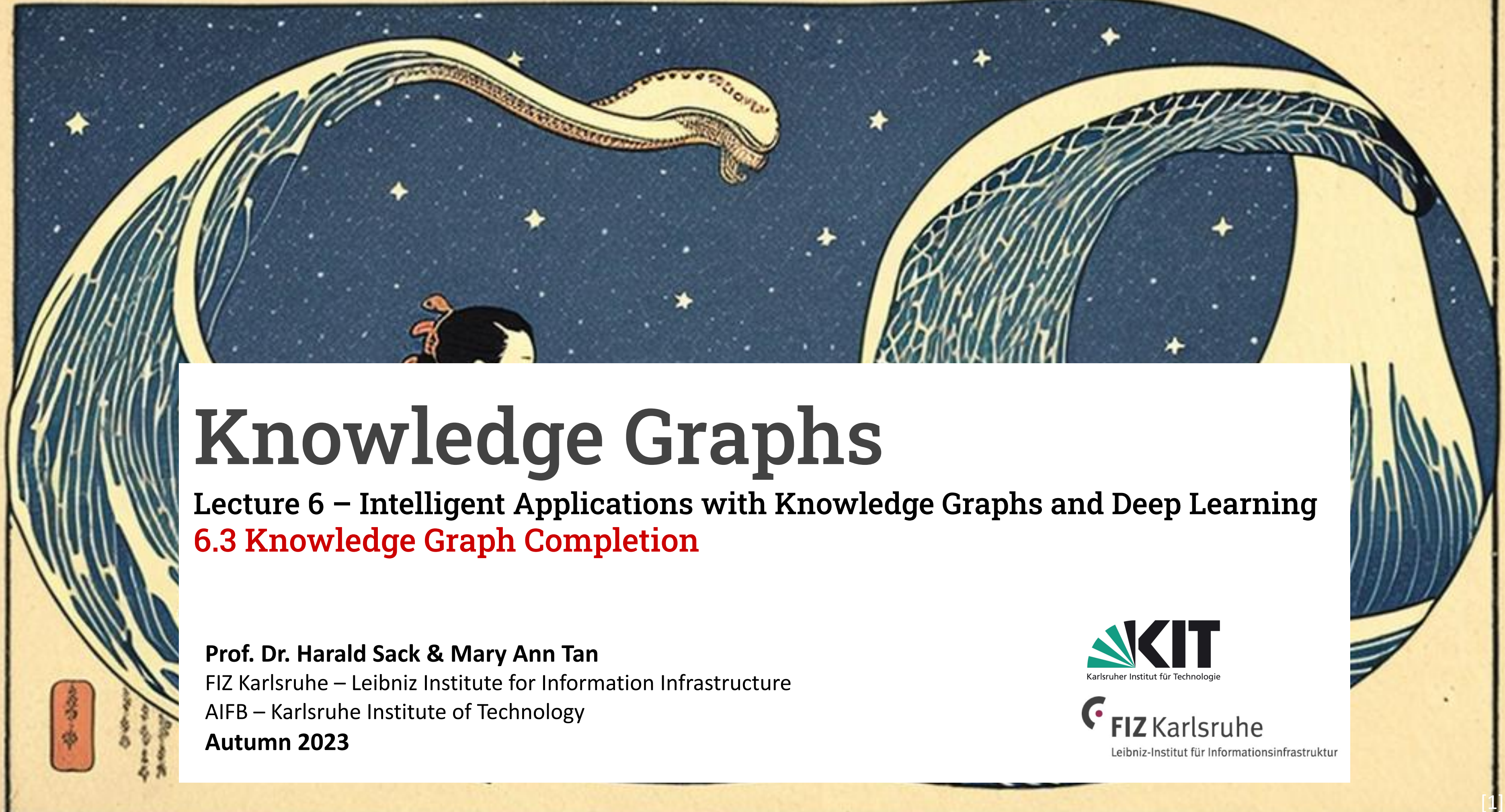


Knowledge Graphs

Lecture 6 – Intelligent Applications with Knowledge Graphs and Deep Learning

6.3 Knowledge Graph Completion

Prof. Dr. Harald Sack & Mary Ann Tan
FIZ Karlsruhe – Leibniz Institute for Information Infrastructure
AIFB – Karlsruhe Institute of Technology

Autumn 2023

Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

[1]

Knowledge Graphs

Lecture 6: Intelligent Applications with Knowledge Graphs and Deep Learning

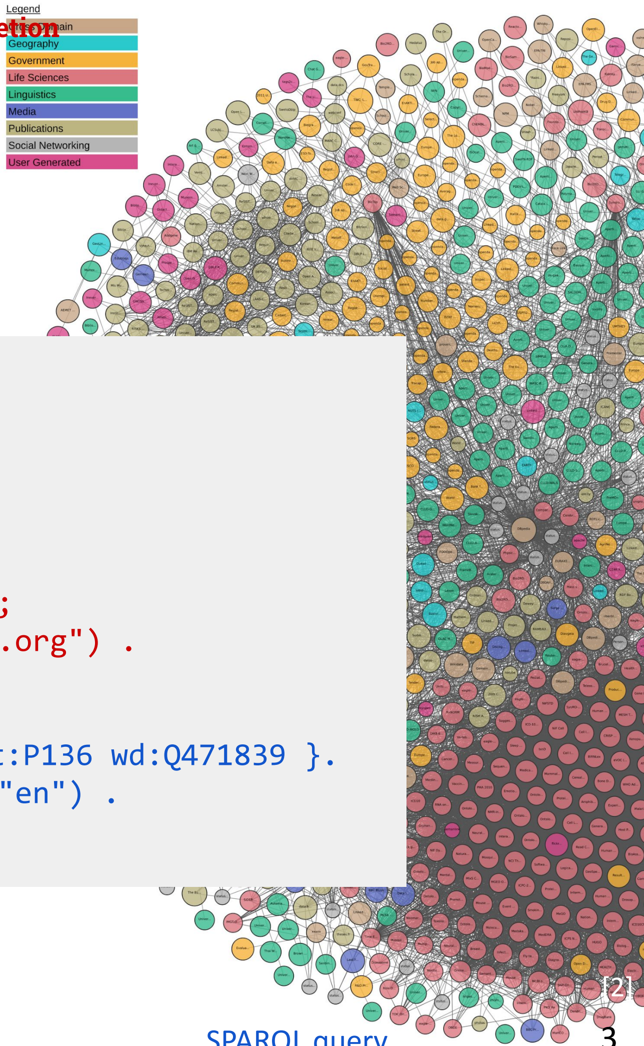
[1]

Can a Knowledge Graph be "complete"?

Check whether all **SciFi films in DBpedia** are also labelled as such in **Wikidata**

```
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX dbc: <http://dbpedia.org/resource/Category:>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT DISTINCT ?wditem ?wditemLabel WHERE {
    SERVICE <http://dbpedia.org/sparql> {
        ?item dct:subject/skos:broader dbc:Science_fiction_films ;
              owl:sameAs ?wditem FILTER regex (?wditem, "wikidata.org") .
    }
    SERVICE <https://query.wikidata.org/sparql> {
       ?wditem wdt:P136 ?occupation FILTER NOT EXISTS {?wditem wdt:P136 wd:Q471839 }.
      ?wditem rdfs:label ?wditemLabel FILTER (LANG(?wditemLabel)="en") .
    }
}
```



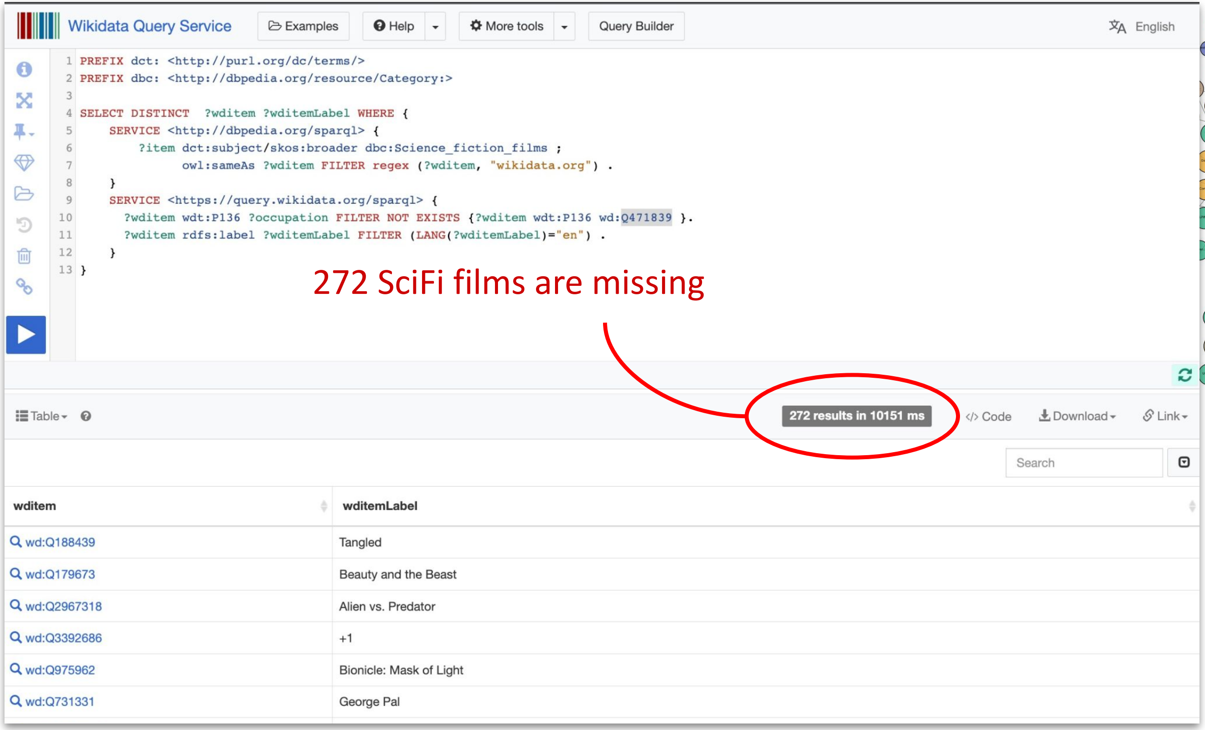
[2]

The Linked Open Data Cloud from lod-cloud.net

[SPARQL query]

Can a Knowledge Graph be “complete”?

Check whether all **SciFi films in DBpedia** are also labelled as such in **wikidata**


Wikidata Query Service | Examples | Help | More tools | Query Builder | English

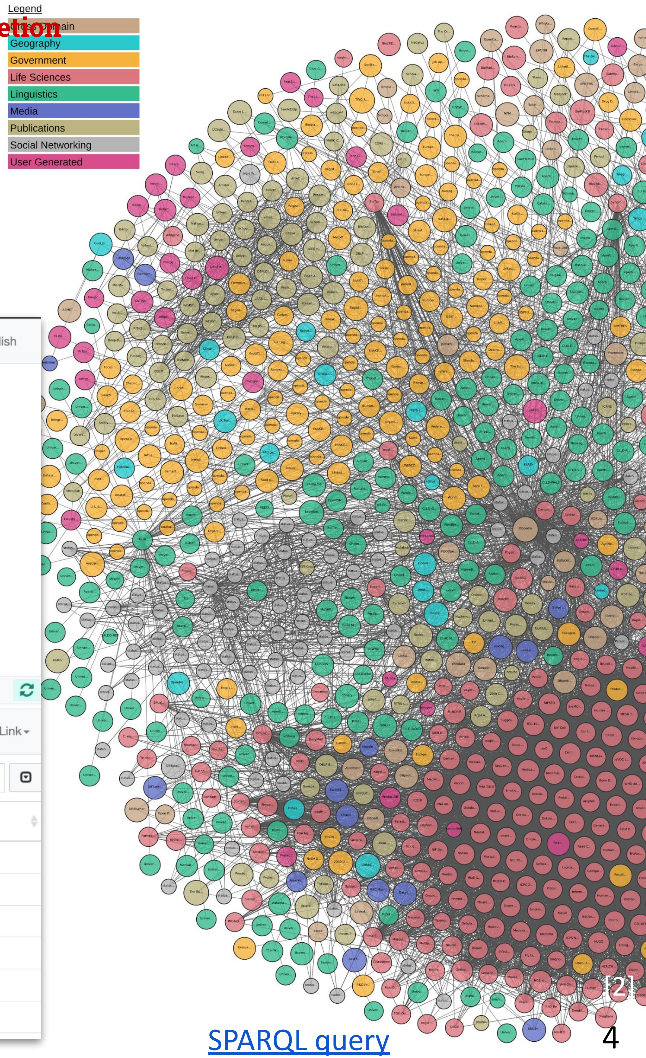
```
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX dbc: <http://dbpedia.org/resource/Category:>

SELECT DISTINCT  ?wditem ?wditemLabel WHERE {
    SERVICE <http://dbpedia.org/sparql> {
        ?item dct:subject/skos:broader dbc:Science_fiction_films ;
              owl:sameAs ?wditem FILTER regex (?wditem, "wikidata.org") .
    }
    SERVICE <https://query.wikidata.org/sparql> {
      ?wditem wdt:P136 ?occupation FILTER NOT EXISTS {?wditem wdt:P136 wd:Q471839 }.
      ?wditem rdfs:label ?wditemLabel FILTER (LANG(?wditemLabel)="en") .
    }
}
```

Table | 272 results in 10151 ms | Code | Download | Link

wditem	wditemLabel
wd:Q188439	Tangled
wd:Q179673	Beauty and the Beast
wd:Q2967318	Alien vs. Predator
wd:Q3392686	+1
wd:Q975962	Bionicle: Mask of Light
wd:Q731331	George Pal

272 SciFi films are missing



[2]

The Linked Open Data Cloud from lod-cloud.net

[SPARQL query](#)

Knowledge Graph Refinement

- As a model of the real world or a part of it, **knowledge graphs cannot reasonably reach full coverage**, i.e., contain information about each and every entity in the universe.
- **It is unlikely,** in particular if heuristic methods are applied to knowledge graph construction, **that the knowledge graph is fully correct**.
- To address those shortcomings, various methods for **Knowledge Graph Refinement** have been proposed, as:
  - Deduplicating entity nodes (entity resolution)
  - Collective reasoning (probabilistic soft logic)
  - **Link prediction** or **Knowledge Graph Completion**
  - Dealing with missing or erroneous values
  - Anything that improves an existing knowledge graph

Graph Completion vs Error Detection

- **Knowledge Graph Completion:**
  Adding missing knowledge to the Knowledge Graph

  E.g. adding a triple:
  <IsaacAsimov, occupation, ScienceFictionWriter>

- **Error Detection:**
  Identifying wrong information in the Knowledge Graph

  E.g. finding inconsistencies:
  <IsaacAsimov, isA, Human>
  <IsaacAsimov, isA, Novel>

Knowledge Graph Completion

- A promising approach for **Knowledge Graph Completion** is
  - to embed Knowledge Graphs into latent spaces (via Knowledge Graph Embeddings) and
  - make inferences by learning and operating on latent representations.

- Such embedding models, however, **do not make use of any rules** during inference, and hence, have limited accuracy.

- E.g. predict that in wikidata the following fact may be complemented:

  (TheMatrix hasGenre ScienceFictionFilm)
  `wd:Q13014087 wdt:P136 wd:Q471839 .`

Tail Prediction

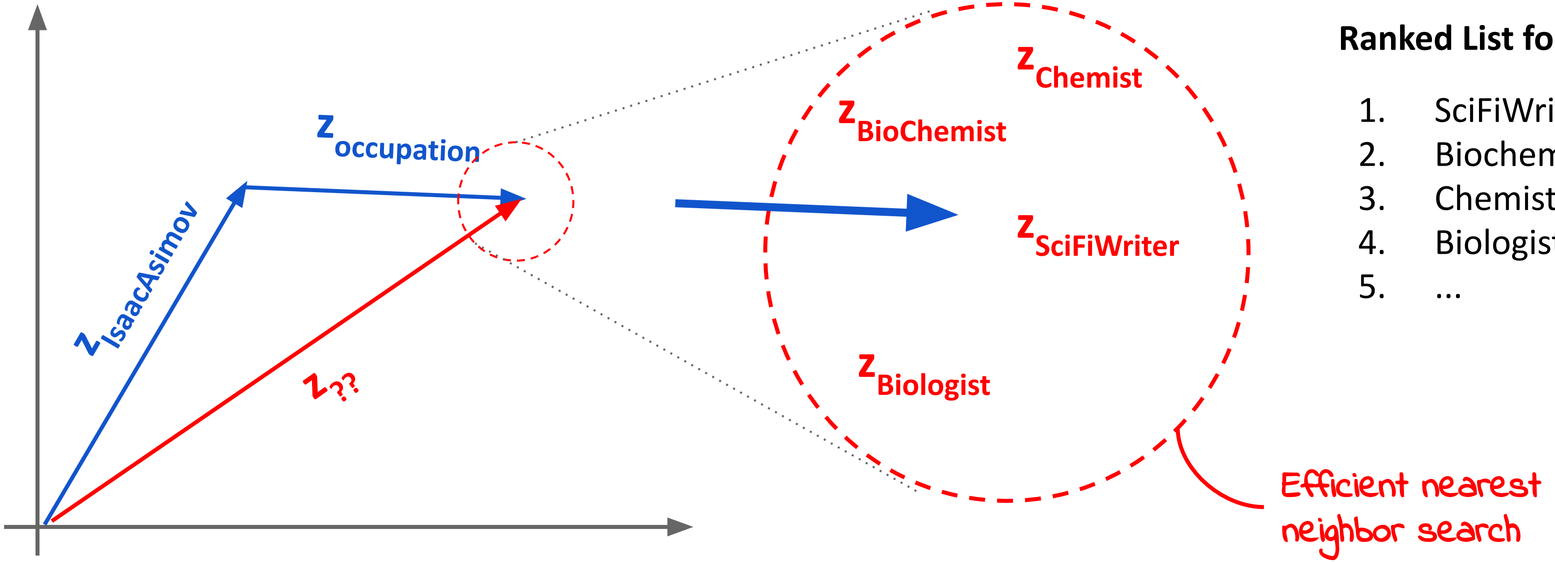
Knowledge Graph Completion

	Task	Example	Result
Link Prediction	**Triple Classification**	(IsaacAsimov, occupation, SciFiWriter)?	(yes, 95%)
	Tail Prediction	(IsaacAsimov, occupation, ?)	(1, SciFi Writer, 0.95), (2, chemist, 0.91) ...
	Head Prediction	(? , occupation, SciFiWriter)	(1, JulesVerne, 0.91) (2, H.G.Welles, 0.90)
	Relation Prediction	(IsaacAsimov, ? , SciFiWriter)	(1, occupation, 0.95)
	Entity Classification (Type Prediction)	(IsaacAsimov, isA, ?)	(1, Person, 0.99) (2, Human, 0.99),...

Link Prediction with KG Embeddings

Use **Translational Embeddings**

- **Unsupervised** methods, e.g. **TransE**, use $z_s + z_p$ to predict z_o
- **Supervised** Methods for prediction, based on embedding vectors



Ranked List for z_o:

1. SciFiWriter, *0.95*
2. Biochemist, *0.89*
3. Chemist, *0.83*
4. Biologist, *0.62*
5. ...

score

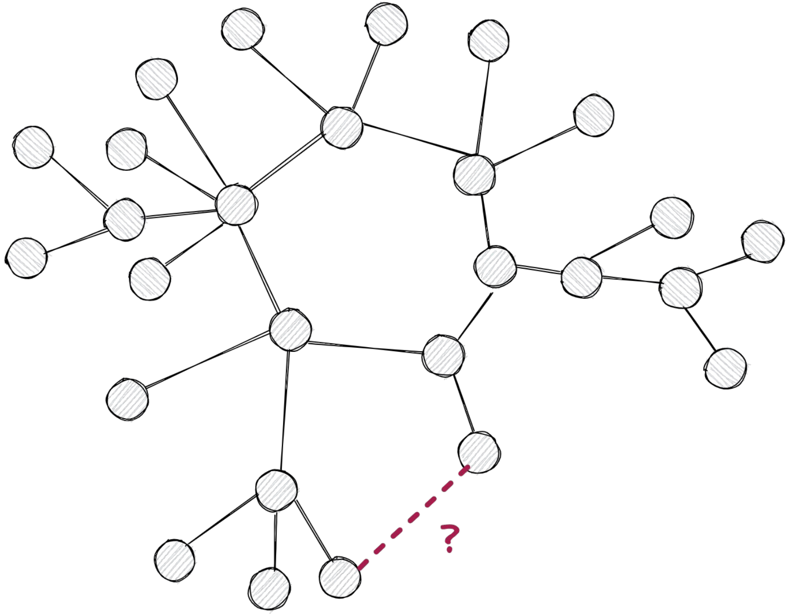
Transductive Link Prediction

Karlsruher Institut für Technologie

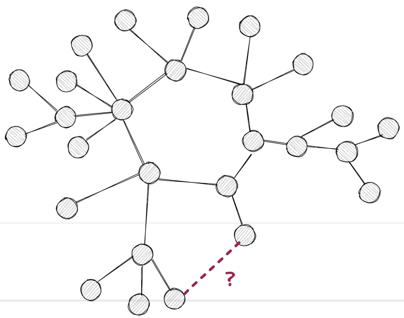
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Transductive Link Prediction:

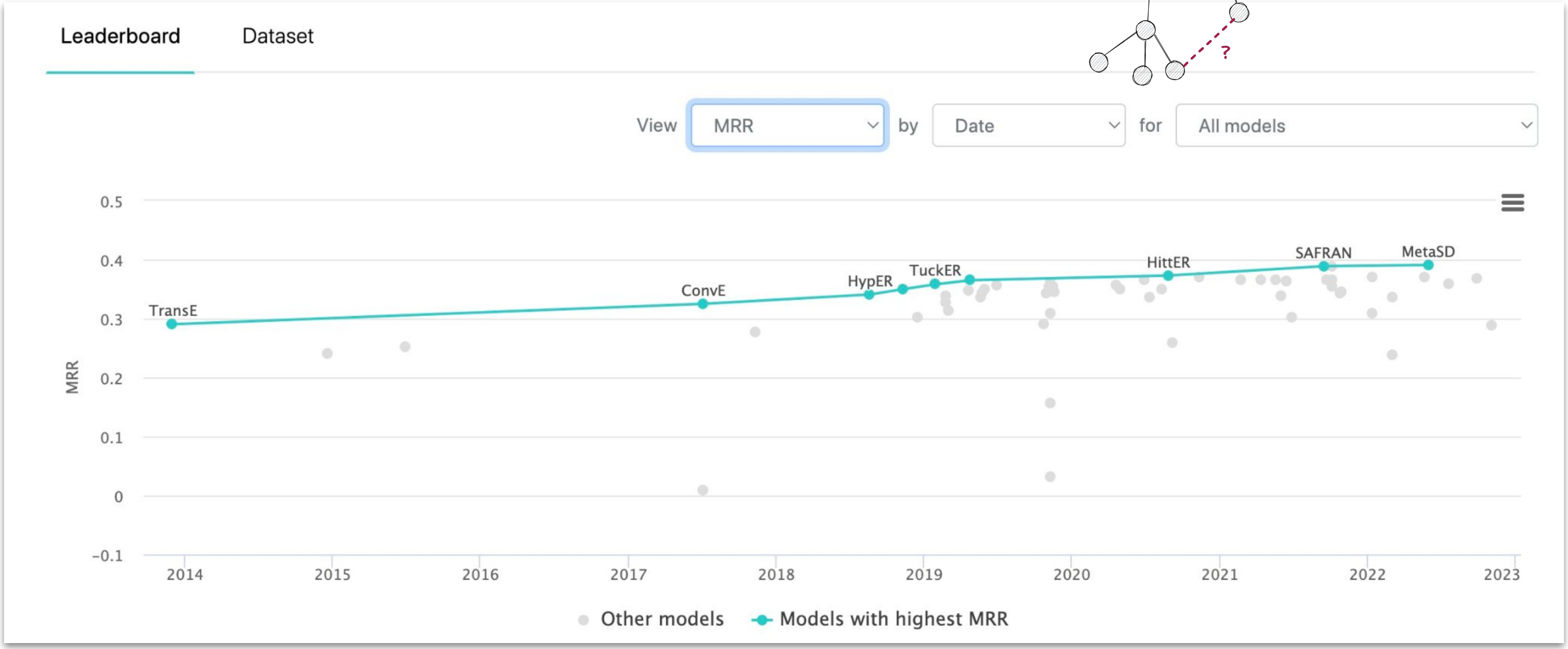
- Predict links in the **same knowledge graph** that has been used as **training data**.
- Entities at **training time** are the same entities used for **prediction**.
- **Cannot operate on unseen entities**, e.g. after a dynamic graph update or in new (sub-)graphs comprised of completely new entities.



Transductive Link Prediction



State-of-the Art in Transductive Link Prediction
(on the example of FB15k-237 benchmark)

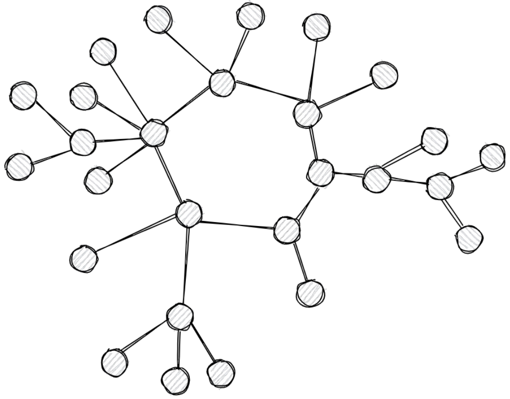


https://paperswithcode.com/

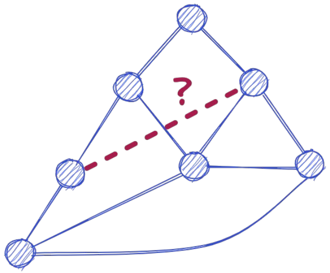
Inductive Link Prediction

Inductive Link Prediction:

- Predict links on a **different knowledge graph** that is not given as **training data**.
- Entities at **training time** are different from entities at **prediction time**.
- **Can operate on unseen entities.**
- We distinguish:
  - Fully-inductive Link Prediction
  - Semi-inductive Link Prediction



Training



Prediction

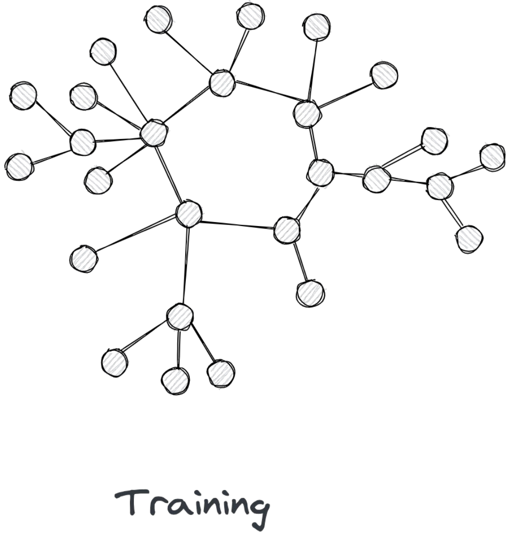
Inductive Link Prediction

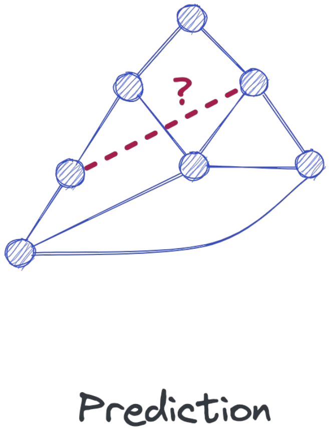
Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Fully-Inductive Link Prediction:

- Prediction knowledge graph is a **new knowledge graph** totally disconnected from the training knowledge graph.
- Link prediction is performed over unseen entities only.
- Pattern: *unseen-to-unseen*

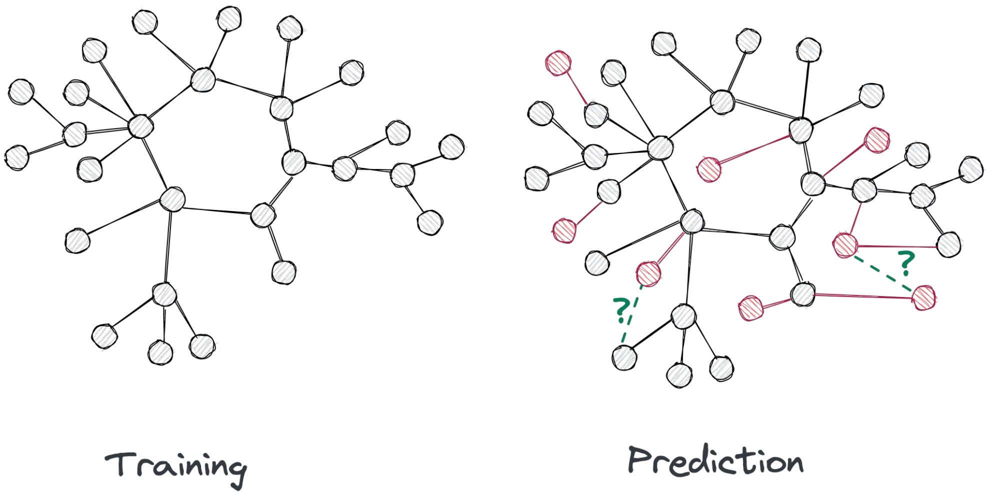




Inductive Link Prediction

Semi-Inductive Link Prediction:

- Prediction knowledge graph is a **larger, updated knowledge graph** that includes and **extends the training knowledge graph**.
- Link prediction can involve both seen and unseen entities.
- Patterns:
  - *seen-to-unseen/unseen-to-seen* and
  - *unseen-to-unseen*



Type Prediction (Entity Classification)

- **Predicting a type or class** for an entity given some characteristics of the entity is a very common problem in machine learning, known as **classification**.

```
<IsaacAsimov, isA, ?>
```

- **Supervised Learning Approach**:
  - Type Prediction can be addressed via a **classification model** based on **labelled training data**,
  - typically the set of entities in a Knowledge Graph which have types attached.

Type Prediction (Entity Classification)

- **Multi-Class Prediction:**
  In Knowledge Graphs, the choice of types/classes for prediction is usually more than two..
  E.g.: SciFiAuthor, Chemist, Climatologists, etc.

- **Single-Label Classification:**
  Only one type/class can be assigned per entity.
  E.g.: <IsaacAsimov, isA, Person>

- **Multi-Label Classification:**
  In Knowledge Graphs, some entities might allow for the assignment of more than one type.
  E.g.: <electron, isA, Particle> and
  <electron, isA, Wave>

Further Applications

Karlsruher Institut für Technologie

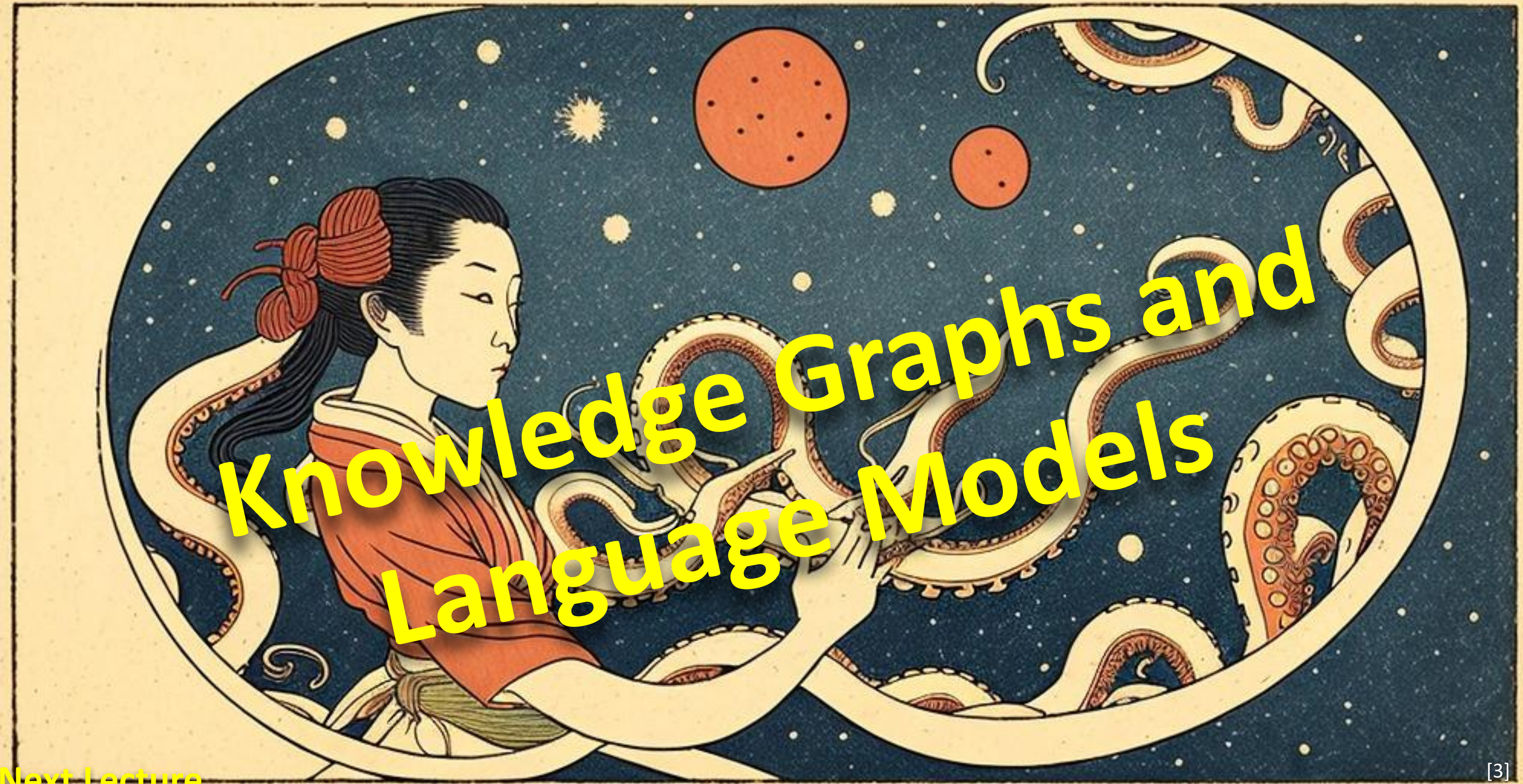
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

- **Identity Prediction**
  Predicting the identity of two entities, i.e. searching for nodes in the knowledge graph that refer to the same entity, but are not stated or entailed to be the same (aka entity matching, record linkage, deduplication, etc.).

- **Fact Checking/Validation**
  Predicting the plausibility of a given fact (RDF Triple Classification).

- **Knowledge Graph Correction**
  Identifying wrong information in the knowledge graph (e.g. via inconsistency detection) and thereby complementing missing information (Knowledge Graph Completion).

Knowledge Graphs and Language Models



Next Lecture...

[3]

Bibliographic References:

- Costabello, L. et al (2020), *ECAI-20 Tutorial: Knowledge Graph Embeddings: From Theory to Practice*.
- Michail Galkin (2022), *Inductive Link Prediction in Knowledge Graphs*, in TowardsDataScience.

Picture References:

[1] “On this colorized woodcut in the style of Hokusai we see a pensive woman together together giant octopus who melancholically entangles the knowledge graph that extends into the vast empty space of the universe to the galaxies and stars.”, created via ArtBot, Deliberate, 2023, [CC-BY-4.0], https://tinybots.net/artbot

[2] John P. McCrae, The Linked Open Data Cloud, [CC-BY-4.0], https://lod-cloud.net/

[3] “On this colorized woodcut in the style of Hokusai we see a pensive woman together together giant octopus who melancholically entangles the knowledge graph that extends into the vast empty space of the universe to the galaxies and stars.”, created via ArtBot, Deliberate, 2023, [CC-BY-4.0], https://tinybots.net/artbot

[3]