Datenmodellierung

Woche 11: Graphdatenbanken.

Institut für Digital Humanities, Historisch-Kulturwissenschaftliche Informationsverarbeitung | Prof. Dr. Øyvind Eide

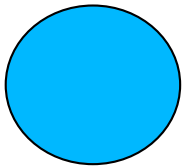
Königsberger Brückenproblem

Frage: Möglichkeit, alle 7 Brücken hintereinander so zu überqueren, dass jede genau *einmal* – also *nicht* mehrmals – überquert wird.

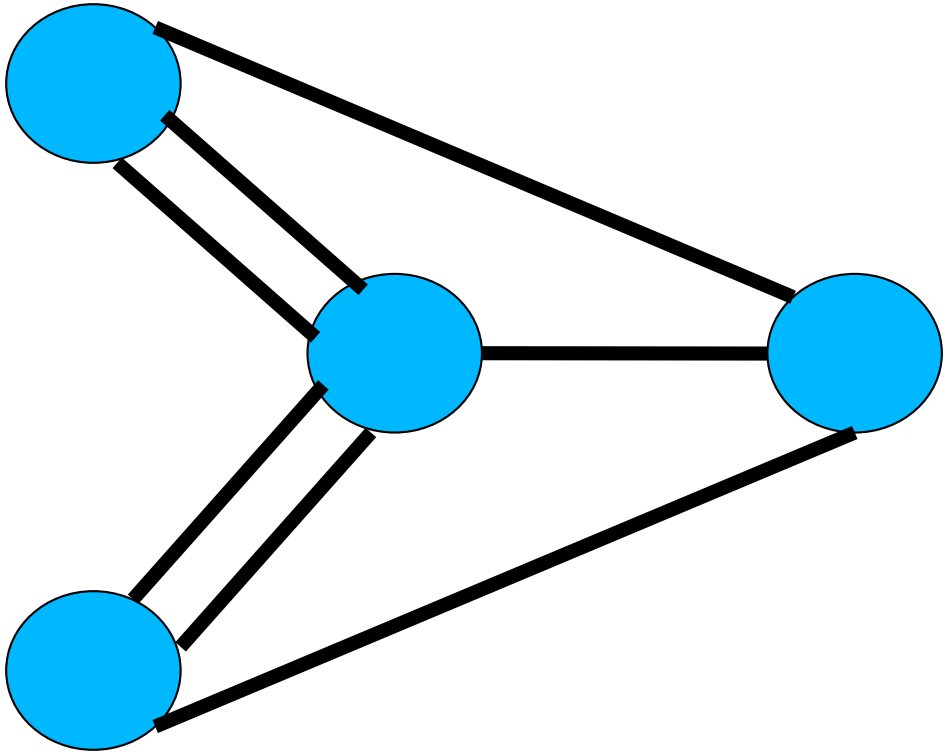
Leonhard Euler (1707–1783).

Universität zu Köln

Ein Graph



Knoten
(Vertex, Nodes)



Kanten
(Edges)

Universität zu Köln

Baum

Ein ungerichteter, zusammenhängender Graph ohne Zyklen heißt Baum.

→ XML

Graphdatenbanken

- Ein Graphdatenbanksystem speichert Daten in Form von Knoten (Entities) und Kanten (Relationships), die miteinander verbunden sind.
- Kernkonzepte:
  - Knoten: Repräsentieren Entitäten wie Personen, Orte, Objekte.
  - Kanten: Beschreiben die Beziehungen zwischen den Knoten.
  - Eigenschaften: Schlüssel-Wert-Paare, die zusätzliche Informationen zu Knoten und Kanten speichern.
- Beispiel: Soziales Netzwerk:
  - Knoten: Personen
  - Kanten: Freundschaften
  - Eigenschaften: Name, Alter, Beziehungsstatus

Vorteile

- **Effiziente Abfragen komplexer Beziehungen:** Optimiert für die Traversierung und Abfrage von vernetzten Datenstrukturen.
- **Flexibilität:** Schemafrei, was bedeutet, dass das Datenmodell leicht angepasst werden kann.
- Leistung: **Skalierbar** und **performant** bei der Verarbeitung von großen vernetzten Datensätzen.
- **Intuitive** Modellierung: Graphen sind natürlicher und intuitiver für die Modellierung von realen Beziehungen.

Graphdatenbanken in den DH

- Komplexe Beziehungen in Geisteswissenschaften
  - Graphdatenbanken sind ideal für die Darstellung und Analyse komplexer Netzwerke und Beziehungen in historischen, literarischen und kulturellen Daten.
- **Datenintegration:** Einfache Integration und Verknüpfung verschiedener Datensätze und Quellen.
- **Visualisierung:** Erleichtern die visuelle Darstellung und Interpretation von Daten durch Netzwerkgrafiken.
- Beispiele
  - Netzwerkanalyse in historischen Daten: Untersuchung von Beziehungen zwischen historischen Persönlichkeiten, Ereignissen und Orten.
  - Visualisierung literarischer Beziehungen: Darstellung von Einflüssen und Verbindungen zwischen Autoren, Werken und literarischen Bewegungen

Crash course:

https://youtu.be/8jNPelugC2s

Neo4j

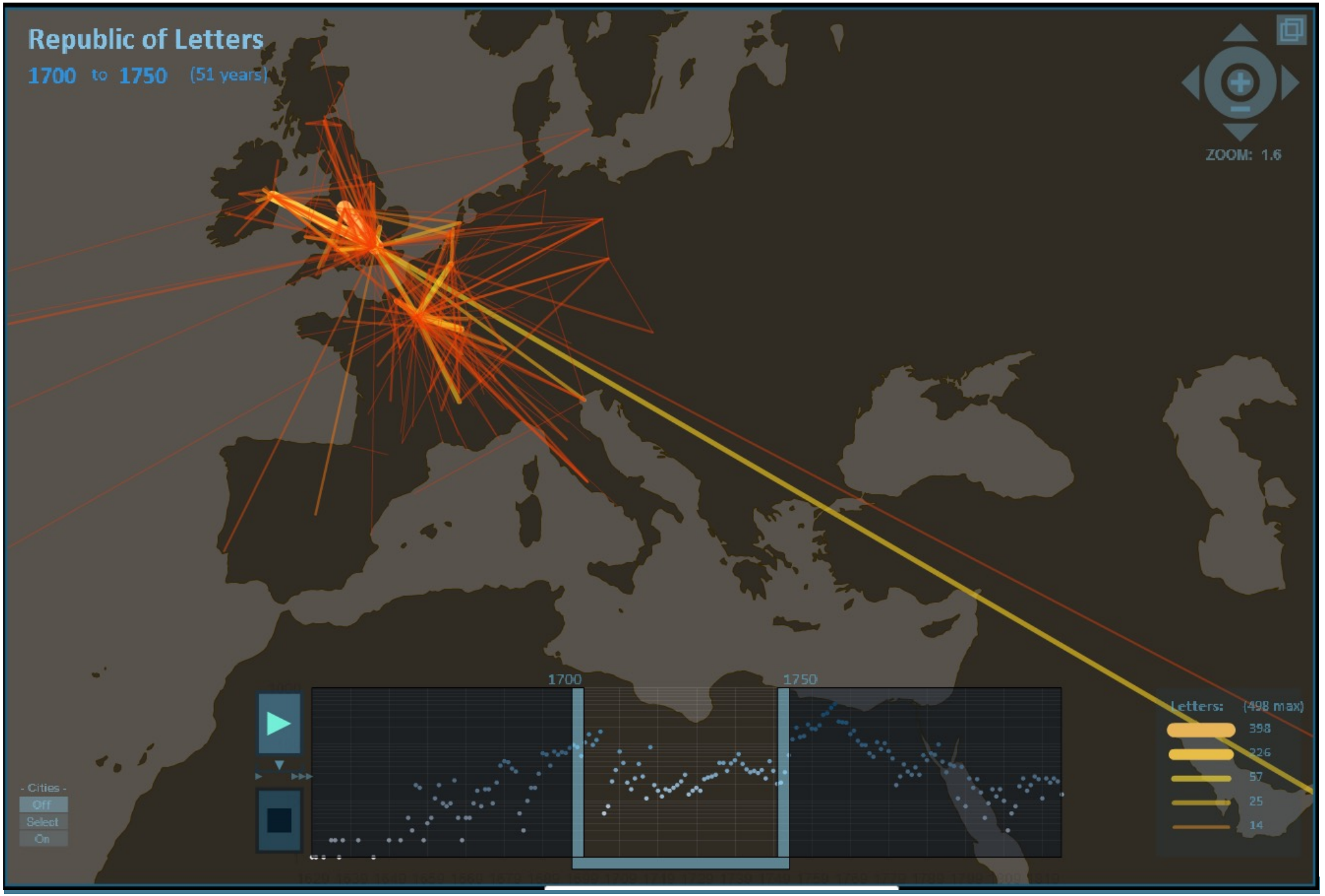
- Neo4j ist eine führende Graphdatenbank, die speziell für die Speicherung, Abfrage und Analyse von vernetzten Daten entwickelt wurde.
- Speichert Daten als Knoten und Kanten mit Eigenschaften.
- Cypher-Abfragesprache: Eine deklarative Sprache, die speziell für das Abfragen von Graphdaten entwickelt wurde
- Beispiel 1:
  - `MATCH (p:Person)-[:KNOWS]->(f:Friend) RETURN p, f`

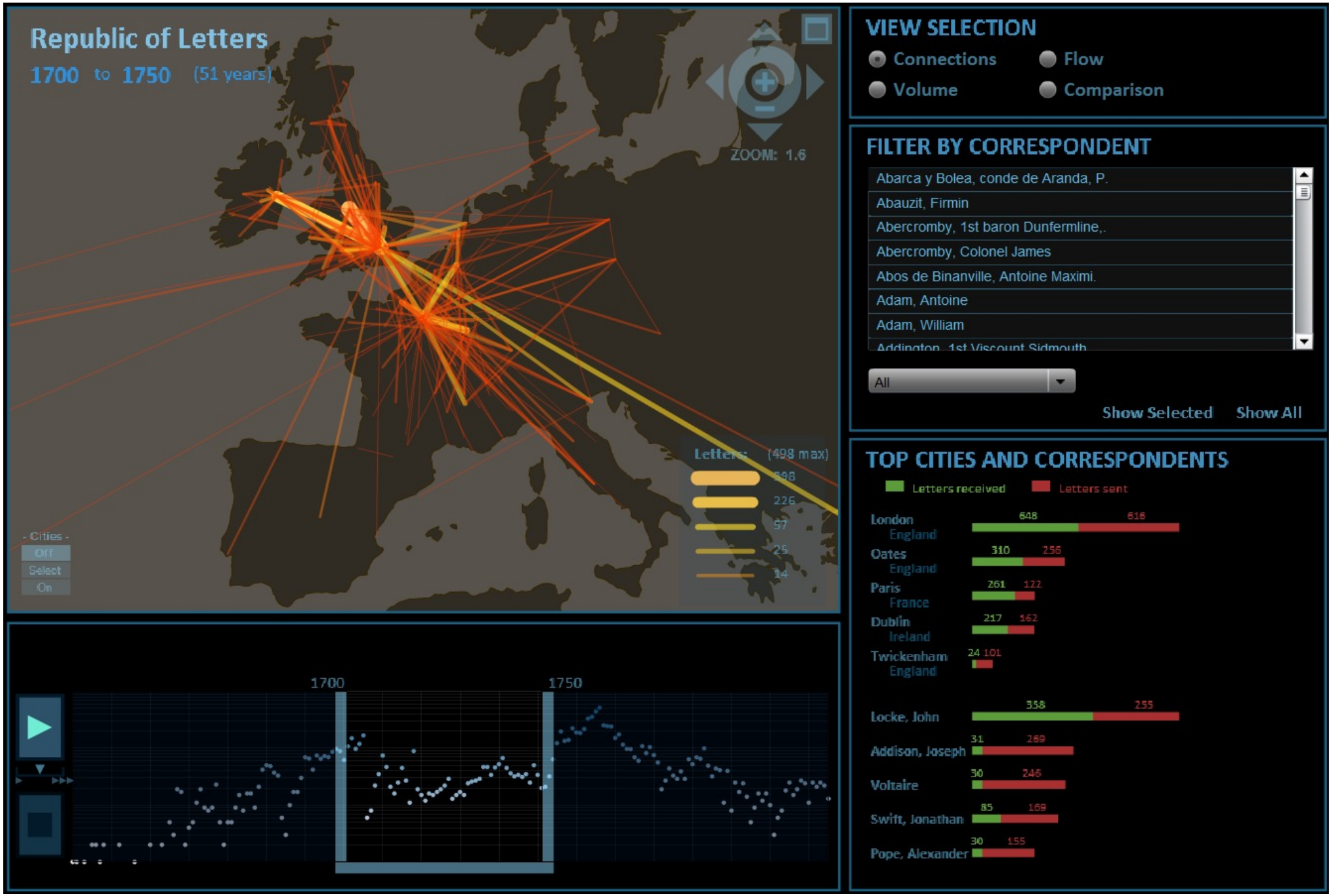
Universität zu Köln

CYPHER

- Deklarative Abfragesprache, die speziell für die Arbeit mit Graphdatenbanken wie Neo4j entwickelt wurde.
- Grundlegende Abfragestruktur
  - MATCH: Mustererkennung in Graphen.
  - WHERE: Filterbedingungen.
  - RETURN: Ergebnisrückgabe.
  - CREATE: Erstellen von Knoten und Kanten.
  - MERGE: Erstellen oder Zusammenführen von Knoten und Kanten.
  - SET: Festlegen oder Aktualisieren von Eigenschaften.
  - DELETE: Löschen von Knoten und Kanten.

www.stanford.edu/group/toolingup/rplviz/

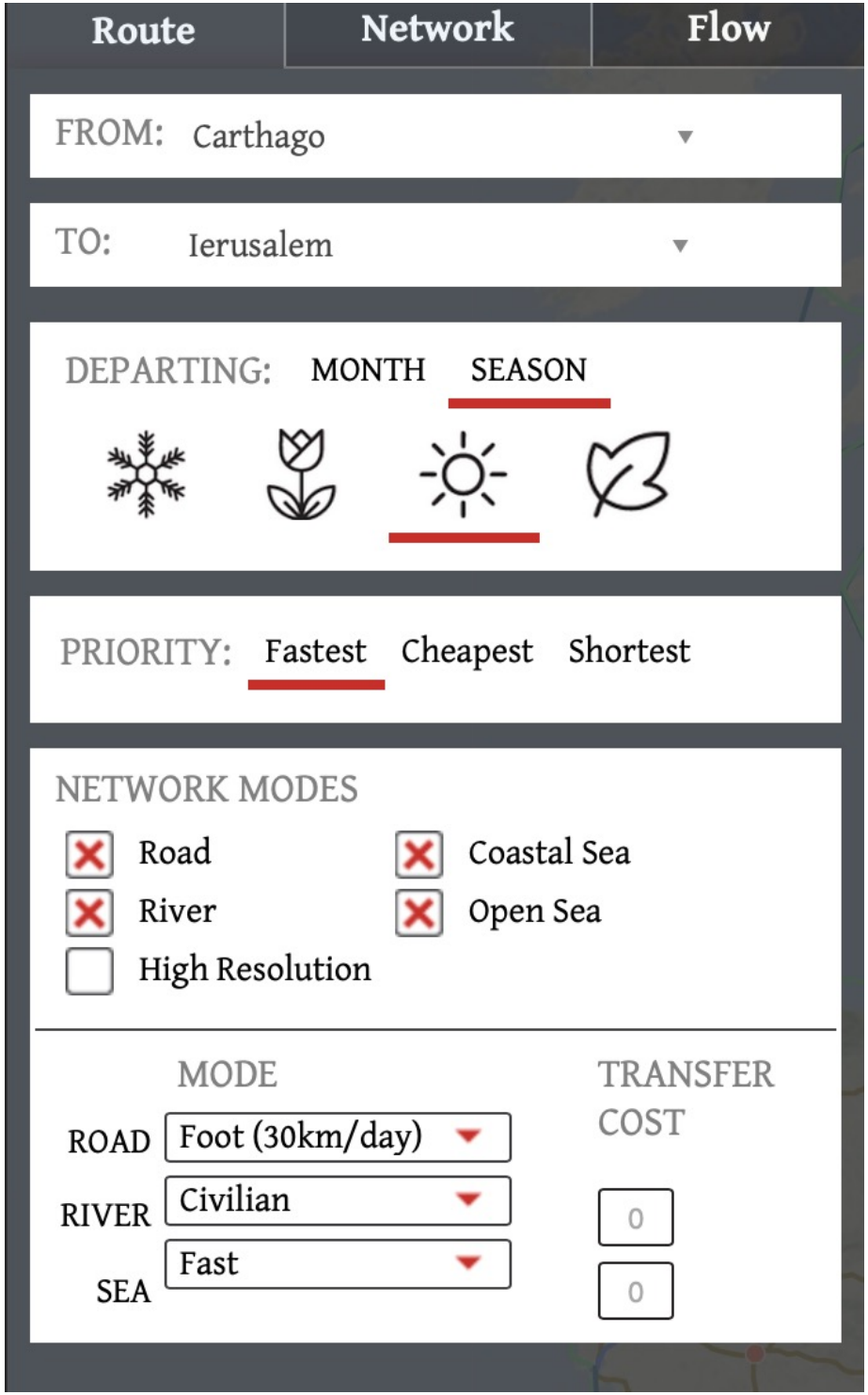


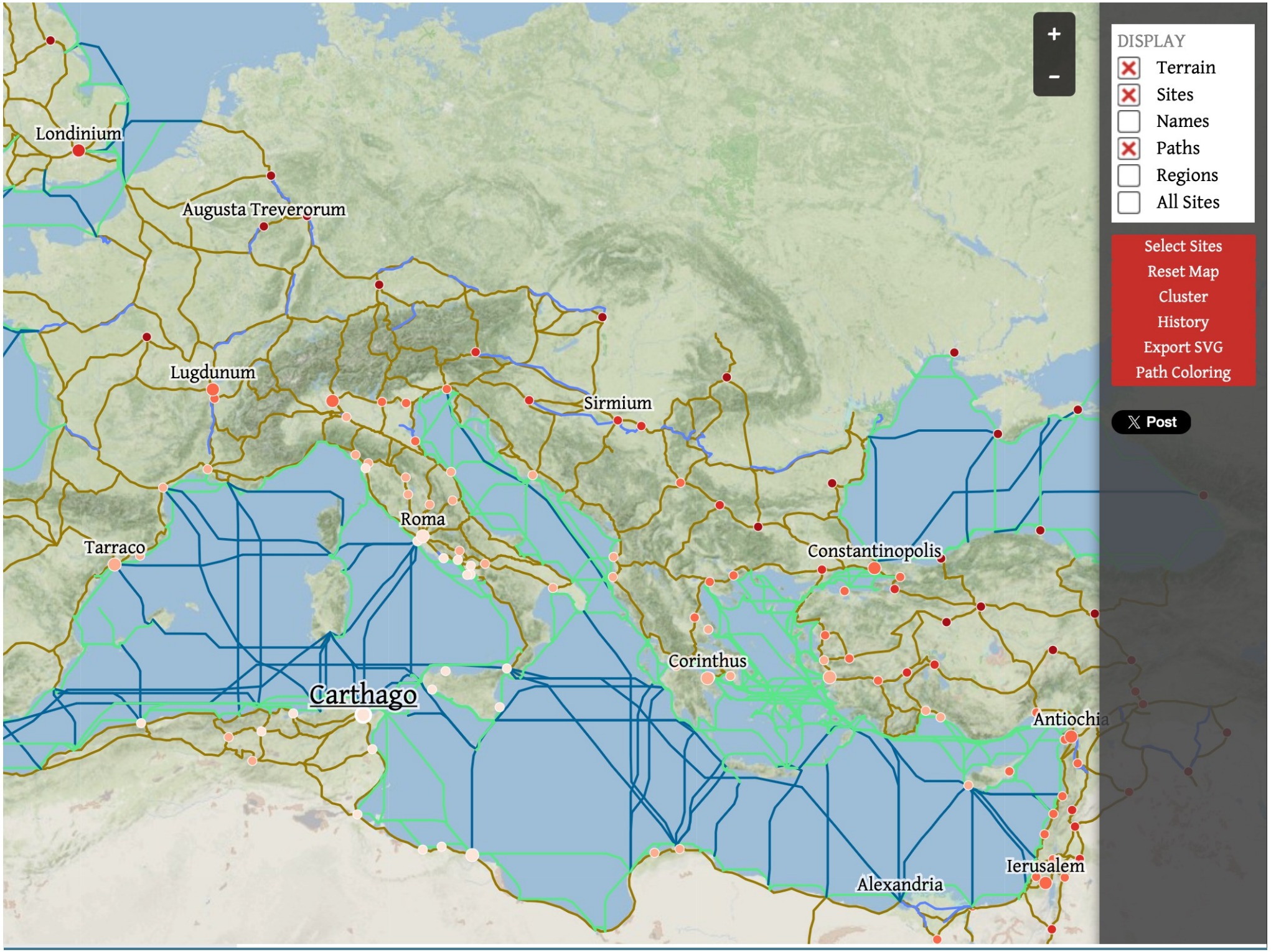

Republic of Letters
1700 to 1750 (51 years)
ZOOM: 1.6
Letters: (498 max)
498
226
97
25
14
- Cities -
Off
Select
On
1700
1750
VIEW SELECTION
Connections
Flow
Volume
Comparison
FILTER BY CORRESPONDENT
Abarca y Bolea, conde de Aranda, P.
Abauzit, Firmin
Abercromby, 1st baron Dunfermline,.
Abercromby, Colonel James
Abos de Binanville, Antoine Maximi.
Adam, Antoine
Adam, William
All
Show Selected
Show All
TOP CITIES AND CORRESPONDENTS
Letters received
Letters sent
London England 648 616
Oates England 310 256
Paris France 261 122
Dublin Ireland 217 162
Twickenham England 24 101
Locke, John 358 255
Addison, Joseph 31 269
Voltaire 30 246
Swift, Jonathan 85 169
Pope, Alexander 30 155

ORBIS

The Stanford Geospatial Network Model of the Roman World

https://orbis.stanford.edu/




Londinium
Augusta Treverorum
Lugdunum
Sirmium
Roma
Tarraco
Constantinopolis
Corinthus
Carthago
Antiochia
Ierusalem
Alexandria
+
−
DISPLAY
Terrain
Sites
Names
Paths
Regions
All Sites
Select Sites
Reset Map
Cluster
History
Export SVG
Path Coloring
Post

Modellierung

- Tabelle + Relationen
  - Spreadsheets
  - Relationsdatenbanken
- Baum
  - XML-Editors
  - XML-Datenbanken
- Graph (Netzwerk)
  - Netzwerksvisualisierung
  - Graphdatenbanken

Text als Graph

- Blogartikel lesen: https://mittelalter.hypotheses.org/7994
- Fragen beantworten:
  1. Welche Probleme treten bei der Verwendung traditioneller XML-Methoden zur digitalen Edition historischer Texte auf, und wie adressieren graphbasierte Datenbanken diese Herausforderungen?
  2. Wie ermöglichen graphbasierte Datenbanken die Erstellung, Suche und den Export multidimensionaler Deutungshierarchien in historischen Texten?