Sobo Research Infrastructure



Data Science for Migration Studies The SoBigData.eu Exploratory

Dino Pedreschi

BIG DATA AND ALTERNATIVE DATA SOURCES ON MIGRATION: FROM CASE STUDIES TO POLICY SUPPORT

European Commission Joint Research Centre (JRC)









A **Multidisciplinary** European **Infrastructure** for **Big Data** and **Social Data Mining** providing an integrated ecosystem for **ethically sensitive scientific discoveries** and advanced applications of social data mining on the various dimensions of social life, as recorded by "big data".









The **pillars** for reaching the goal

- a **distributed data ecosystem** for procurement, access and curation of big social data
- a platform of interoperable, social data mining methods and associated skills: tools, methodologies and services for mining, analysing, and visualising complex and massive datasets
- a community of multidisciplinary scientists, innovators, public bodies, citizen organizations, SMEs, as well as data science students at any level of education, brought together by extensive networking and innovation actions

Big Data Ecosystem

- Open Data

- Restricted Data
- Virtual Collections

Social Mining

- Text & Social Media Mining
- Social Network Analysis
- Human Mobility Analytics
- Web Analytics
- Visual Analytics
- Social Data

Ethical and Legal Framework

Social Mining & Big Data Ecosystem

Virtual Access

E-infrastructure



Transnational Access

Open calls Exploratory projects



Networking

Training Dissemination Innovation Accelerator

Building the e-infra & boosting joint research: exploratories

Our first exploratories: i.e. the Research Environments tailored to specific multidisciplinary domains

- different resources will made be available (and discoverable) by exploratories: data, methods and results/publications
- We have now four **driving scenarios (more to come)**:
 - societal debates
 - societal well-being and economic performance
 - city of citizens
 - migration studies

Exploratory: Big Data for City of Citizens

Ô



xploratory:

0

Big Data for Well Being and Economic Performance



Deprivation Index (in France) predicted with Mobile Phone traces

Exploratory: Big Data for Societal Debates

0



Polarization, controversy and topic trends on societal debates through social media





Migration studies

- Use <u>unconventional big datasets</u> (social networks, mobile phones, publications, supermarkets) to study migration (flows, stocks, impact on countries of origin and destination)
- Use <u>official datasets</u> to validate results from unconventional datasets
- Use both types of data to evaluate impact of policies on migration
- We are conducting several projects

Big datasets

Social network and web data

- Twitter Streaming data: various twitter datasets from project partners, in various languages, with geolocation
- GDELT Knowledge Graph database. a Big Data repertoire of online news articles.
- Mobile phone data
 - Orange dataset: mobile calls between Senegal and the rest of the world (country to country, 2012).
- Highly educated migrants
 - Company data (Estonia and Italy): members of the governing boards of companies (with place of birth).
 - Publication data: DBLP (computer science) and APS (physics)
- <u>Supermarket transaction data</u>
 - Itemized shopping transactions (over 10 years, with place of birth)







The story: Migration stages

- GO: Understanding migration stocks and flows
 - Nowcasting migration through the Twitter lens
 - Brain-drain and scientific migration
 - Policy and migration
- STAY: Evaluating migrant integration
 - Sentiment related to migration topics
 - Migration and language
 - Multi-culturality and sentiment
 - Migrant start-uppers
- **RETURN: Return of migrants**
 - Data journalism approach



Some examples

GO

Migration stocks



- <u>Ongoing</u> analysis
 - Large curated collection of geo-localised tweets
- Estimate user residence
 - monthly country from which the user posted in most days
- Estimate user nationality
 - Language and network most used
- Compare with official data



Brain drain



<u>Question:</u>

- what is the extent of migration of highly education migrants and what is the effect on the receiving community
- Ongoing Analysis:
 - quantify scientific migration in various scientific communities
 - evaluate success of immigrants in science
 - quantify <u>entrepreneurship</u> of immigrants (company data) and success in society
 - underline <u>achievements</u> of immigrants in an attempt to understand whether migration is beneficial both for individuals and receiving society

Scientific migration

<u>Question:</u>

- What features affect scientists migration?
- Ongoing Analysis:
 - The features that most distinguish migrant from non-migrant scientists are the number of collaborators (positive influence) and the ratio between collaborators internal and external to the current institution (negative influence).



dblp

Scientific migration

• Question:

- When do scientists migrate, and where do they go?
- Ongoing Analysis:
 - Using machine learning we can predict whether a scientist will move or not in the next 5 years
 <u>model</u> ACC recall precision F1 AU

model	ACC	recall	precision	F.T	AUC
tree	0.87	0.84	0.86	0.85	0.87
logit	0.83	0.89	0.65	0.75	0.85

dblp

• We can also predict whether the scientist will move to the institution with which they collaborated most or not (F1=0.72)

Policy and migration

<u>Question:</u>

- how does policy affect immigration, particularly the ratio between legal and illegal migrants?
- Planned Analysis:
 - quantify legal and illegal migration using official data but also alternative datasets (roaming data, tweets around hot areas, such as the "jungle of Calais")
 - identify trend changes
 - <u>Challenge 1:</u> identify policy changes and other (economical, historical, legal) factors that could have cause observed trends
 - Challenge 2: identify possible policies to curb illegal migration



Istat





Some examples

STAY

Sentiment on migration topics: Perception of the Mediterranean Refugee Crisis

- What is the <u>evolution of the discussions</u> about refugees migration in Twitter?
- What is the <u>sentiment of users across Europe</u> in relation to the refugee crisis?
- What is the <u>evolution of the perception</u> in countries affected by the phenomenon?
- Are users more <u>polarised</u> in countries most impacted by the migration flow?

Sentiment on migration topics: Perception of the Mediterranean Refugee Crisis

• European country mentions

Africa & Middle East country mentions





Sentiment on migration topics: Perception of the Mediterranean Refugee Crisis

- Internal and external perception by country
 - Index $\rho\,$ the ratio between pro refugees users and against refugees users
 - Red means a higher predominance of positive sentiment, higher ρ
 - Yellow means a higher predominance of negative sentiment, lower p



Multiculturality and sentiment



twitt

- Question:
 - how does migration affect overall <u>sentiment</u> and <u>language</u> of a community?
- Ongoing Analysis:
 - quantify sentiment in tweets coming from different countries (geolocalised) and in different languages
 - compare sentiment of <u>various languages in the same location</u>
 - compare sentiment of <u>the same language in various locations</u> across the world.
 - compare sentiment across areas with different levels of immigration

Integration from sentiment



- Question:
 - Can we extract an integration index based on sentiment and language usage on twitter?
- Ongoing Analysis:
 - Sentiment spreading annotate a large dictionary of words starting from a small annotated seed (ANEW, Bradley&Lang dictionary)
 - The correlation between original and modelled valence of words is an indicator of how language is used in multicultural regions of the world – a superdiversity index
 - By analysing superdiversity in subpopulations on Twitter we aim to underline the effects of the presence of migrants on language and sentiment at the same time.

Food and migration

• Data

- customer, country of birth, association date,
- shopping baskets, products details,
- time and location of shopping, amount spent
- Objective
 - While the time passes is the shopping profile of foreign customers becoming more similar to those of Italian customers?
 - Can we efficiently quantify integration of various groups of immigrants?
- Collaboration with Paris School of Economics







Some examples

RETURN

Return of migrants



- Demal te niew (Go and Come Back)
- Documentary interviews with migrants returning to Senegal from Italy
 - Featured in Espresso, El Pais
 - Presented at Ethnographic Film Festival, Amsterdam and International Day of Migrants, Dakar.
- <u>http://journalismgrants.org/projects/demal-te-niew-go-and-come-back</u>

Conclusions

- A research direction
 - Building an integration index based on various Big Data types:
 - Evaluation of Local Shopping Habits Adoption
 - Ethnical/Local Products Adoption
 - Analysis of Possible Recipes Consumption
 - Evaluation of Economic Condition through the Observation of Shopping Offers Adoption
 - Superdiversity index and sentiment variability across communities
- Collaboration through SoBigData Transnational Access



Sobo Research Infrastructure

www.sobigdata.eu H2020 excellent science research infrastructure



Ph.D. in Data Science

- Started: academic year 2017-2018
- Call for 2018-2019 opens in December 2017
- http://phd.sns.it/data-science/

