Building a Graph Database on Historical Charters in *Neo4j*

Mario Hinderhofer

Advanced Software Practical

30.04.2018

Outline

Motivation

Neo4j and Graph Databases

- Cypher Query Language

Implementation

- First ER Model
- Major Changes
- Current Model
- Difficulties

Outlook

Motivation

Our goal is to:

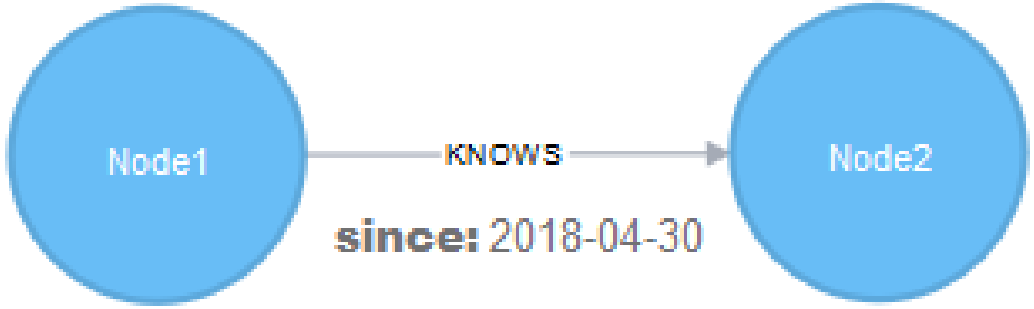
- Link data that's already existing in the charters
- Gather reliable (hand-sourced) information on the relations of persons

***Neo4j* allows us to:**

- Put the focus on the relations between all our entities
- Get a better understanding of the connections and thereby power/influence the Focus Group had

Neo4j and Graph Databases

- Nodes, (directional) Edges and Attributes
- Relations (edges) are the main focus
- Visualization is a core feature
- Flexibility: Adding new data and changing the structure is no problem
  - Models don't have to be done ahead of time
- Substantially faster than relational databases for highly connected data



Cypher Query Language

Structured after SQL

- **SELECT** col **FROM** table **WHERE** … **AND** …
- **MATCH** (p:PERSON) **WHERE** … **AND** … **RETURN**

Aimed to be easy to use for anyone

- (p:Person)-[:HAS_LOCATION]->(loc:Location)

Example Query:

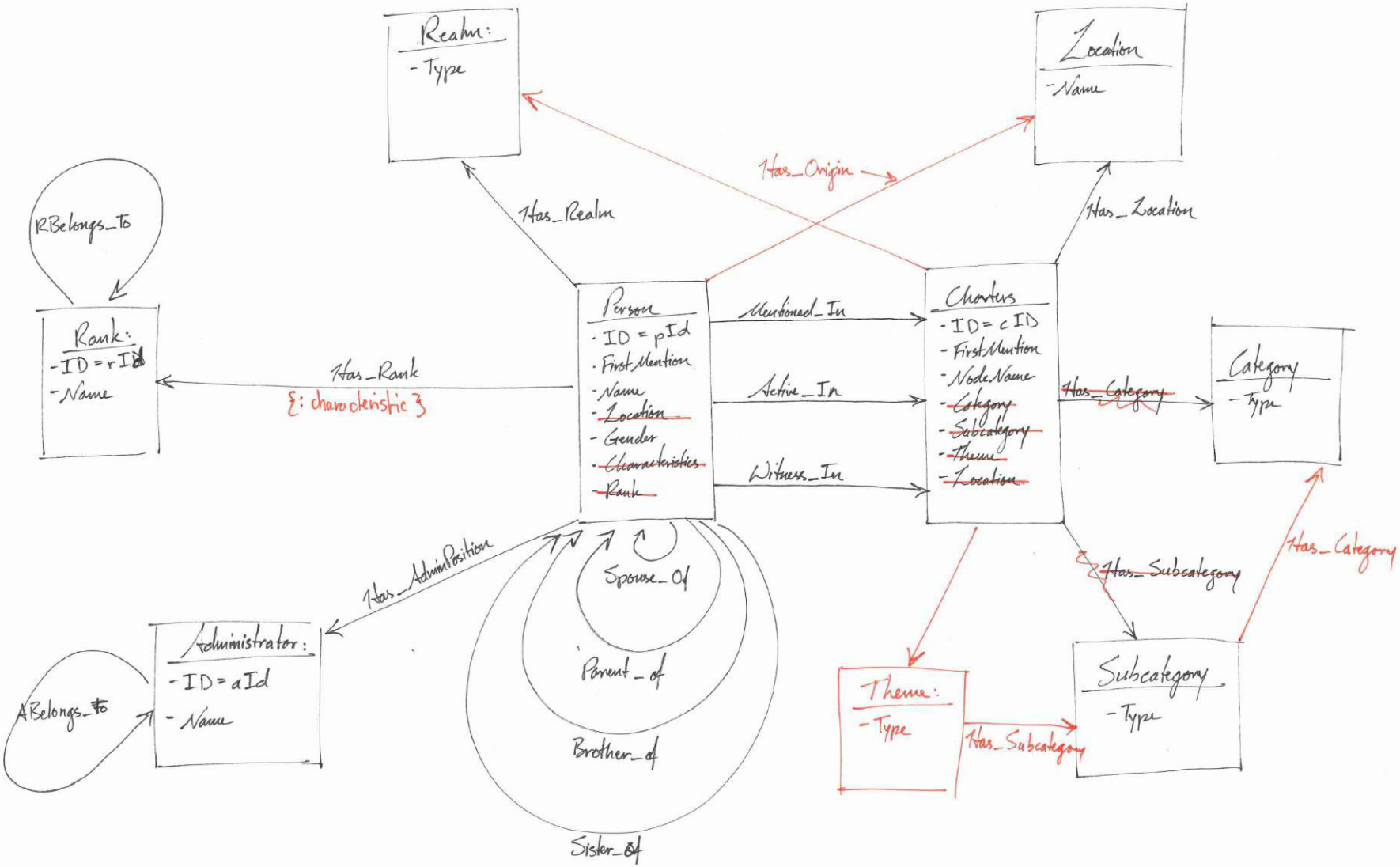
MATCH (p:Person), (r:Rank)

WHERE r.name = “Emperor“

AND (p)-[:HAS_RANK]->(r)

RETURN p;

First ER Model Iteration



First Iteration: Design issues

Nodes and attributes were used the same way

Some data was duplicated

- Increased risk of errors/typos

Complicated queries

- Spouse_Of, Parent_Of, ...
- Mentioned_In, Active_In, Witness_In
- Charters → Theme → Subcategory → Category

Data was split into persons and charters

First Iteration: Missing information

Information captured was too generic

- Person P was active in Charter C
  - What was the charter about? What role did he have?
- Database on its own only showed basic relations
  - More precise information had to be looked up in charters
- Charters might contain more information
  - A charter often contains more than one event

No representation of the buildings/sites of interest other than locations

Major Changes

Some “property nodes” have become attributes

Focus Group has been defined

- Persons have an attribute focusGroup: {yes, no}
  - Focus Group persons have more detailed relations
- Points of interest have been defined as buildings that can be active in events

Persons have appearances

Important charters are split into events

- Events have sites and topics

Why do we need a Focus Group?

Extracting all data from 1431 persons in 161 charters is out of scope, however

- Without detailed information we're not able to draw any conclusions
- A lot of the data is only indirectly relevant for us

Our Focus Group allows us to add additional data to persons of interest and acts as a filter for queries

It includes persons that are strongly connected to the case studies (especially families of Ministerialis)

Appearance

Allows us to store data on a person in relation to a charter/event

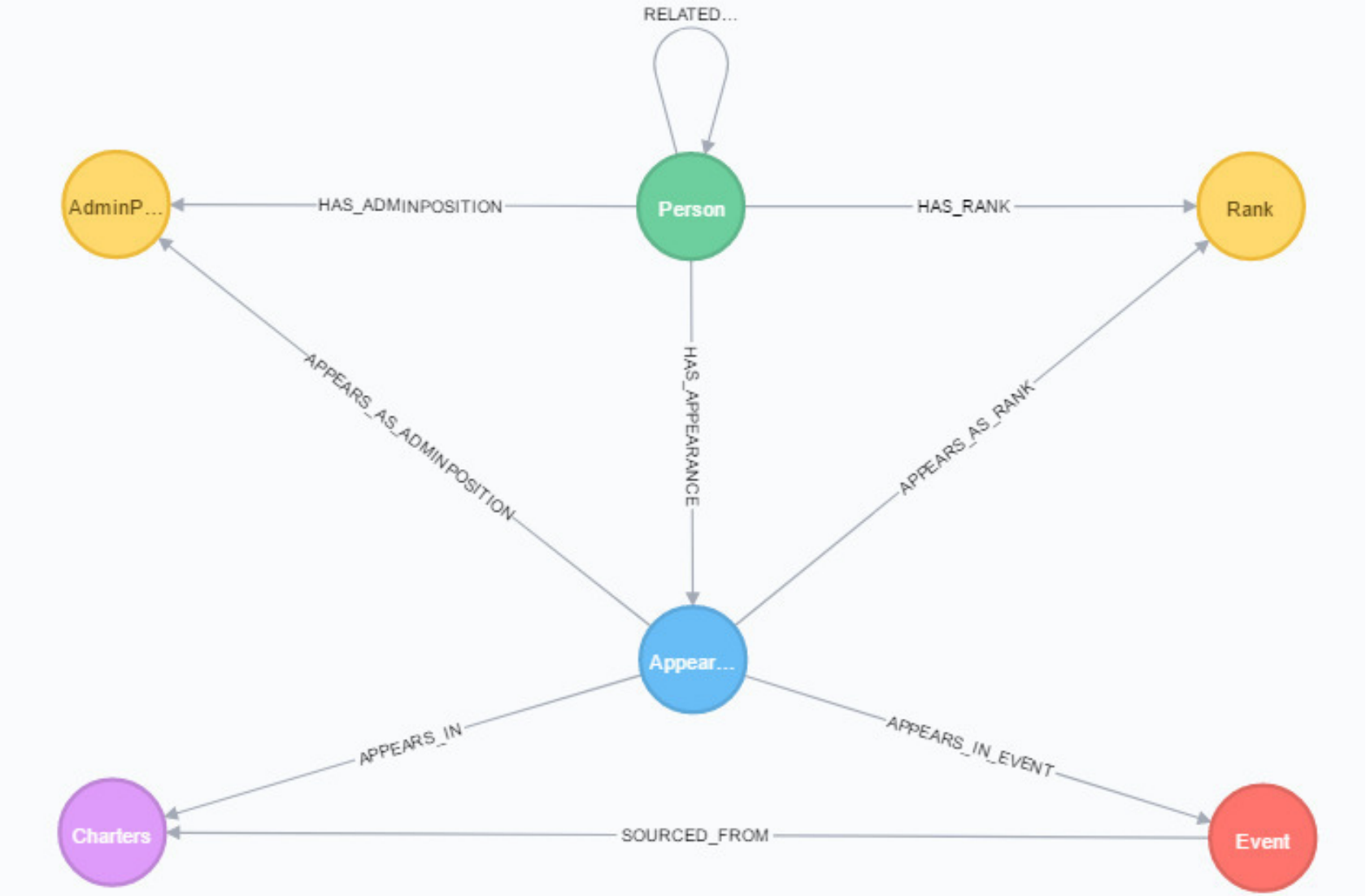
- e.g. ranks might change between charters

Persons can have several appearances per charter with different roles

Every person has appearances, but:

- Extensive data is only available for the Focus Group

Appearance


RELATED...
AdminP...
HAS_ADMINPOSITION
Person
HAS_RANK
Rank
APPEARS_AS_ADMINPOSITION
HAS_APPEARANCE
APPEARS_AS_RANK
Appear...
APPEARS_IN
APPEARS_IN_EVENT
Charters
SOURCED_FROM
Event

Events in Charters

Persons and Buildings of Focus Group have detailed relations to Events

- Role: Giver, Recipient, Confirmer, …
- Amount: 10 Heller
- Type: e.g. Confirmation

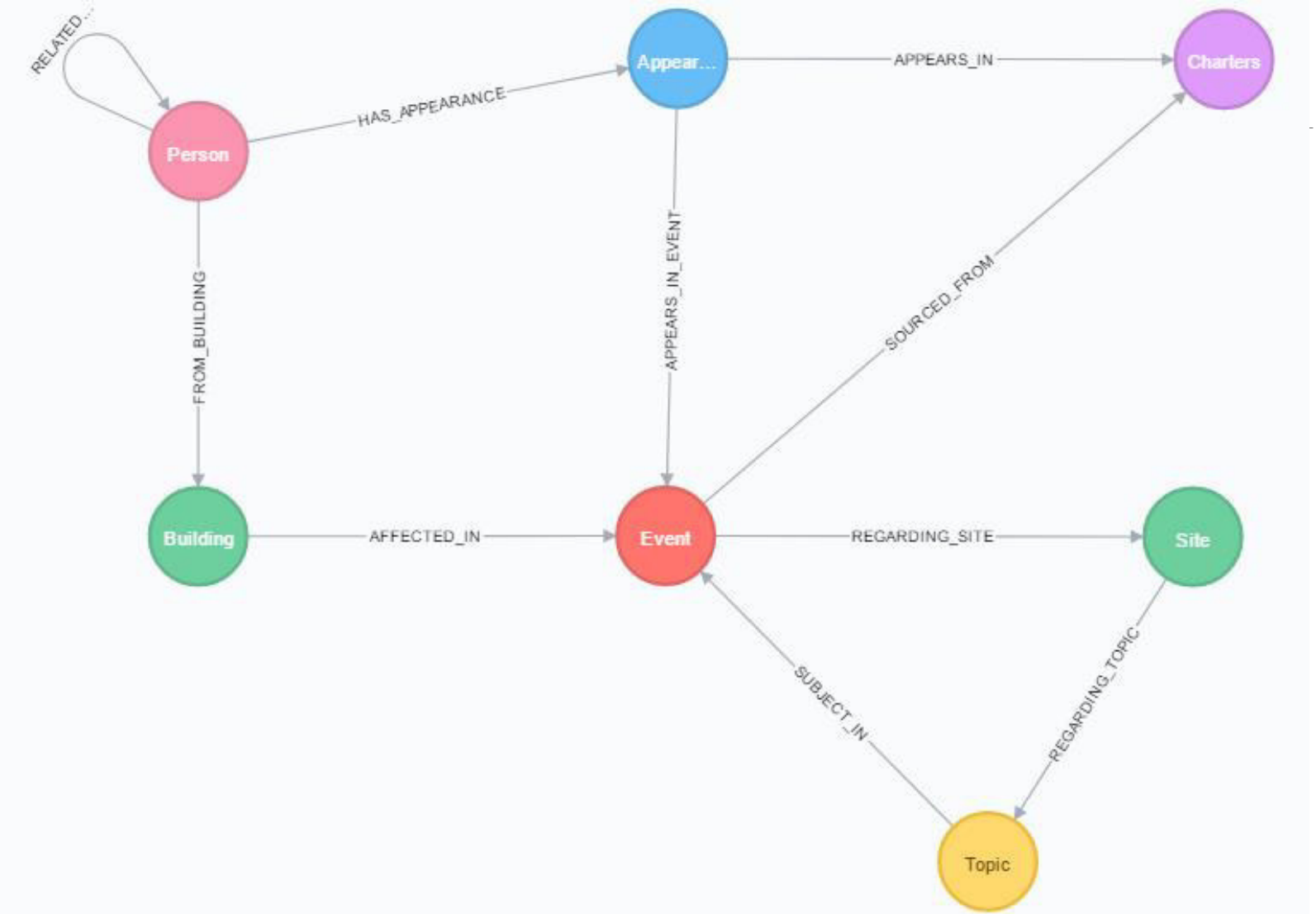
Sites allow us to see who owns what and when

- e.g. fields around Lampertsmühle

Topics

- In how many sales was Person P involved?
- How often has Site S been sold?

Events in Charters



Events in Charters

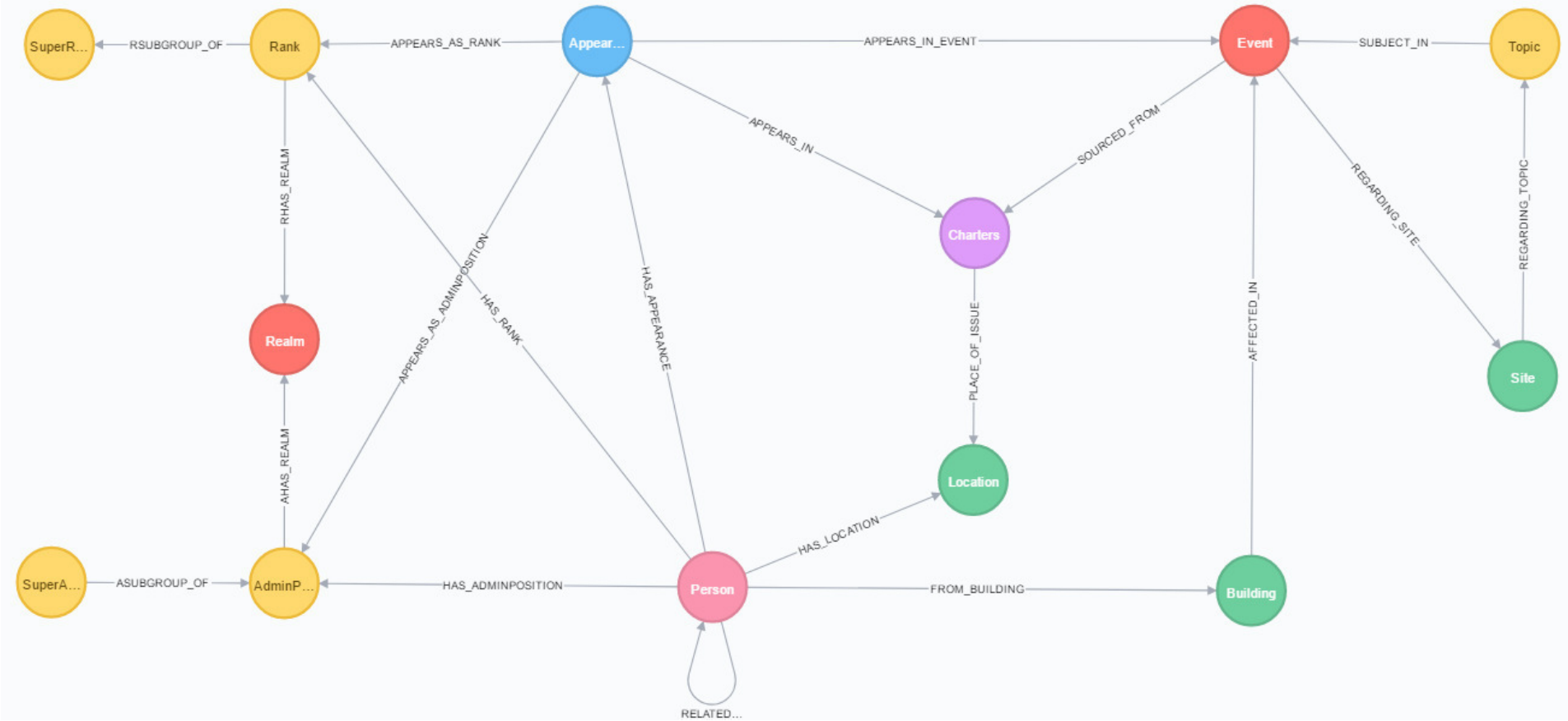
Events are important because they

- give us the most detailed information
- Allow us to track who did what with which person or building
- Can give us information about the six case studies and their surroundings

Topics

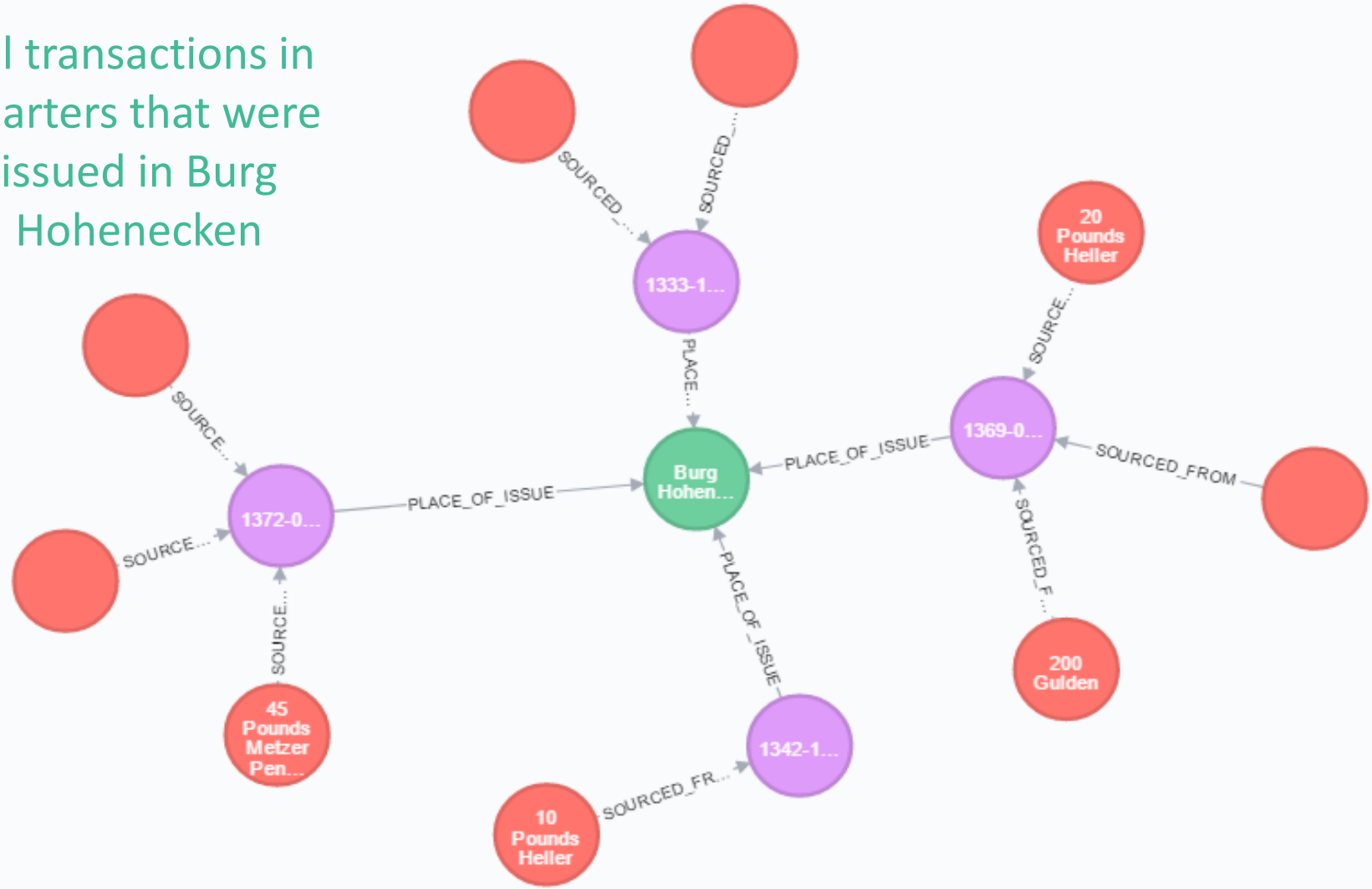
- Enable us to filter events more thoroughly in combination with the event type

Current Model



Example Query

All transactions in charters that were issued in Burg Hohenecken



Difficulties with Modelling

Charters contain more data than we can afford to model

- Choosing the right cut-off is important

Fitting the design to our questions / queries

- e.g. Focus Group vs. Non Focus Group, Location vs. Building

Not all information in the model is equally important

Getting a broader image of the focus groups without going out of scope

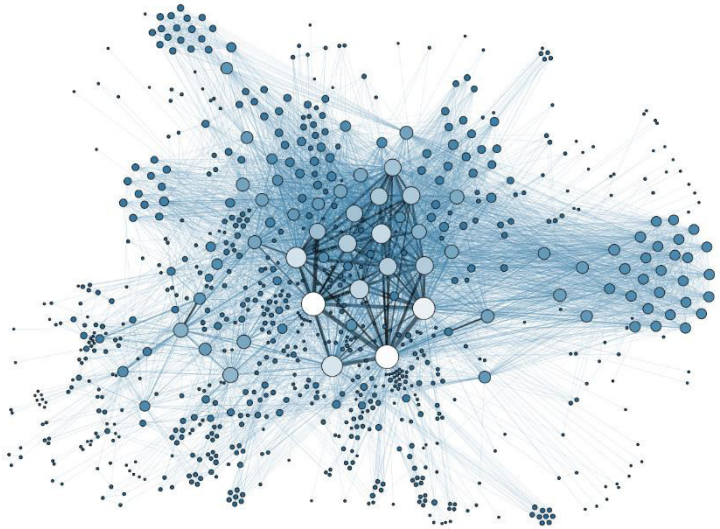
Outlook

Georeference the Location nodes in GIS

Improve the Visualization

- *Neo4j* Browser doesn't allow for much customization
- Use *GEPHI* (visualization and network analysis software)

Combine the Neo4j network analysis with the GIS results in the Unity Game Engine



Thank you for listening!

Supervisor: Dr. Susanne Krömker, IWR Heidelberg
Partner: Aaron Pattee

Contact:
Mario Hinderhofer
mariohinderhofer@gmail.com