Data visualization for SORMAS

10 November 2020

Stéphane Ghozzi

Hemholtz Centre for Infection Research (HZI)

stephane.ghozzi@helmholtz-hzi.de

Software

SORMAS is an infectious-disease surveillance and response tool for recording:

- **cases** (persons infected)
- their **contacts**
- **events** and their **participants**

open source

deployed in France, Germany, Ghana, Nigeria, Switzerland

HZI + partners: development, training, support

current development focus on **COVID-19**

https://www.sormas.org/
https://www.sormas-oegd.de/
https://github.com/hzi-braunschweig/SORMAS-Project

Data

High dimensional:

- person (address, phone number, sex, age, occupation, ...)
- disease course (symptoms, laboratory tests, ...)
- context (setting of event, ...)
- workflow of public-health workers (contact in quarantaine, ...)

... and much more!

N.B. strong data privacy

In progress: generation of a **credible synthetic data set** for

- software testing
- training
- **analyses, visualizations, collaborations**

Visualizations

The usual

- distributions
- time series
- choropleth maps
- . . .

Indicators

- reproduction number R_e
- dispersion factor K
- anomaly detection
- . . .

Networks

the most interesting... and challenging!

Idea/requirement:
- graphs of **infections**, **contacts** and **event participation**
- both **abstract** and in geographical **space**
- highlight necessary collaboration between adminstartive units

Difficulty:
- quickly difficult to read, e.g. ~ 13,000 persons and events in COVID-19 test data for only 3 counties
- high-dimensionality of relevant information

$\implies$ need to hierarchize and navigate information!

Networks: first idea

In progress, rough sketch... still messy!

❶ Build graph:

- nodes = persons and events
- edges = infection, contact and participation

❷ Filter nodes:

- by time, e.g. of reporting
- by **components** (connected subgraphs) ~ clusters

N.B. COVID19 component sizes skewed: most cases infect no one, a few larger clusters

❸ Aggregation at regional level:

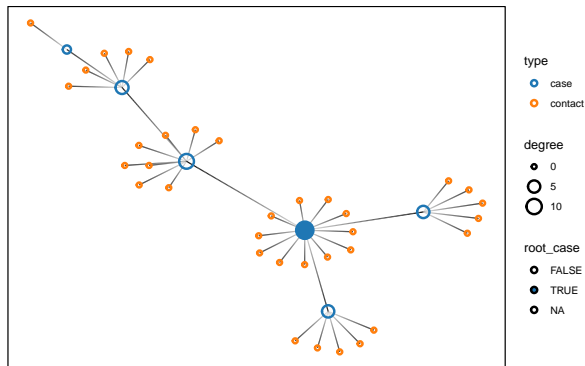- highlight *trans-regional* components

❹ Visualization:

- node color = type of person or event
- node size = degree or betweenness (aggregated: number of persons)
- edge type = type relation (aggregated: number of relations)
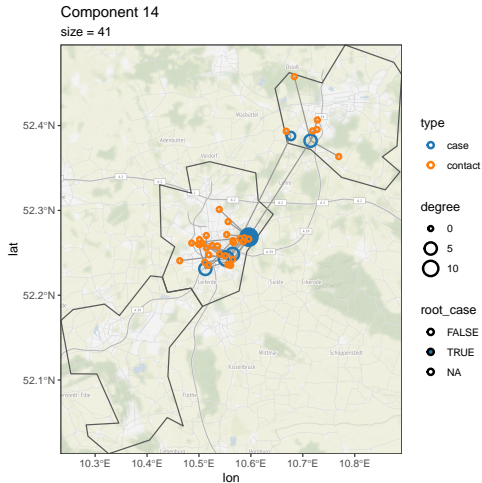- **interactive**

Example 1: component with infection chain
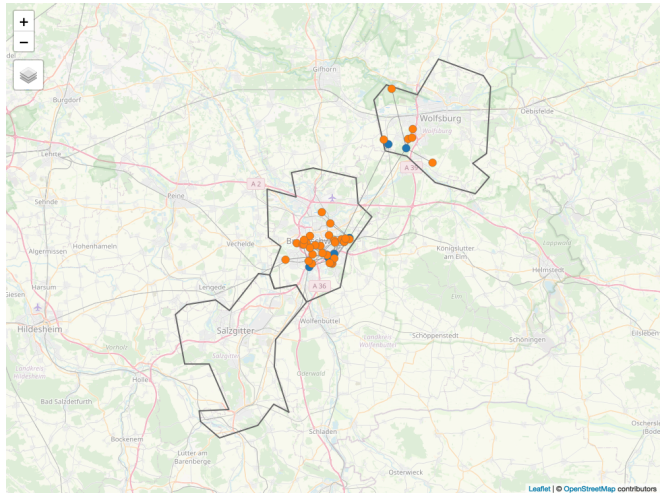
abstract visualization

Component 14

size = 41

static visualization in space

Component 14
size = 41

interactive visualization in space

Example 2: component with event
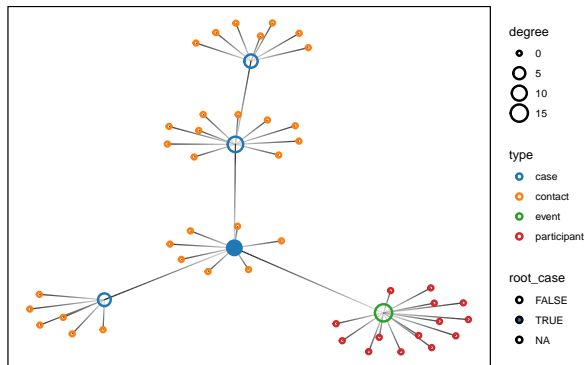
abstract visualization

Component 55

size = 49

static visualization in space

Component 55
size = 49

interactive visualization in space