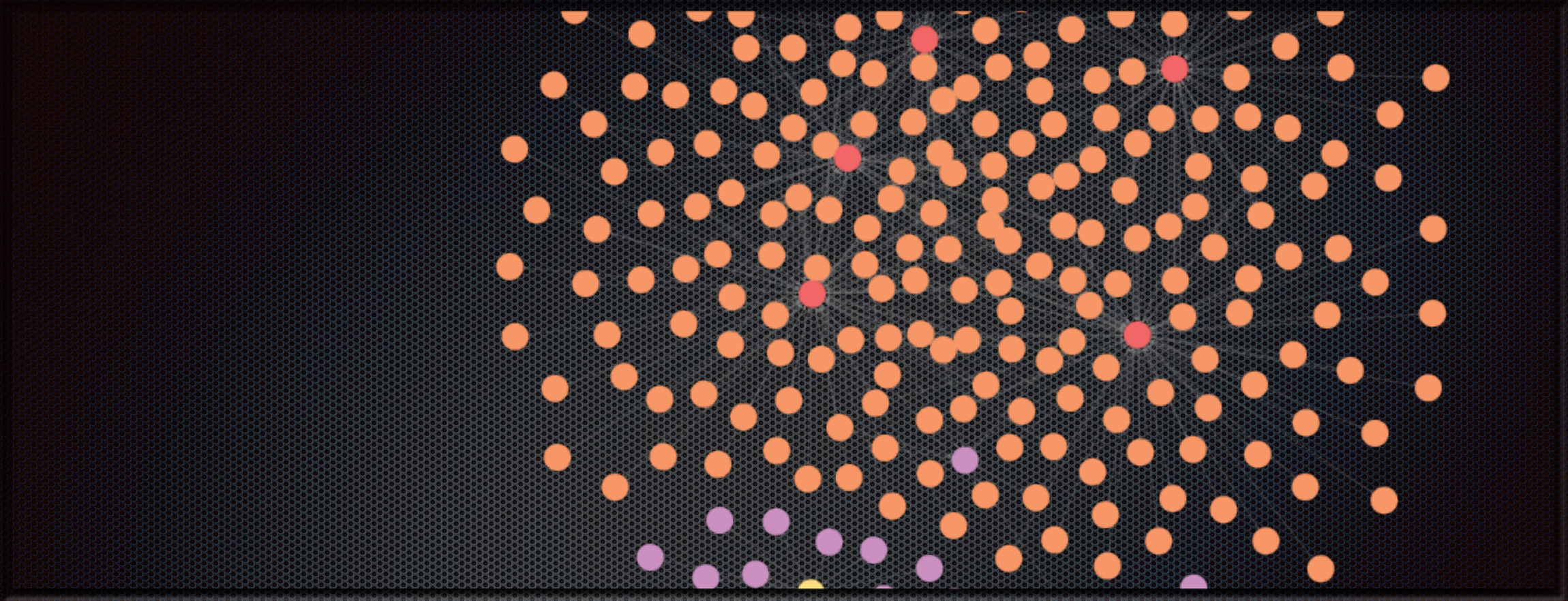


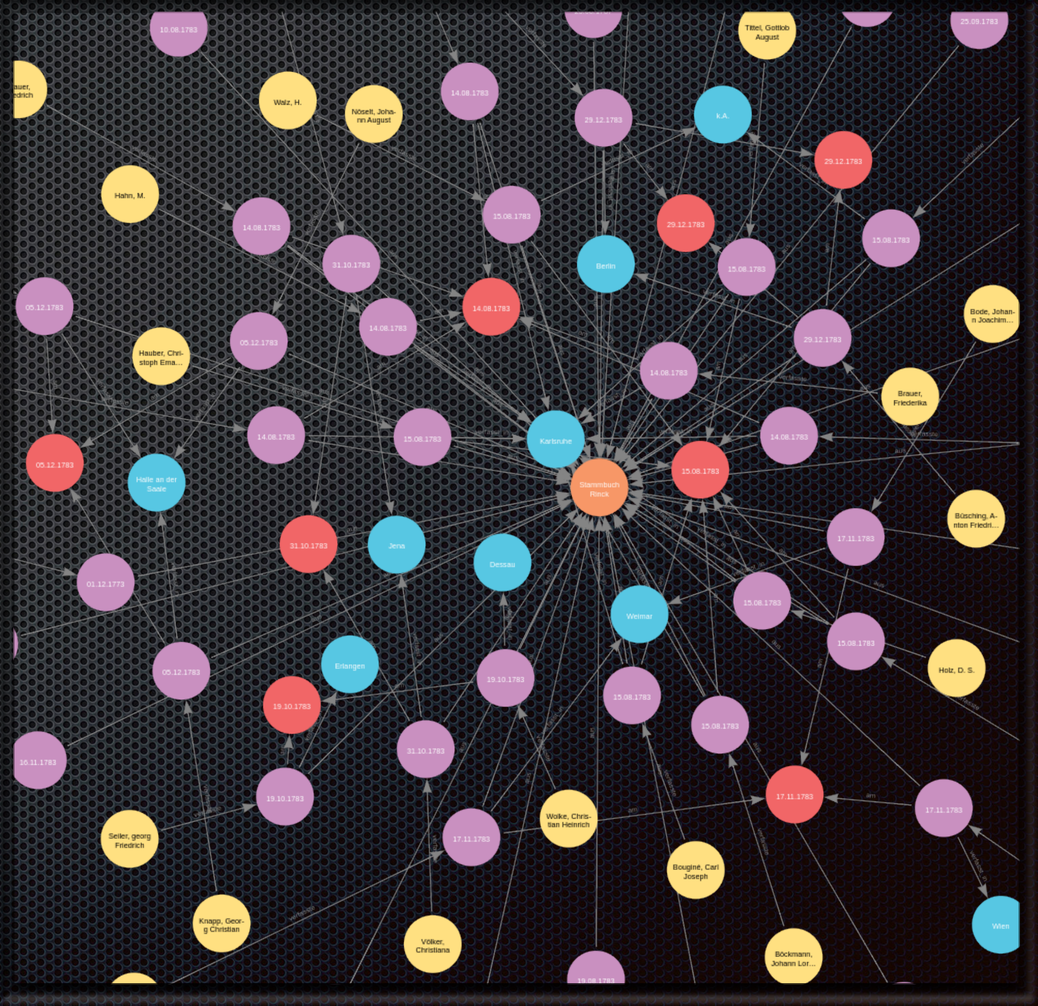
Von Graphen und Netzwerken.
Ein kurzer Einblick in Graphtechnologien und die Graphentheorie

Dr. Franziska Klemstein

Vortrag im Rahmen der Veranstaltungsreihe "Digital Humanities – Wie geht das?"

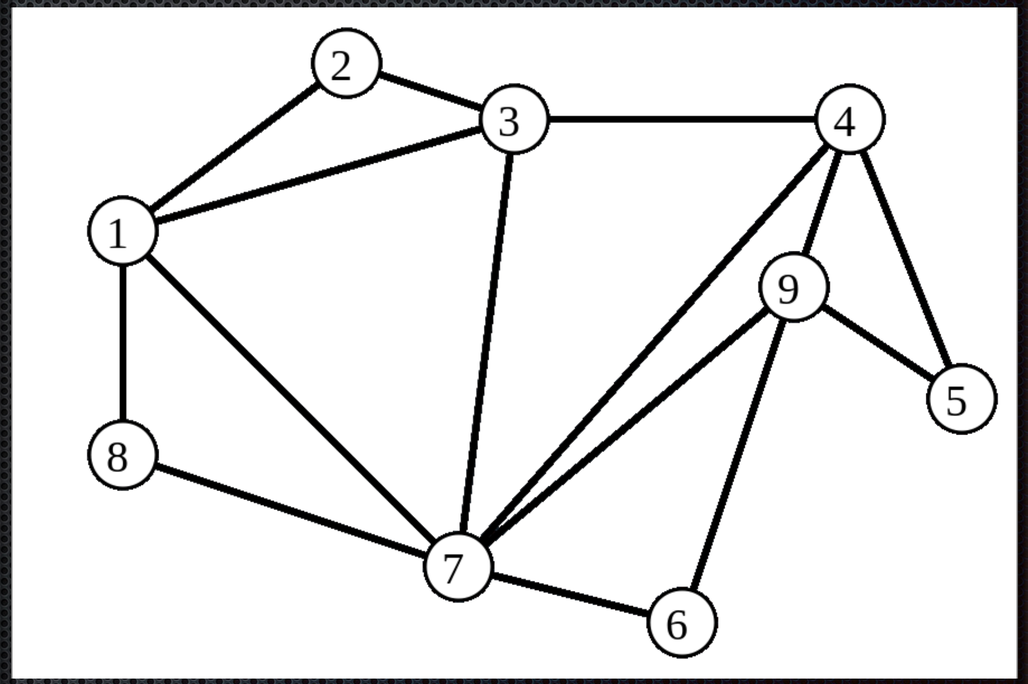
Überblick

- Begrifflichkeiten: Graph und Netzwerk
- Graphentheorie / Netzwerktheorie
- Graphtechnologie: Datenbanken und Analysetools

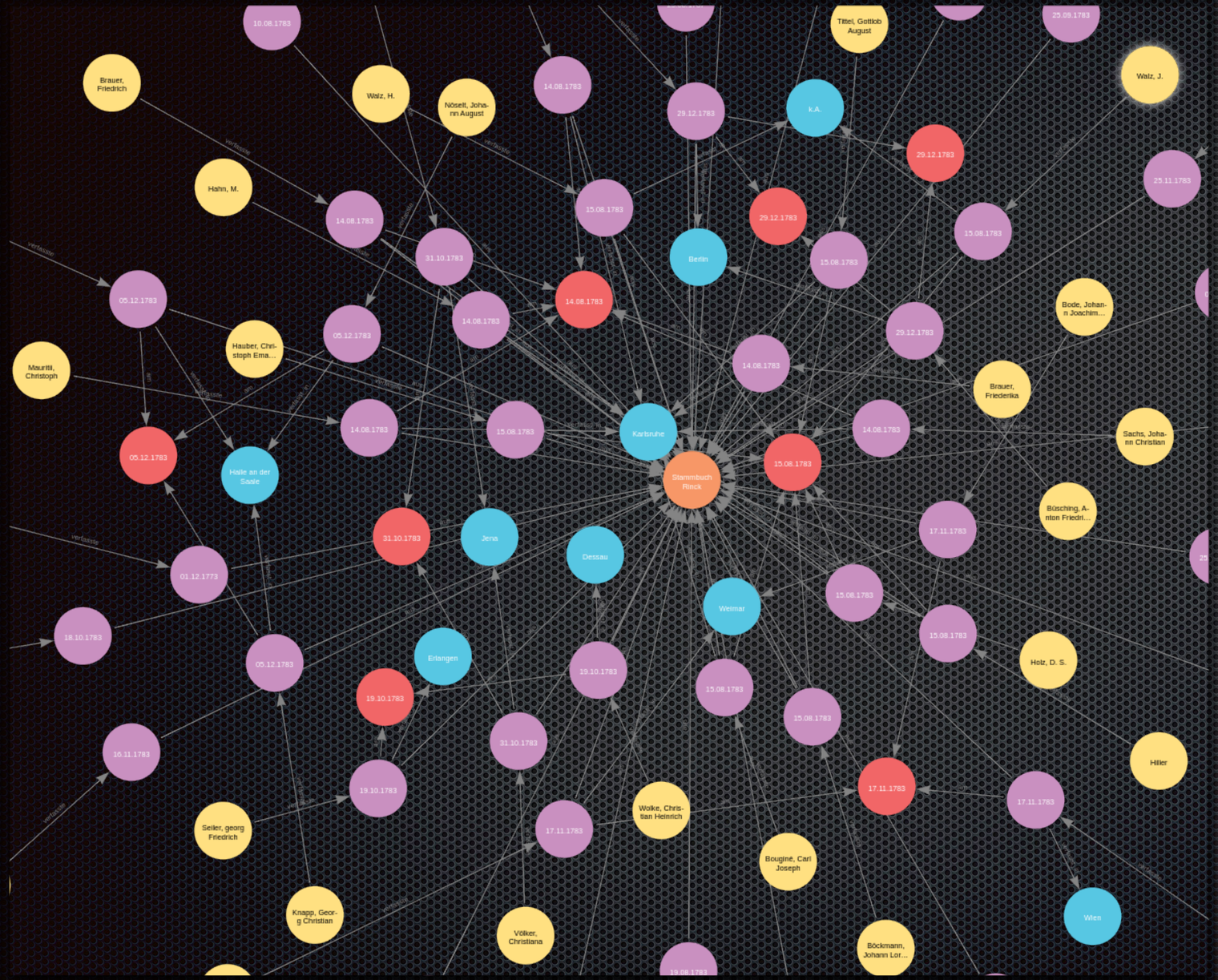


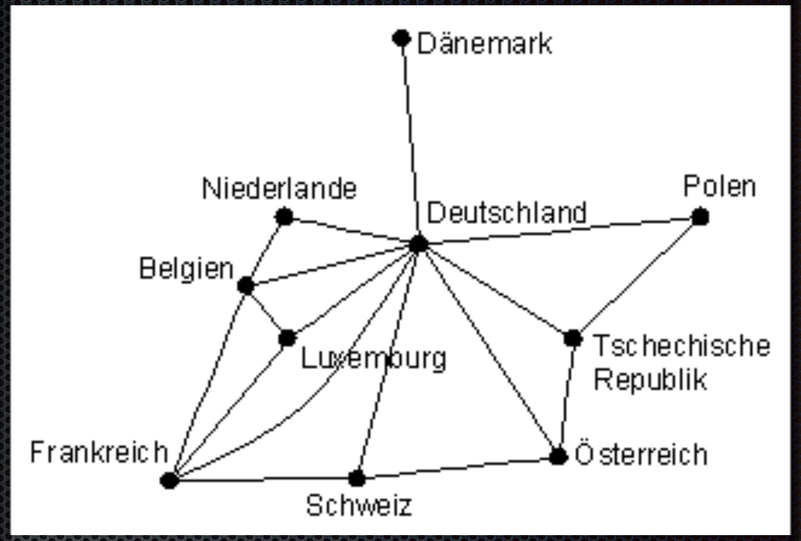
Begrifflichkeiten

Graph & Netzwerk

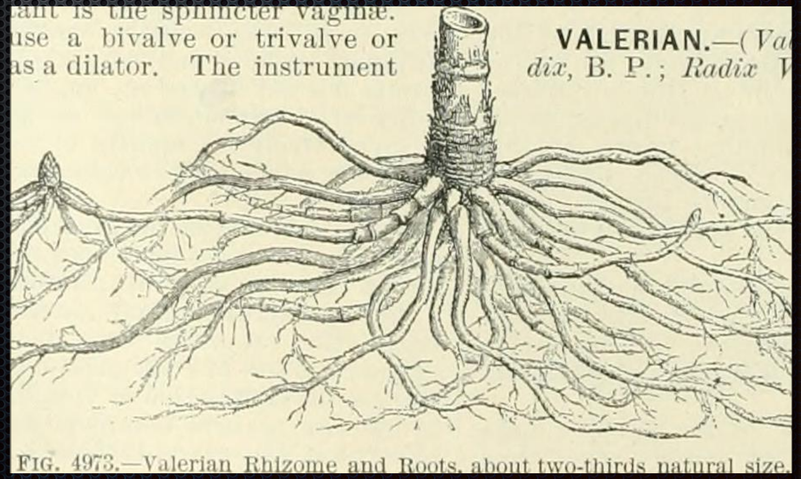


Bildquelle: https://de.wikipedia.org/wiki/Eulerkreisproblem#/media/Datei:Hierholzer(1).svg





Deutschland und Nachbarländer



Valerian rhizome and roots

Bildquellen: https://de.wikipedia.org/wiki/Einfacher_Graph#/media/Datei:Graph-BRD-Nachbarn-1.png (oben rechts), https://commons.wikimedia.org/wiki/File:A_Reference_handbook_of_the_medical_sciences_embracing_the_entire_range_of_scientific_and_practical_medicine_and_allied_science_%281900%29_%2814780151741%29.jpg (unten rechts).

Theorien

Graphentheorie und Netzwerktheorie

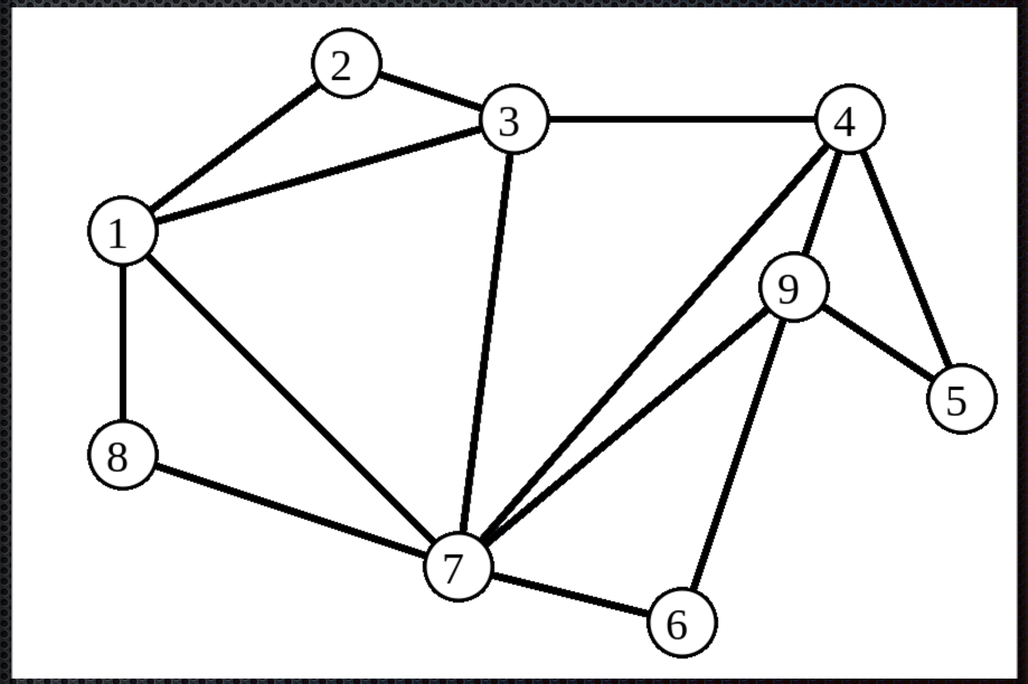
Bildquelle: https://de.wikipedia.org/wiki/Problem_des_Handlungsreisenden#/media/Datei:TSP_Deutschland_3.png

Graphentheorie

- Teilgebiet der Mathematik
- Betrachtungsgegenstand sind Graphen (Menge von Knoten und Kanten), deren Eigenschaften und ihre Beziehungen

$$N=(V,E) \text{ mit } V=\{a,b,c,d,e\}$$

Eulerkreisproblem



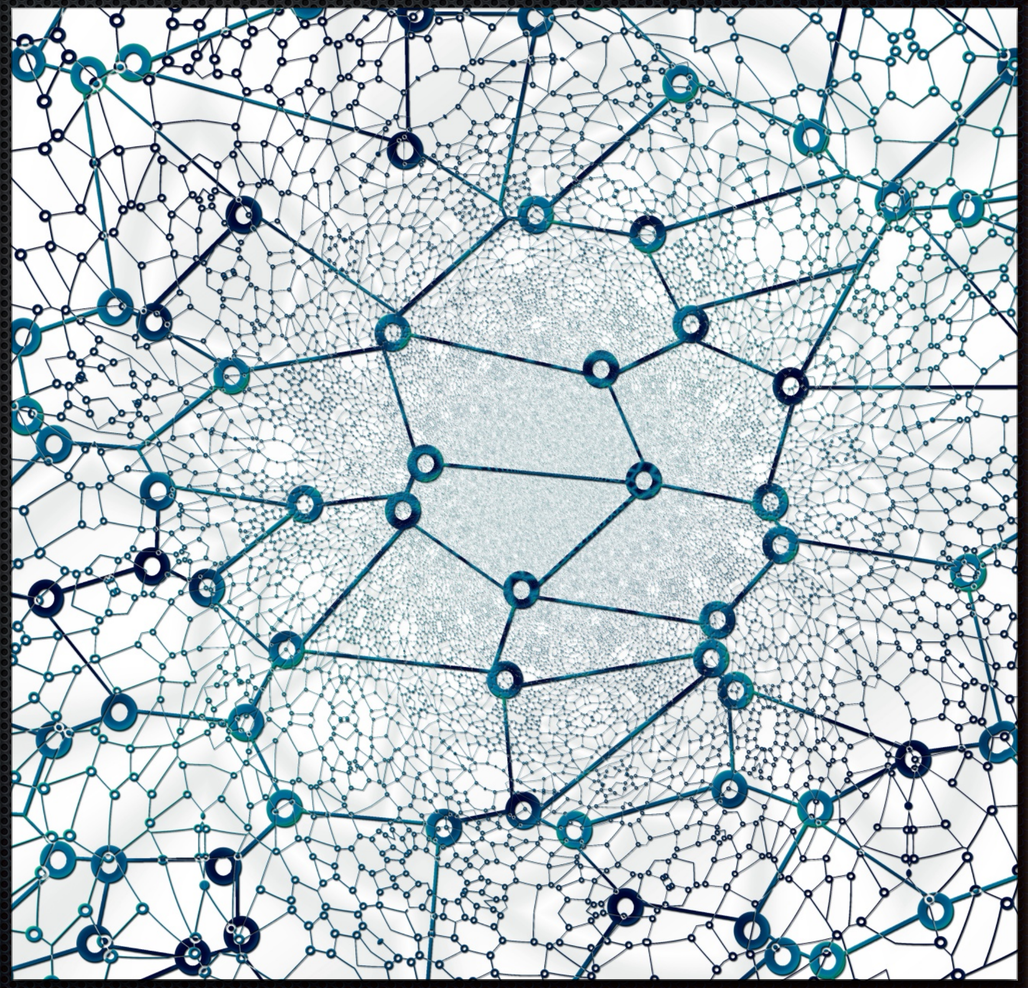
Bildquelle: https://de.wikipedia.org/wiki/Eulerkreisproblem#/media/Datei:Hierholzer(1).svg

Problem des Handlungsreisenden

Bildquelle: https://upload.wikimedia.org/wikipedia/commons/c/c4/TSP_Deutschland_3.png

Netzwerktheorie

- Netzwerktheorie / Netzwerkforschung / "social network analysis"
- Systemtheorie
- Akteur-Netzwerk-Theorie



Bildquelle: Gerd Altmann "Netzwerk total", CC0 Public Domain. https://www.publicdomainpictures.net/de/view-image.php?image=266331&picture=netzwerk-total

Technologien

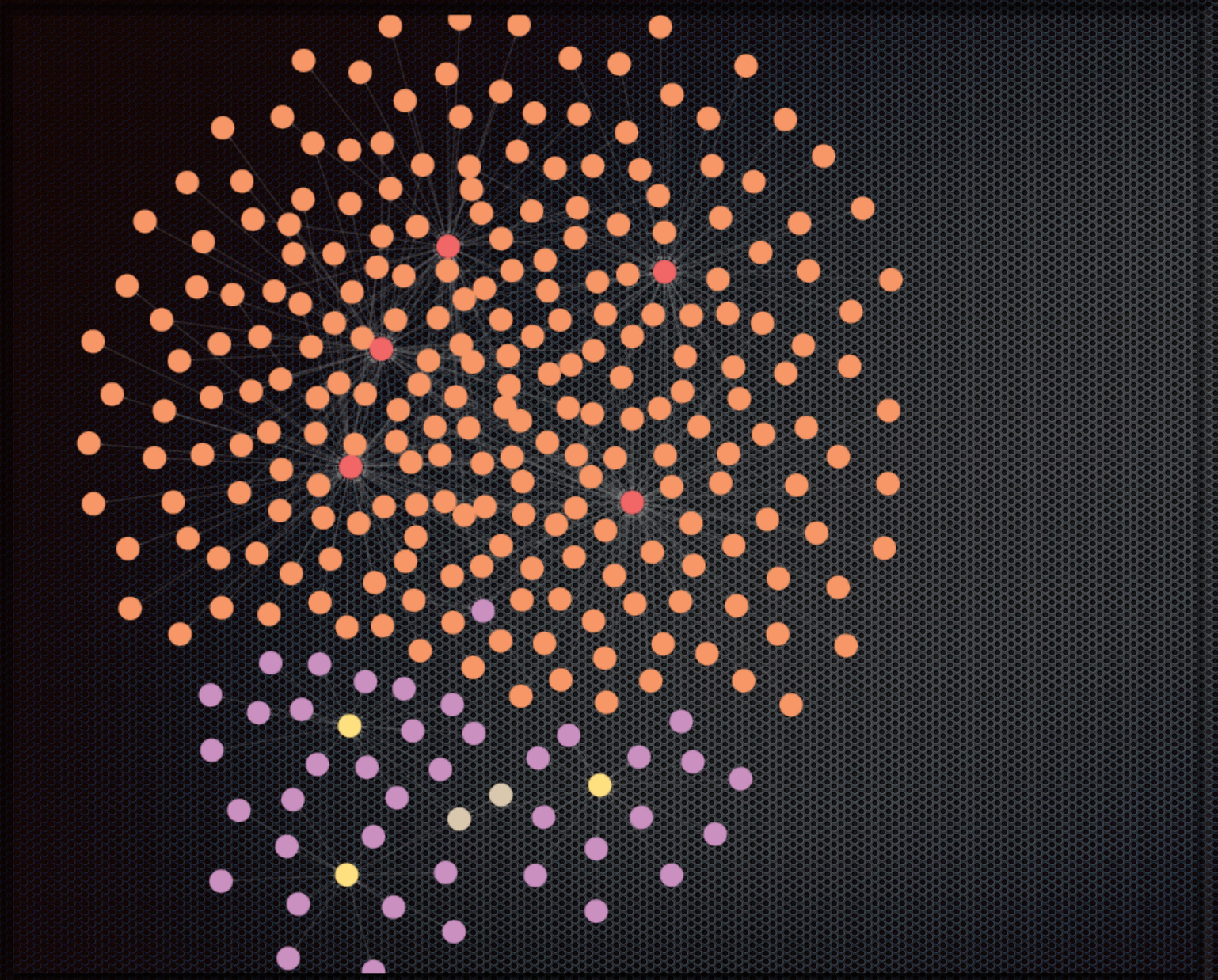
Datenbanken und Analysetools

Bildquelle: Gerd Altmann, https://www.publicdomainpictures.net/en/free-download.php?image=global-networking&id=297102

Graphdatenbanken

- Datenbank, die Graphen nutzt
- Graph besteht aus Knoten und Kanten
- Konzepte: Resource Description Framework (RDF), Labeled-Property Graph (LPG)
- Abfragesprachen: SPARQL, Gremlin, Cypher

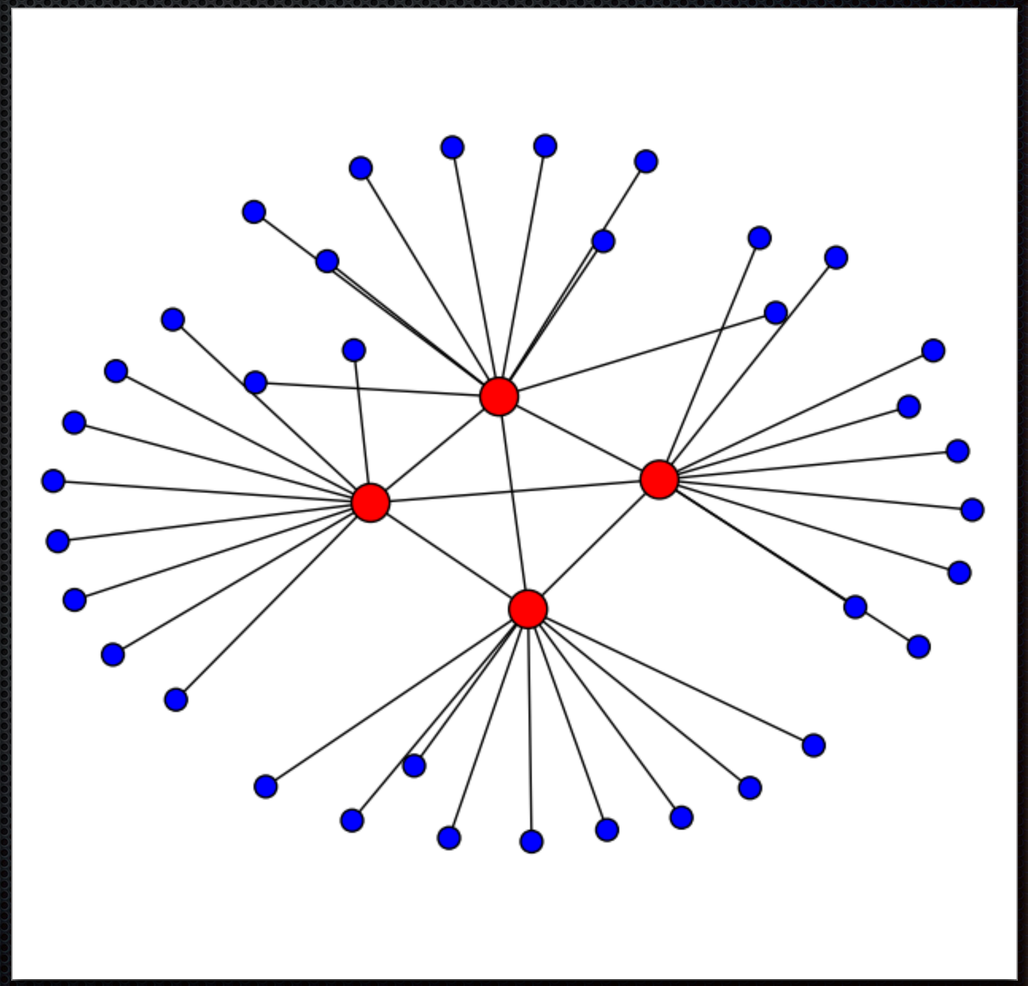
Bildquelle: https://upload.wikimedia.org/wikipedia/commons/e/e5/Neo4j-logo_color.png



Bildquellen: https://en.wikipedia.org/wiki/Blazegraph#/media/File:Blazegraph_parallelogram_logo.png (oben rechts), https://de.wikipedia.org/wiki/OrientDB#/media/Datei:OrientDB_Logo_2014_280x177.jpg (unten rechts).

Analysetools

- NetworkX
- OrientDB
- IBM Graph
- Amazon Neptune

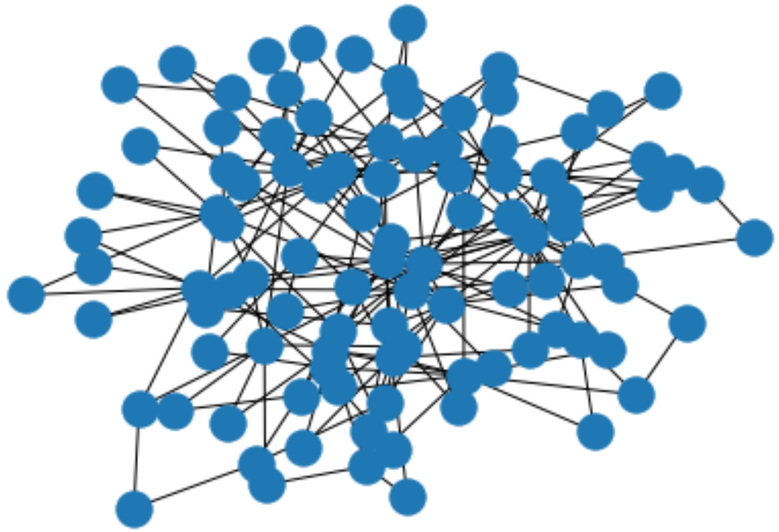


Bildquelle: https://upload.wikimedia.org/wikipedia/commons/e/ea/Disassortative_network_demonstrating_the_Rich_Club_effect.png

Analysetool: NetworkX

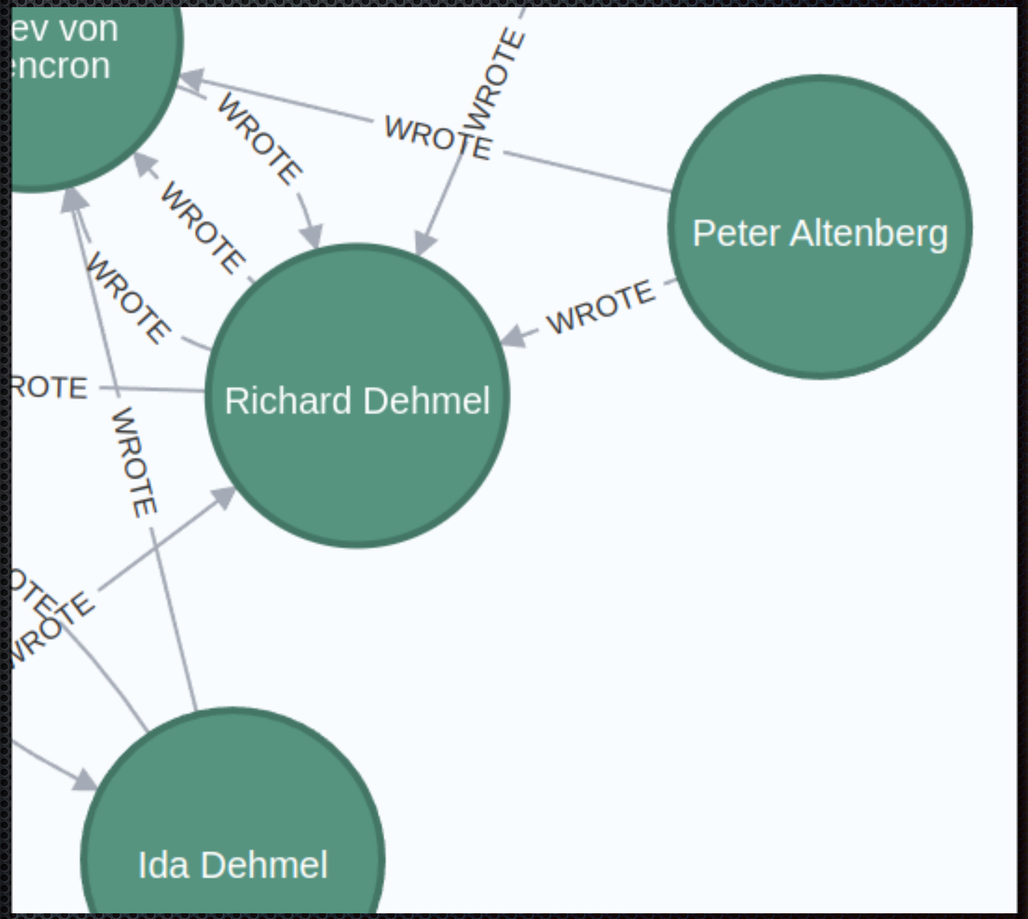
- Python-Bibliothek
- Erstellung, Bearbeitung und Untersuchung komplexer Strukturen
- gerichtete und ungerichtete Graphen
- Bereitstellung von verschiedenen Algorithmen zur Untersuchung der Netzwerke/Graphen

```
In [6]: import networkx as nx

In [7]: G = nx.barabasi_albert_graph(100,2)

In [8]: nx.draw_spring(G);
```

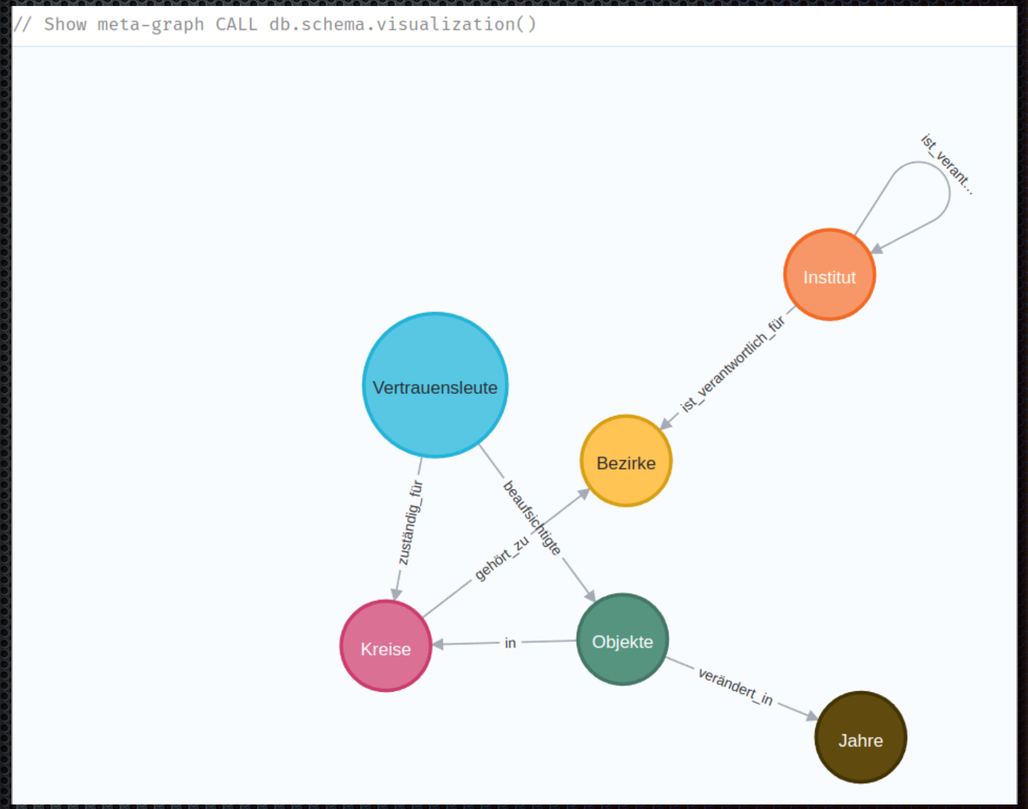


Neo4j

Einblick & Ausblick



Von der Quelle zum Modell und vom Modell zum Graphen



Erste(r) Knoten

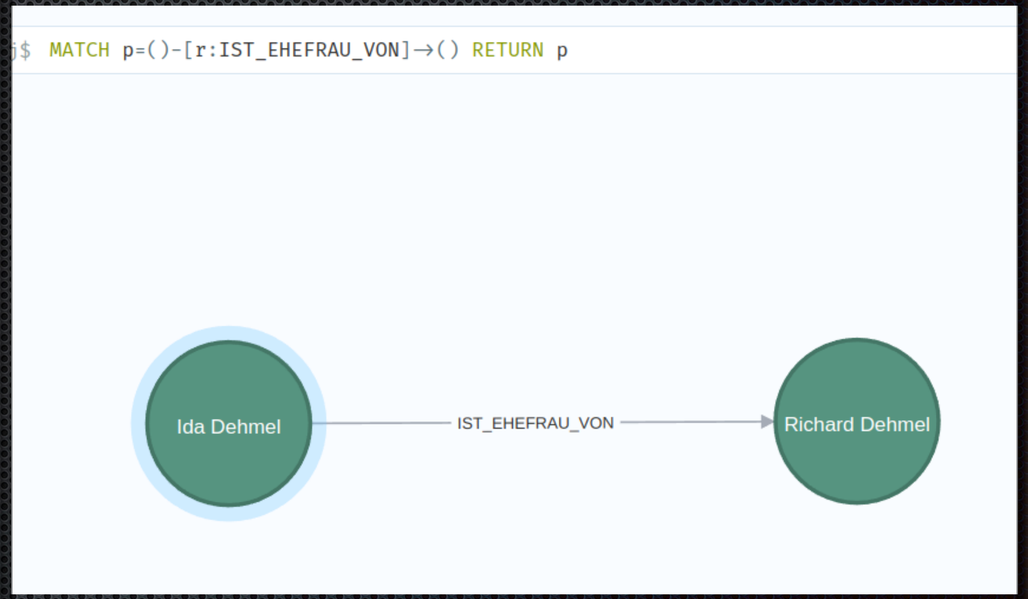
```
CREATE
  (a:Person {name: 'Richard Dehmel'}),
  (b:Person {name: 'Ida Dehmel'}),
  (c:Person {name: 'Adolf Loos'})
```

Added 3 labels, created 3 nodes, set 3 properties, completed after 101 ms.

Ergebnis

Erste Beziehung

MATCH (a:Person), (b:Person)
WHERE a.name='Ida Dehmel' AND
b.name='Richard Dehmel'
    CREATE (a)-[:IST_EHEFRAU_VON]->(b)



Abfragen mit Cypher

- Graph Query Language
- von SQL inspiriert
- ermöglicht aussagekräftige und effiziente Abfragen

CREATE clauses can create many nodes and relationships at once.

```
MATCH (ee:Person) WHERE ee.name = "Emil"
CREATE (js:Person { name: "Johan", from: "Sweden", learn: "surfing" }),
(ir:Person { name: "Ian", from: "England", title: "author" }),
(rvb:Person { name: "Rik", from: "Belgium", pet: "Orval" }),
(ally:Person { name: "Allison", from: "California", hobby: "surfing" }),
(ee)-[:KNOWS {since: 2001}]→(js),(ee)-[:KNOWS {rating: 5}]→(ir),
(js)-[:KNOWS]→(ir),(js)-[:KNOWS]→(rvb),
(ir)-[:KNOWS]→(js),(ir)-[:KNOWS]→(ally),
(rvb)-[:KNOWS]→(ally)
```

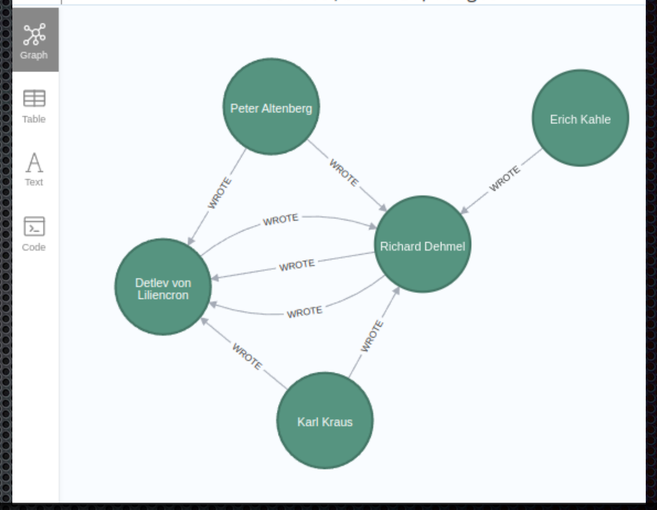
4 / 9

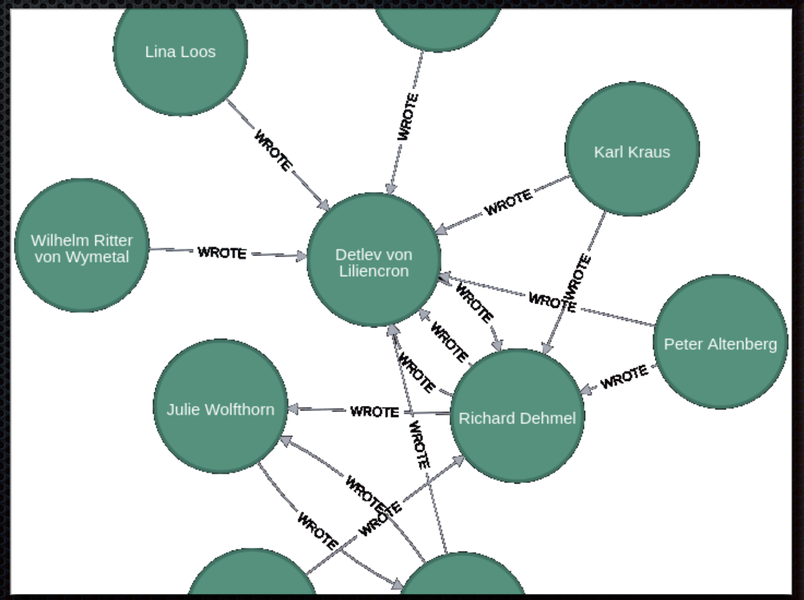
```
MATCH (a:Person)-[w:WROTE]→(b:Person)
WITH a,count(w) as Relation, collect(w) as WROTE WHERE Relation >1
RETURN a as Person, Relation as Anzahl ORDER BY Anzahl DESC
```

Graph | Table | Text | Code

"Person"	"Anzahl"
{"Name":"Richard Dehmel"}	3
{"Name":"Peter Altenberg"}	2
{"Name":"Karl Kraus"}	2
{"Name":"Ida Dehmel"}	2

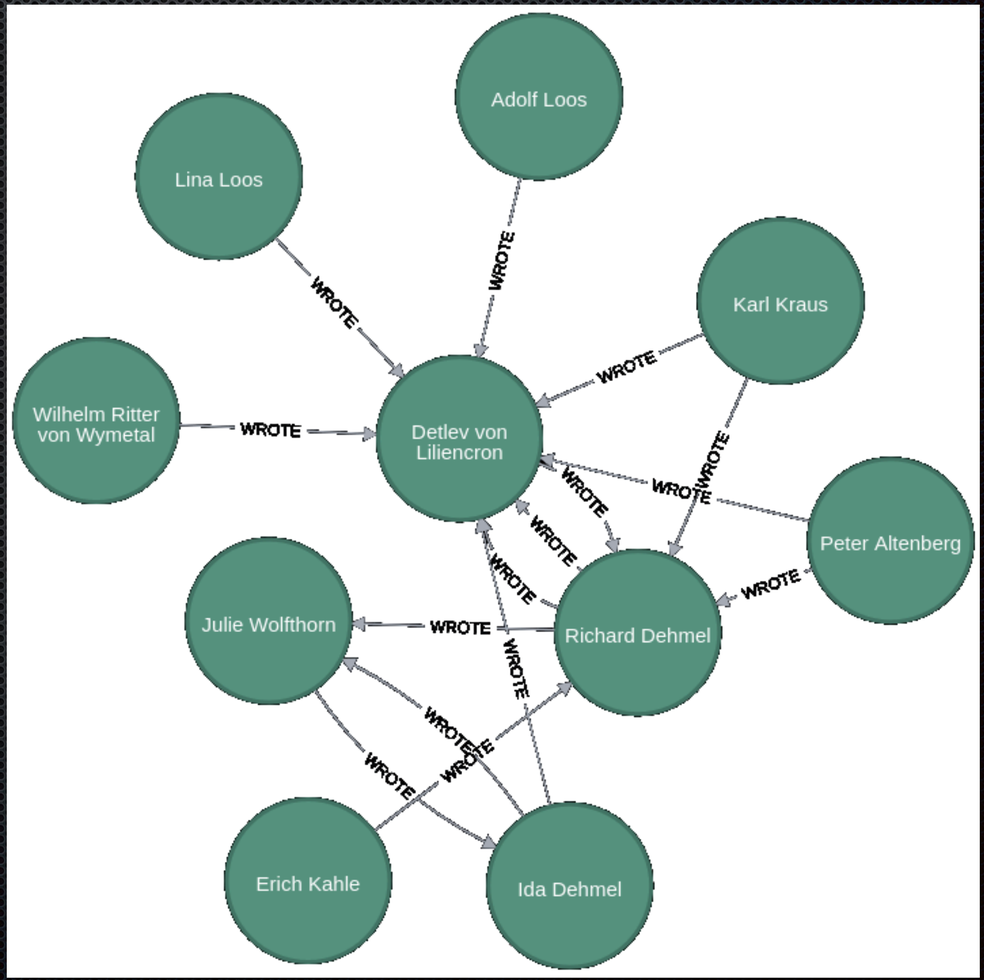
```
MATCH (a:Person)-[:WROTE]→(b:Person)
WHERE b.Name='Richard Dehmel'
    RETURN a as Absender,b as Empfänger
```





Triangle Count in neo4j

- Algorithmus
- zählt die Anzahl der Dreiecke für jeden Knoten >> 3-Clique
- Ziel: Zusammenhalt von Gemeinschaften erkennen und analysieren



nodeA	nodeB	nodeC
Detlev von Liliencron	Peter Altenberg	Richard Dehmel
Detlev von Liliencron	Karl Kraus	Richard Dehmel

Name	triangleCount
Detlev von Liliencron	2
Richard Dehmel	2
Peter Altenberg	1
Karl Kraus	1
Julie Wolfthorn	0
Adolf Loos	0
Lina Loos	0
Wilhelm Ritter von Wymetal	0
Erich Kahle	0
Ida Dehmel	0

Clustering Coefficient in neo4j

- Algorithmus zur Berechnung des lokalen Clusteringkoeffizienten
- Clusteringkoeffizient ist ein Maß für die Cliquenbildung in einem Netzwerk

Name	localClusteringCoefficient
Peter Altenberg	1.0
Karl Kraus	1.0
Richard Dehmel	0.2
Detlev von Liliencron	0.09523809523809523
Julie Wolfthorn	0.0
Adolf Loos	0.0
Lina Loos	0.0
Wilhelm Ritter von Wymetal	0.0
Erich Kahle	0.0
Ida Dehmel	0.0

Vielen Dank für Ihre Aufmerksamkeit

@FKlemstein@digitalcourage.social
@FKlemstein

Literaturempfehlungen

- Rik Van Bruggen: Learning Neo4j, Birmingham 2014.
- Bettina Hollstein: Strukturen, Akteure, Wechselwirkungen. Georg Simmels Beiträge zur Netzwerkforschung, in: Christian Stegbauer (Hrsg.): Netzwerkanalyse und Netzwerktheorie. Ein neues Paradigma in den Sozialwissenschaften, Wiesbaden 2008, S. 91–103.
- Manfred Nitzsche: Graphen für Einsteiger. Rund um das Haus vom Nikolaus, Wiesbaden 2005.
- Dirk Wintergrün: Netzwerkanalysen und semantische Datenmodellierung als heuristische Instrumente für die historische Forschung, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2019.