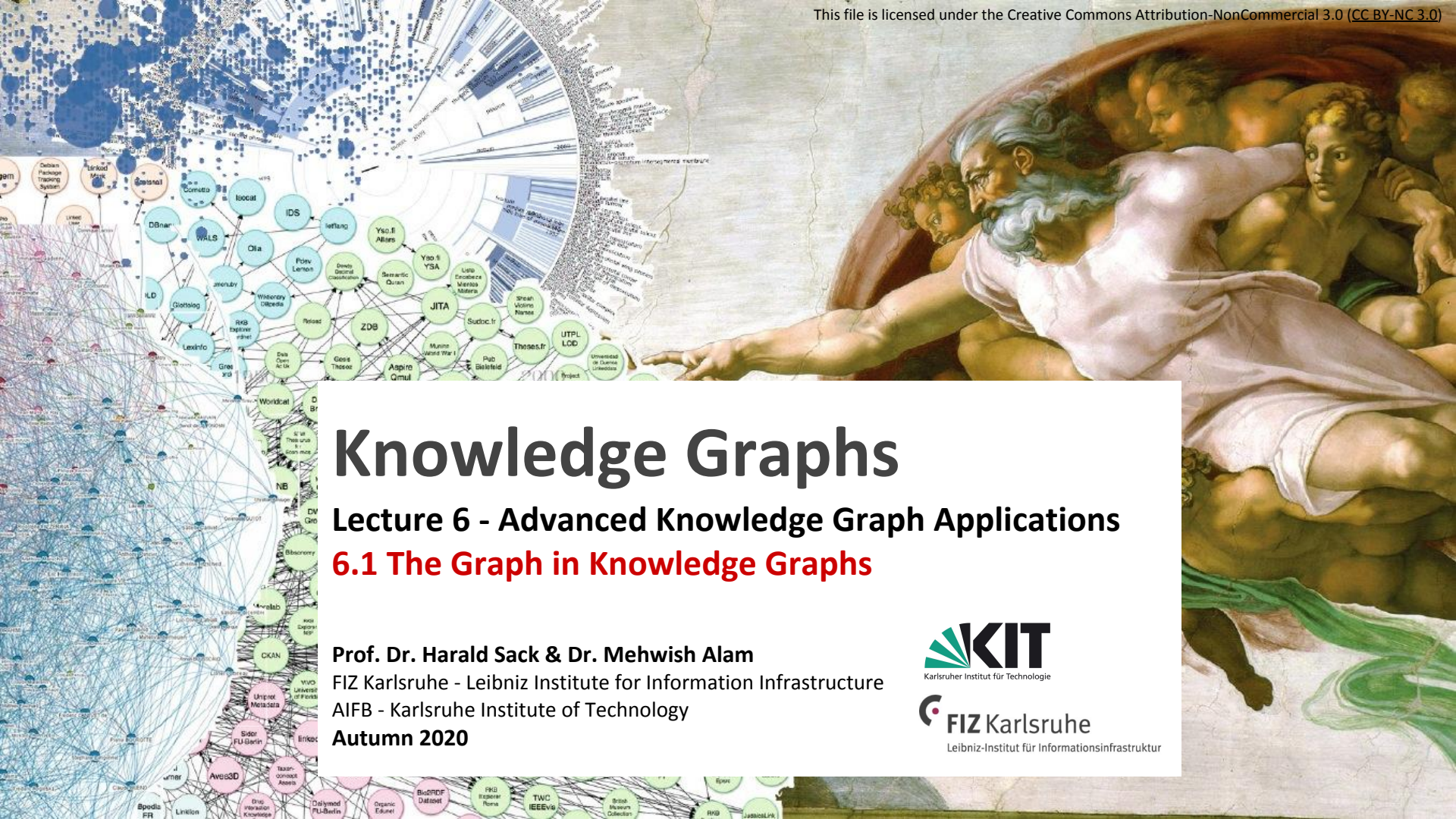


Knowledge Graphs

Lecture 6 - Advanced Knowledge Graph Applications

6.1 The Graph in Knowledge Graphs

Prof. Dr. Harald Sack & Dr. Mehwish Alam
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure
AIFB - Karlsruhe Institute of Technology
Autumn 2020

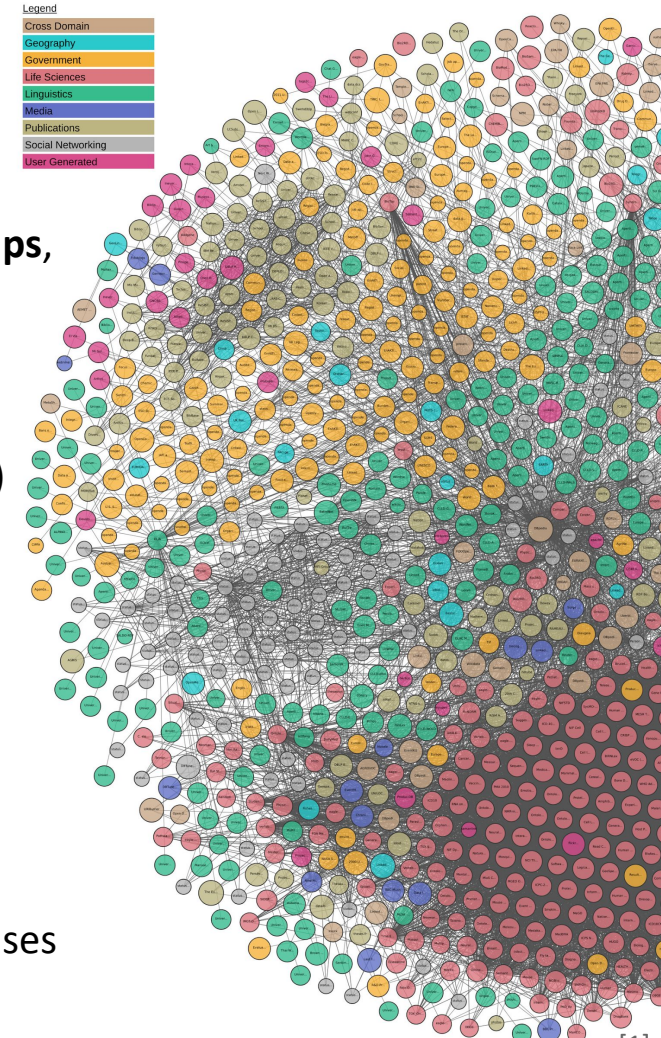
Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Knowledge Graphs

Lecture 6: Advanced Knowledge Graph Applications

Knowledge Graph Recap



- 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 to **FAIR Data principles***

The Linked Open Data Cloud from lod-cloud.net

[1]

Knowledge Base Definition

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

A **knowledge base** (**KB**) is a technology used to store complex structured and unstructured information used by a computer system. The initial use of the term was in connection with expert systems which were the first knowledge-based systems.

Wikipedia

knowledge base

Free Online Dictionary of Computing

‹*artificial intelligence*›

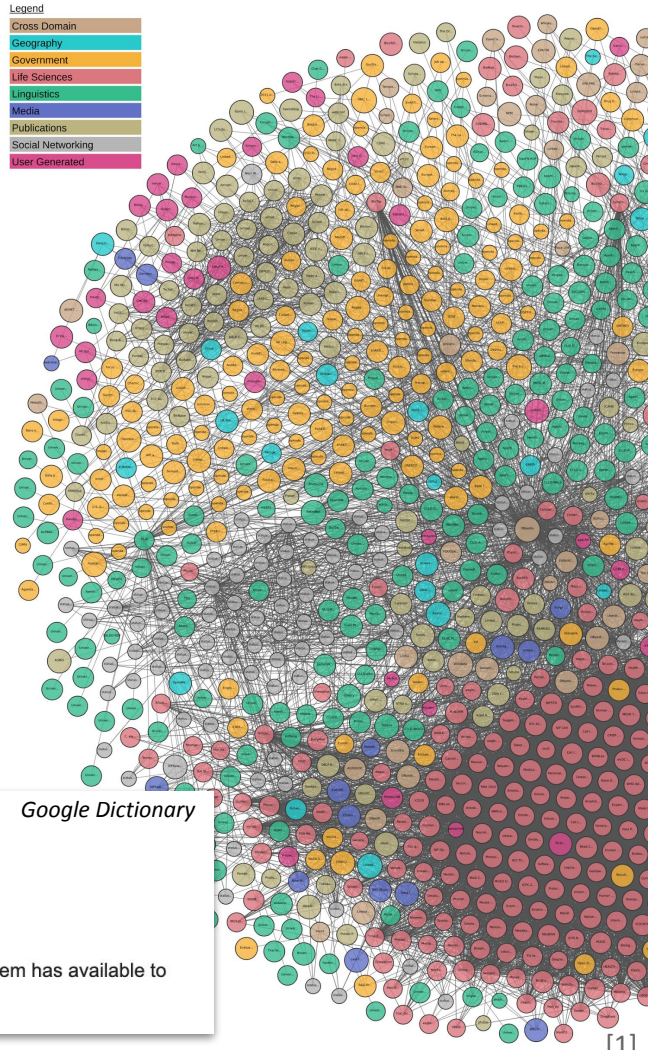
A collection of knowledge expressed using some formal knowledge representation language. A knowledge base forms part of a knowledge-based system (KBS).

knowledge base

Google Dictionary

noun

1. a store of information or data that is available to draw on.
2. the underlying set of facts, assumptions, and rules which a computer system has available to solve a problem.



[1]

Graph Definition

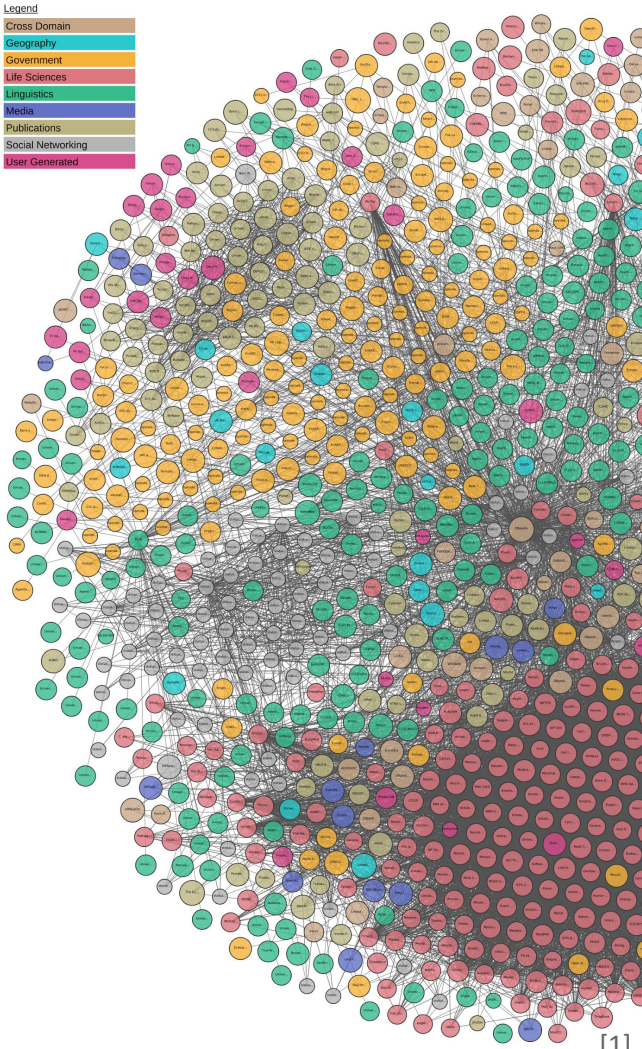
A Knowledge Graph is a Knowledge Base that is a **Graph**.

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).



[1]

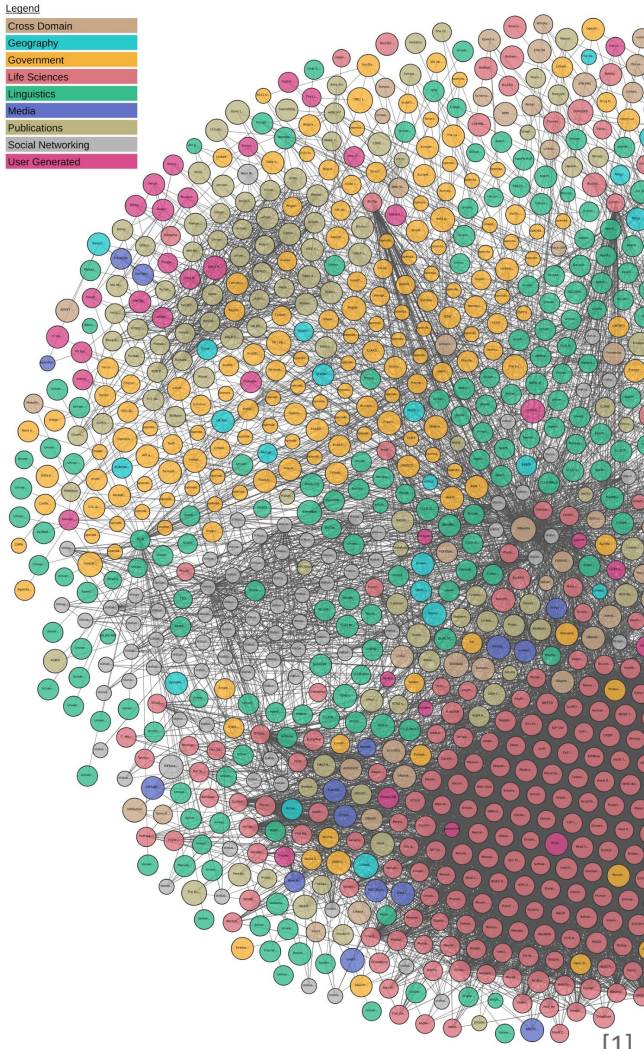
Graph Definition (cont.)

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 $(e_1, e_2, \ldots, e_n)$ 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 any other vertex.



[1]

How Can You Characterize a Knowledge Graph?

"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

[1]

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?

[1]

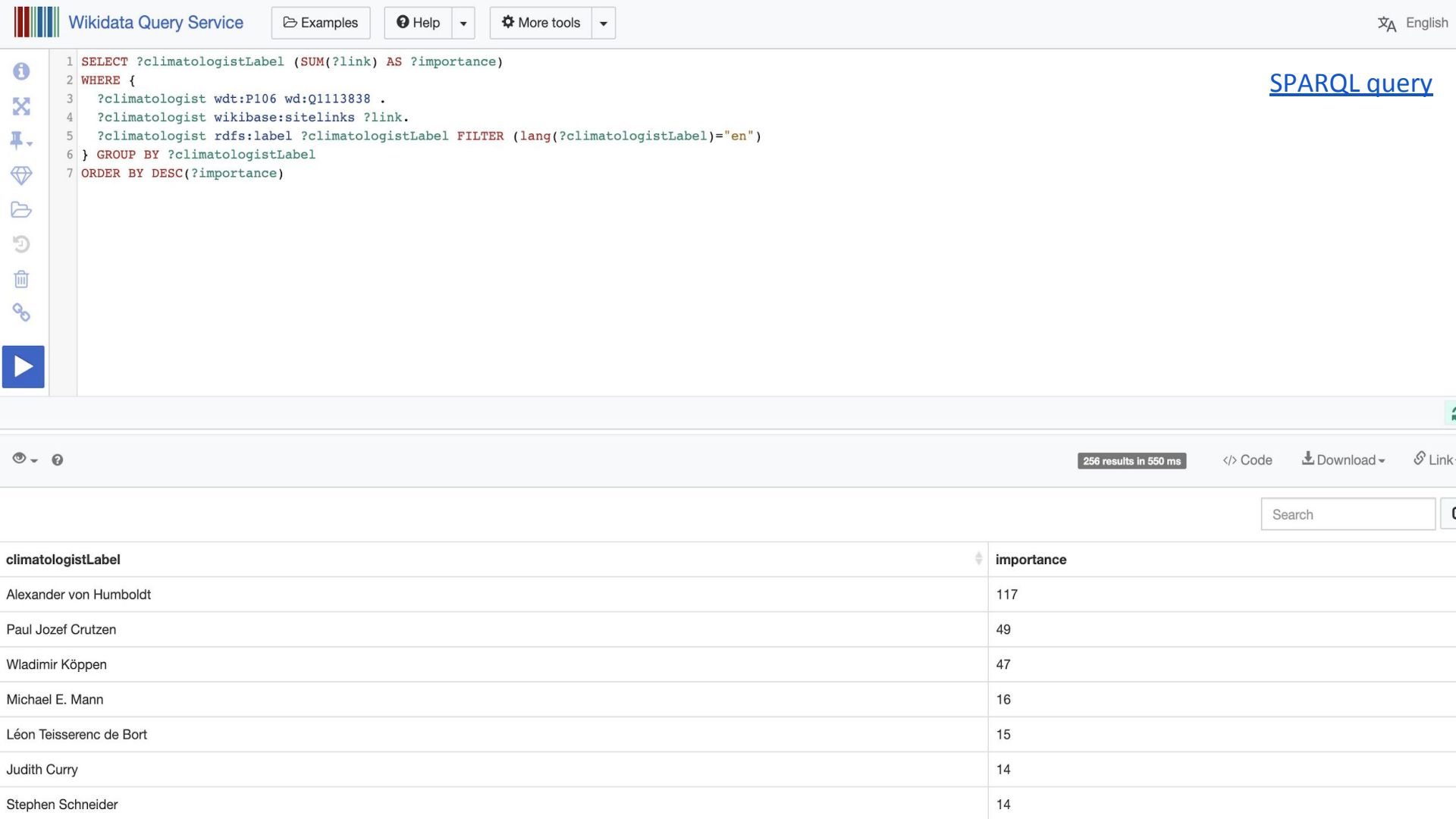
What makes a Node important?

- 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)

What makes a Node important?

- **Wikidata Example:**
  - A Wikidata entity (node) might be important, if it is referenced by many Wikipedia pages
  - what are the most important Climatologists?

```
SELECT ?climatologistLabel (SUM(?link) AS ?importance)
WHERE {
  ?climatologist wdt:P106 wd:Q1113838 .
  ?climatologist wikibase:sitelinks ?link.
  ?climatologist rdfs:label ?climatologistLabel
   FILTER (lang(?climatologistLabel)="en")
} GROUP BY ?climatologistLabel
ORDER BY DESC(?importance)
```



climatologistLabel	importance
Alexander von Humboldt	117
Paul Jozef Crutzen	49
Wladimir Köppen	47
Michael E. Mann	16
Léon Teisserenc de Bort	15
Judith Curry	14
Stephen Schneider	14

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 e.g.
  - Eigenvector centrality, Katz centrality, PageRank, etc.

Further Centrality Measures

- Further Measures to characterize a Knowledge Graph
  - Sizes
    - number of nodes
    - number of facts
    - avg 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**.

[1]

Further Centrality Measures

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

Definition 1.8

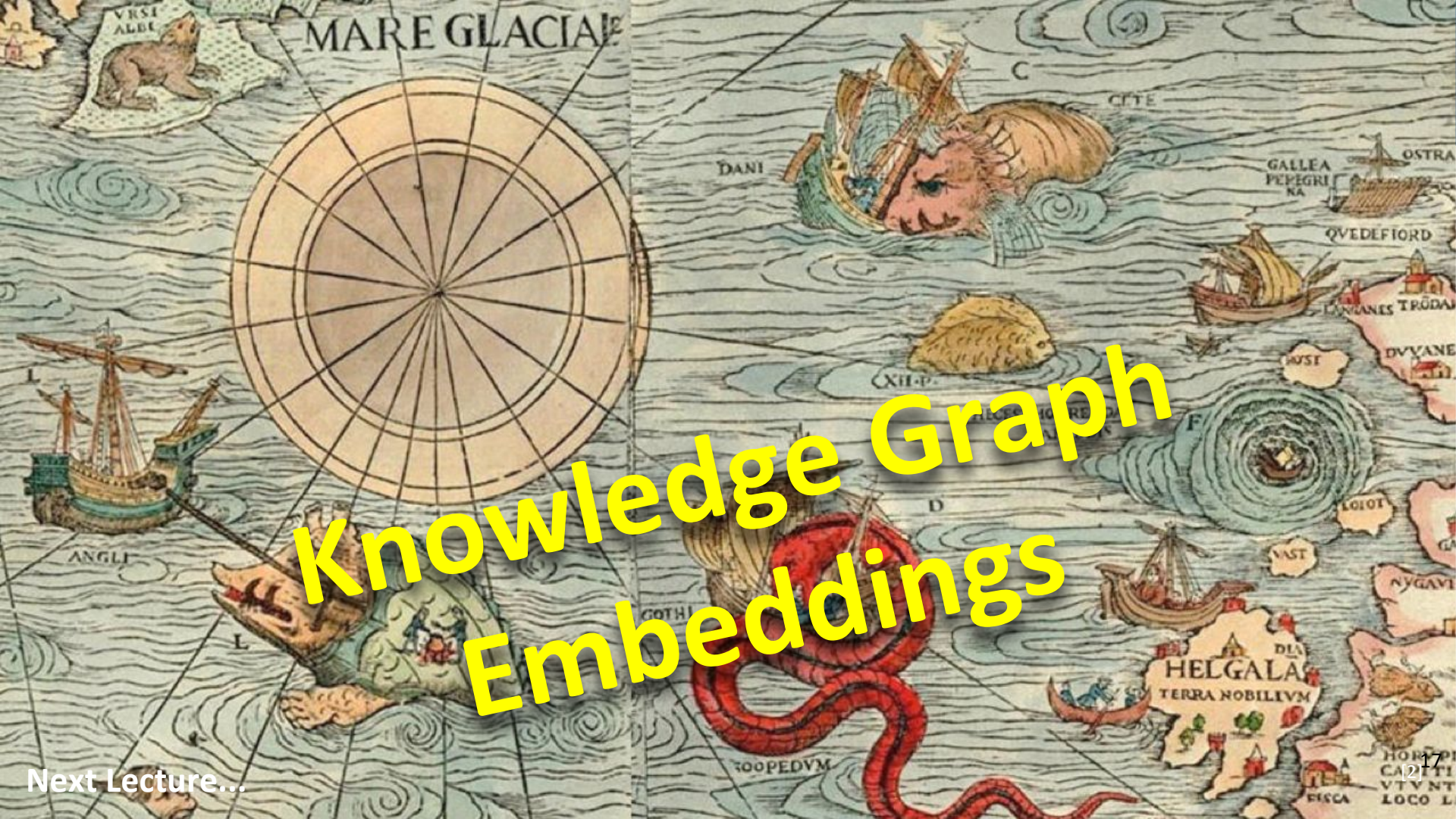
- The **radius** of a graph is the minimum eccentricity of a graph, i.e. the shortest of the maximum distances between any pair of nodes.

Further Centrality Measures

- Further (structural) measures to characterize a Knowledge Graph:
  - Sizes
    - number of nodes
    - number of facts
    - avg number of facts per node
  - KG diameter
  - KG radius
  - avg in/out degree
  - avg path length
  - and many more...

Knowledge Graphs and Important Nodes

- In **Knowledge Graphs**, the importance of nodes might further be depending on
  - the properties (i.e. edge attributes)
  - the node labels (i.e. further attributes of nodes)
  - specific nodes or edges might be ignored, as 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



Next Lecture...

Karlsruher Institut für Technologie
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Picture References:

- [1] John P. McCrae, The Linked Open Data Cloud, [CC-BY-4.0]
  https://lod-cloud.net/
- [12] Carta marina, a wallmap of w:Scandinavia, by Olaus Magnus. The caption reads : Marine map and Description of the Northern Lands and of their Marvels, most carefully drawn up at Venice in the year 1539 through the generous assistance of the Most Honourable Lord Hieronymo Quirino.
  https://en.wikipedia.org/wiki/File:Carta_Marina.jpeg