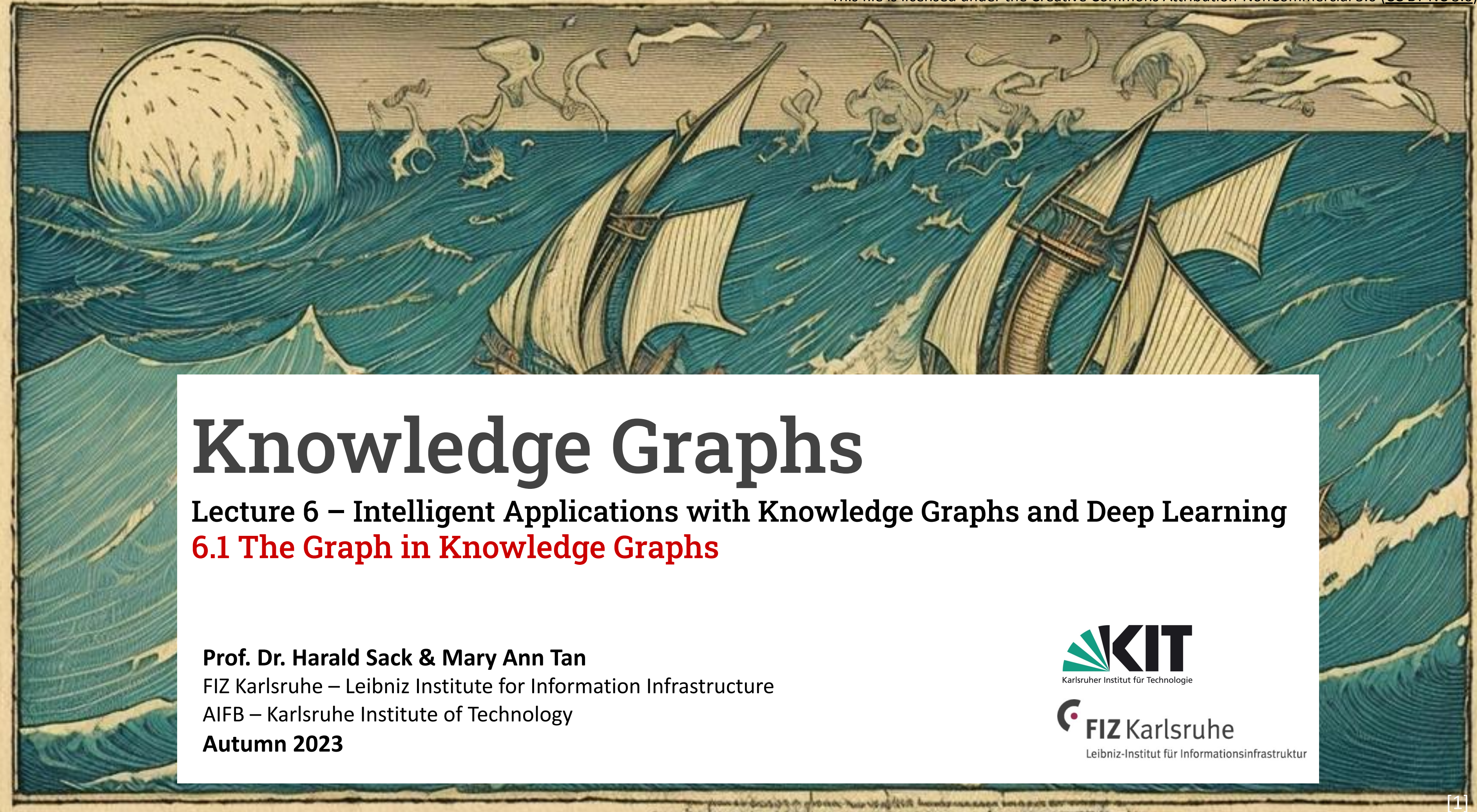


Knowledge Graphs

Lecture 6 – Intelligent Applications with Knowledge Graphs and Deep Learning

6.1 The Graph in Knowledge Graphs

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

Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

6.1 The Graph in Knowledge Graphs

Excursion 8: Distributional Semantics and Language Models

6.2 Knowledge Graph Embeddings

6.3 Knowledge Graph Completion

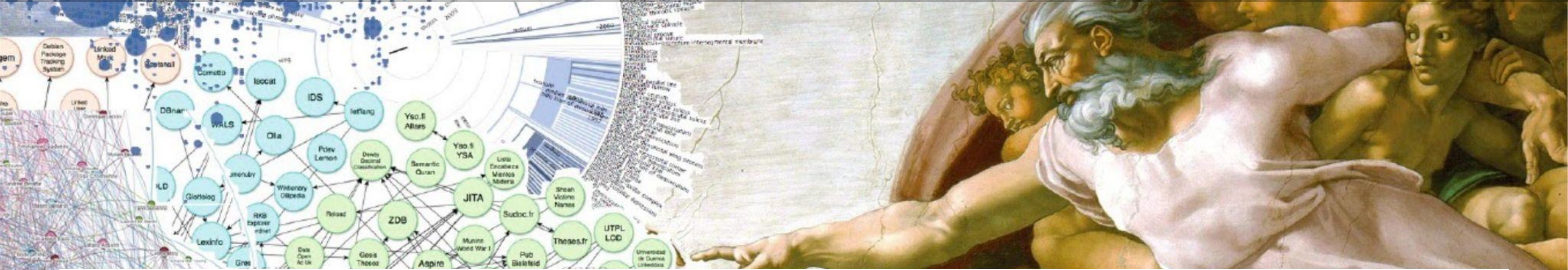
6.4 Knowledge Graphs and Language Models

6.5 Semantic Search

6.6 Exploratory Search and Recommender Systems

[1]

Knowledge Graph Recap

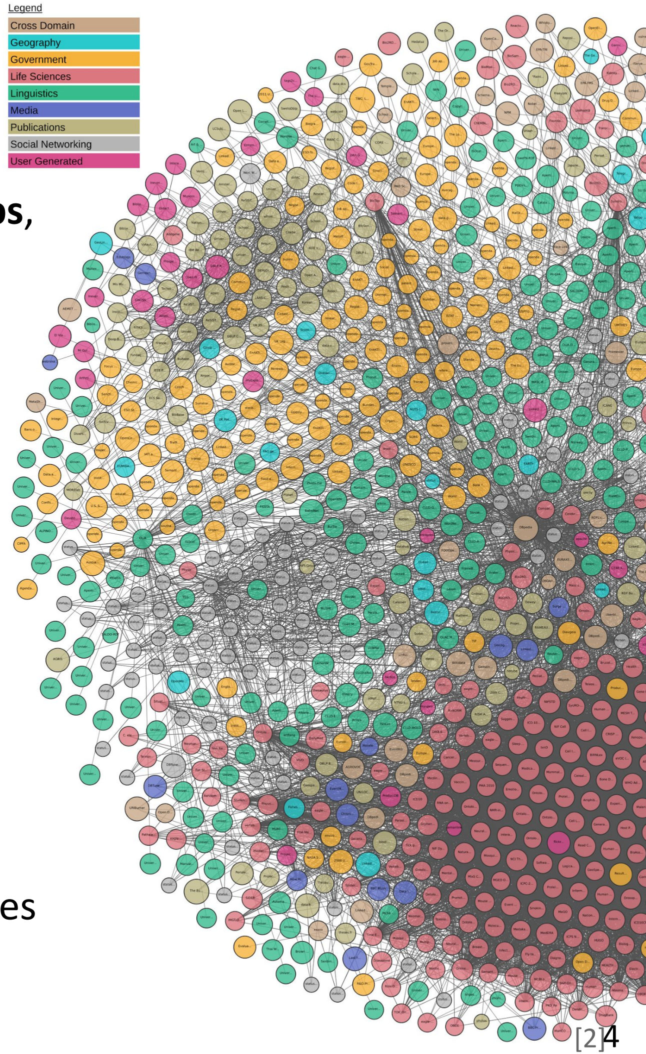


A knowledge graph

1. mainly describes real world entities and their interrelations, organized in a graph,
2. defines possible classes and relations of entities in a schema,
3. allows for potentially interrelating arbitrary entities with each other and
4. covers various topical domains.

Knowledge Graphs

- A **Graph** consisting of **concepts**, **classes**, **properties**, **relationships**, and **entity descriptions**
- Based on **formal knowledge representations** (RDF(S), OWL)
- Data can be **open** (e.g. DBpedia, Wikidata), **private** (e.g. supply chain data), or **closed** (e.g. product models)
- Data can be **original**, **derived**, or **aggregated**
- We distinguish
  - **instance data** (ground truth),
  - **schema data** (vocabularies, ontologies)
  - **metadata** (e.g. provenance, versioning, licensing)
- **Taxonomies** are used to categorize entities
- **Links** exist between internal and external data
- Including **mappings** to data stored in other systems and databases
- *Fully compliant with **FAIR Data principles***



The Linked Open Data Cloud from lod-cloud.net

[2]

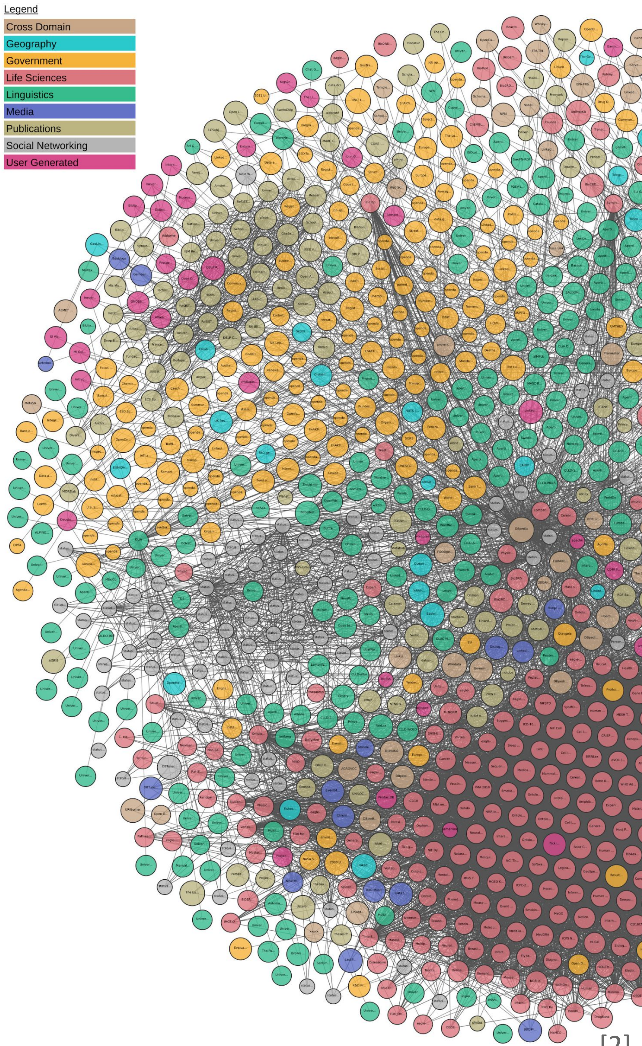
A Knowledge Graph is a Knowledge Base that is a **Graph**.

Definition 1.1

A **simple directed graph G=(V,E)** consists of a set **V** of **vertices**, |V|=n, and a set **E** of **directed edges**, $E \subseteq V \times V$, where each edge $e_i=(v_k, v_l)$, $e_i \in E$ is an ordered pair of two vertices (v_k, v_l) with $v_k, v_l \in V$.

Definition 1.2

- A **graph with self-loops** is a graph extended with the option of having edges that relate a vertex to itself.
- A **multi-graph** is a graph that may have multiple edges with the same vertices.
- An **edge-labelled graph** is a graph that has an additional **labelling function** $\lambda : E \rightarrow L$ that maps each edge in E to an element in a set of labels L (similarly for vertex-labelled graphs).



The Linked Open Data Cloud from lod-cloud.net

[2]

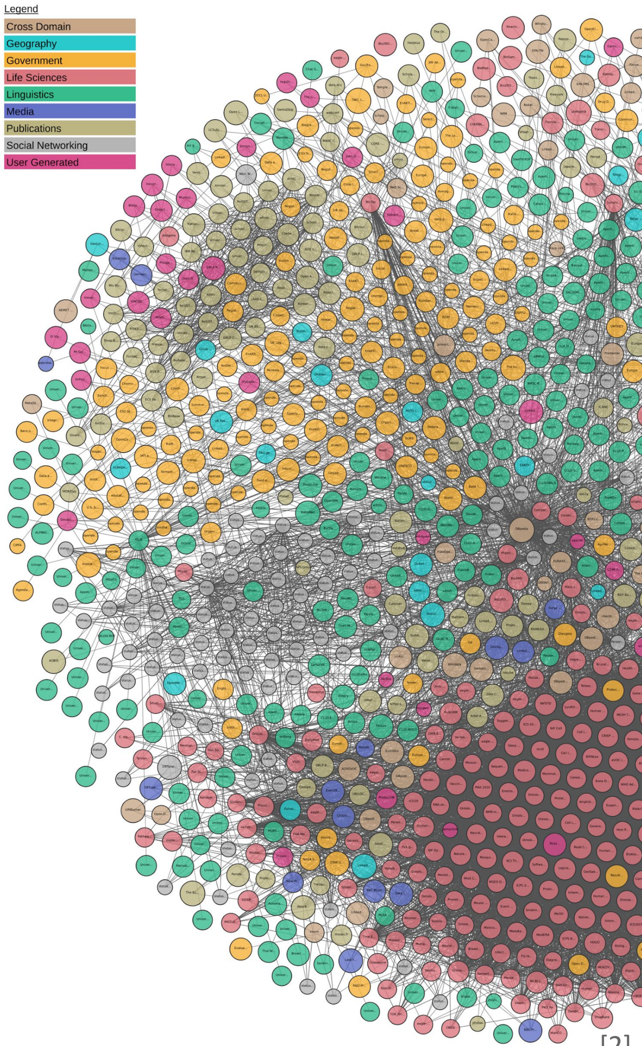
Graphs (contd.)

Definition 1.3

- An edge is said to be **incidental** to the vertices it connects.
- The **degree** of a vertex is the number of edges that are incidental to it.
- In a directed graph, the **in-degree** of a vertex is the number of edges pointing towards it; analogously for **out-degree**.

Definition 1.4

- A **directed path** in a directed graph is a sequence of consecutive edges $(e_1, e_2, \dots, e_n)$ with $e_i=(v_l,v_k)$ and $e_{i+1}=(v_k,v_m)$.
- A directed graph is **strongly connected** if there is a directed path from any vertex to every other vertex.



[2]

The Linked Open Data Cloud from lod-cloud.net

How Can You Characterize a Knowledge Graph?

Karlsruher Institut für Technologie
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

"Should I use Knowledge Graph A or Knowledge Graph B to solve my problem?"

- How to compare two Knowledge Graphs?
  - Size
  - Coverage
  - Completeness
  - Level of Detail
  - Accuracy
  - Reliability
  - etc.

- **Idea**: **Structural Comparison** by just comparing the Graphs

[2]

Graph Centrality Measures

- **Network analysis** has developed methods for **finding the most important vertices in a graph**.
- **Vertex importance** based on the structure of such graphs is called **centrality.**
- But, what makes a node important?

[2]

What makes a Node important?

- Many networks can be considered to describe a **flow** of something (goods, information, etc.)
- A node might be **important**, if
  - a lot flows from it (in a supply chain),
  - to it (in a network of links), or
  - through it (in a communication network)
- Flow might be modelled by (weighted) paths, possibly factoring in their length and/or number
- Paths might be more important if they pass through important nodes
- In knowledge graphs, the importance of edges and nodes may also depend on more complex features (e.g., edge or vertex labels)

[2]

What makes a Node important?

Wikidata Example:

- A Wikidata entity (node) might be important, if it is referenced by many Wikipedia pages
- Who is the most important Science Fiction author?

```
SELECT ?authorLabel (SUM(?link) AS ?importance)
WHERE {
  ?author wdt:P106 wd:Q18844224 .
  ?author wikibase:sitelinks ?link.
  ?author rdfs:label ?authorLabel FILTER (lang(?authorLabel)="en")
} GROUP BY ?authorLabel
ORDER BY DESC(?importance)
```

[2]

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

[Link to Query]

```
SELECT ?authorLabel (SUM(?link) AS ?importance)
WHERE {
  ?author wdt:P106 wd:Q18844224 .
  ?author wikibase:sitelinks ?link.
  ?author rdfs:label ?authorLabel FILTER (lang(?authorLabel)="en")
} GROUP BY ?authorLabel
ORDER BY DESC(?importance)
```

Table | 5208 results in 3356 ms | Code | Download | Link

Search

authorLabel	importance
Mark Twain	236
Voltaire	234
Edgar Allan Poe	206
Rudyard Kipling	176
Bertrand Russell	176
Arthur Conan Doyle	174
Jules Verne	174
Steven Spielberg	158
H. G. Wells	157

Degree Centrality

- A simple form of centrality restricts to incoming/outgoing paths of length one

Definition 1.5

- The **in-degree centrality** of a directed graph is given by the in-degree of each node.
- The **out-degree centrality** and the **degree centrality** (for undirected graphs) are defined analogously.

- There are more sophisticated forms of centrality, as:

  Eigenvector centrality, Katz centrality, PageRank, etc.

More Centrality Measures

- Further Measures to characterize a Knowledge Graph
  - Sizes
    - number of nodes
    - number of facts
    - average number of facts per node
  - KG diameter

Definition 1.6

- The **eccentricity** of a node is the maximal distance between a certain node and any other node.
- The **diameter** of a graph is the maximum **eccentricity** of a graph, i.e. the greatest distance between any pair of nodes.
- To find the diameter of a graph, first find the **shortest path** between each pair of nodes. The greatest length of any of these paths is the **diameter of the graph**.

[2]

More Centrality Measures

- Further Measures to characterize a Knowledge Graph
  - Sizes
    - number of nodes
    - number of facts
    - average number of facts per node
  - KG diameter
  - KG radius

Definition 1.7

The **radius** of a graph is the minimum eccentricity of a graph, i.e. the shortest of the maximum distances between any pair of nodes.

[2]

More Centrality Measures

Further (structural) measures to characterize a Knowledge Graph:

- Sizes
  - number of nodes
  - number of facts
  - average number of facts per node
- KG diameter
- KG radius
- average in/out degree
- average path length
- and many more...

[2]

Important Nodes in Knowledge Graphs

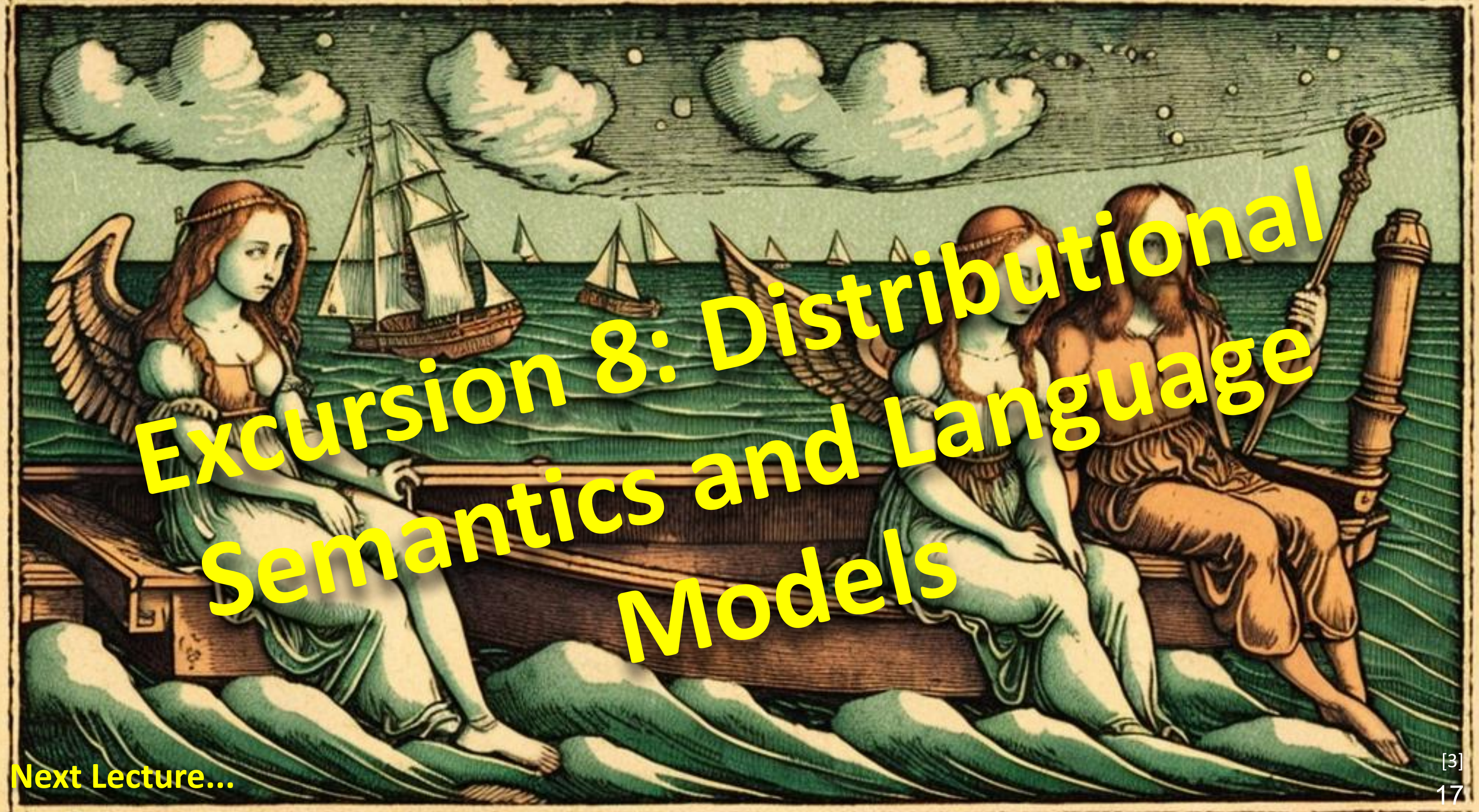
Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

In **Knowledge Graphs**, the importance of nodes might further depend on:

- the properties (i.e. edge attributes)
- the node labels (i.e. further attributes of nodes)
- specific nodes or edges might be ignored, e.g.
  - Basically for every entity in a (OWL encoded) knowledge graph the following fact holds:
    `:entity rdf:type owl:Thing`
  - Therefore, we might ignore this fact if we want to determine the importance of nodes

[2]


Excursion 8: Distributional Semantics and Language Models
Next Lecture...
[3]

Bibliographic References:

Diestel, R. (2010). *Graph Theory (Vol. 173)*. Heidelberg; New York: Springer.
Chap. 1 Basics, pp. 1–17.

Picture References:

[1] “On this colorized Renaissance woodcut we see two sailing ships driven towards the edge of flat Earth. Underneath the waves there lures a fierce dragon. The ocean's waters are pouring down from the edge of flat Earth..”, 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 Albrecht Dürer we see a pensive cupid together with a beautiful female angel, bothe are melancholically watching two sailing ships on the vast ocean of flat Earth driven towards the edge of the world, where the waters are pouring down in the empty universe.”, created via ArtBot, Deliberate, 2023, [CC-BY-4.0], https://tinybots.net/artbot