ERHVERVSSTYRELSEN

Fraud Detection with Graphs at the Danish Business Authority

Dan Saattrup Nielsen

20210309

About Me

- Research Associate in Machine Learning, University of Bristol
- Former Machine Learning Consultant at the ML Lab, DBA
- PhD in Mathematics, University of Bristol

The Danish Business Authority

ERHVERVSSTYRELSEN

- Responsibilities of the authority:
  - Good framework conditions for business development
  - Efficient and professional business service for companies
  - Effective business regulation and enforcement
  - COVID-19 compensation schemes
- All Danish businesses have to register at the authority

The Problem

The state is divided into areas of operation, each checking for adherence to rules and regulations within their area of responsibility.

Fraudsters often operate across these barriers because they are motivated by financial gain.

No one is able to connect the dots...

The Knowledge Graph

Knowledge Graph

ERHVERVSSTYRELSEN

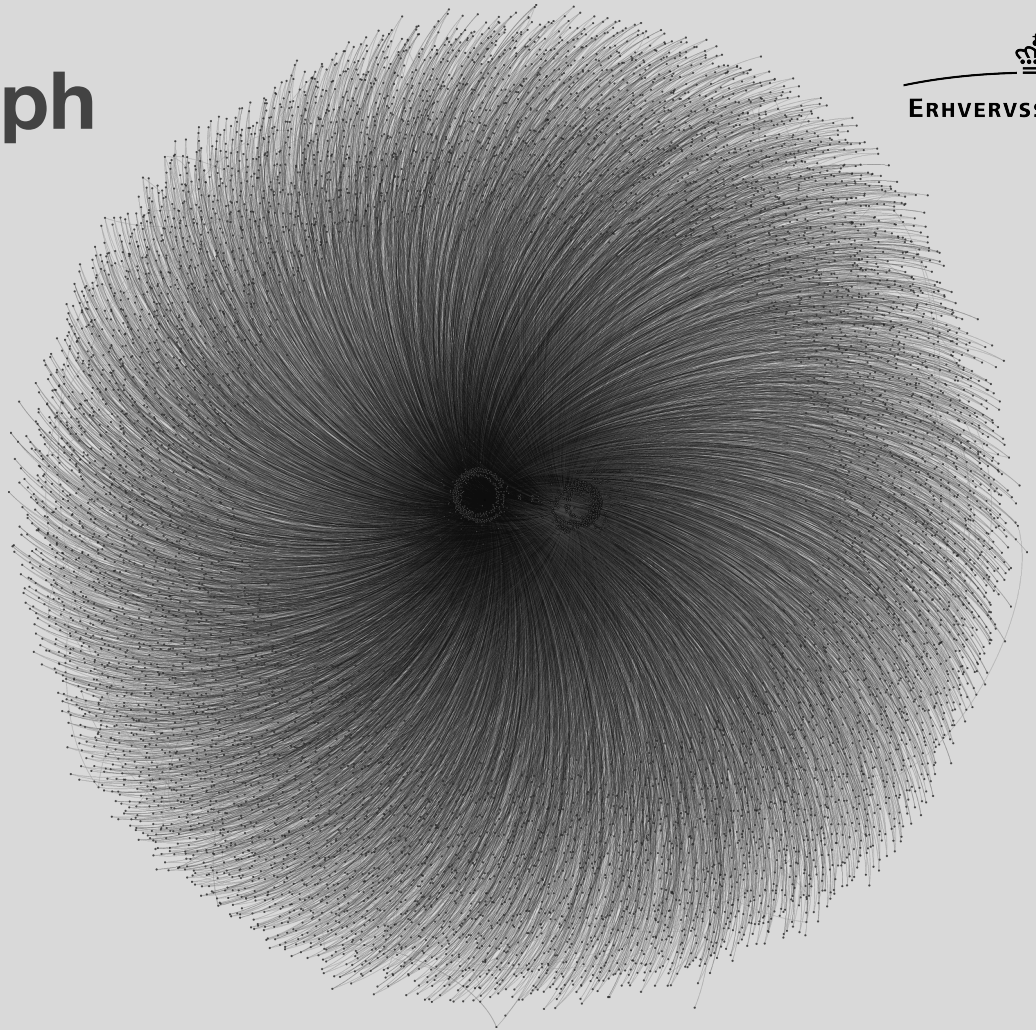
800,000+ companies

400+ million nodes

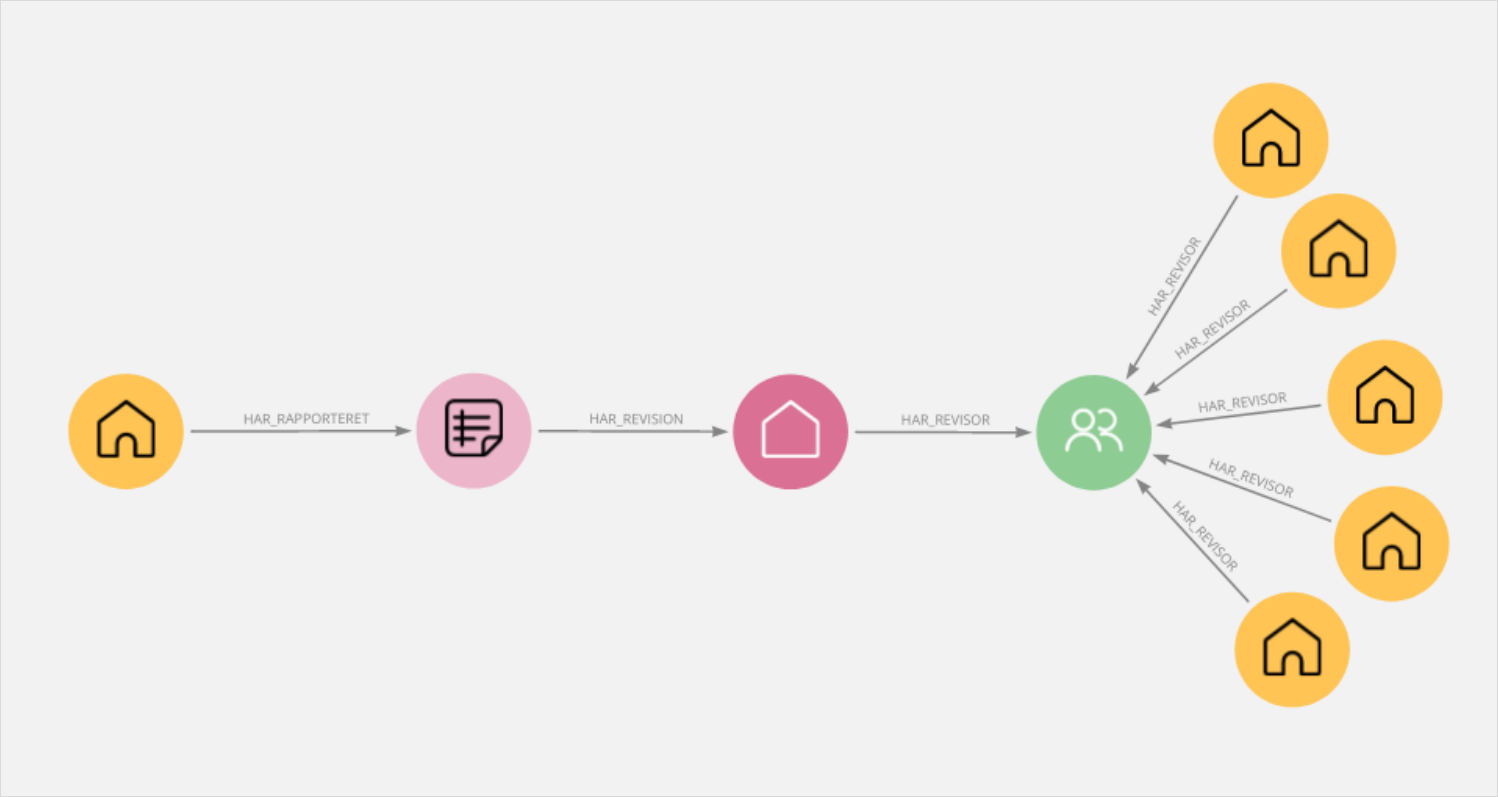
700+ million relations

Includes:

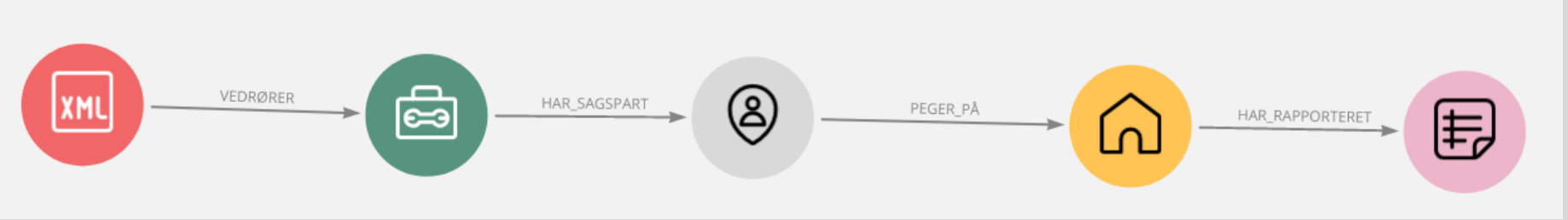
- Ownership
- Management
- Addresses
- Accounting
- Tax & VAT
- Case handling
- IP addresses
- And more!



Network Example: Accountants



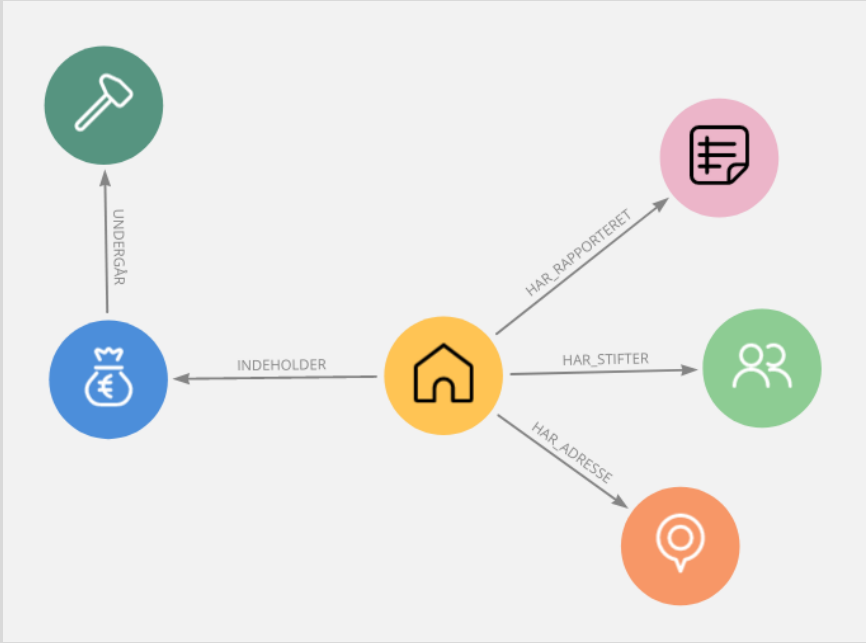
Network Example: Application to Annual Report



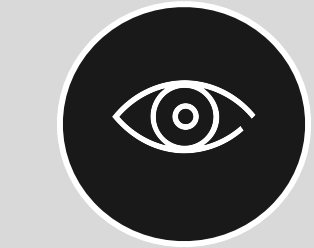
A Data-Driven Risk Score

Multiple features are used in order to calculate a risk score:

- Changes to company activity
- Preliminary VAT determinations
- Location history
- Key figures from annual accounts
- Changes to management
- Solidity



Model to Model



Identity

Control of identity papers for foreign actors in management

Assets

Fictitious or erroneous asset declarations in companies

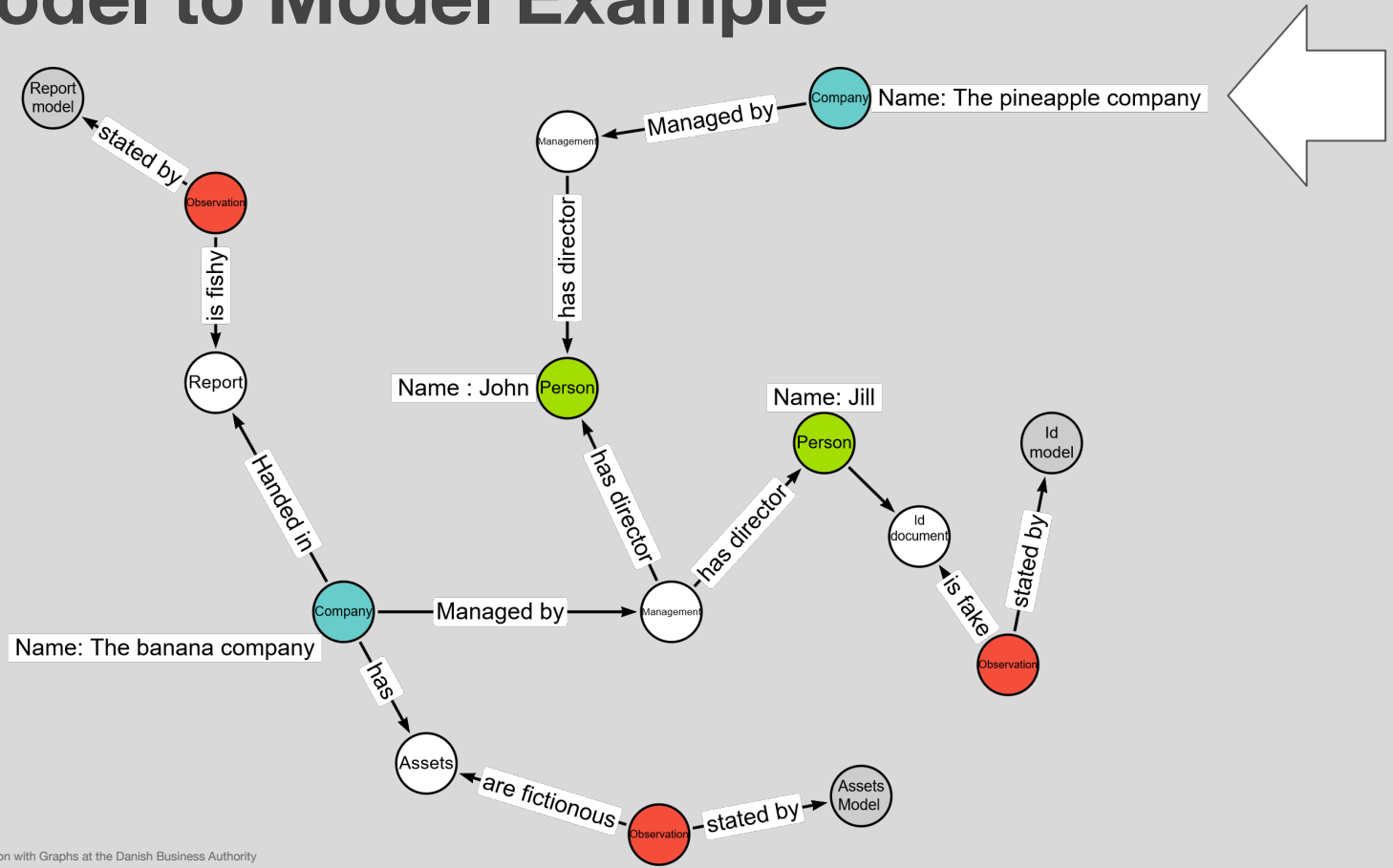
Reports

'Weaponising' unstructured text from audits

1st line

Assess the risk of fraud for a company

Model to Model Example



COVID-19 Compensation Schemes

COVID-19 Lockdown

March 2020
4 compensation areas

230,000+ applications
3.5+ billion EUR

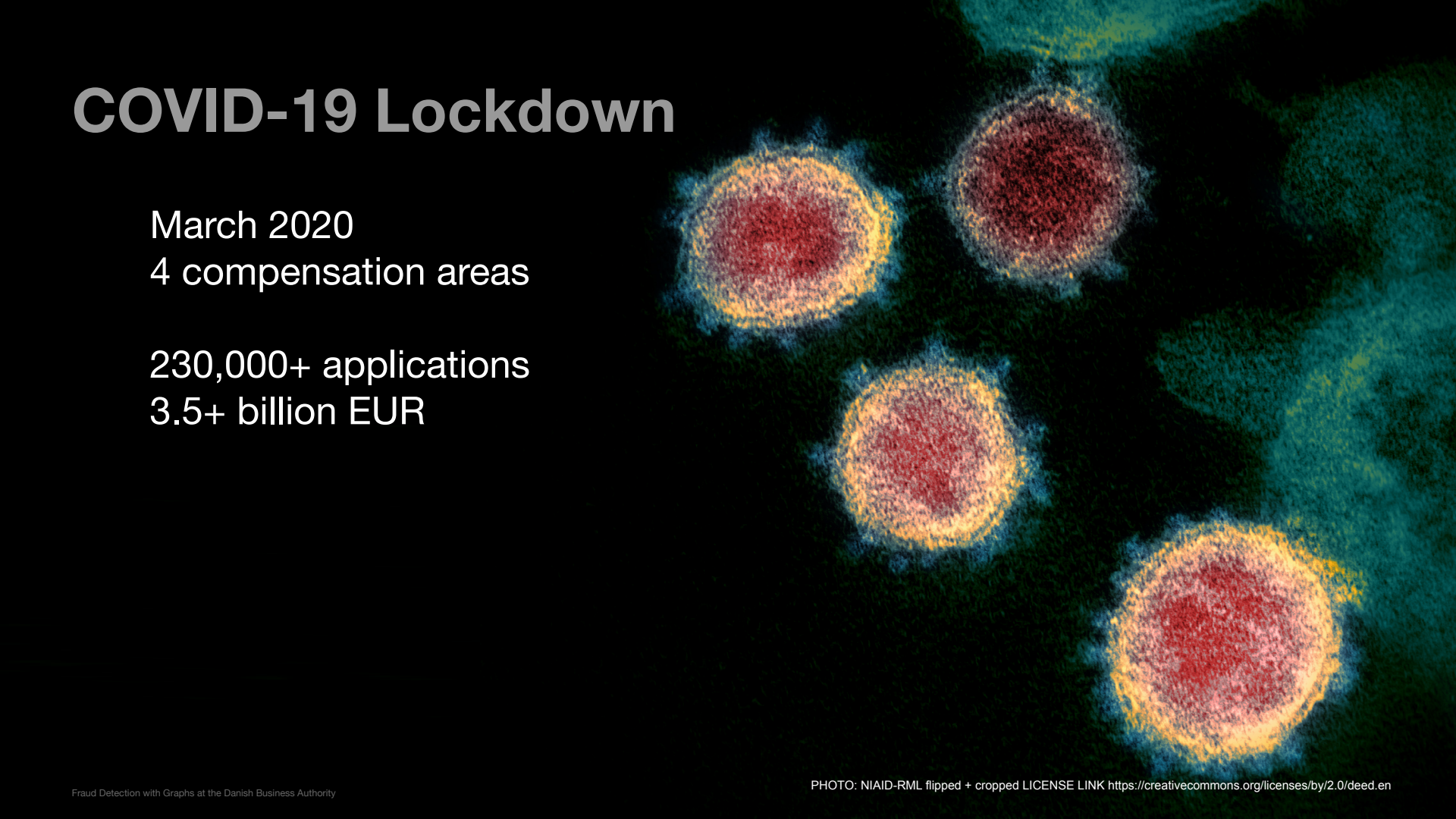
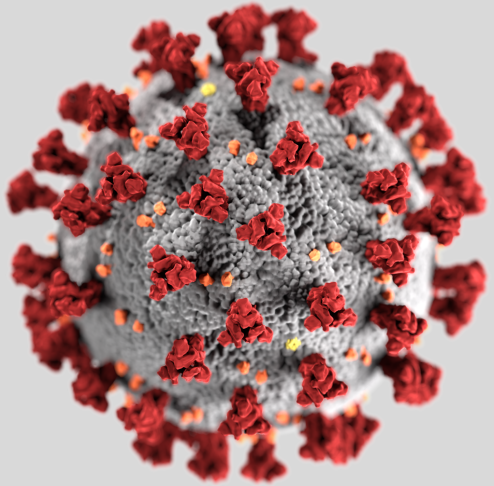


PHOTO: NIAID-RML flipped + cropped LICENSE LINK https://creativecommons.org/licenses/by/2.0/deed.en

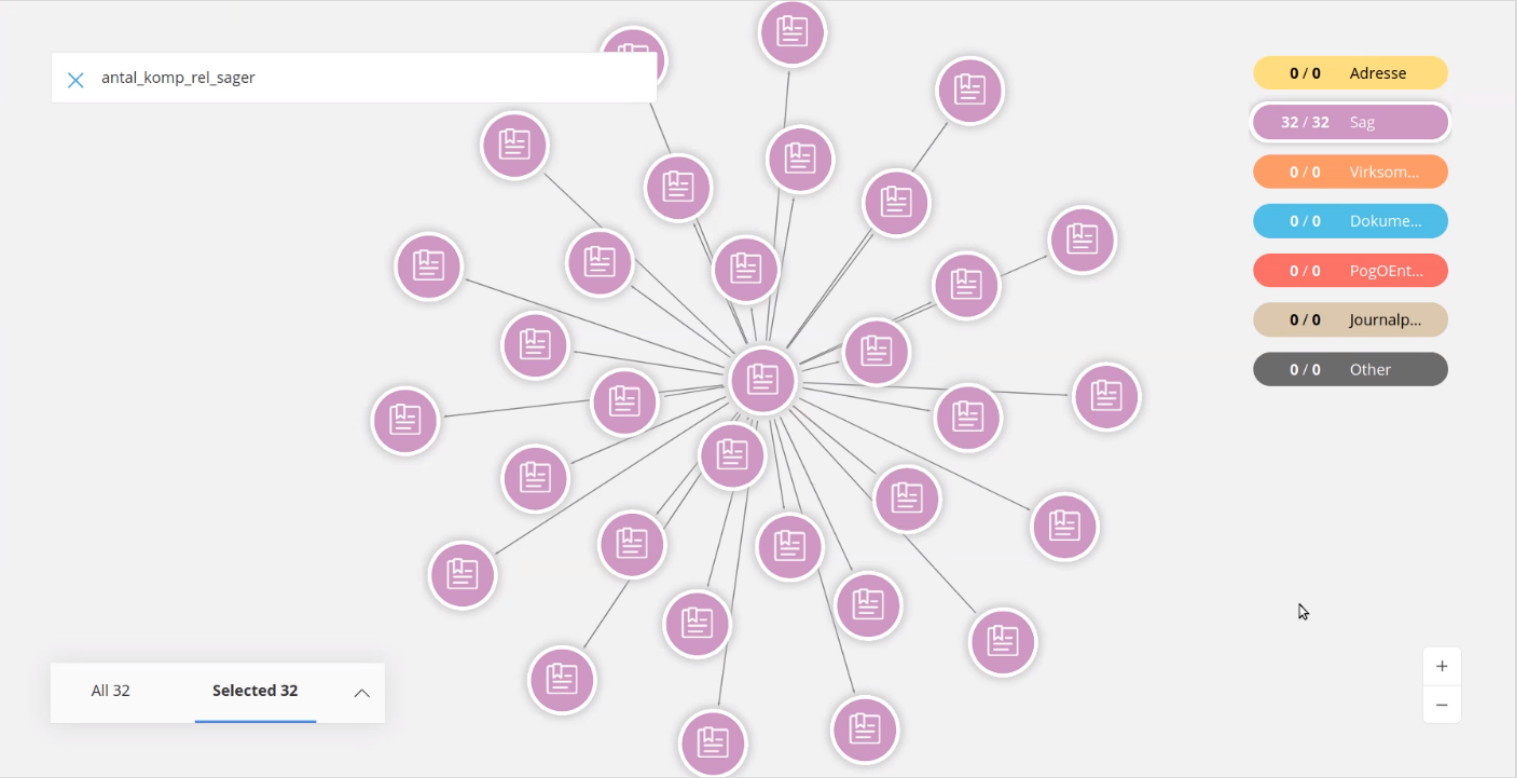
COVID-19 Application Data

Features include:

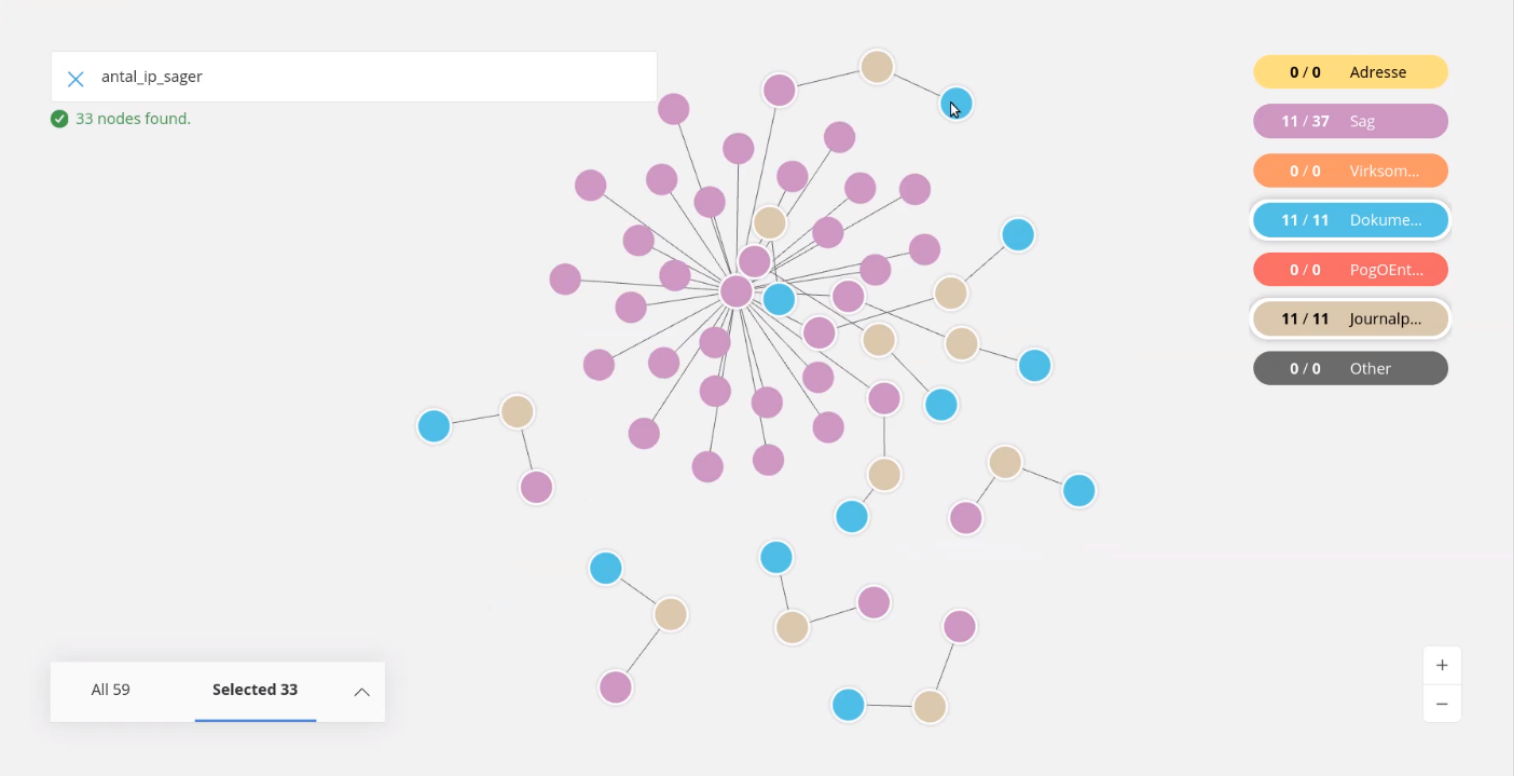
- Related compensation cases
- IP-addresses used in compensation
- Application deadlines
- Compensation history of the applicant
- Data from compensation applications



COVID-19 Application Data: Related Cases



COVID-19 Application Data: IP-addresses



Ethics

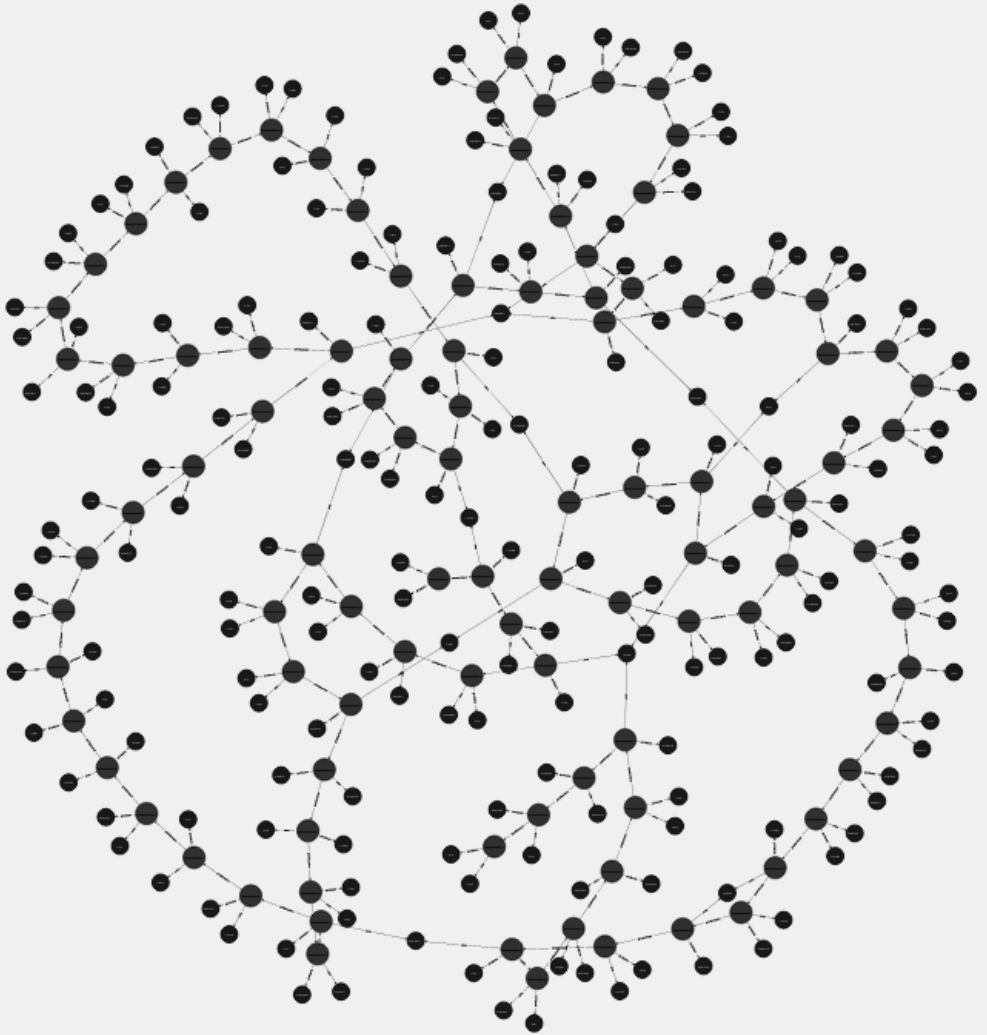
Ethical Principles (informal)

ERHVERVSSTYRELSEN

- Do not make lists of people
- No calculations without a concrete business purpose
- Check that we are hitting what we are aiming for
- Data driven ethics, do not rely on good intentions
- Always be able to answer: What models use which data for what purpose?

Data Driven Ethics

- (Near) real time events
- All event messages kept
- Evaluation data



Key Takeaways

Principles

ERHVERVSSTYRELSEN

- **Connect** authority data
- Event driven, so we may react in **near real time**
- **Network analysis** – understanding company lifecycles
- Metadata strategy - **data from data**
- ML – **machine insight**
- Transparency and traceability - so we may **explain**
- Evaluation – for continuous **improvement**
- **Collaboration** with domain experts

Effect

ERHVERVSSTYRELSEN

- All companies are subject to control
- Automatically with every change in data
- Full transparency and traceability
- Human in the loop

ERHVERVSSTYRELSEN

Thank you for your attention!


saattrupdan


saattrupdan.github.io

Globe icon by Vitaly Gorbachev