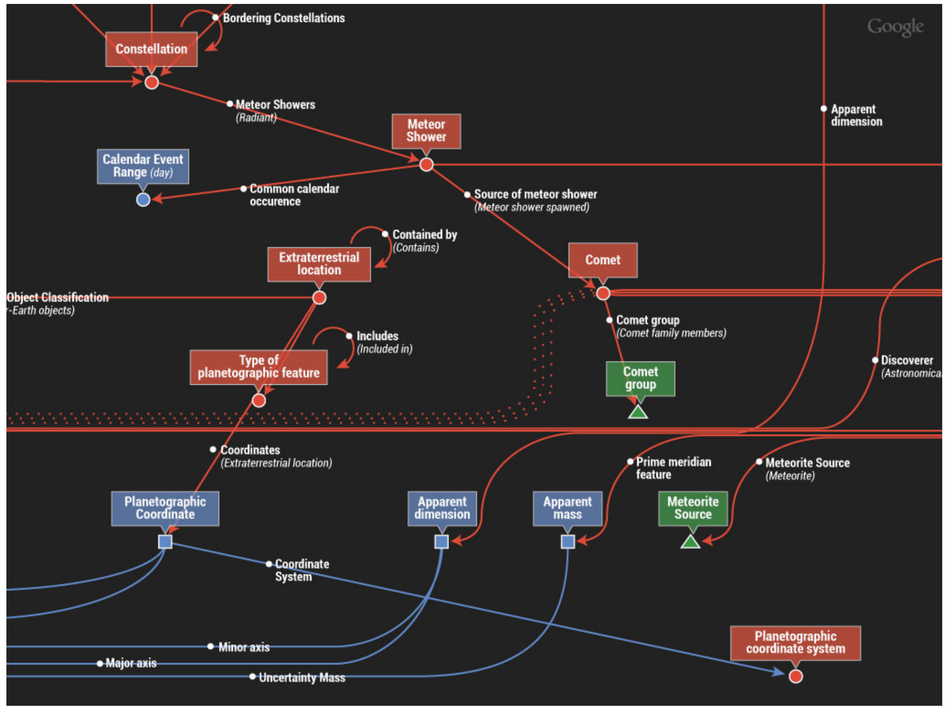
images by Wikipedia

Domain Specific KGs

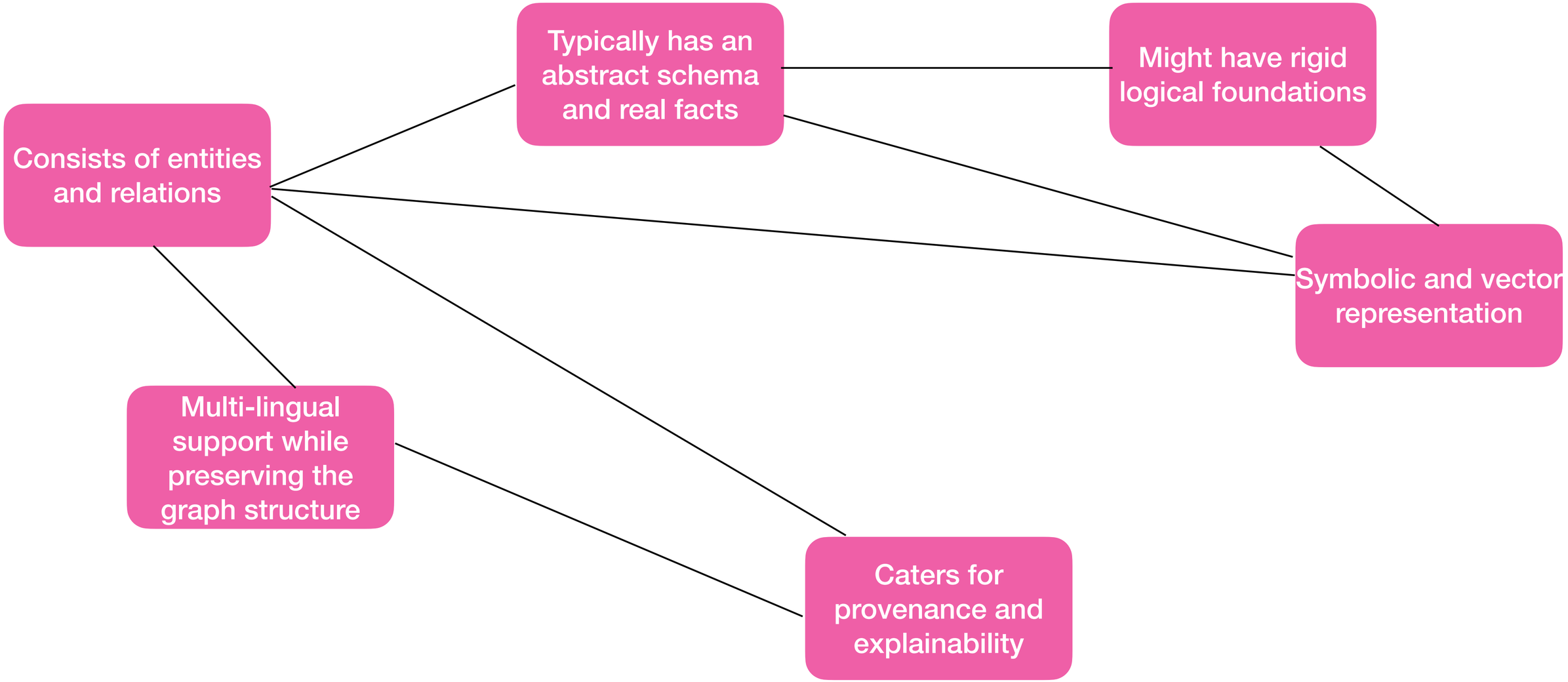
- Other knowledge graphs are specialized in a domain.
  - E.g. WordNet is a linguistic knowledge graph that gathers semantic relations between words (e.g., hypernyms, hyponyms).
- Several knowledge graphs cover the domain of Life Sciences, typically because databases in other formats (e.g., relational databases) are transformed in Linked Open Data:
  - BIO2RDF
- Specific ontologies model (part of) the life sciences domain.
  - Genomic CDS models and reason knowledge about genomic variations of patients to match these patients with appropriate guidelines and clinical decision support messages.
- Often part of dedicated repositories, e.g. Bioportal

What is a knowledge graph

- Given a set of entities E, and relations R, a KG is a directed multi-relational graph G that contains triples of type (s, p, o)
  - $G \subseteq E \times R \times E : (s, p, o) \in G$
  - describes entities and relations
  - defines a schema interrelating arbitrary entities on various topical domains



Characteristics of KG



COMP318

Ontologies and Semantic Web

End Knowledge Graphs - Part 1

Dr Valentina Tamma

V.Tamma@liverpool.ac.uk