Urban resilience, social exclusion, and poverty – (big) data challenges

Maria-Cristina Marinescu, Barcelona Supercomputing Center

Smart Cities

GREEN

EQUITABLE

SUSTAINABLE

SELF-SUFFICIENT

LIVEABLE

RESILIENT

SECURE

Cities can`t be smart without reflecting cross-domain dependencies

NEW HOUSING IN PREVIOUSLY RAN DOWN NEIGHBOURHOOD
… housing may become unaffordable for previous neighbours
… neighbourhoods loose identity, friends/family move apart
… existing local issues spread globally (crime, STDs – Baltimore housing projects, ...)

TRAM TO AVOID TRAFFIC JAMS AND RUN DIRECTLY
… may collapse car /public transport. traffic that runs in different direction

BETTER, FASTER MAIN AVENUE
… may split neighbourhoods

WALKABLE CITIES
… commercial traffic worsens, no place to stop
… less parking space for neighbours
… possibly in places where people don`t usually walk! (e.g. steep hills)

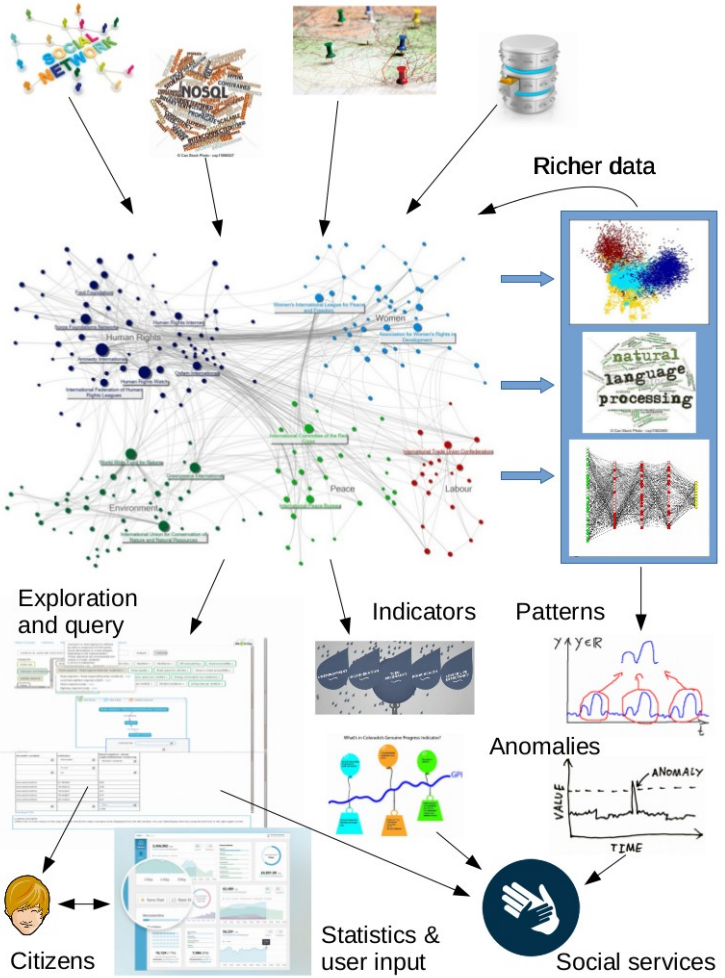
CHANGING / DAMMING WATER COURSE FOR CITIES
… ecosystem changes
… displaced people

CITY FOCUSES ON LOCAL PRODUCE / RURAL REGION SPECIALIZES IN FEW HIGHLY DEMANDED PRODUCTS
… draught or plague may hit a crop → there`s no backup plan for rural region, and delayed response in city

Concrete problems in urban environments

- Need to plan, respond, optimize, predict, asses risk, etc
- Use case 1: *individuals at risk of poverty or social exclusion*
  - Water consumption data
  - Padron
  - Income / taxes (hacienda)
  - Education, school attendance
  - Health services
  - Etc…

  What if no access to data?



Concrete problems in urban environments

- Need to plan, respond, optimize, predict, asses risk, etc, which requires:

Instrument and monitor the city

Integrate heterogeneous data sets

Space, time, unstructured data, etc – much more data, new types of data

Analyze and query

Understand complex patterns from big data

Compute and optimize metrics

Understand the effect of changes/decisions beforehand

Predict, including (timed) events

Simulate

Visualize

Why integration via modeling?

- ***Simplifies the development*** of applications that require integrated access to city data sources (cross-domain)
- Enables *solution reuse* as we move from one city to the next
- The world is open, changing, incomplete, and data may be faulty. Models are ***easy to evolve and maintain*** (reuse, repurpose, naturally models change) without modifying the application or the data sources.
- Standardized indicators for cities may be implemented as part of the model.

Ontology = Class + Relations + Constraints

Knowledge Base = Ontology + instances + (Standard) Inference and rules

Interoperability challenges

Data sharing - LOD

Data cleaning

Data enrichment (metadata)

Data integration (via modeling)

Building the model: evolving ontologies, competence queries
- need complex queries!!
- need tools to make complex models user fiendly

Models as a way to integrate data:
- often lots of missing data + some data can`t be monitored
- data may not "naturally" fit the model´s relationships - semantic approximations? e.g. Legal entities / address in a block
- granularity of data issues

Format / mapping issues (data to model mapping)

Analysis

Prediction:
- garbage in / garbage out - data skew or poor quality
- correlation NOT causality; common sense is the hardest to learn by a machine (semantics can capture common sense)
- e.g. semantic helps to disambiguate (St. George on a Bike)
- understand propagating effects? e.g. earthquake

Reasoning (w. Probabilities)

Computing indicators:
- expert formulas vs via data mining
- can standardize city indicators as part of the model

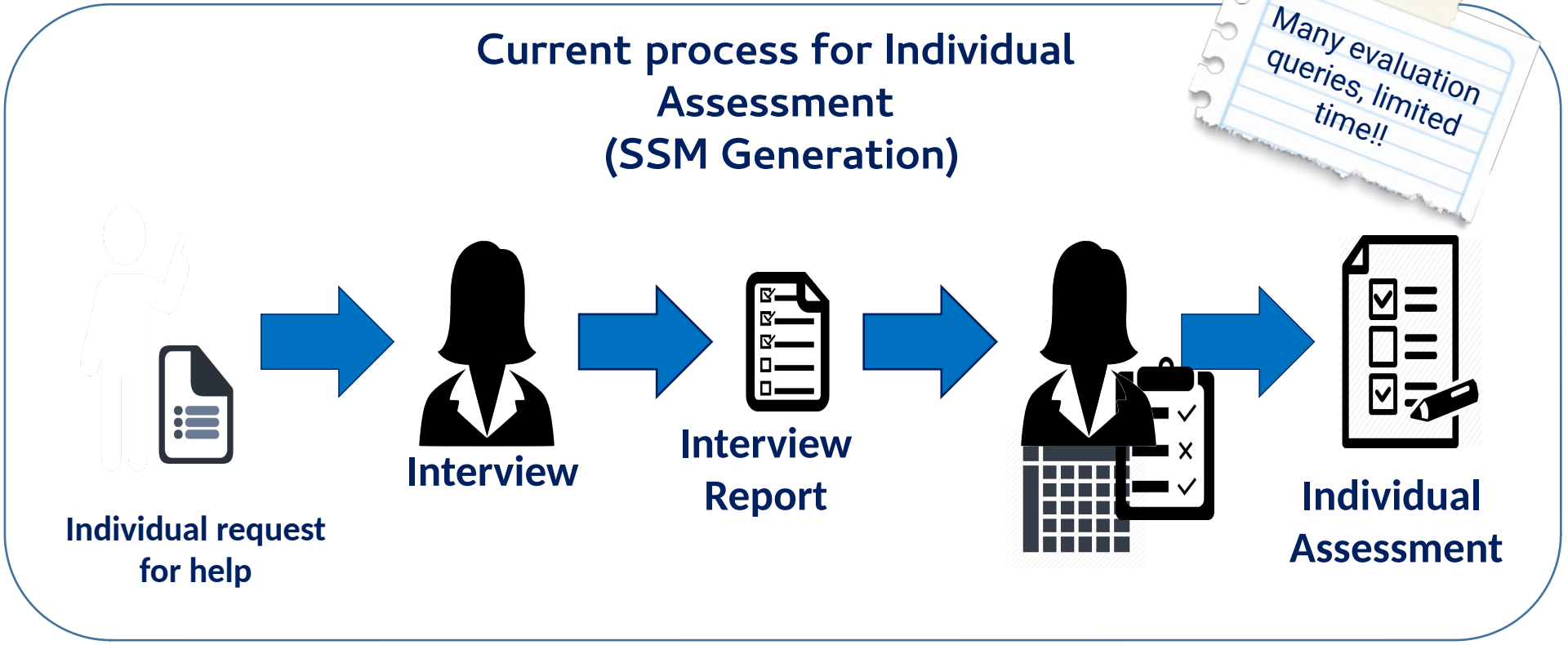
Data quality

Model constraints / programatically
Depends on the application
Some data is subjective

Use case 1: Social exclusion and poverty

- Data from very many different sources, but some is totally missing and can “break” the model graph
- Lots of data is statistical – not enough to learn patterns, e.g. “not socially integrated may → addiction”, “drinking and living alone may → skip meds”, “drinking and small kids may → domestic violence”, “unstable job and chronic disease may → in danger of eviction” etc
- Rights to access the data + anonimization (laws different in different places at different times)
  - People don´t ask for help (it´s cultural; need to change the way people percieve the profile of the poor or excluded)
  - Example in Barcelona - farmacies raise alarms when people that take meds don´t show up - locally things are possible that i can`t do globally
- Data is not always clean: **incompletenes ok** - inconsistencies not ok, constraint checking as part of the model vs programatically (+ and -), Open World Assumption!!
  - same apt_id but different addresses (use additional info to understand where the error is - use ids of people living there? other data?)
  - 2 people/ 3 ids in same appt (are 2 of them the same person?) vs 4 people/ 3 ids (are there underage individuals living there?)
  - addresses with MANY inhabitants – usually social services locations!
  - married underage (or under 16)

Self-sufficiency matrix



Why use SSM?

- Holistic view of an individual
- "Standard" way of evaluation
- Accountability of decision making
- Evaluation of policies' impact
- Proven usefulness / already in use

- Social workers could take advantage of an approximate initial snapshot to focus on real causes
- Semantic model and indicators can help!

Use case 2: Urban resilience

- Projects:
  - Model water system, contamination, transportation, energy, integrated infrastructures – in Barcelona (mostly w. City council / depts)
    - Data is very basic (sensors, components: taxi stands, electric vehicles, smart towers, …)
    - People tend not to stress cross-domain interconnections in the model because it´s something that was not possible before
    - Only with statistical data it´s not enough, but people don`t want to publish data (how to encourage sharing?)
  - Urban planning (w. IBM and the Barcelona Urban Ecology agency): we received cleaned data (problem examples: intersecting geometrical shapes, building heights in meters and floor numbers didn´t match) - they decided beforehand which source to believe!
  - UNHabitat – urban resilience model: modeling, no data, they wanted to (1) compute indicators based on survey data rather than (open) data, (2) simulate cascaded effects.

E.g. Urban resilience

- Long-term
  - Policy
  - Education
  - Studies / Analysis
- Short-term
  - Prediction
  - Early warning
  - Response

E.g. Urban resilience

Response

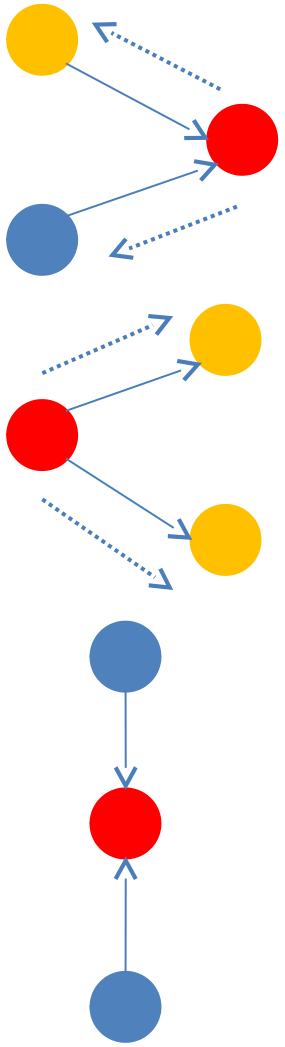
Diagnose the problem:

– What created it? Is it a shock or a stress? Can it be fixed?

– What does it affect? Generate alerts, notify, provide alternative ways to satisfy the dependency

– What (else) has an effect on the experienced stress or shock?

It´s about *understanding and formalizing connections*!



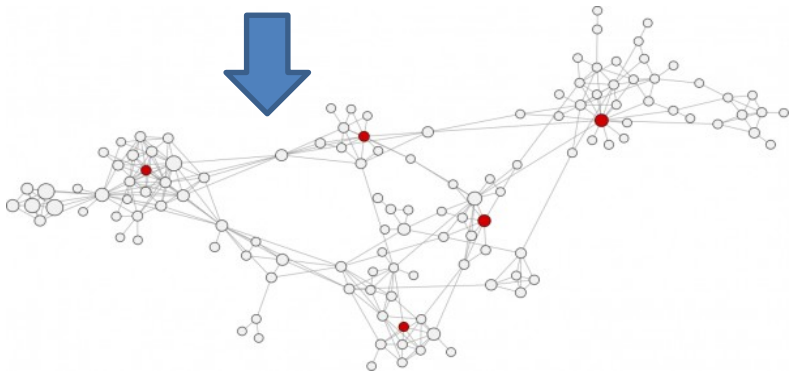
E.g. Urban resilience

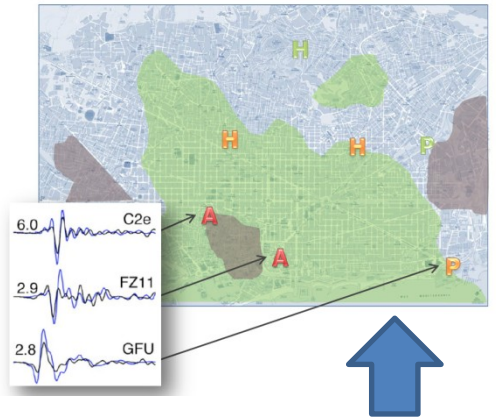
Prediction

- Pre-event: which are the most vulnerable aspects that affect city resilience?

  - Can be used for early warning or in long-term studies

  - Technology: network analysis, e.g. loosely connected clusters, hubs, etc.





Shaking map + vulnerability

Critical infrastructures

Damage maps need detailed modeling of substrate and mesh generation

- Post-event: where is immediate response most needed when not evident, too expensive, time-consuming, or infeasible to find out?

  - Used for response

  - Technology: Probablistic approach or deterministic approach – simulation, e.g. where to send rescue first in case of earthquake?

E.g. Urban resilience

Building for resilience

- Analyze via what-if scenarios

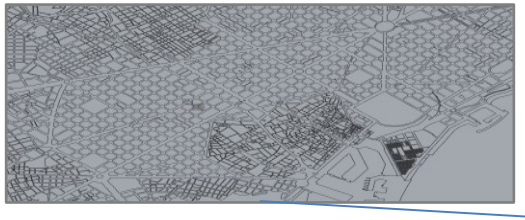
  Assume a change, what are all the aspects that are affected, and how do indicators change values? E.g. Baltimore house projects

- Network analysis

- Predict using machine learning, probabilistic techniques, etc.

- Simulation, e.g. polution dispersal

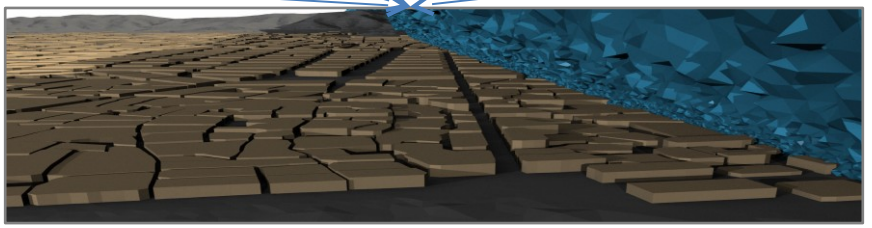
Cadastre (geometry)



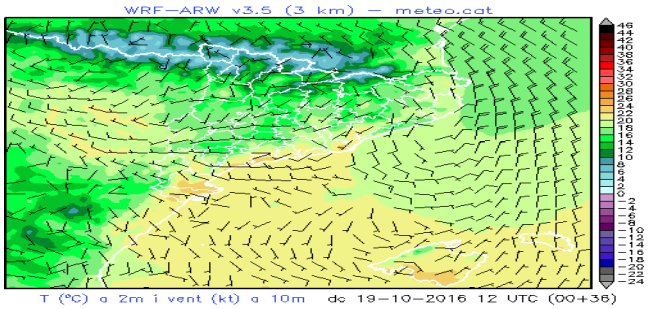
LIDAR (height)

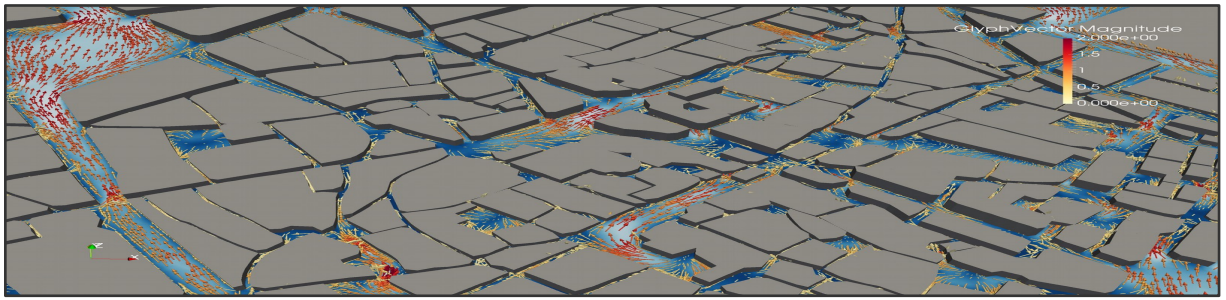
Topo elevation

Wind initial conditions at meso-scale level

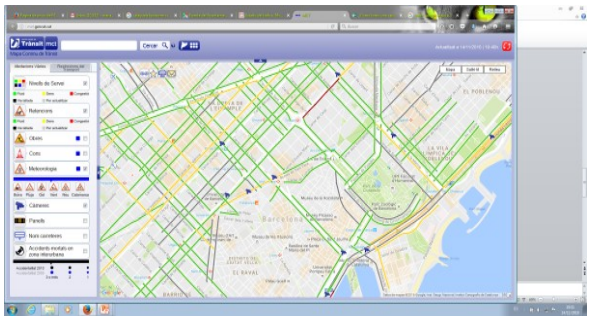


City mesh at block level

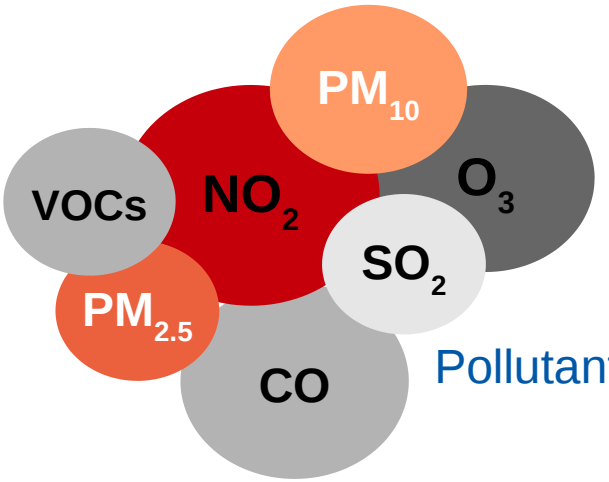




Weather



Traffic



Pollutants

Other examples of data issues

- Data can also be just wrong:
  - student declares living alone to get scholarship; he lives with parents
  - registering at different address for school district reasons / …
  - illegal work
  - businesses that change area, type of service provided, ... but don´t declare it. businesses that are not declared at all (quantity of garbage produced can tell a story). AirBnB – gentrification.
  - renovation year recent but no license registered
- Very hard to detect, but without crossing data virtually impossible.

Questions?

maria.marinescu@bsc.es