



# Building an Urban Theft Map by Analyzing Newspaper Crime Reports

---

LAURA PO AND FEDERICA ROLLO

UNIVERSITY OF MODENA AND REGGIO EMILIA

ITALY

Are you worry about thefts?

---

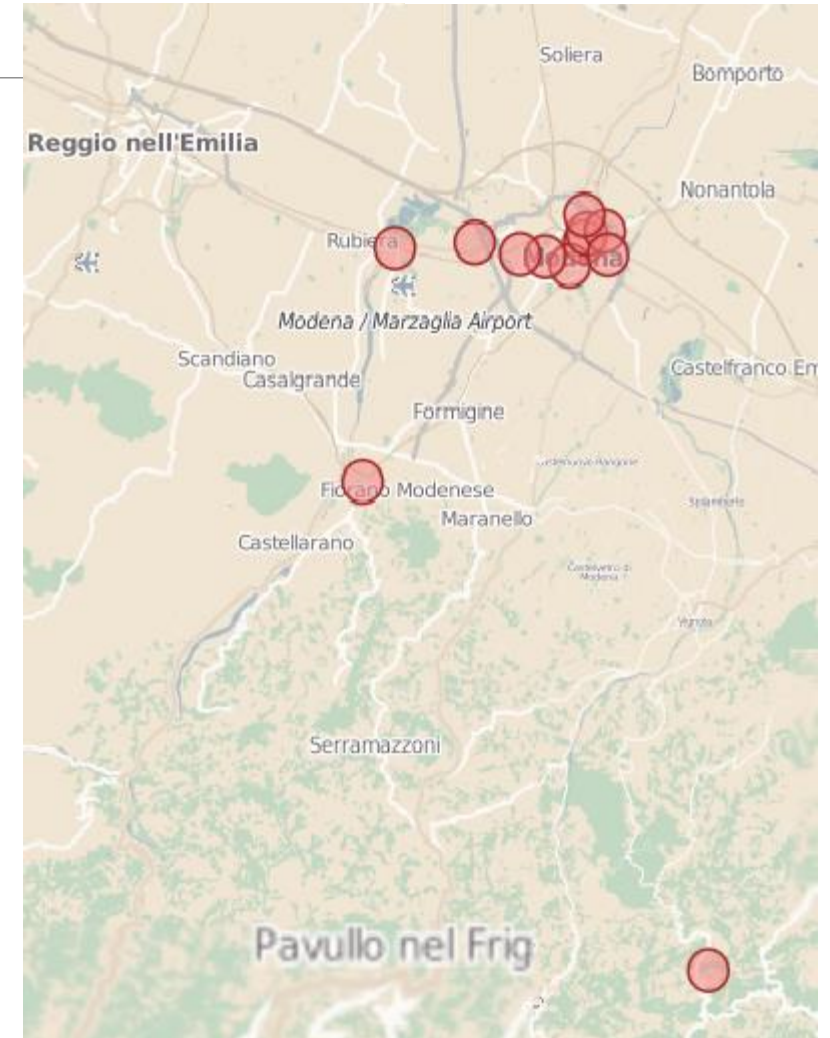


# Crime Mapping

**Crime analysis** is a set of systematic, analytical processes directed at providing timely and pertinent information relative to crime patterns and trend correlations in order to assist the police in crime reduction, prevention, and evaluation.

**Crime mapping** is the process of using a geographic information system (GIS) to conduct spatial analysis of crime problems.

Mapping crime allows **police** analysts to identify crime hot spots, moreover it increases public confidence and **citizen** engagement and promotes transparency.



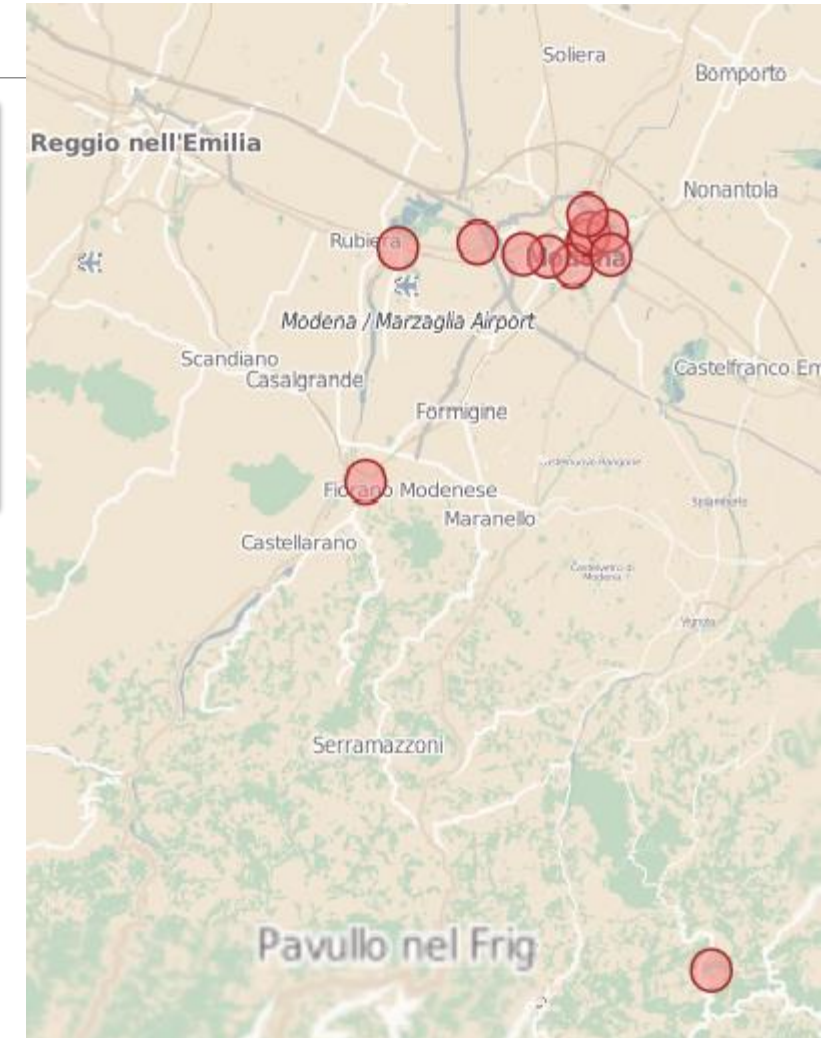
# Crime Mapping

Crime analysis is an analytical procedure and pertinent patterns and trends to help the police in crime evaluation.

This research is focused on analyzing and mapping thefts by crawling data from local newspapers using text mining techniques for a medium Italian city, Modena.

Crime mapping [2] is the process of using a geographic information system (GIS) to conduct spatial analysis of crime problems.

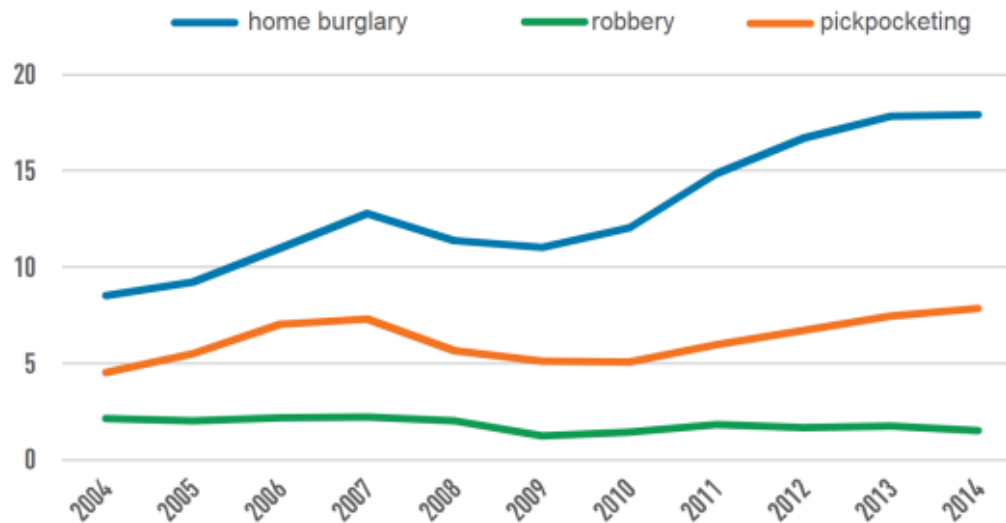
Mapping crime allows police analysts to identify crime hot spots, moreover it increases public confidence and citizen engagement and promotes transparency.



# Theft in Italy

The National Institute of Statistics, ISTAT, studies the crime rates across the country.

Regarding the thefts and robberies reported, Italy ranks fifth in the EU ranking after Denmark, Belgium, The Netherlands, and Sweden.



ITALY  
**4.105,3**  
number of theft reports every 100 thousand inhabitants

- 1.000 - 2.000
- 2.000 - 3.000
- 3.000 - 4.000
- 4.000 - 5.000
- 5.000 - 6.000
- 6.000 - 7.000
- > 7.000



Within the Italian country, the data from CENSIS shows that in 2016 the reported crimes in Italy were 2,5 millions, the 8.2% less than in 2008.

# Theft reports as Open Data

---

USA, Canada, Denmark, Finland, France, Germany, and UK countries published government crime open datasets as listed by the *Open Data Barometer* in its 4th report.

Crime data is most often available in the form of **thematic reports**: the data contained in most of them are **not in open or machine-readable formats**.

Italian crime information becomes available only after the ISTAT's collection, and analysis, this means only after some months/years, and in an aggregate manner.

Few Italian cities provide open, and updated crime datasets: Torino, Trento.



**OPEN  
DATA**

# Theft reports from Newspapers

---

**Newspapers** are a source of (mostly) authentic and timely information and many of them make an electronic version available online. There is a large amount of information such as crimes or accidents.

Is crucial to take into consideration that **newspapers are looking for crime stories** that have some **news value** and that are of interest to the community. Therefore only a few of the police reports will be transformed in news.

In Italy, newspapers have been analyzed for crime mapping in Milan, Livorno, Venice.



# Modena

---

Modena is a medium city of 185,000 inhabitants and a province that counts 700,000 citizens on the south side of the Po Valley, in the Emilia-Romagna region of northern Italy.





# Crime Analysis on Modena Newspaper

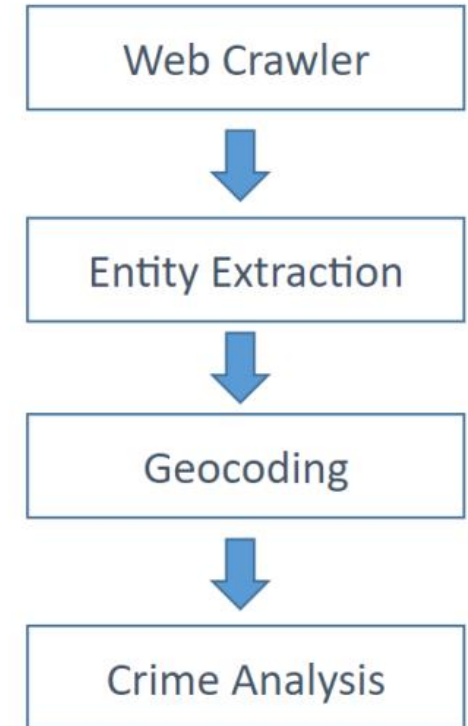
---

Newspaper articles are mined using a *Java Web Crawler: jsoup*.

The application parses the content of each page, detects the useful information and stores them in a database.

The main information stored are:

- the URL of the news,
- the title of the news,
- the description,
- the address where the theft took place (municipality, area, and street),
- the date and the time of publication of the news,
- the stolen object (this item is extracted from the description and the title by using a text analysis tool).



# Web Crawler

MODENATODAY Sezioni **Furti**



## Mirandola, furto di rame nel cantiere del Centro Nuoto

Spariti nel fine settimana ben 300 metri di cavi dall'impianto fotovoltaico

lunedì, 13 agosto



## Piscine di Carpi, fermato un ladro di biciclette

Denunciato per furto durante un controllo mirato della Municipale per contrastare un fenomeno fin troppo frequente

sabato, 11 agosto



## Carpi, un mendicante fa arrestare due ladri dopo un colpo al supermercato

E' stato uno straniero che chiedeva l'elemosina all'uscita del negozio a fornire l'identikit di due giovani marocchini che poco prima avevano rubato il portafogli di un cliente. Droga a casa di uno dei due

venerdì, 10 agosto



## Ruba gli zaini da auto e pulmino dei turisti davanti al Mef, arrestato

Un 24enne già noto per numerosi precedenti è stato bloccato dalla Volante grazie alla segnalazione di un cittadino attento. Aveva preso di mira i mezzi dei visitatori del Museo in via Ferrari

<http://www.modenatoday.it/tag/furti/>

**Title:** Mirandola, furto di rame nel cantiere del Centro Nuoto Spariti nel fine settimana ben 300 metri di cavi dall'impianto fotovoltaico

**URL:** <http://www.modenatoday.it/cronaca/furto-rame-piscina-mirandola-13-agosto-2018.html>

**Date:** 13 August 2018 12:34

**Description:** Approfittando del fine settimana e delal conseguente sospensione dei lavori, alcuni ladri sono riusciti ad introdursi nel cantiere del Centro Nuoto di Mirandola, che da mesi è soggetto a profondi lavori di rinnovo. L'obbiettivo dei malviventi era ben preciso: i cavi di rame degli impianti. I criminali sono infatti riusciti a trafugare bel 300 metri di fili elettrici, "spolpando" il sistema fotovoltaico dell'impianto. Sul caso indagano i Carabinieri. [...]

## Web Crawler

MODENATODAY Sezioni Furti



### Mirandola, furto di rame nel cantiere del Centro Nuoto

Spariti nel fine settimana ben 300 metri di cavi dall'impianto fotovoltaico

lunedì, 13 agosto



### Piscine di Carpi, fermato un ladro di biciclette

Denunciato per furto durante un controllo mirato della Municipale per contrastare un fenomeno fin troppo frequente

sabato, 11 agosto



### Carpi, un mendicante fa arrestare due ladri dopo un colpo al supermercato

E' stato uno straniero che chiedeva l'elemosina all'uscita del negozio a fornire l'identikit di due giovani marocchini che poco prima avevano rubato il portafogli di un cliente. Droga a casa di uno dei due

venerdì, 10 agosto



### Ruba gli zaini da auto e pulmino dei turisti davanti al Mef, arrestato

Un 24enne già noto per numerosi precedenti è stato bloccato dalla Volante grazie alla segnalazione di un cittadino attento. Aveva preso di mira i mezzi dei visitatori del Museo in via Ferrari

<http://www.modenatoday.it/tag/furti/>

## Entity Extraction

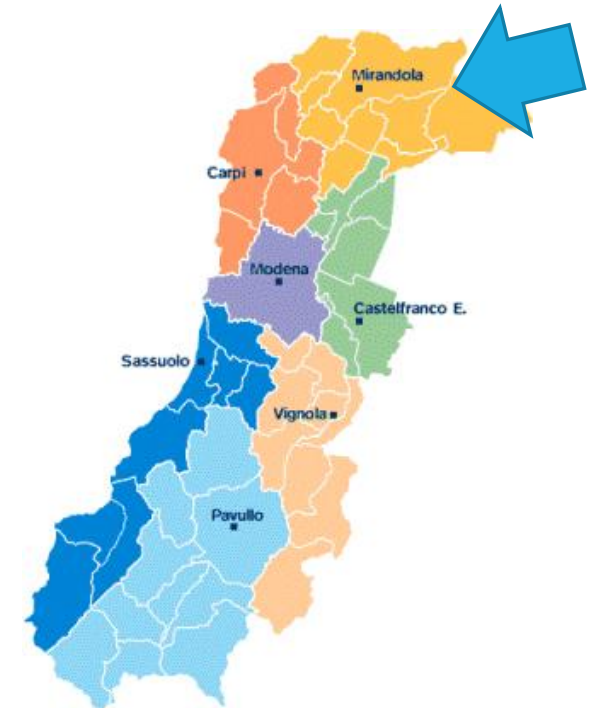
**Title:** **Mirandola**, furto di rame nel cantiere del Centro Nuoto Spariti nel fine settimana ben 300 metri di cavi dall'impianto fotovoltaico

**URL:** <http://www.modenatoday.it/cronaca/furto-rame-piscina-mirandola-13-agosto-2018.html>

**Date:** 13 August 2018 12:34

**Description:** Approfittando del fine settimana e del conseguente sospensione dei lavori, alcuni ladri sono riusciti ad introdursi nel cantiere del Centro Nuoto di Mirandola, che da mesi è soggetto a profondi lavori di rinnovo. L'obiettivo dei malviventi era ben preciso: i cavi di rame degli impianti. I criminali sono infatti riusciti a **trafugare** ben **300 metri di fili elettrici**, "spolpando" il sistema fotovoltaico dell'impianto. Sul caso indagano i Carabinieri. [...]

## Geocoding



Modena  
Thefts DB

# Some data

---

The database is updated with news until August 31<sup>st</sup>, 2018

The total theft reports extracted from May 2011 to August 2018 are **1760**.

- **100%** have the **date**
- **78%** have the **geolocation**
- **83%** have an **identified stolen object** - However, the reliability of the stolen object identification method is not proved

# Issues with Geolocation

---

Sometimes in the content of the news, the address is not present or there is an address for which the Google API can not find latitude and longitude.

- Some simplifications/assumptions:
  - in case, there is only the name of the street, we assume the city is Modena
  - in case, there is only the name of the municipality we use the city center
- However,
  - we still have 174 cases with neither the city nor the address
  - and 221 cases with city and municipality that cannot be solve by the Google API

# Issues with Stolen Object Identification

---

The extraction/identification of the Stolen Object is a syntactical extraction process:

- delete the stopwords from the text (title + description)
- identify relevant words that follow expressions like 'theft', 'steal', 'stolen', ...

“the two thefts have stolen a bike to a boy in the city center“



“two thefts have stolen bike boy city center“



'bike'

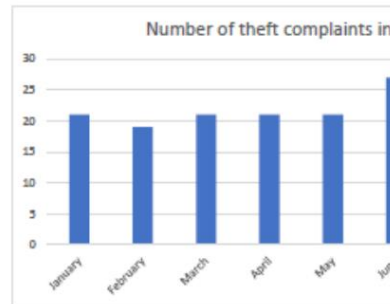
# Crime Analysis

- Crime Map analysis - locating on a map the crimes happened in a specific month;
- Density Crime Map visualization - displaying of a heat map that localize critical area for a specific year;
- Hot spot detection - identifying the number of crimes per district within the city or cities within the province;
- Crime comparison - comparing the number of crimes to get an idea about the growth or decrease of thefts over the time.



## Modena Theft Map

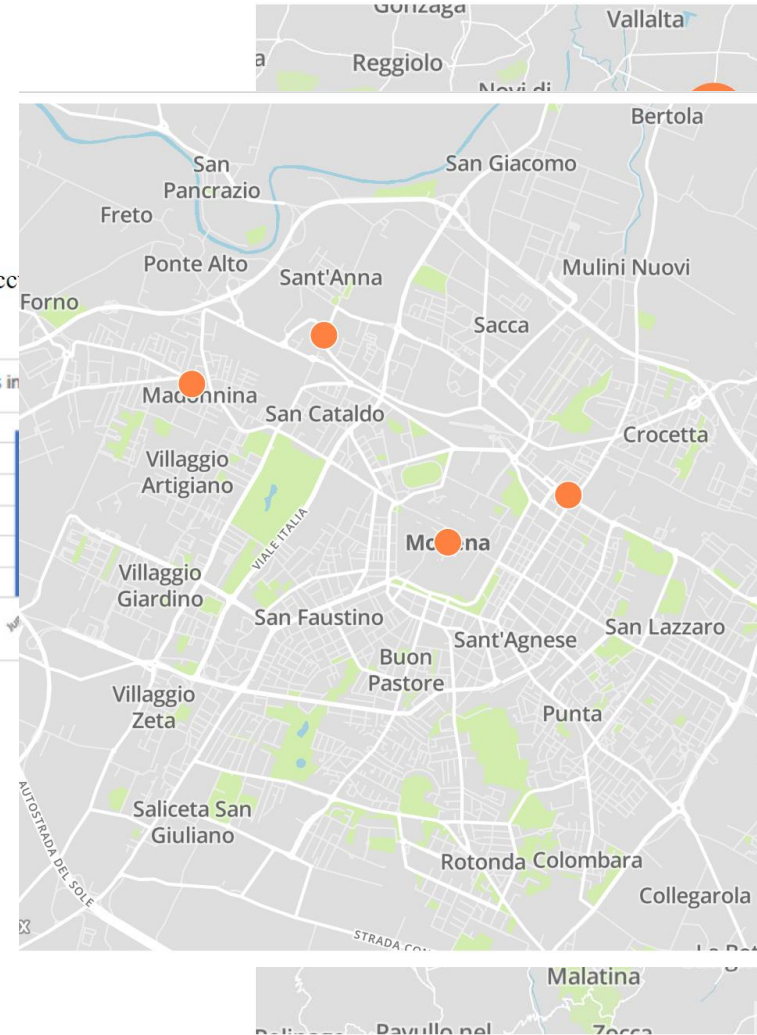
The map on the right shows the thefts occurred



Select another interval time (from May 2011 to August 2018).

August ▼ 2018 ▼

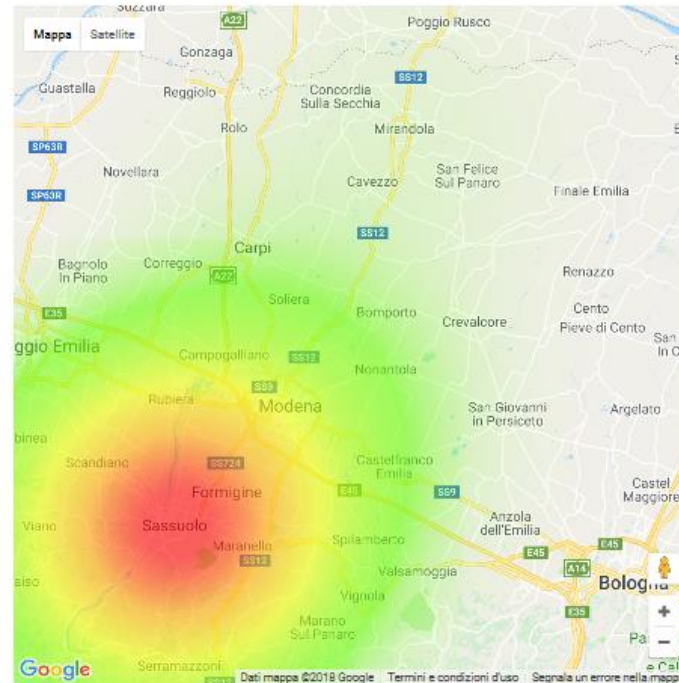
Show



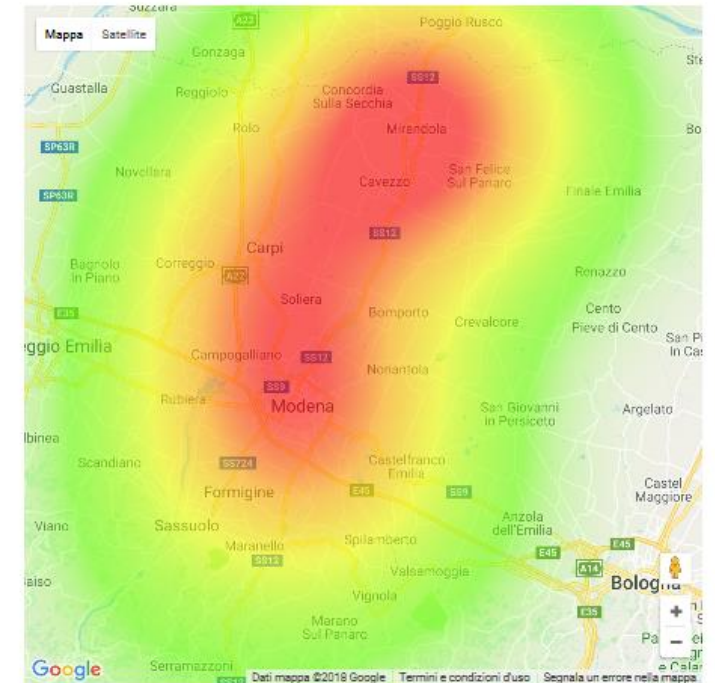
# Crime Analysis

- Crime Map analysis - locating on a map the crimes happened in a specific month;
- Density Crime Map visualization - displaying of a heat map that localize critical area for a specific year;
- Hot spot detection - identifying the number of crimes per district within the city or cities within the province;
- Crime comparison - comparing the number of crimes to get an idea about the growth or decrease of thefts over the time.

Theft complaints in 2015



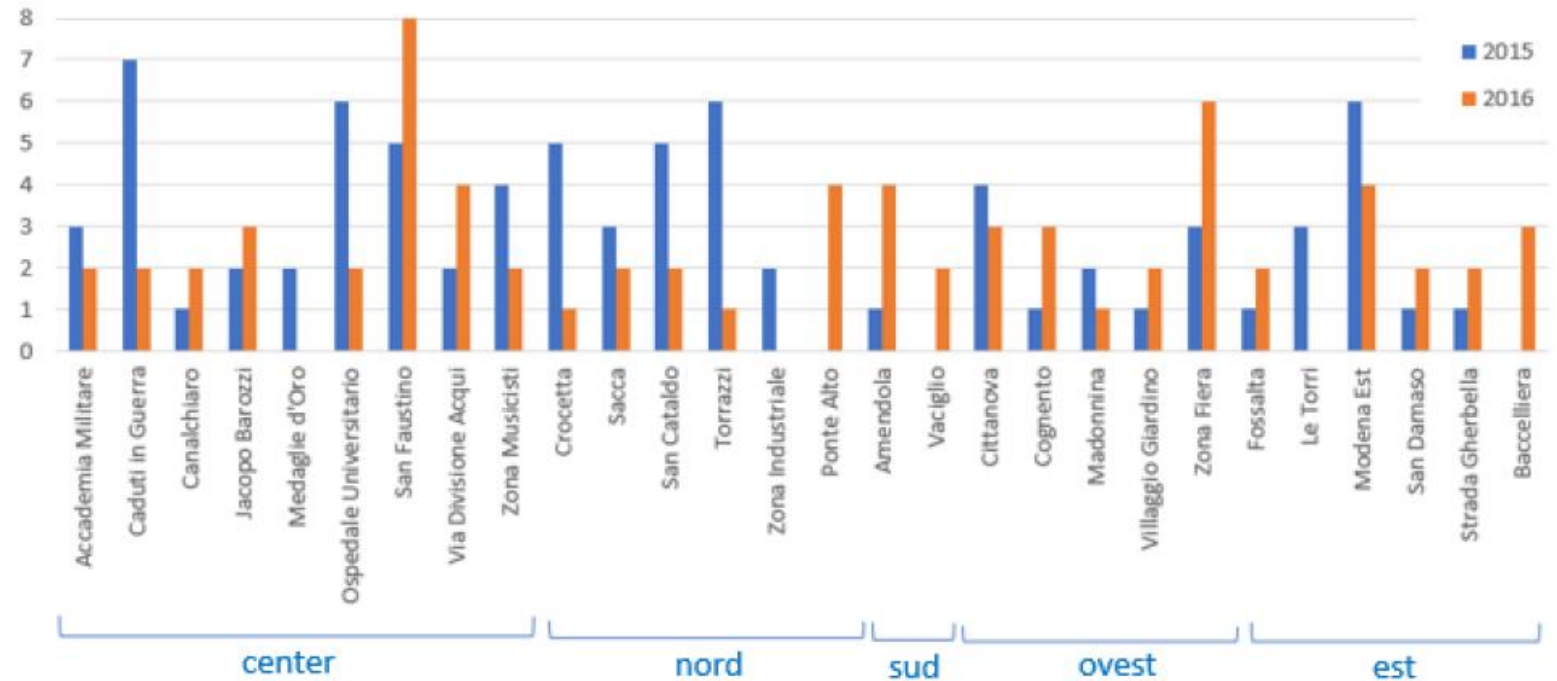
Theft complaints in 2016





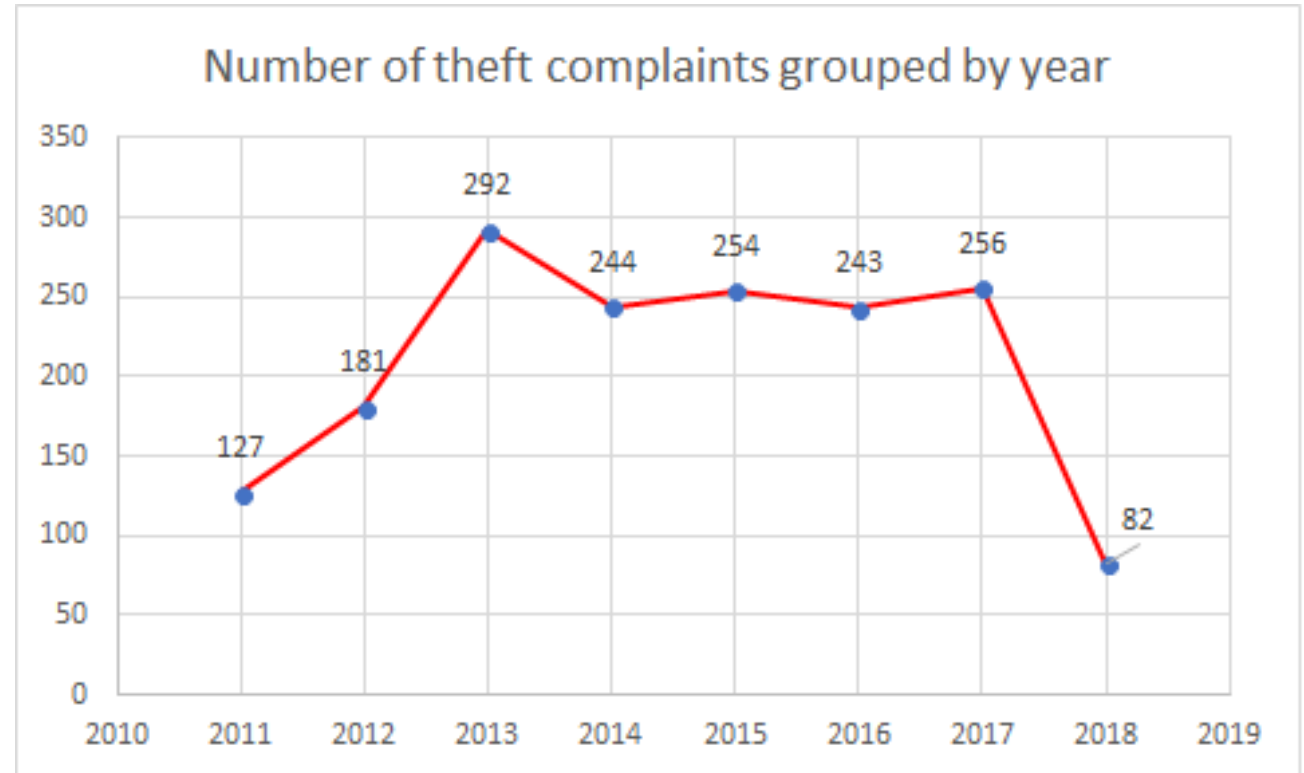
# Crime Analysis

- Crime Map analysis - locating on a map the crimes happened in a specific month;
- Density Crime Map visualization - displaying of a heat map that localize critical area for a specific year;
- Hot spot detection - identifying the number of crimes per district within the city or cities within the province;
- Crime comparison - comparing the number of crimes to get an idea about the growth or decrease of thefts over the time.



# Crime Analysis

- Crime Map analysis - locating on a map the crimes happened in a specific month;
- Density Crime Map visualization - displaying of a heat map that localize critical area for a specific year;
- Hot spot detection - identifying the number of crimes per district within the city or cities within the province;
- Crime comparison - comparing the number of crimes to get an idea about the growth or decrease of thefts over the time.



# Conclusion

---

- We built an interactive visually oriented web application providing summarized information about thefts in Modena
- Information are extracted from newspaper articles and builds a structured database of thefts.
- *We shown the great opportunity of crime mapping in a medium-size city as Modena.*

# On line web app

---

The web application is available at

<https://www.dbgroup.unimo.it/theftmap>



# Open issues

---

- Identify the exact date and time of an event
- Extract the stolen objects (topic detection/semantic analysis)
- Access all the theft reports

# Future Work

- Collaborate with law enforcement agencies to manage crime data in real time
- Integrate crime news from other local newspapers (de-duplication)
- Crime prediction
- Publication of crime information as **Linked Open Data** (to empower data integration, data reuse and data discovery)

# Projects

## TRAFAIR - Understanding Traffic Flows to Improve Air quality

Start: November, 1st 2018 - Duration: 2 years

Co-Funded (75%) by: INEA (EU)

The TRAFAIR project will set up or improve a traffic and air quality sensor network in 6 European cities of different size: Zaragoza, Florence, Livorno, Santiago de Compostela

It will provide real time estimates of traffic in the city on a urban scale and a service for prediction of urban weather forecast and traffic (using smart technologies)

## H-BOLD - Building high level visualizations on Big Open Linked Data

Started: January, 1st 2018 - Duration: 18 months

Funded by UNIMORE

This project aims to develop a generic tool for the exploration and multilevel analysis of Linked Open Data by addressing the challenge of managing large datasets.

# Contacts

- <https://sites.google.com/view/laurapo>
- [laura.po@unimore.it](mailto:laura.po@unimore.it)
- <https://www.dbgroup.unimore.it>