Exploiting Forensic DNA data to draw pictures of ‘organized’ or transnational crime(s)

Patrick **Jeuniaux**, Bertrand **Renard** & Vanessa **Vanvooren**

NICC INCC

Mathilde **Huet** & Antony **Leplingard**

POLICE NATIONALE · DIRECTION CENTRALE DE LA POLICE JUDICIAIRE

Wim **Bernasco**

nscr

VU University Amsterdam

Sabine **De Moor**, Stijn **Van Daele** & Tom **Vander Beken**

UNIVERSITEIT GENT

The Stockholm Criminology Symposium

Stockholm, 10th of June 2015

Overview

1. Context
2. Objectives
3. Methodology
4. Results
5. Summary
6. Future activities

Institutional & personal **context**

- My immediate context:
  - Forensic and Criminology Institute in Belgium
    - National Forensic DNA database
      - Data analysis issue
- Our partners:
  - BE : Univ. Gent / Vrije Univ. Brussel / Kathol. Univ. Leuven
  - FR : French Judiciary police.
  - NL : NFI / NSCR / Univ. Leiden / Vrije Univ. Amst.
  - UK : Northumbria Univ., SCJS.
  - Europol.

General notions: **DNA data**

Sample of biological material (ex: blood)

↓

Useful for unique identification

DNA profile data

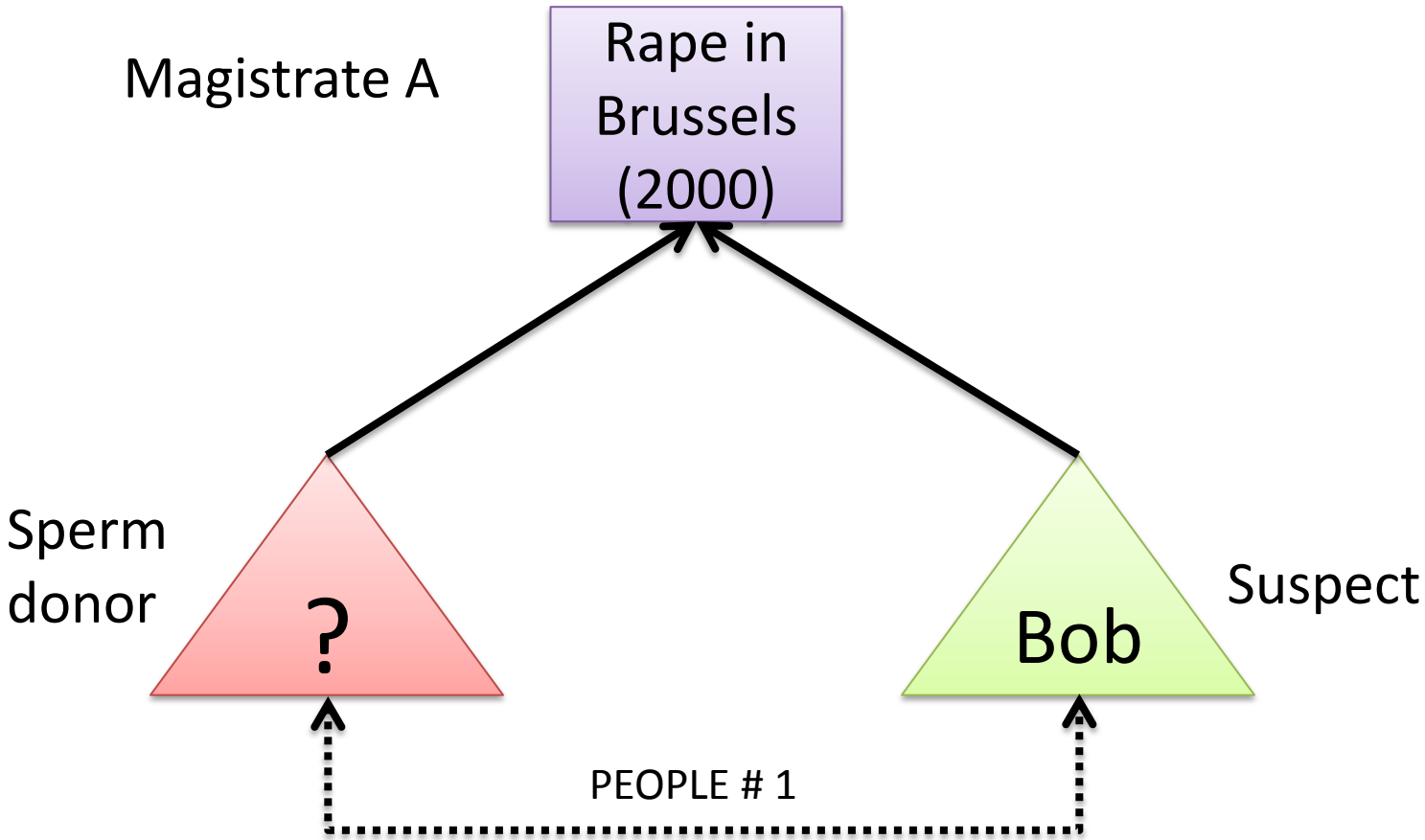
- Locus 1 : 12, 13
- Locus 2 : 8, 10
- Locus 3 : 2, 3
- (…)
- Locus 10 : 1, 15

↓

Useful for comparing DNA profiles

DNA data base : store of DNA profiles

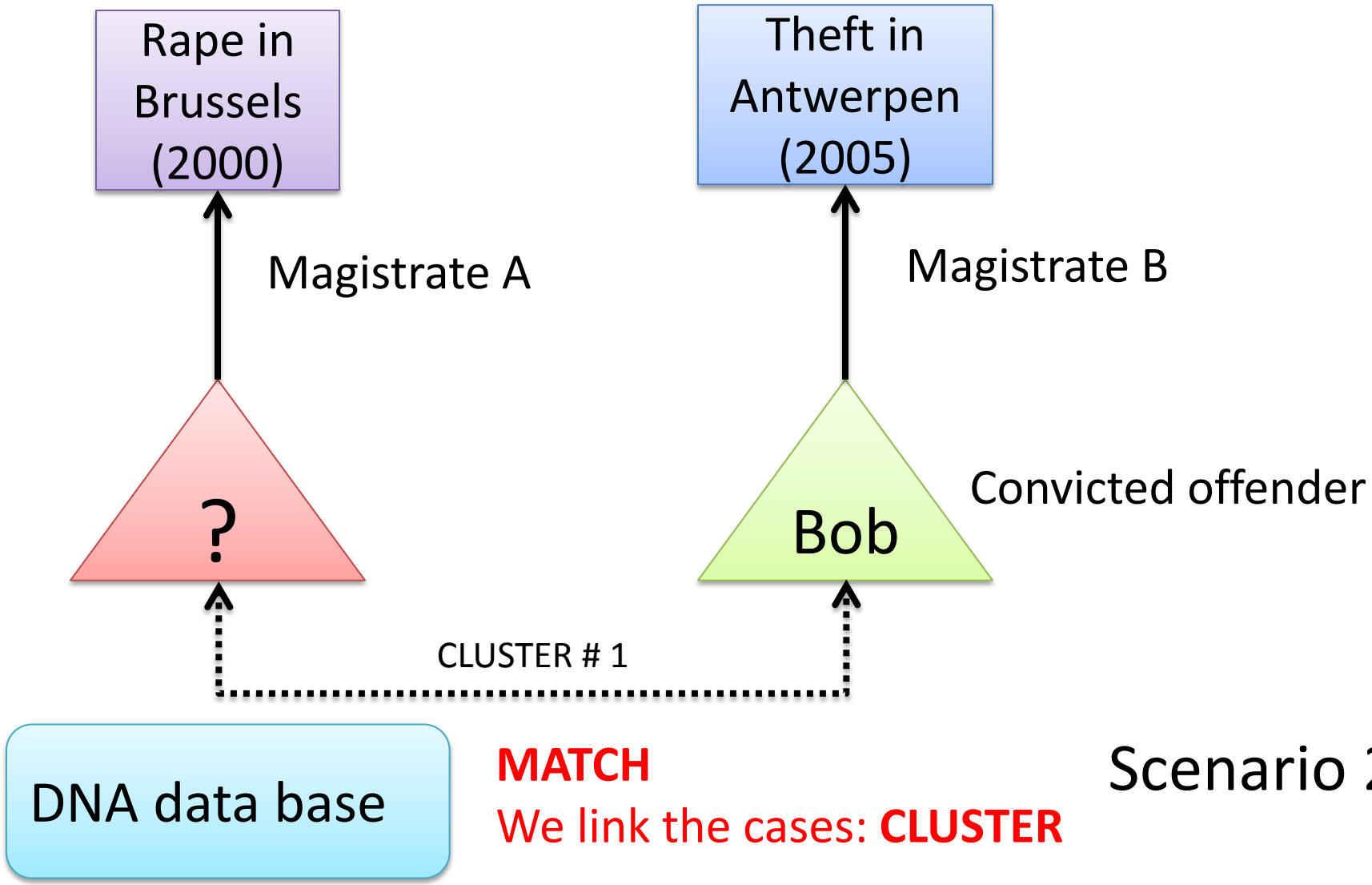
General notions: **comparing** DNA profiles



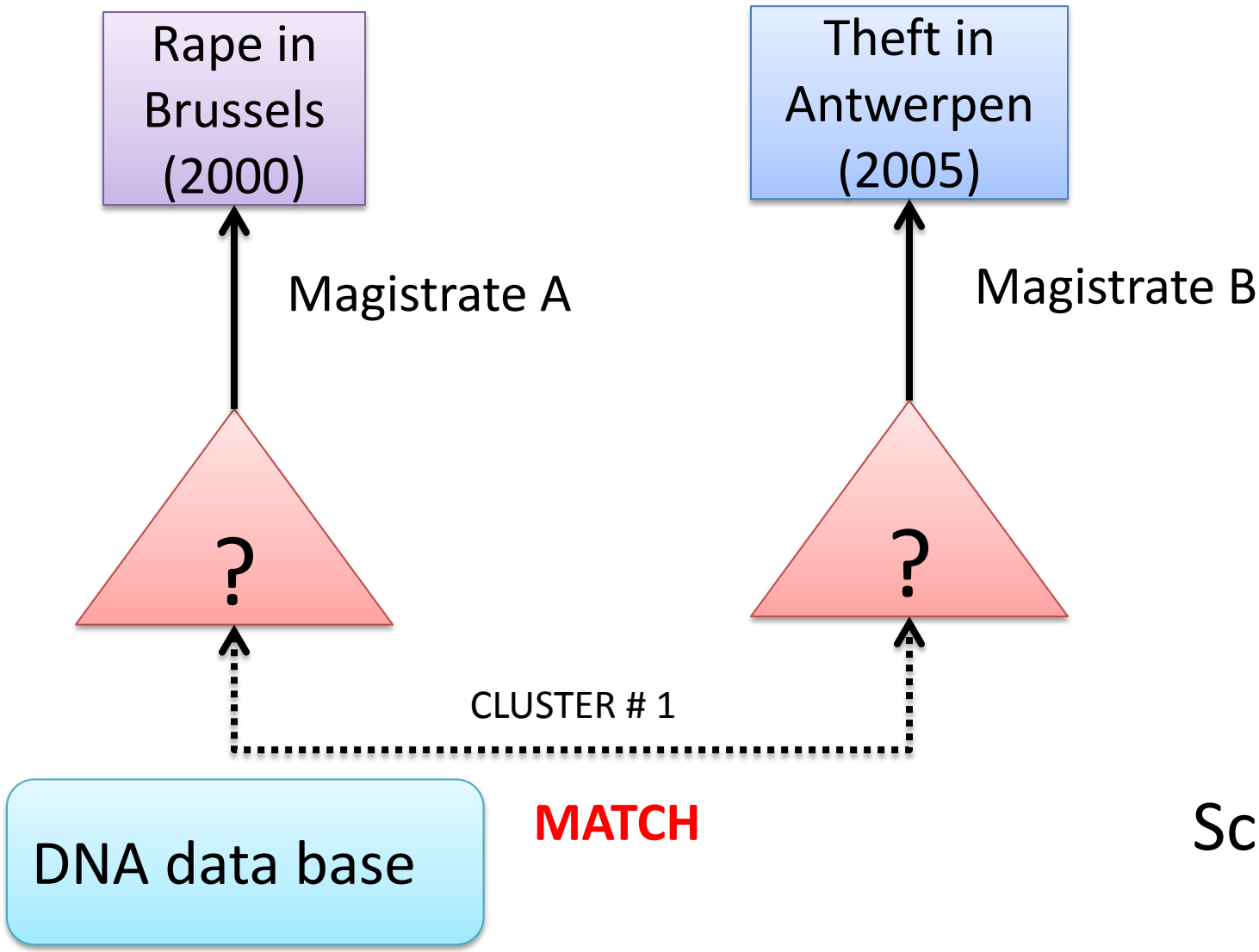
Result of the comparison: **MATCH**

Scenario 1

General notions: comparing through a **database**

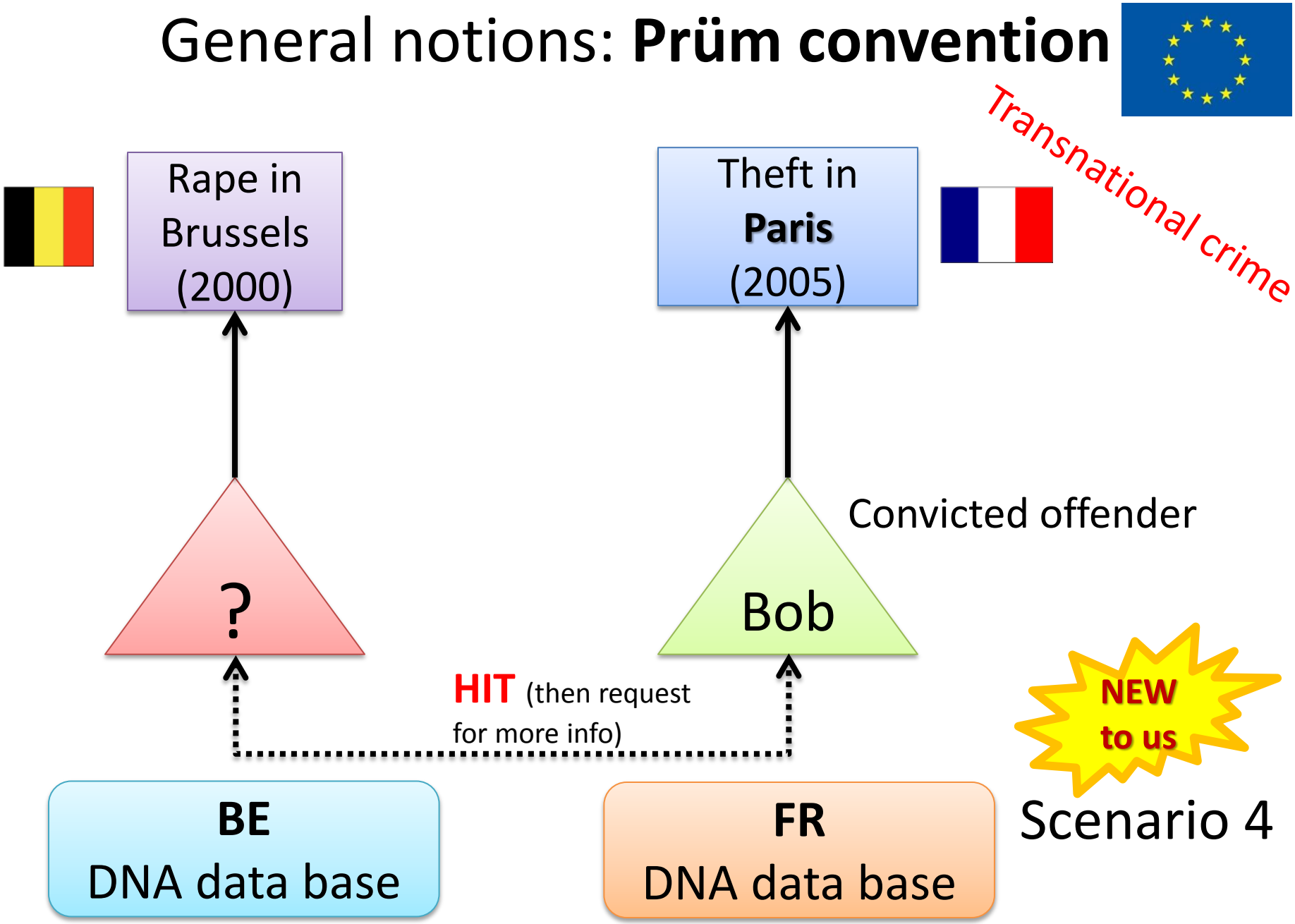


General notions: **linking** cases



Scenario 3

General notions: **Prüm convention**

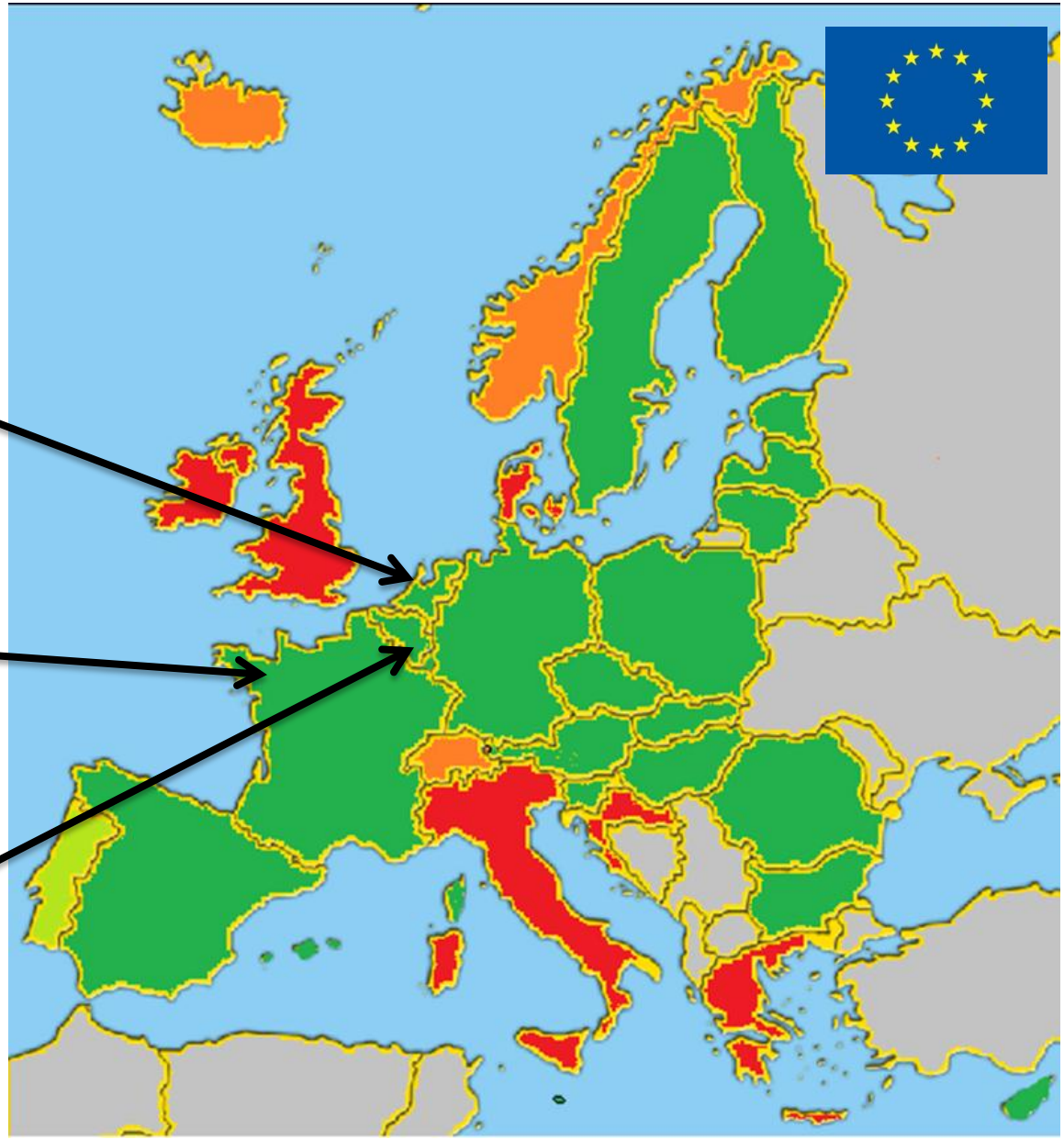


22 EU countries are Prüm operational (in green)

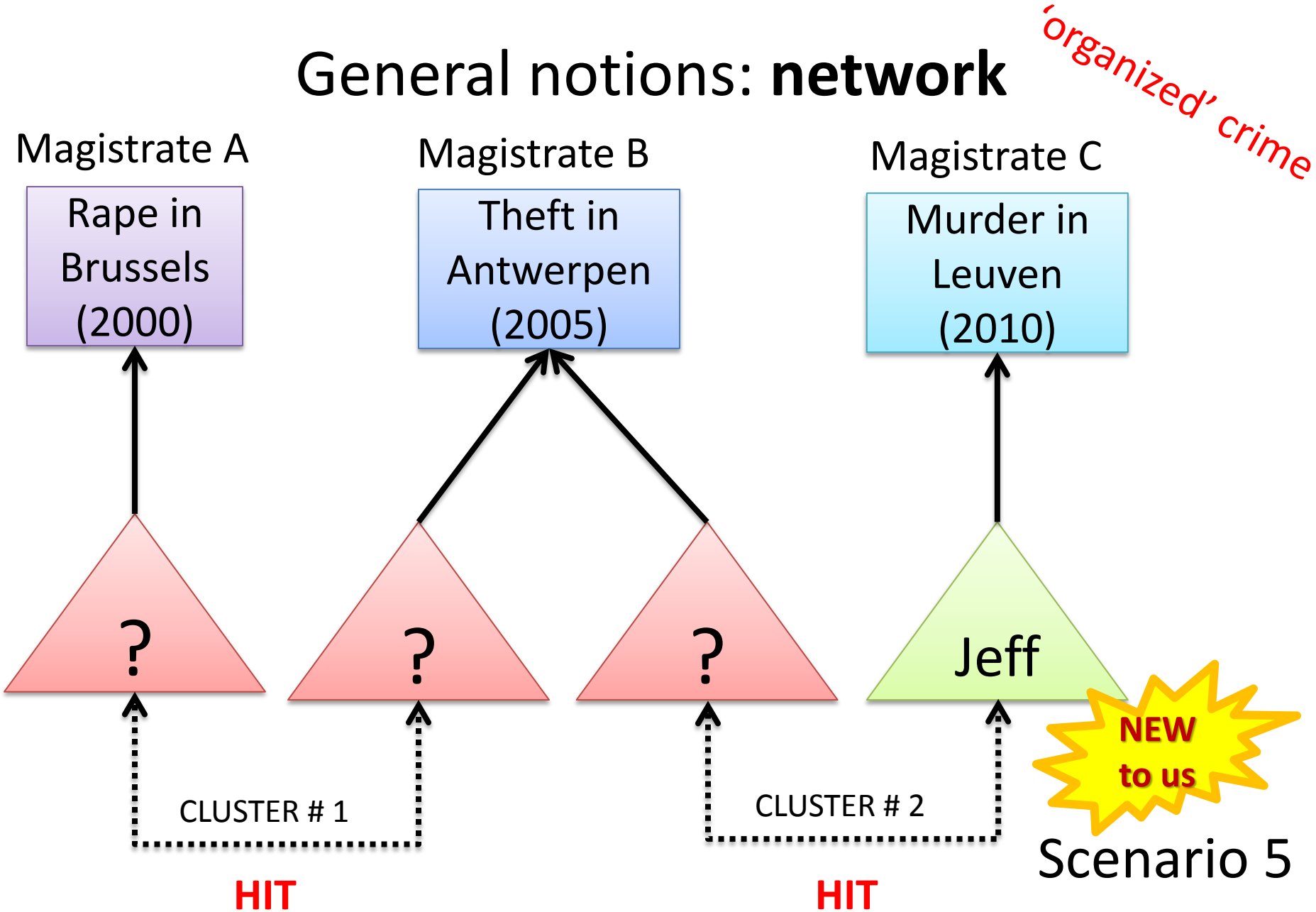
NL
operational
since 2008

FR
operational
since 2009

BE
operational
since 2014



General notions: **network**



Context : A quick **chronology**

- 1999 : DNA law in Belgium
- 2005 : Prüm convention
- 2008 : Prüm EU council decisions
- 2011 : new BE law
- 2012 : start of PIES project
  - Co-funded by European Commission
- 2014 : BE becomes Prüm operational
- 2014 : start of Be-Gen project
  - Co-funded by BE science policy office

Context: DNA-based **issues**

IMPORTANT
but not discussed today

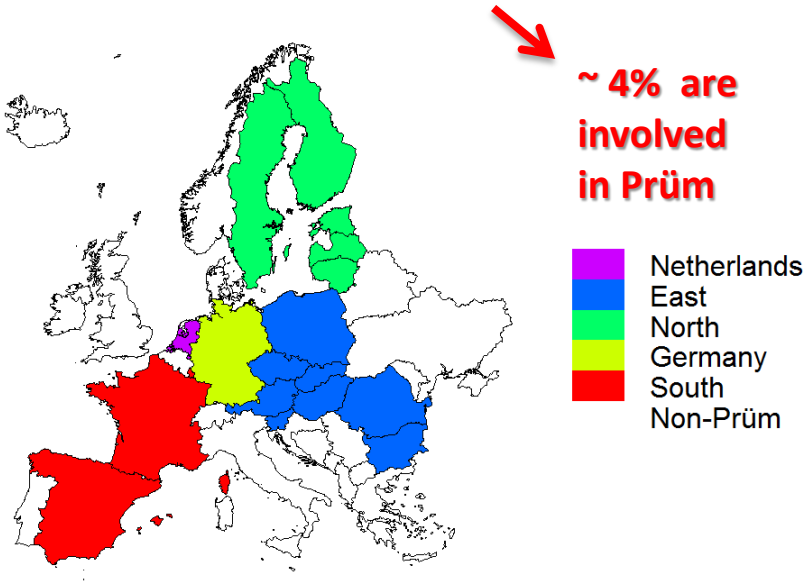
- Better **knowledge** of database **impact**.
  - **Evaluation** of DNA-based evidence and Prüm by looking at **concrete** cases (not only statistics).
- Better knowledge of database **content**.
  - NEW Better **communication** / **reporting**
    - Issue of **understandable** and **uniform** statistics (Prüm).
  - NEW **Unexploited** information?
    - Networks: forensic intelligence, judicial cooperation
- Two types of **uses**
  - Operational (main stakeholders: magistrates)
  - Strategic (criminologists, national or EU policy makers)

Objectives of this study

1. Gaining a better **understanding** of the BE DNA database.
2. Drawing useful **graphics**
3. Calculate reliable **statistics**.
4. Draw pictures of '**organized**' crime (networks).
5. Draw picture of **transnational** crime (Prüm).
6. Improve our **reporting**.
7. Prepare a **method** to help other EU member states to produce similar graphics and statistics.
8. Conducting **criminology** research.

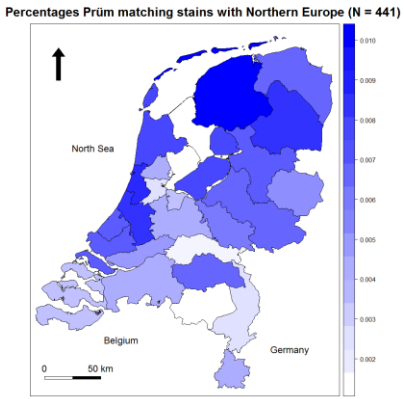
Bernasco, Lammers, & van der Beek (2014)

The Netherlands: n = **79 171** unidentified stains



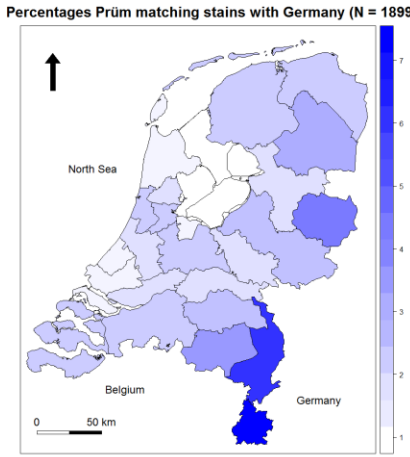
~ 4% are involved in Prüm

0.1 %



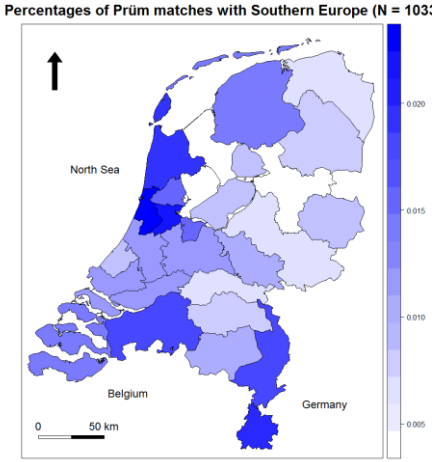
North, n = 441

Percentage of NL **stains** that **match** with Prüm regions



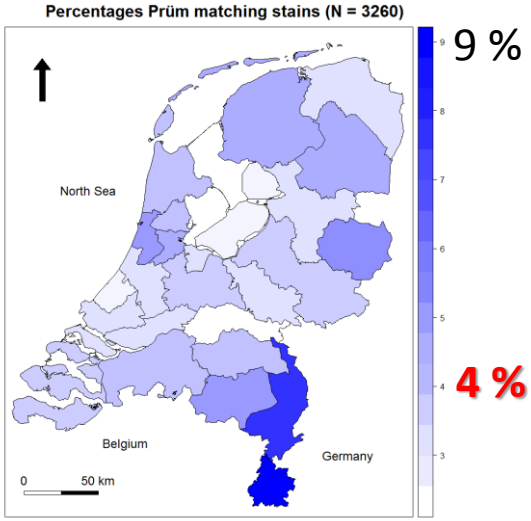
7 %

Germany, n = 1899



0.2 %

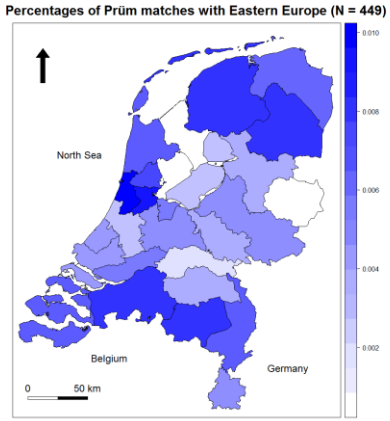
South, n = 1033



9 %

4 %

All, n = 3260



0.1 %

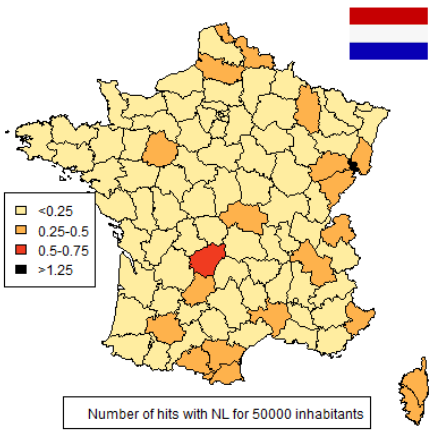
East, n = 449

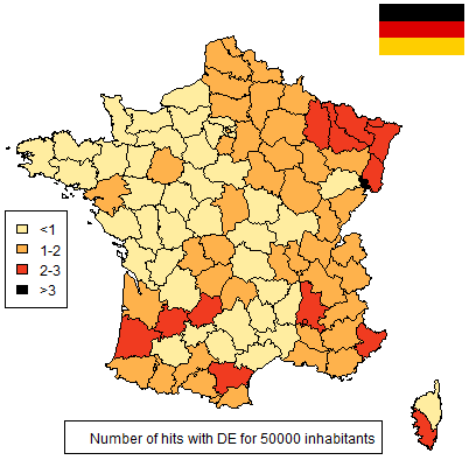
Huet & Leplingard (2014)

Analysis of : 3362 hits

Number of unique hits per 50 000 **inhabitants**

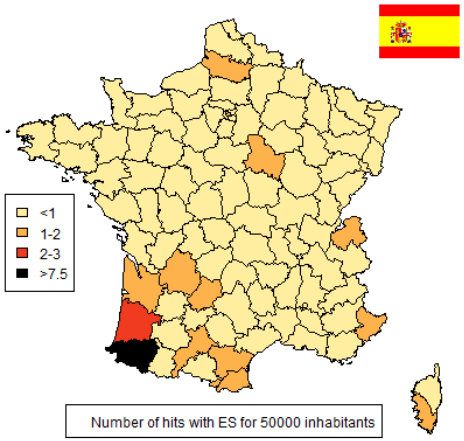
NL : 170 hits

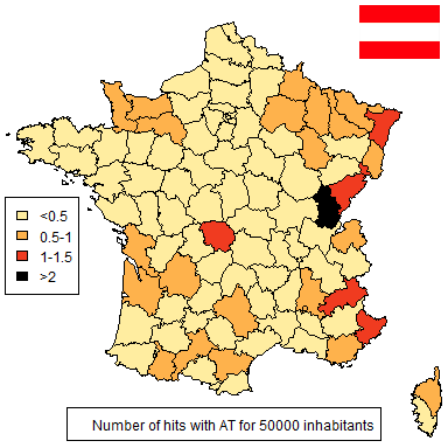




DE : 1534 hits

ES : 865 hits

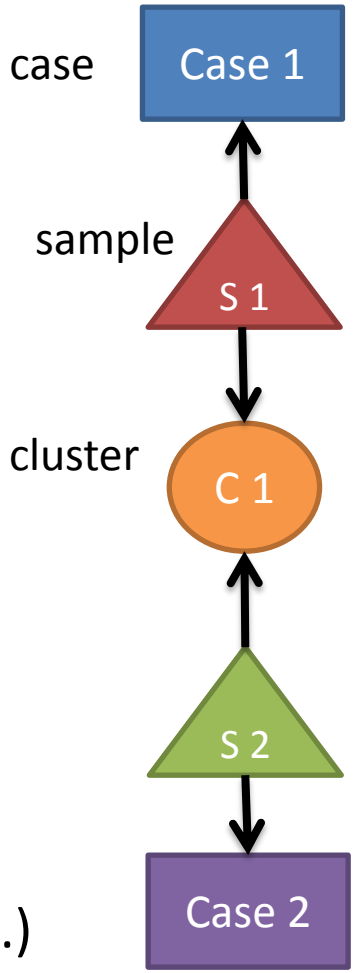




AT : 534 hits

Methodology

- Source of the data: BE admin. database
  - SQL extractions
- Time period: 1999-2015 (17 years)
  - Prüm operation since 2014.
- Size and type of the data:
  - > 50 000 admin. **cases**
    - ~ criminal cases (type of crime)
    - Belgian districts
  - > 100 000 references to **samples**
    - < DNA profiles (stain, person, suspect, offender, …)
    - Country (BE, FR, NL)
  - > 5000 **clusters**
    - ~ people (identified or not)

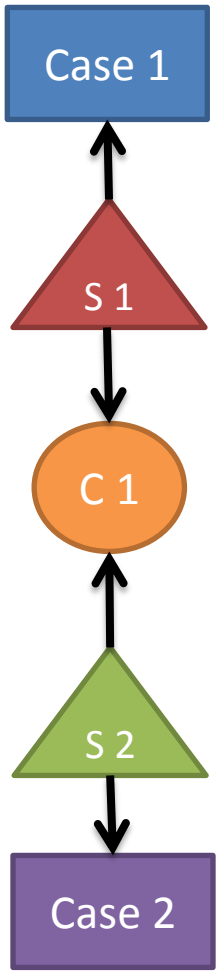


Methodology

- WARNING ⚠
  - An **administrative** case is **not identical** to a **criminal** case or a **crime**.
  - Administrative cases and samples result from **various circumstances**
    - Occurrence of crime
      - + crime being detected and reported
    - Use of forensic DNA data
      - Some district have specific policies
    - Transmission of data to the database

Methodology

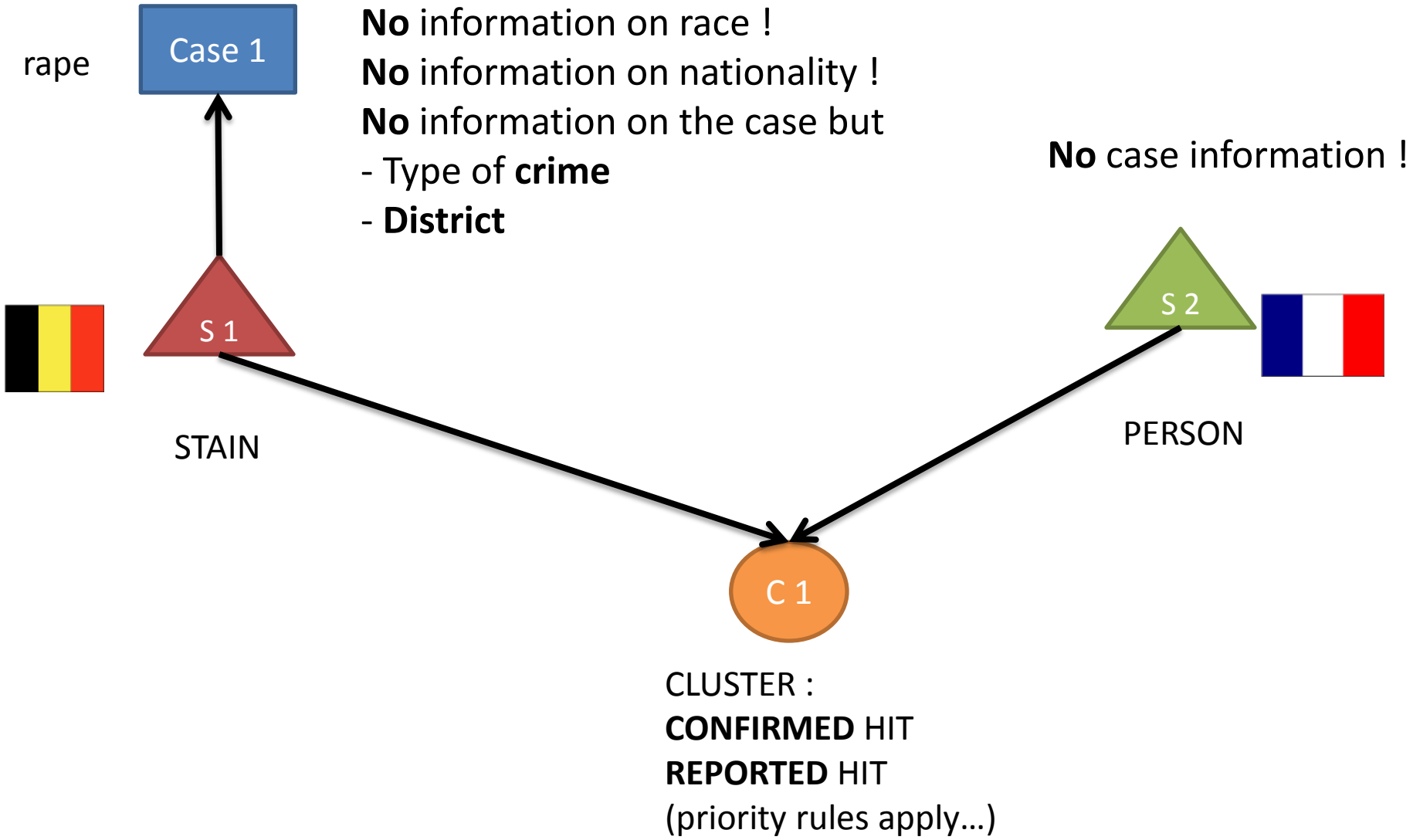
Example of
a component:



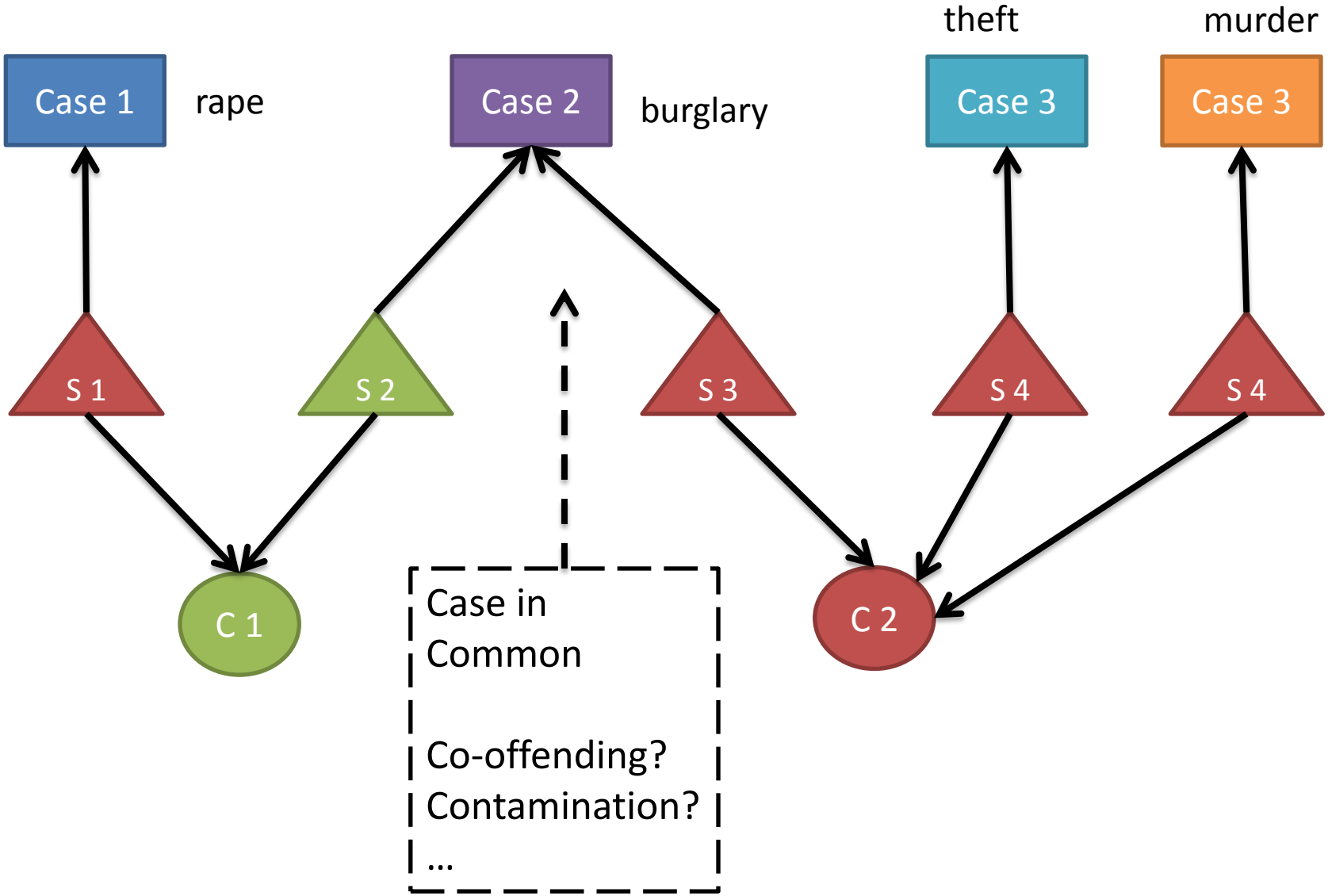
- **Software**
  - Database cleaning/selection/anonymization :
    - Python
  - Relationship recovery/visualization
    - Python + igraph
  - Statistical analysis
    - R + igraph, ca

Recovery of **components** of the graph

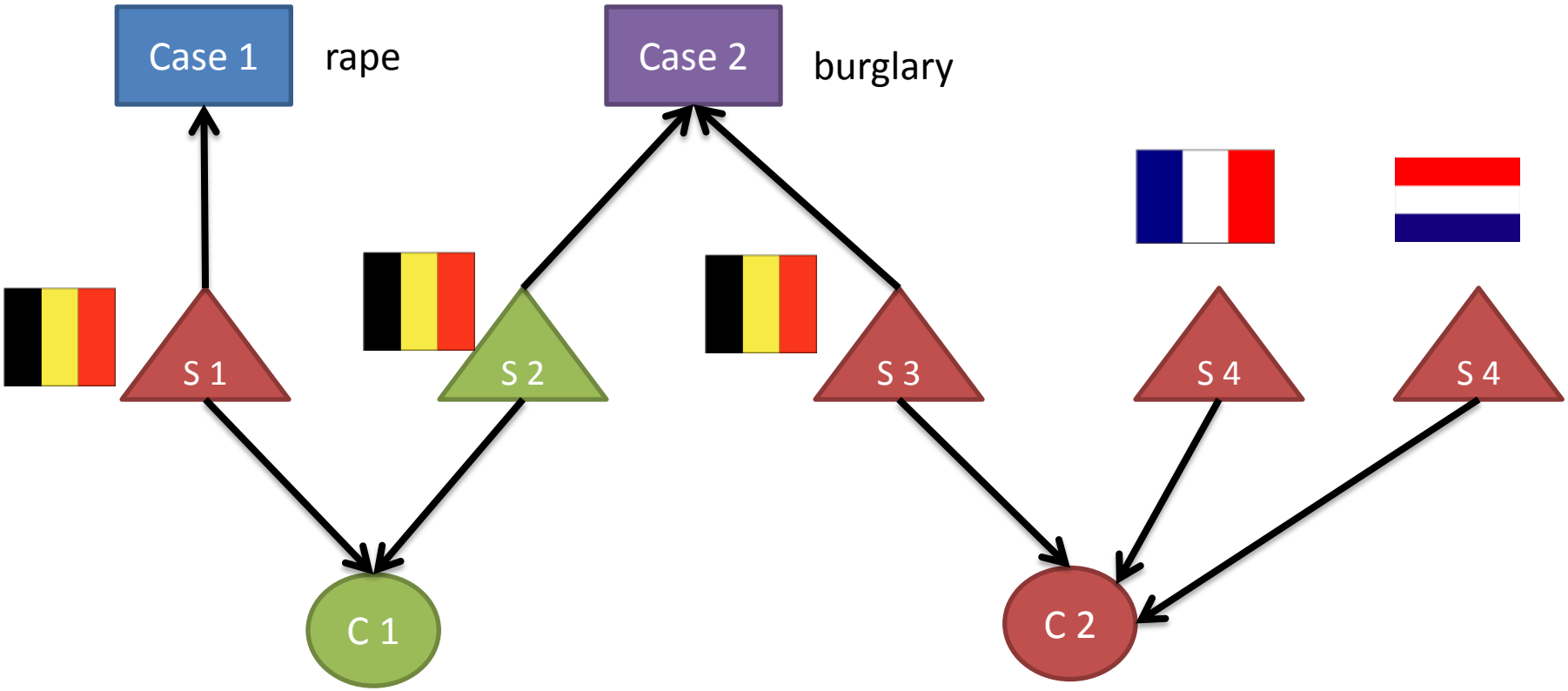
Methodology: transnational crime



Methodology: 'organized' crime



Methodology: organized/transnational



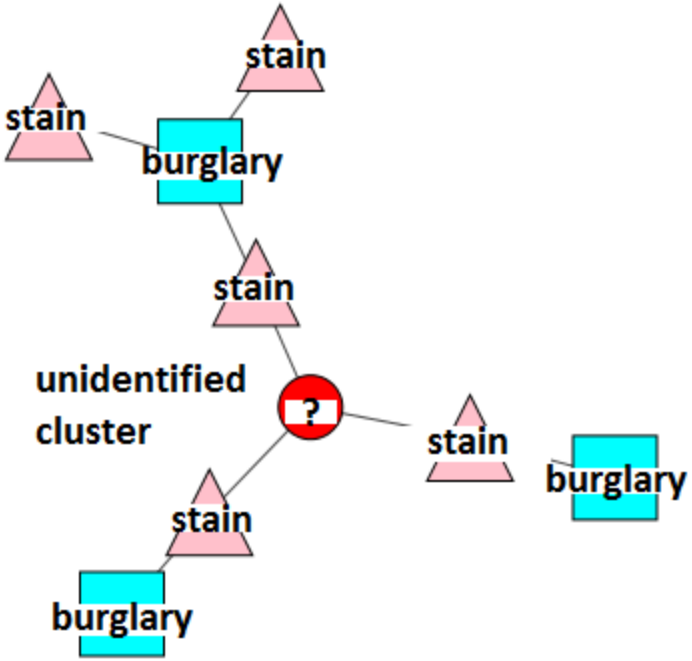
Results

- What **kind** of components do we have?
  - Let's look at **concrete output** of the program.
  - 10 examples !

- **A few statistical observations** on the properties of the objects that make the components.

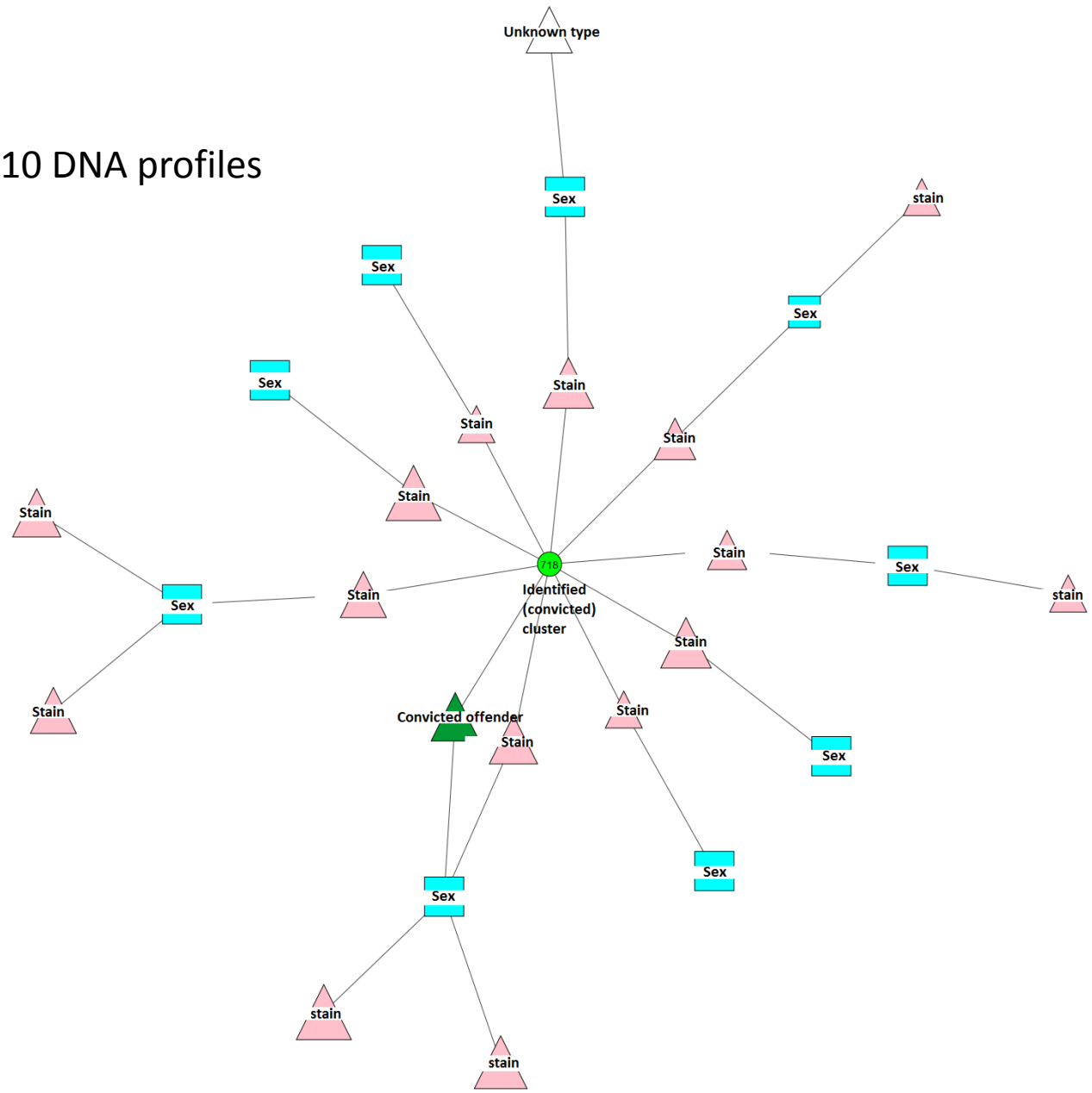
Example 1

- 1 (unidentified) cluster of 3 DNA profiles
- Only national data
  - Only one district.
- Cases (squares)
- Profiles (triangles)
- Clusters (circles)



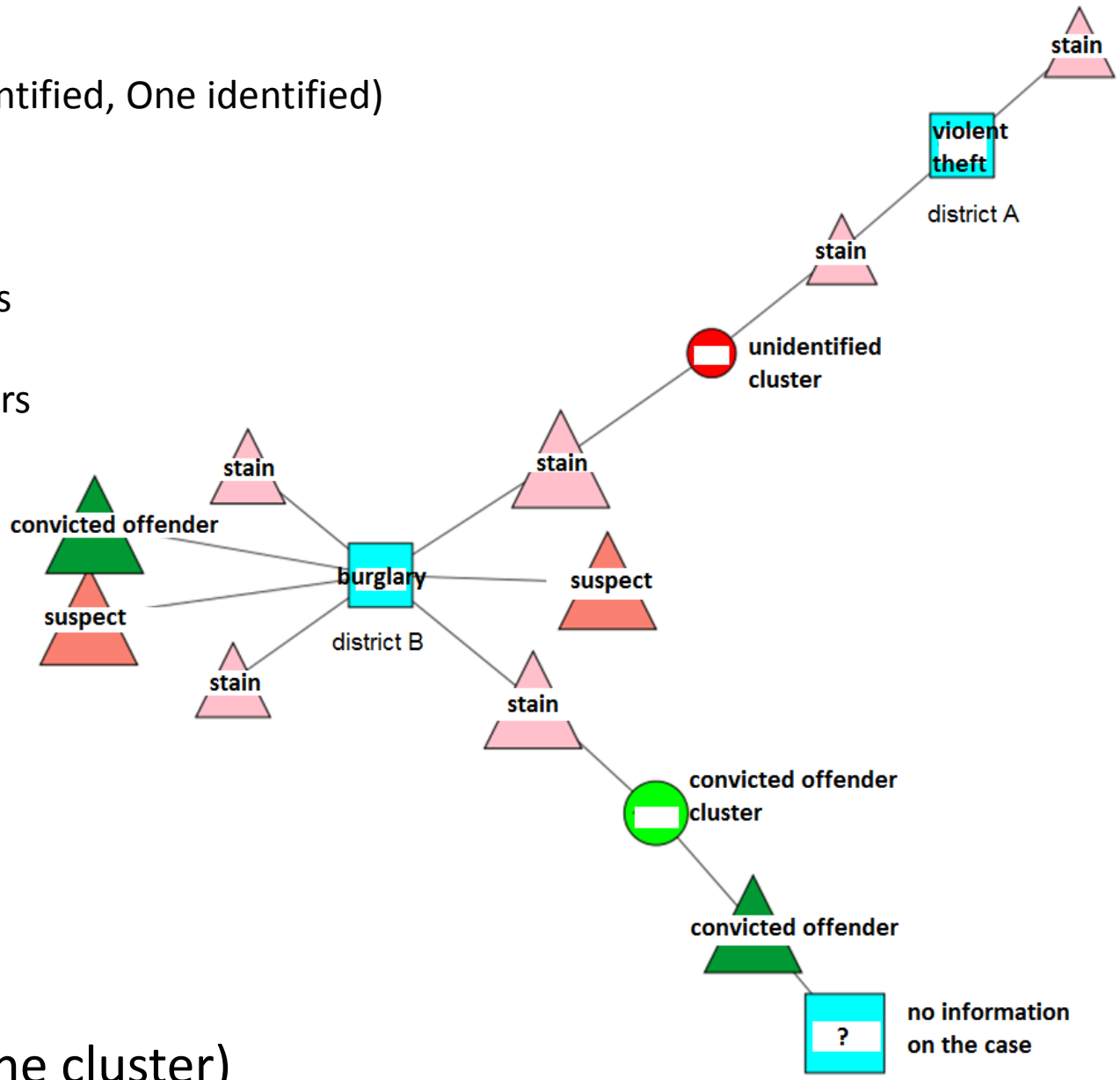
Example 2

- 1 (identified) cluster of 10 DNA profiles
- Only national data
  - Only one district.
- Cases (squares)
- Profiles (triangles)
- Clusters (circles)



Example 3

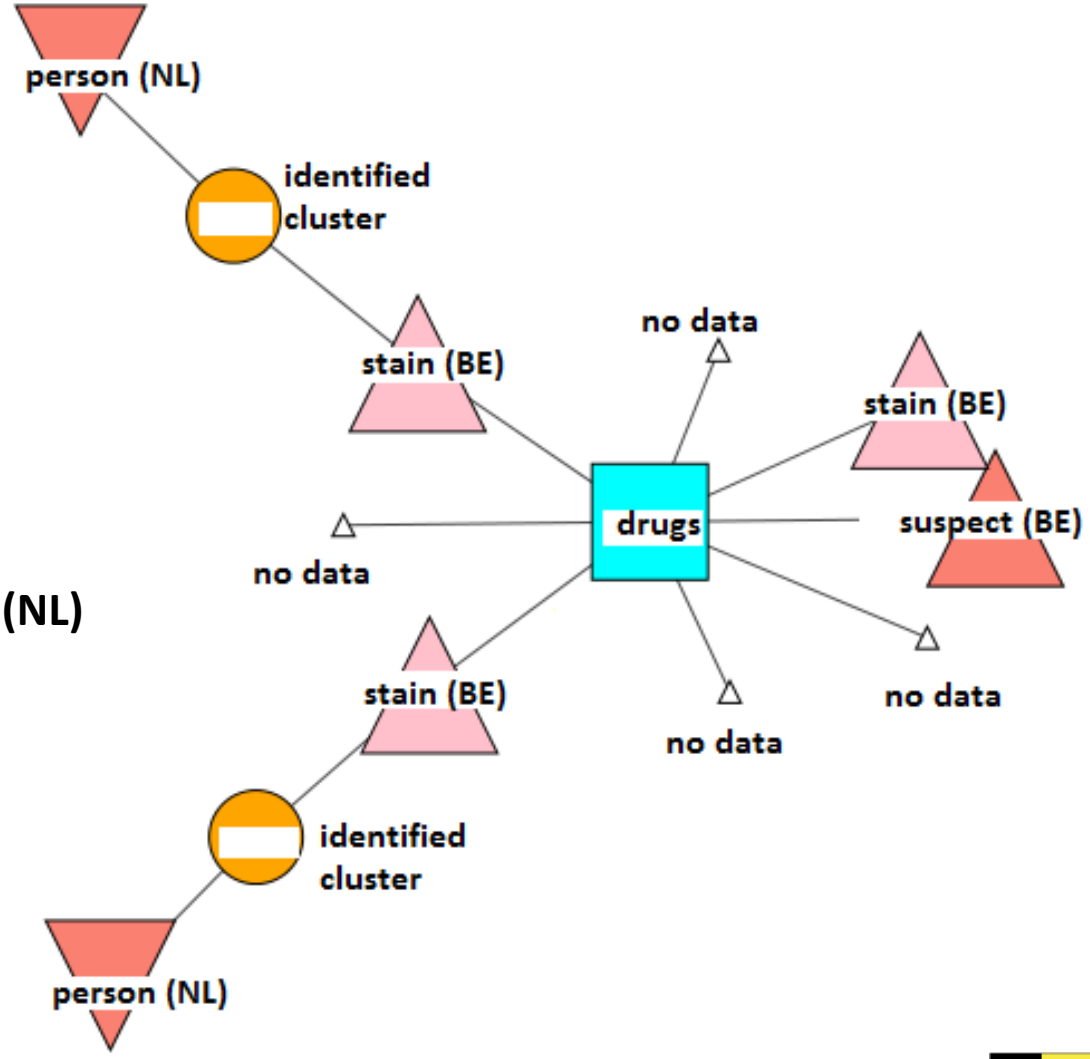
- 2 clusters (One unidentified, One identified)
- Only national data
  - Different districts
- Cases, Profiles, Clusters



Network
(with more than one cluster)

Example 4

- 2 clusters (identified)
- National data (BE) **+ Prüm data (NL)**
- Cases, Profiles, Clusters

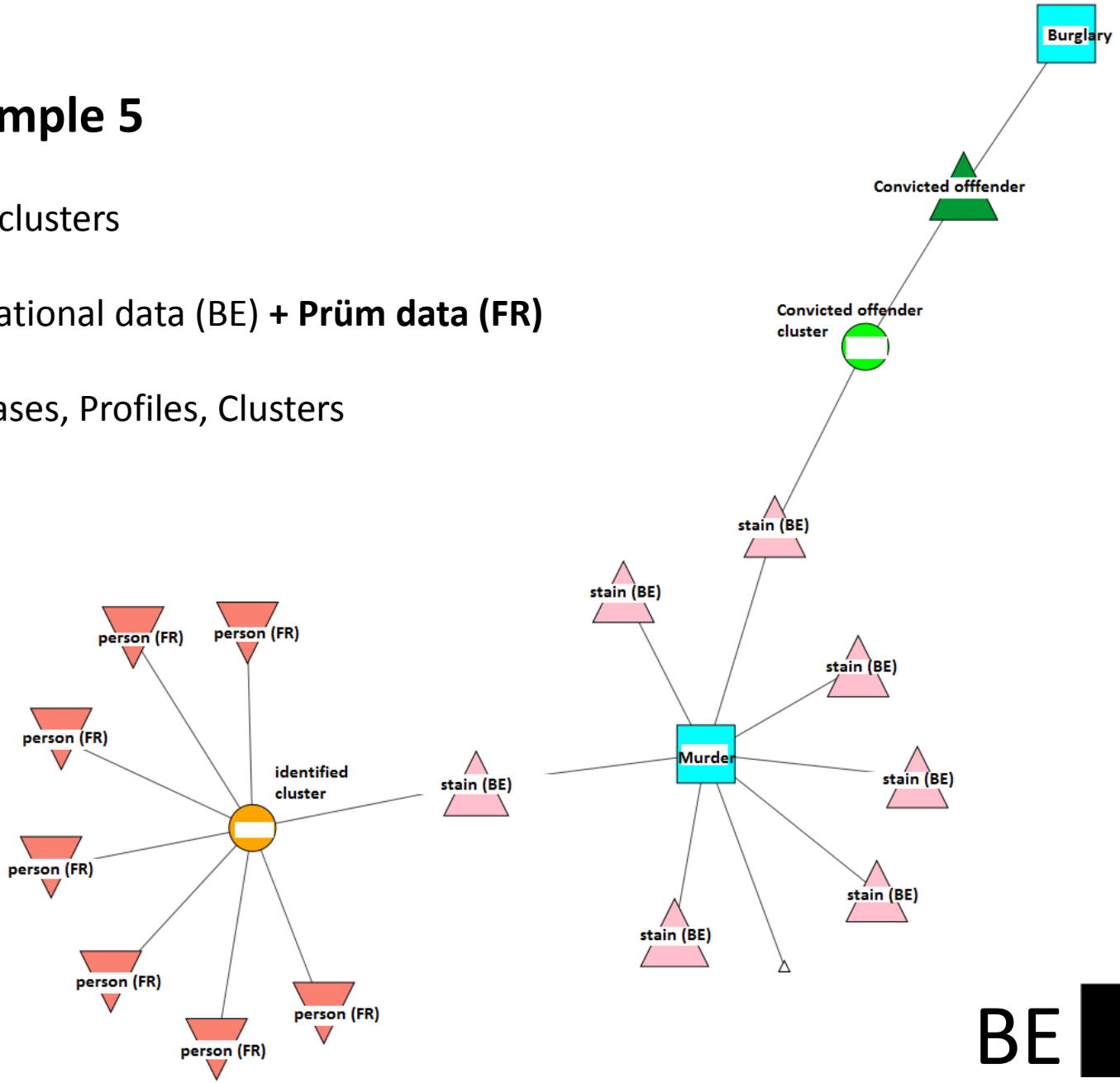


NL

BE

Example 5

- 2 clusters
- National data (BE) **+ Prüm data (FR)**
- Cases, Profiles, Clusters

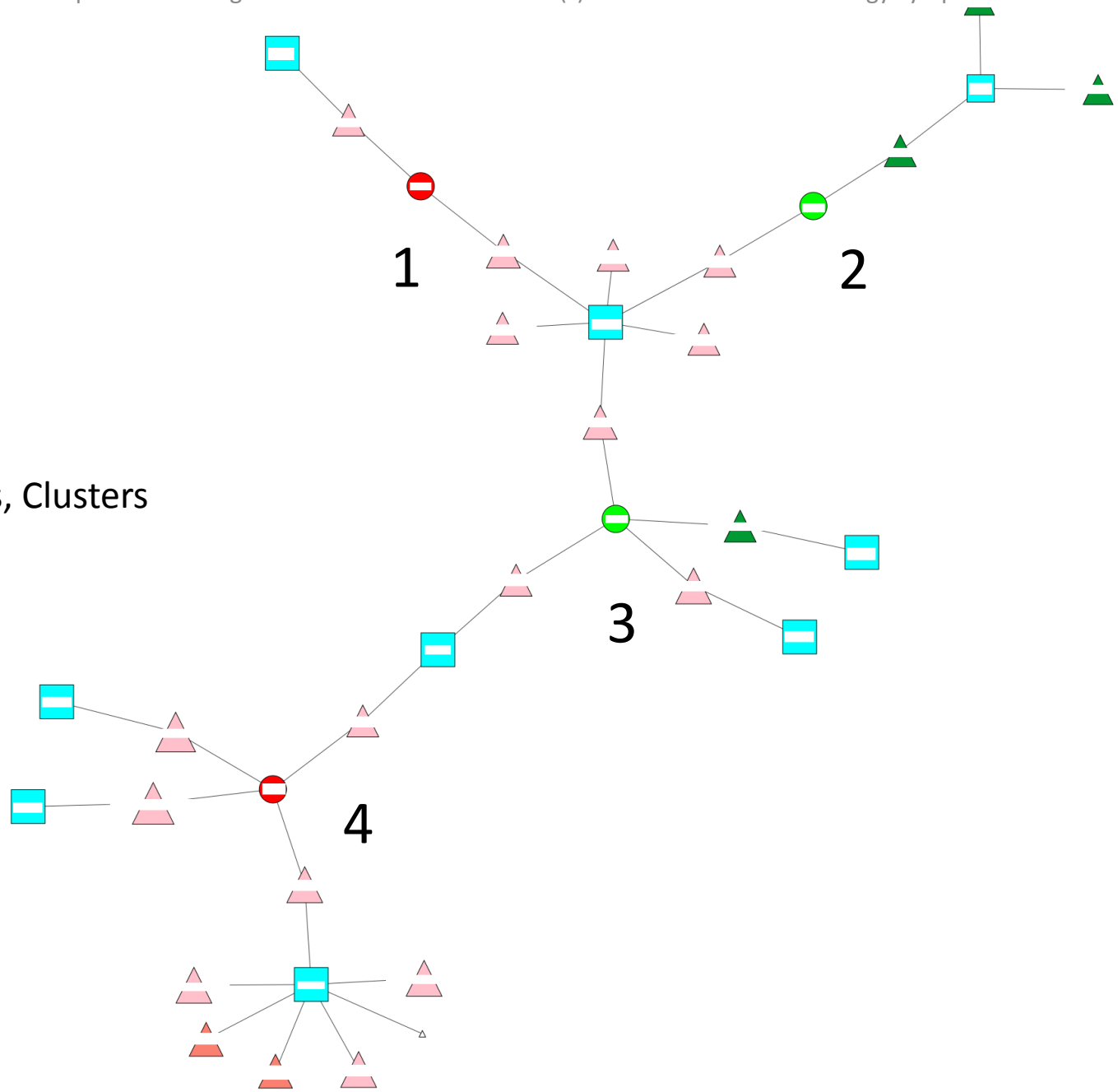


FR

BE

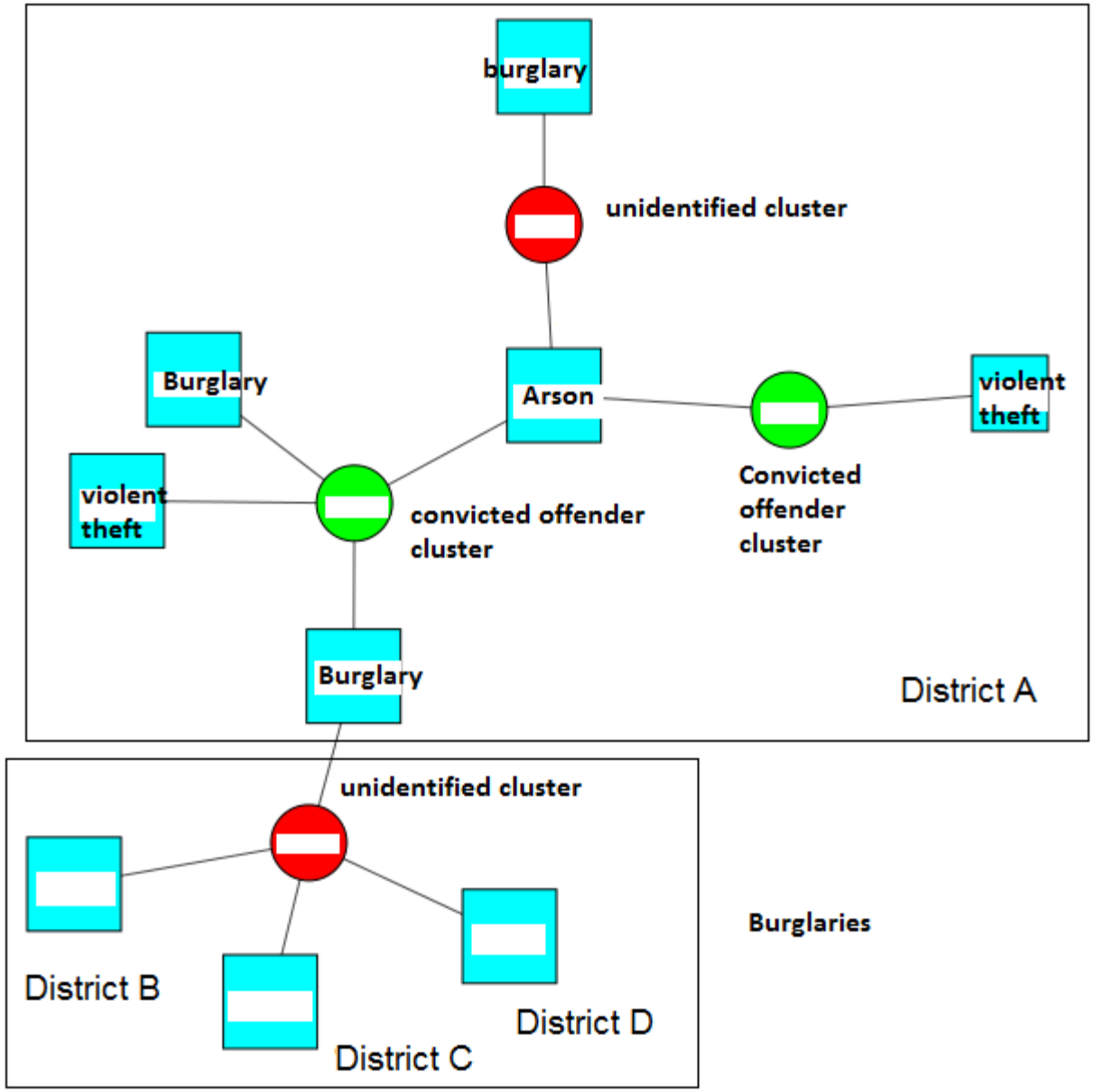
Example 6

- **4** clusters
- National data
- Cases, Profiles, Clusters



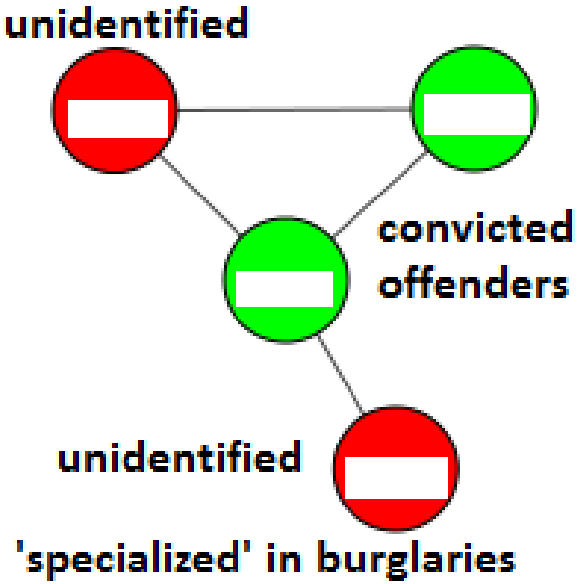
Example 6 (con't)

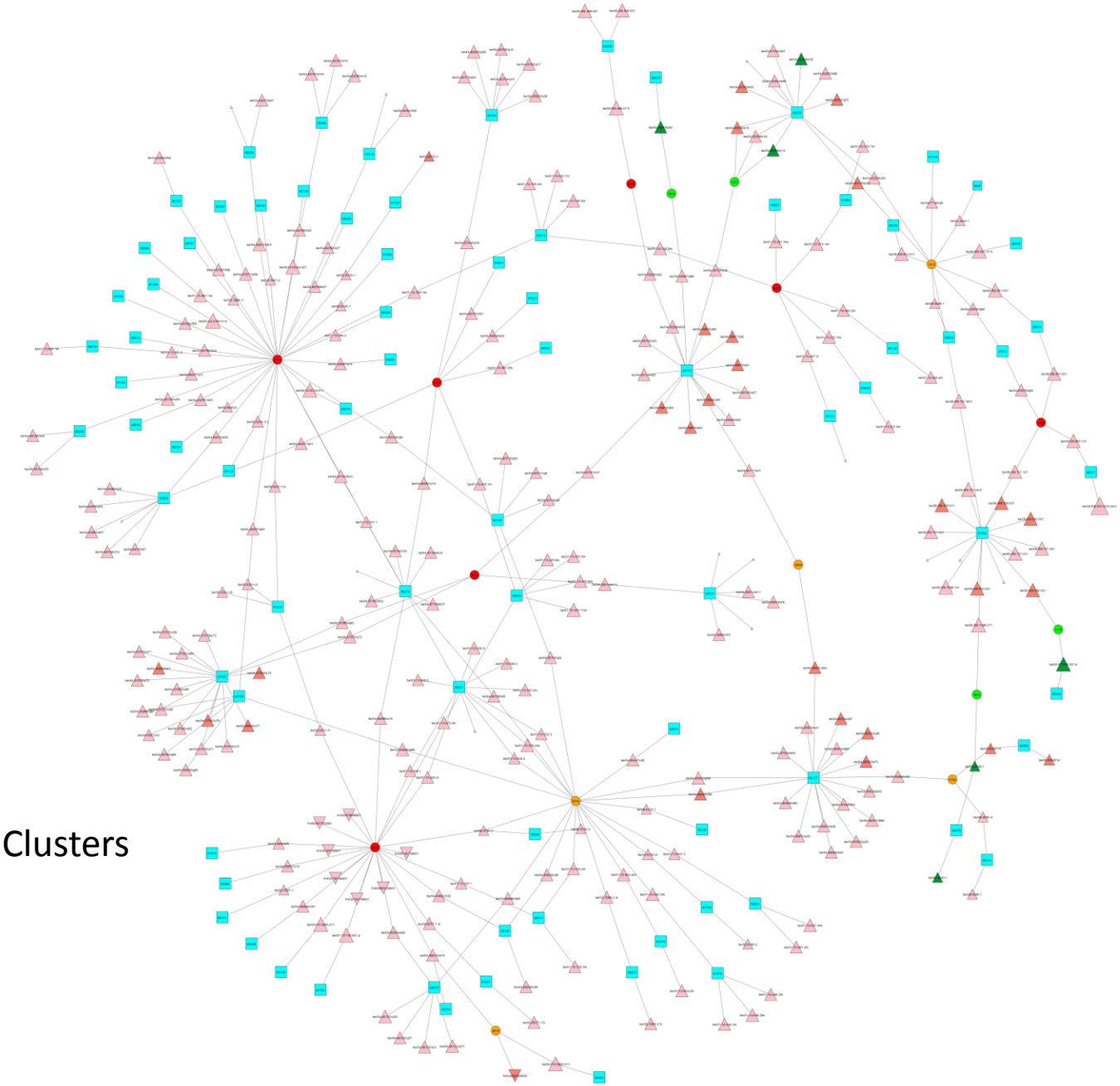
- 4 clusters
- National data
- Cases, Clusters
  - **(profiles removed)**



Example 6 (con't)

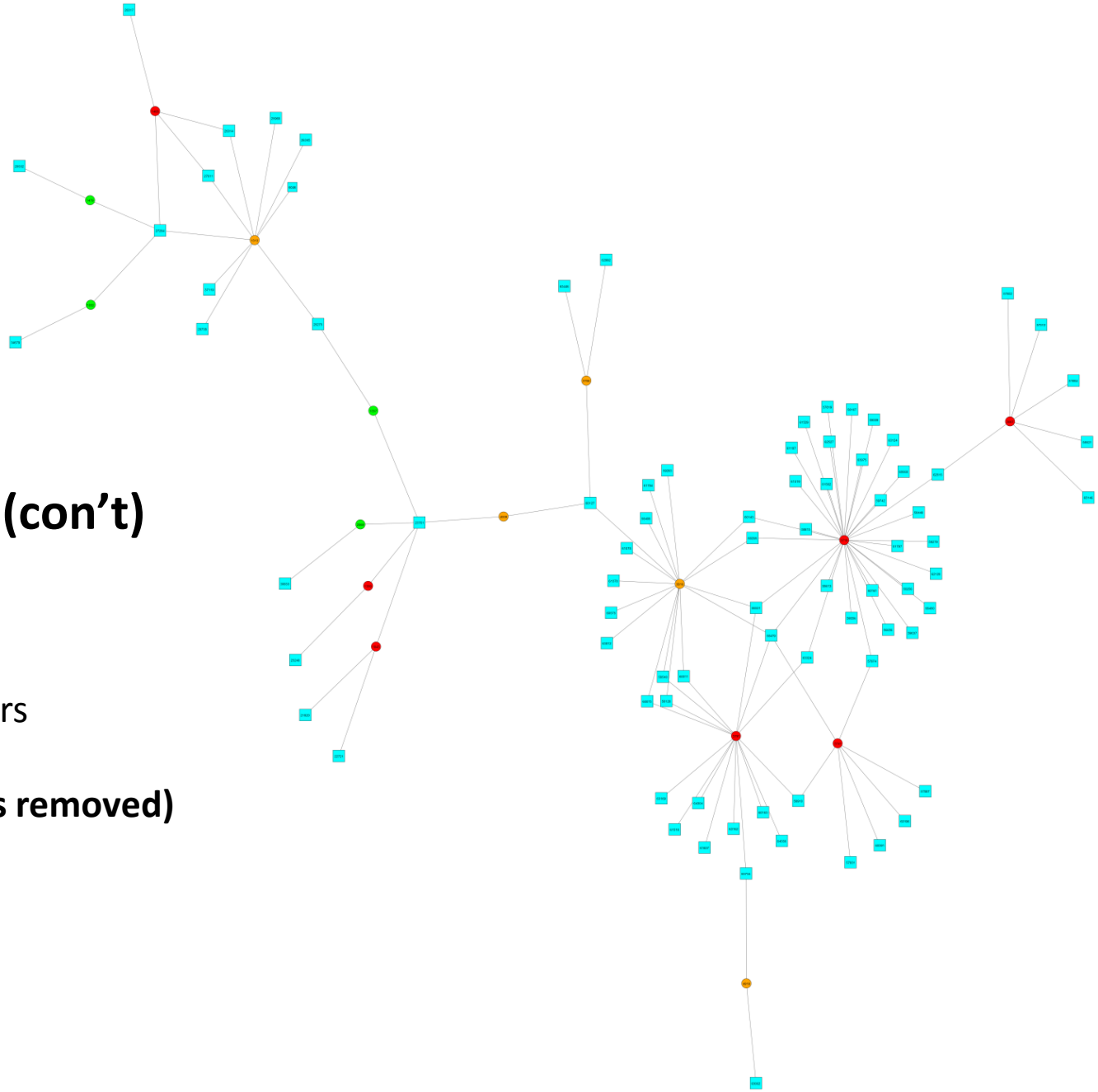
- 4 clusters
- National data
- Clusters **only**
  - (profiles removed)
    - **(cases removed)**





Example 7

- **16** clusters
- Cases, Profiles, Clusters

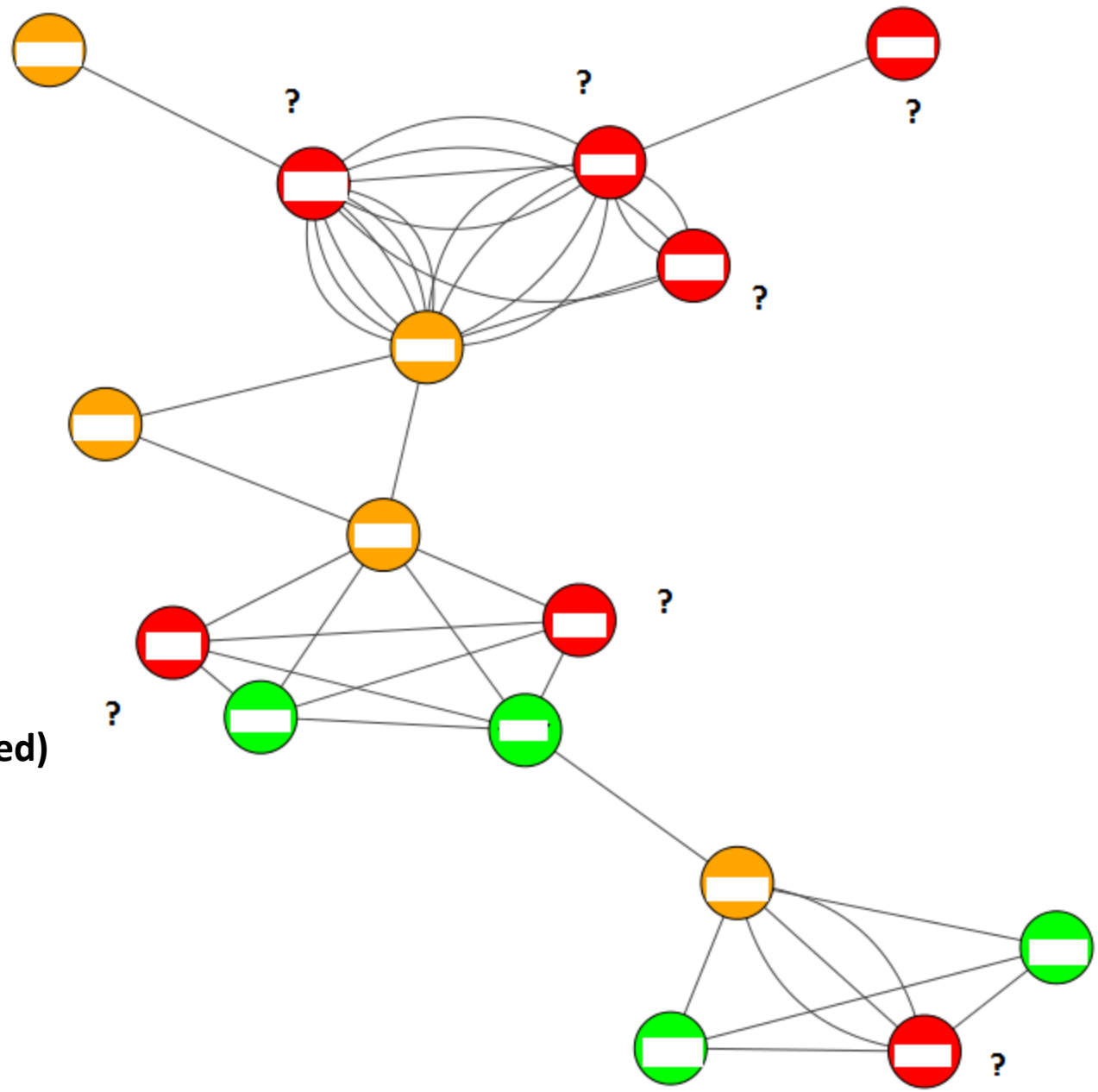


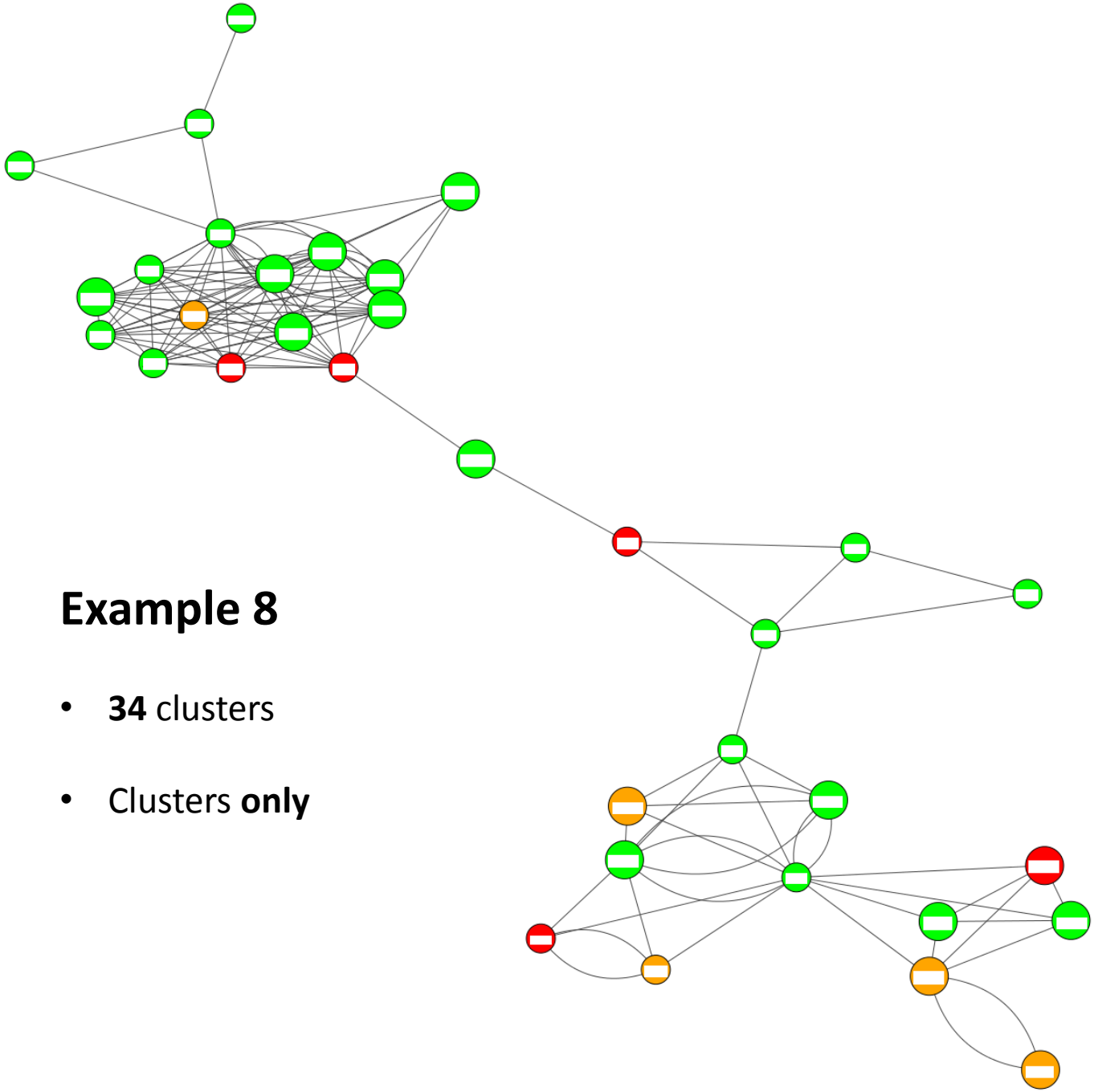
Example 7 (con't)

- **16** clusters
- Cases, Clusters
  - **(profiles removed)**

Example 7 (con't)

- **16** clusters
- Clusters **only**
  - (profiles removed)
    - **(cases removed)**



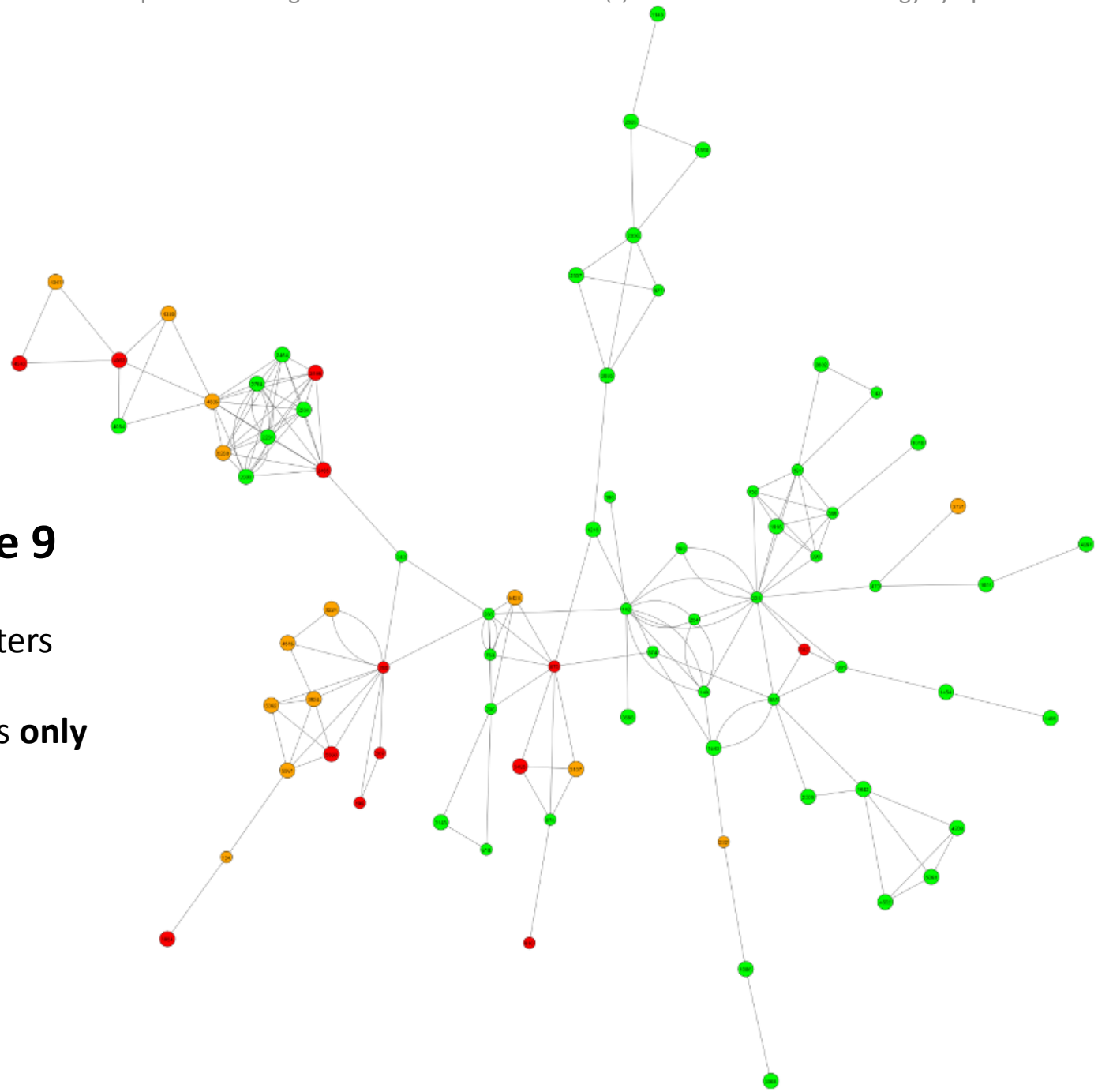


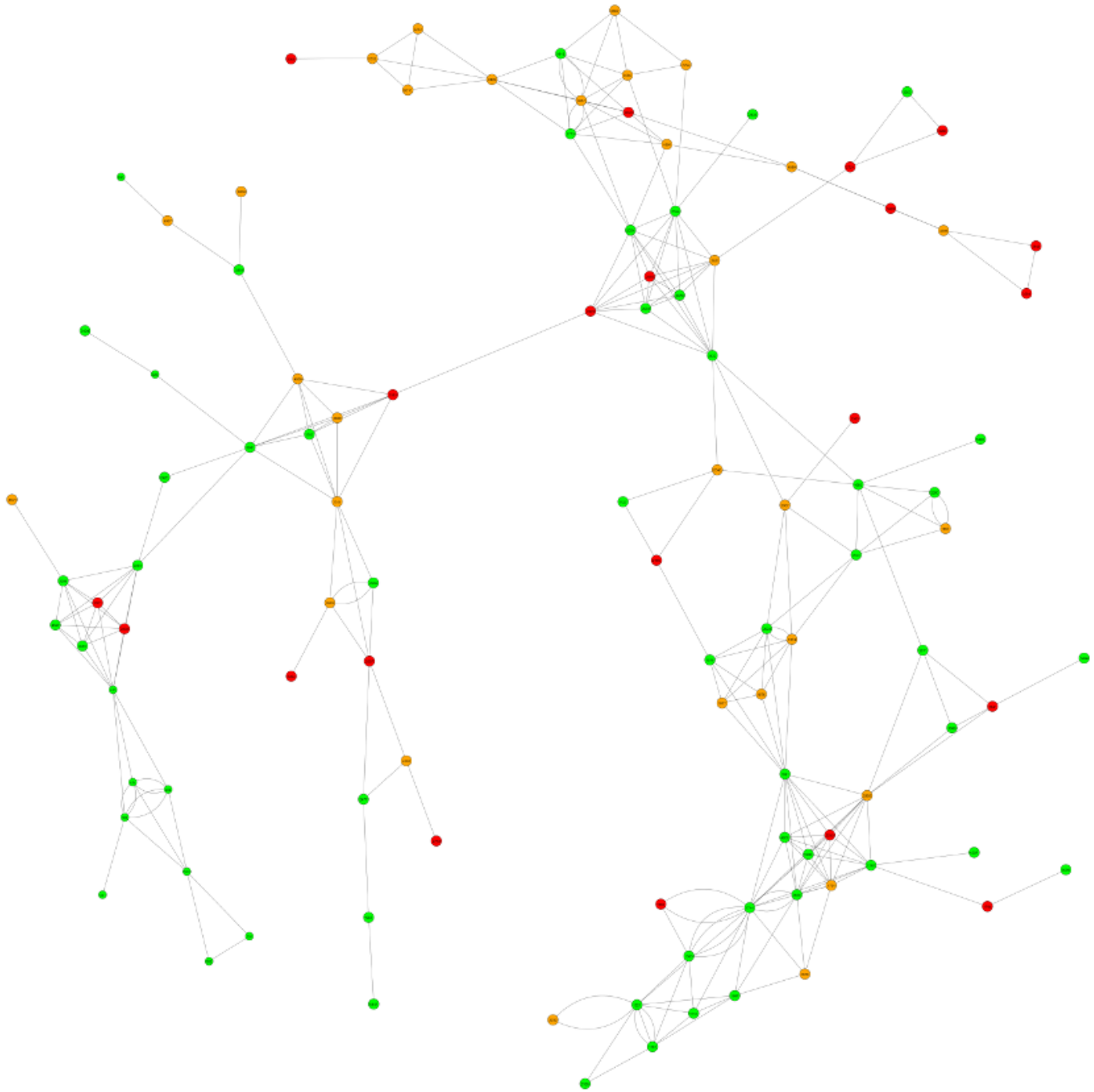
Example 8

- **34** clusters
- Clusters **only**

Example 9

- **79** clusters
- Clusters **only**

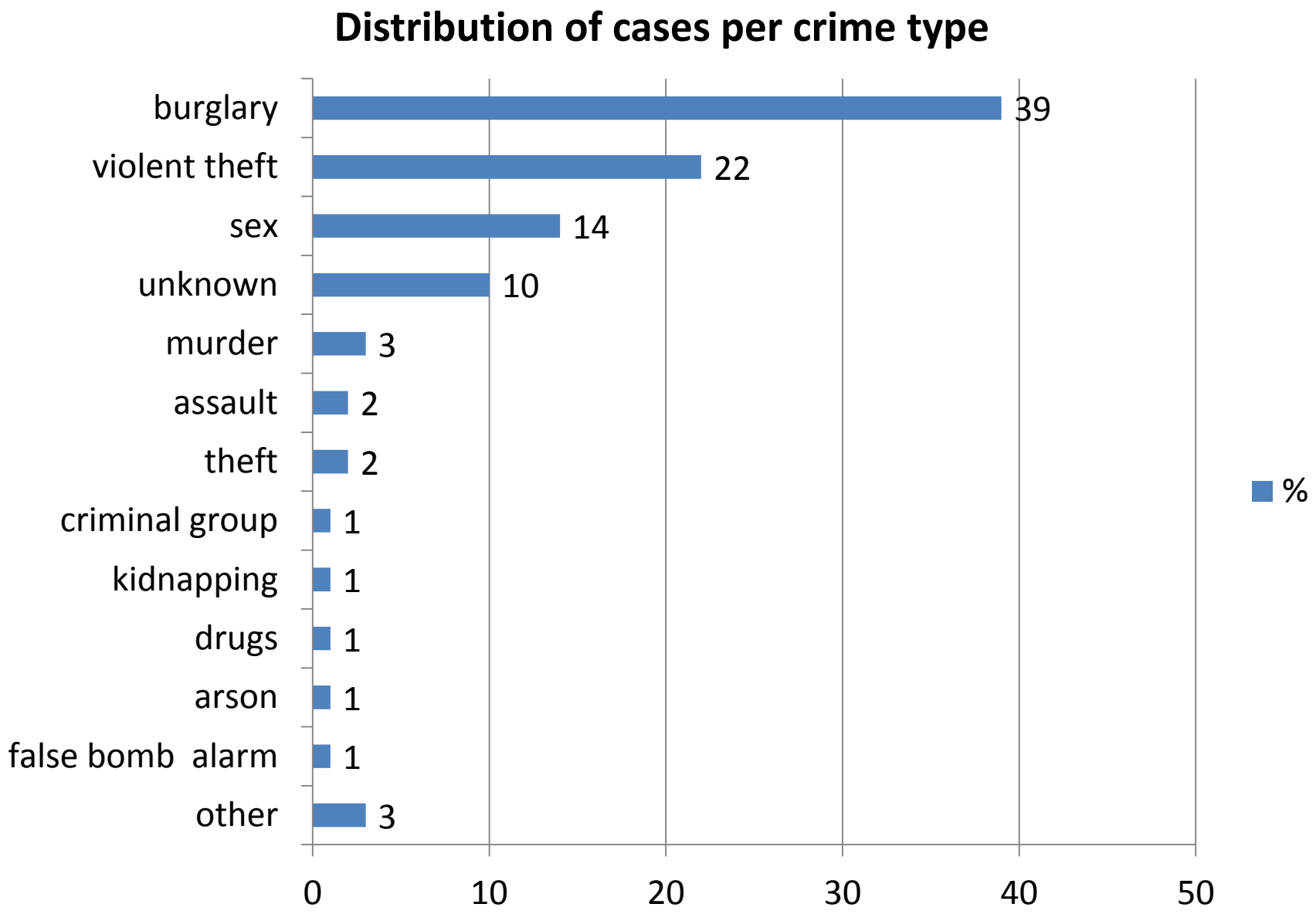




Example 10

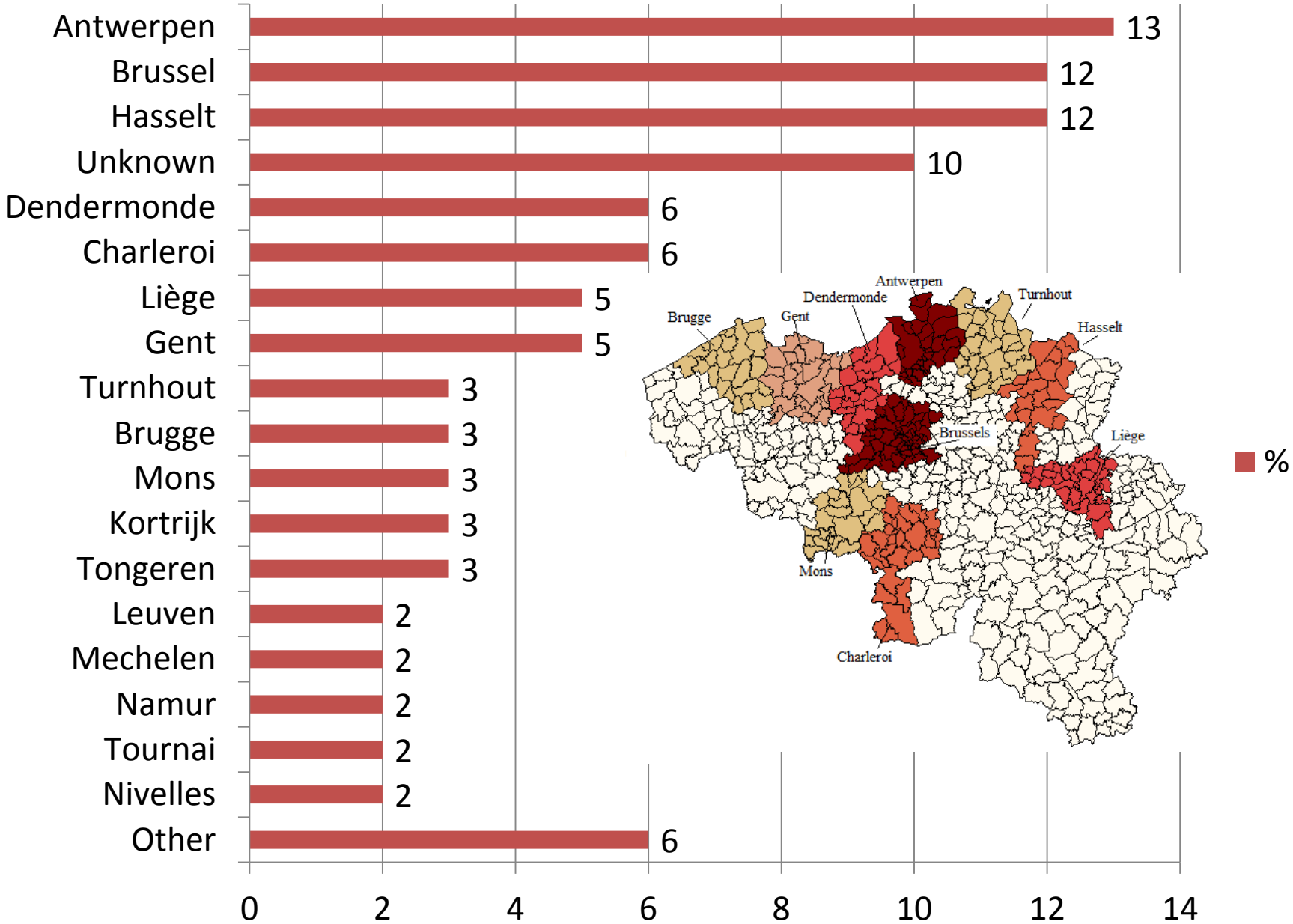
- **107** clusters
- Clusters **only**

A few observations…


Distribution of cases per crime type
burglary
violent theft
sex
unknown
murder
assault
theft
criminal group
kidnapping
drugs
arson
false bomb alarm
other
39
22
14
10
3
2
2
1
1
1
1
1
3
0
10
20
30
40
50
%

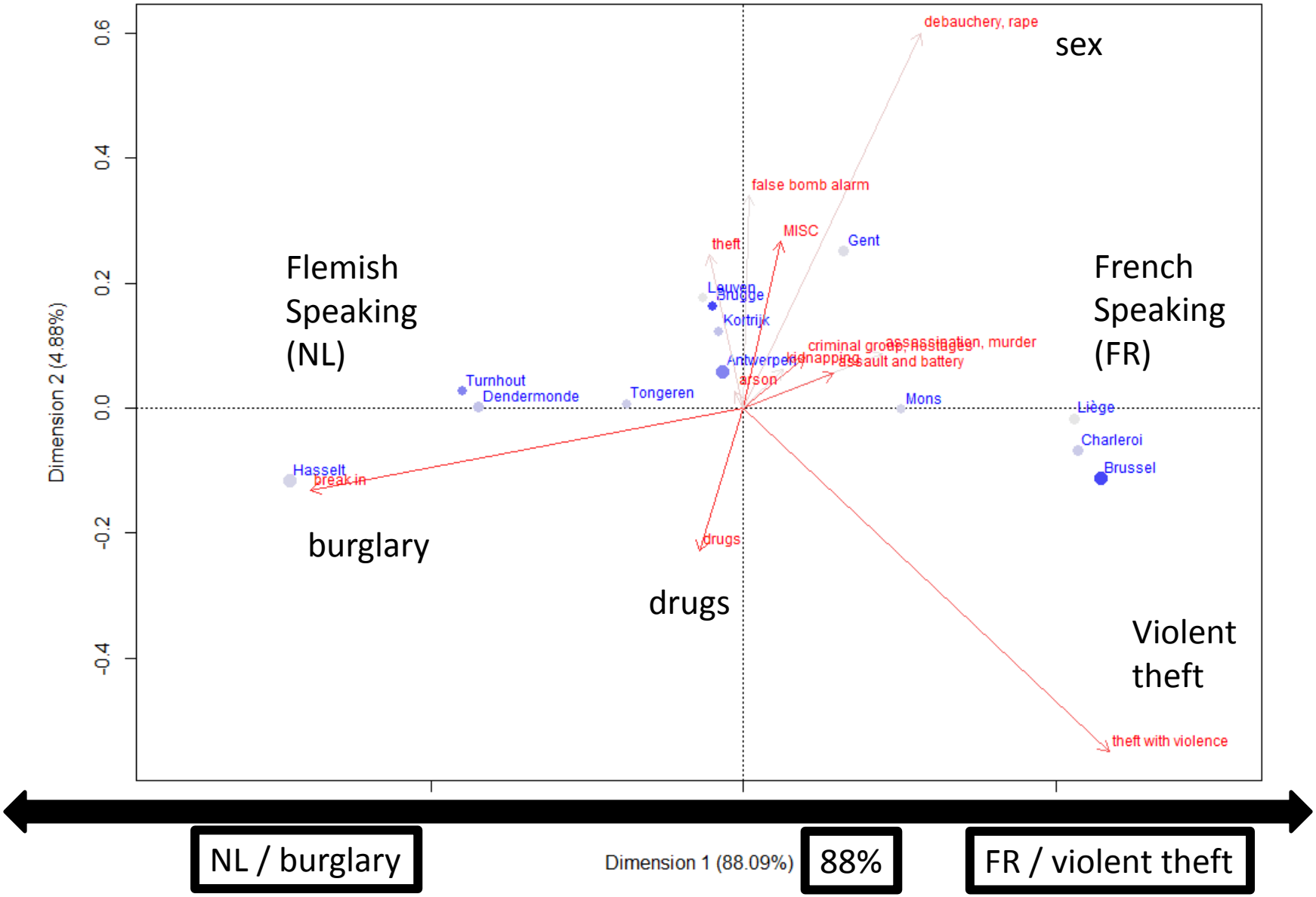
> 50 000 cases

Distribution of cases per district



Crimes – districts association

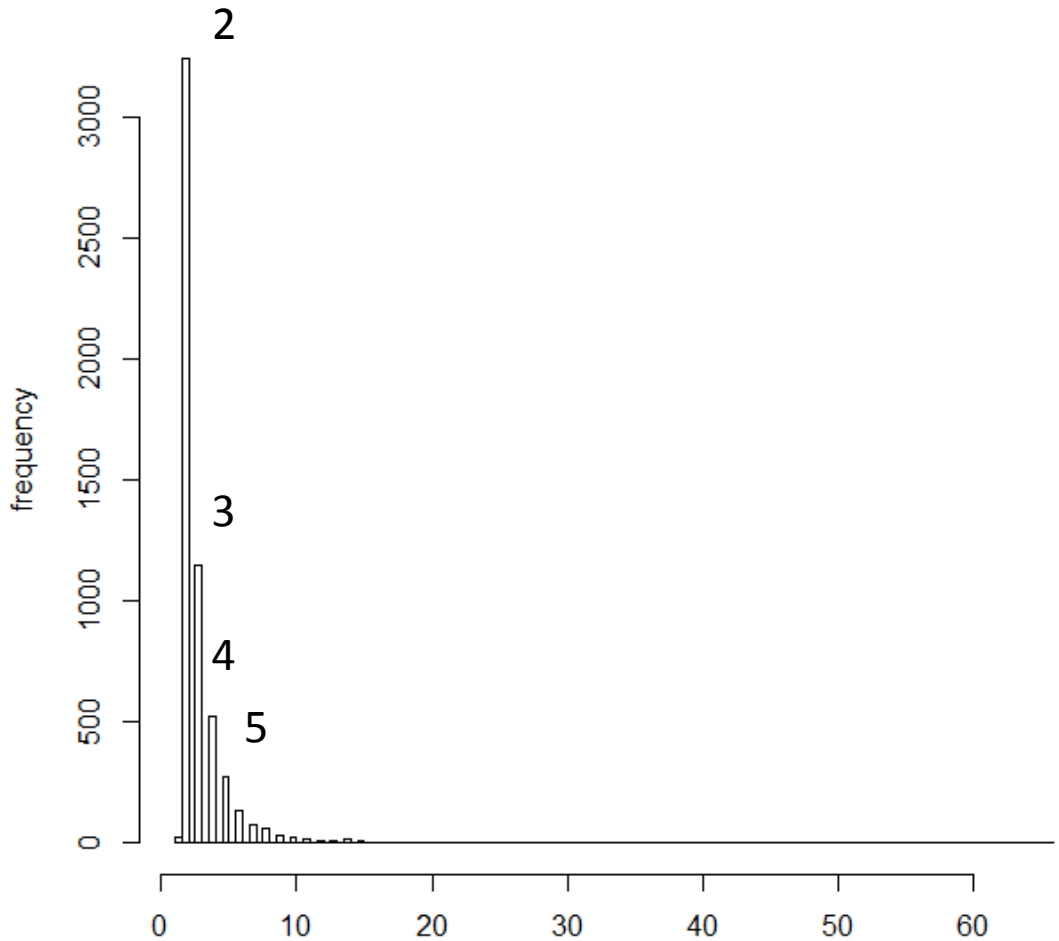
Visualization through correspondence analysis



Connectivity between cases through clusters

- > 50 000 cases
  - > 10 000 **cases** (**18%**) are connected through clusters.
- > 100 000 **samples**
  - > 15 000 (**15%**) in a cluster
- > 5000 clusters
  - 58% : **2 profiles**
  - 21% : 3 profiles
  - 9% : 4 profiles
  - 5% : 5 profiles
  - Max = 66 profiles

Frequency of cluster size



(Number of profiles in a cluster)

Degree of **identification** of clusters?

- > 5000 clusters (people)
  - **> 4000 (75** %) are **identified** (!)
    - 25% : stain only
    - 25% : at least one suspect (no convicted offender)
    - 50% : convicted offender

- Where does the identification come from?
  - **60**% : already identified thanks to **BE** data
    - > 3000 people
  - **15**% : identified thanks to **Prüm only**
    - > **700** people ! (50% FR, 50% NL)

Relationship between **identification** of clusters and **size** of clusters?

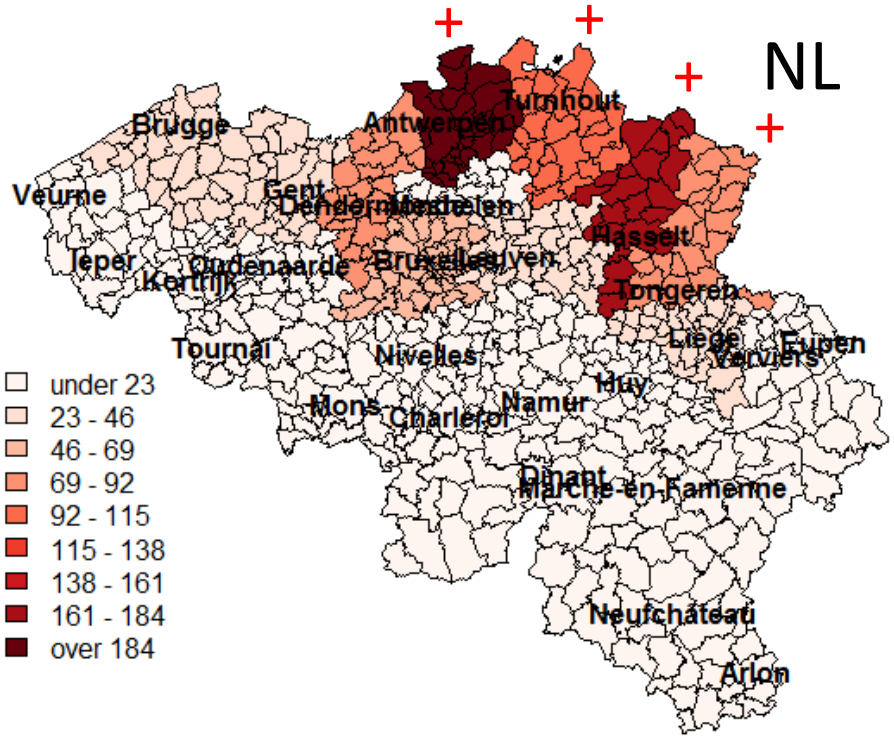
- Logit model
  - Identification ~ size of clusters

- For a 1 unit increase in size,
  - the log odds of being identified increases by 0.05
  - X^2 (1) = 7.3, $p < .01$

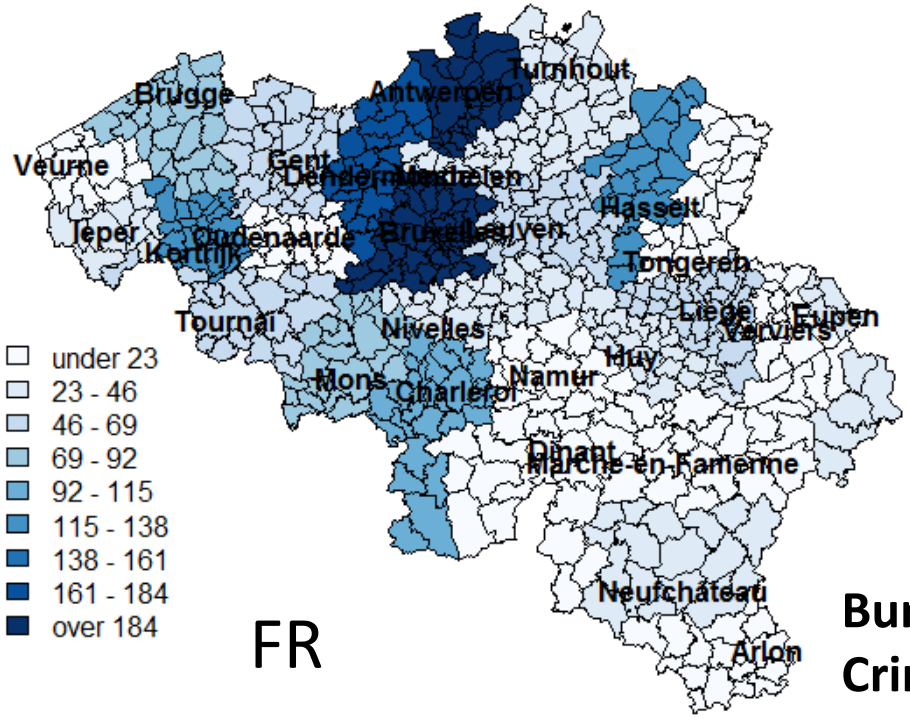
- In other words, the more profiles in a cluster, the higher the probability of being identified.

Transnational offending

- > 2000 Prüm profiles
- > 700 Prüm *clusters*



Burglary: 39 % => 62%
Drugs: 1% => 12%
Criminal group: 1% => 4%



Burglary: 39 % => 66%
Criminal group: 1% => 5%

Are **clusters** usually found in **networks**?

- Relationship between clusters
  - 57% : isolated cluster
    - (not in a network)
  - 43% : cluster in relation to other cluster(s)
    - (in a network)

Characteristics of **networks**?

- > 3000 components with a cluster
  - 83% components: isolated clusters
  - 17% components : > **600** networks of clusters

- Among the +600 networks
  - 59% are **couples** 
  - 22% are **trios** 
  - Largest network is made of **107** clusters.

- **Burglaries**: 39% => 51%
- **Criminal group**: 1% => 4%

Summary

- We wanted to improve our **knowledge** of the Belgian DNA database (old and new intelligence).
- We can produce **pictures** and **statistics** on transnational and 'organized' offending by exploiting the **links** in such a database.
- **Transnational** offending: Prüm is potentially helpful. Since 2014, already **+700** people identified thanks to it.
- '**organized**' crimes: **+600** networks that could be used for operational and strategic purposes.

Future activities

- **Further** analysis of components (clusters & networks).
  - Sequential & survival analysis.
  - Social network analysis.
- Analysis of **validity**
  - Cross/ comparison to other data – police data/ judiciary files.
- Analysis of **similar datasets**
  - Ex: Dutch DNA database.
- **Upcoming conference** on September 29th 2015
  - See next slide.

http: // NICC . fgov . BE / PIES - 2015

Source of funding

The writing and research for this presentation were supported by the **PIES project** and the **Be-Gen project**.

The PIES project "The Prüm Implementation, Evaluation, and Strengthening of Forensic DNA Data Exchange", project number HOME/2011/ISEC/AG/PRUM/4000002150 received financial support of the ISEC Programme "Prevention of and Fight against Crime" (**European Commission – Directorate-General Home Affairs**), grant agreement number 30-CE-0498536/00-03.

https://nicc.fgov.be/**PIES**

The Be-Gen project "Understanding the Operational, Strategic and Political Implications of the National Genetic Database" received financial support of the BRAIN-be Programme "Belgian Research Action through Interdisciplinary Networks" (**Belgian Science Policy Office**), contract number BR/132/A4/Be-Gen.

http://nicc.fgov.be/**Be-Gen**

*The sole **responsibility** lies with the authors.*
The European Commission and the Belgian Science Policy Office
are not responsible for any use that may be made of the information contained in this paper.