Open Data + Big Data for a Better Mexico
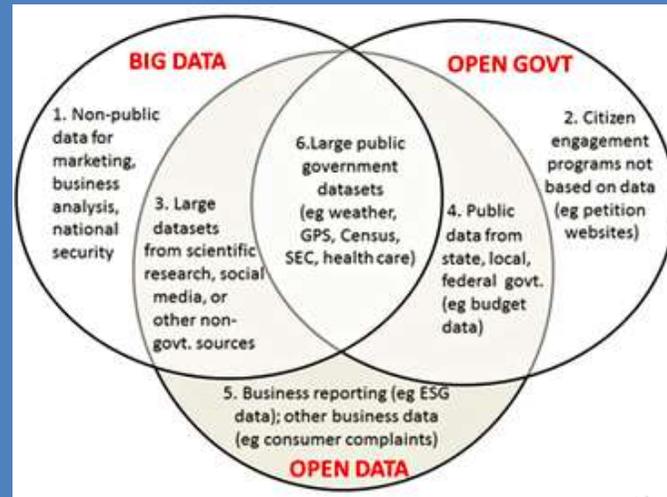
Jose Tam

jose.tam@itesm.mx

Tecnológico de Monterrey

April 23, 2015

+ Contents

- What do we undertand by Open Data?
- How can Open Data be used?
- Linked Data Methodologies
- Big Data with Open Data in Mexico
- Case Study: Natural Disaster Plan Improvement
- Links

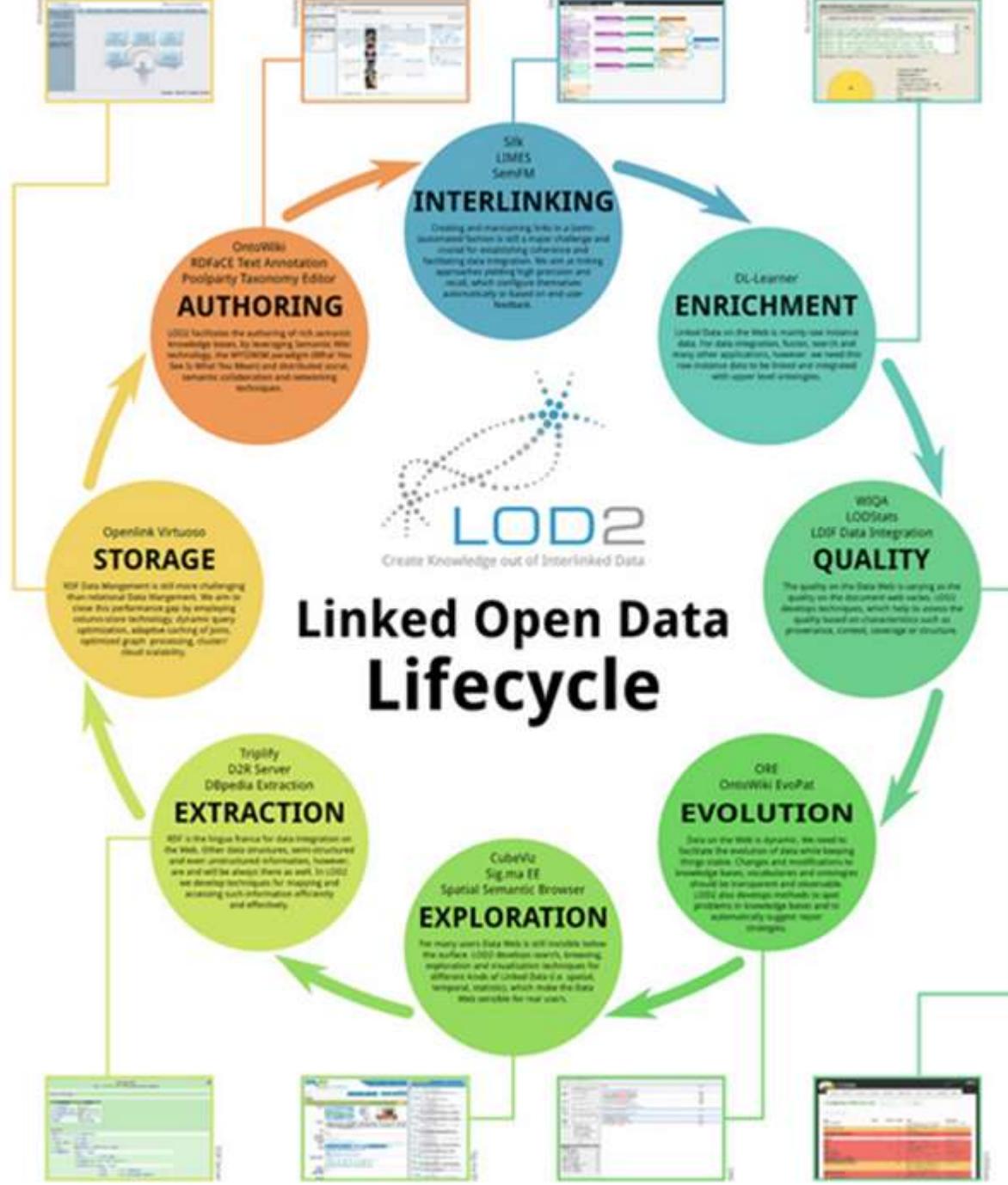# What do we understand by Open Data?

- Social media text
- Cell phone locations
- Channel click information from set-top box
- Web browsing and search
- Product manuals
- Communications network events
- Call detail records (CDRs)
- Radio Frequency Identification (RFID) tags
- Maps
- Traffic patterns
- Weather data
- Mainframe logs
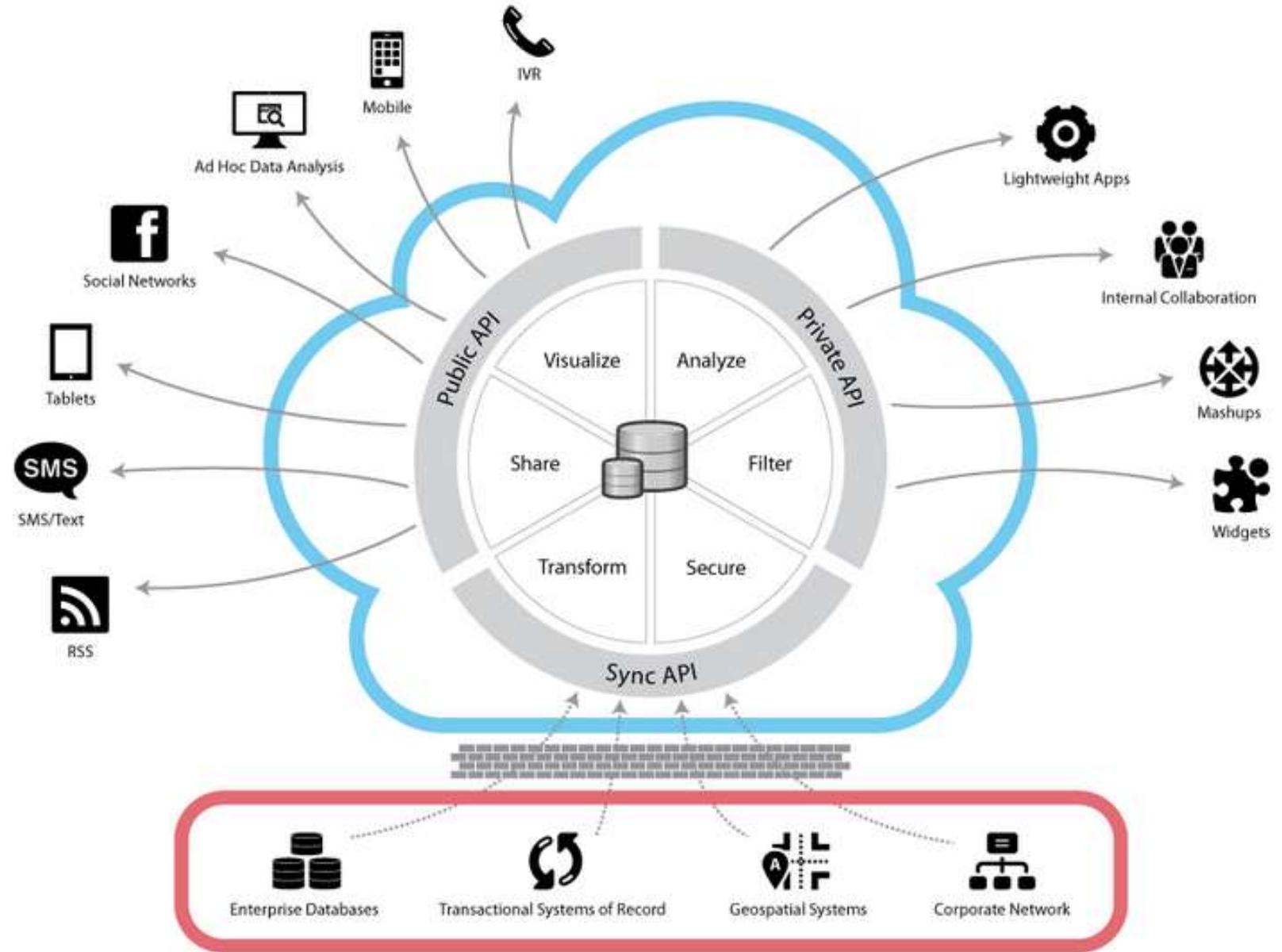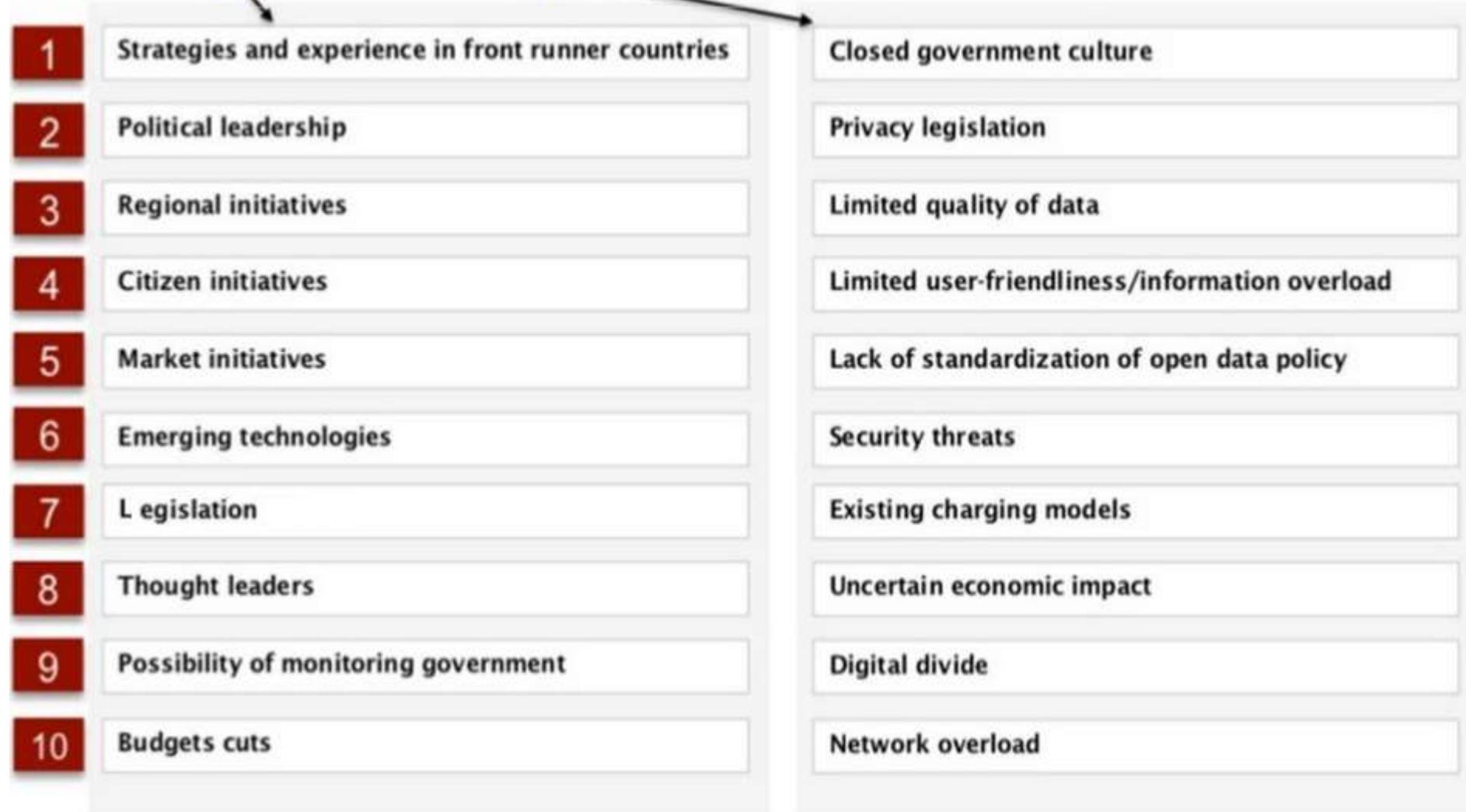
How can we use Open Data?

The Linked Open Data Lifecycle

The importance of an Open Data Policy

Drivers and barries of open data policy implementation

#	Drivers	Barries
1	Strategies and experience in front runner countries	Closed government culture
2	Political leadership	Privacy legislation
3	Regional initiatives	Limited quality of data
4	Citizen initiatives	Limited user-friendliness/information overload
5	Market initiatives	Lack of standardization of open data policy
6	Emerging technologies	Security threats
7	L egislation	Existing charging models
8	Thought leaders	Uncertain economic impact
9	Possibility of monitoring government	Digital divide
10	Budgets cuts	Network overload

How can we use Open Data in Mexico?

A National Data Initiative in Mexico

The first space in which citizens use open data to the public and private sector to help the government make decisions that improve the quality of life of citizens.

The objective of this experiment is to exploit the creativity of the participants and find information and innovative solutions that provide the government with information it needs to generate better public policies.

In this exercise foundation provided by the government of the municipality of Zapopan, the Federal Government, in addition to databases of private initiative, donated by Telefonica Movistar will gather.

The teams generate a project that addresses a public issue in Zapopan and produce useful information for decision makers.
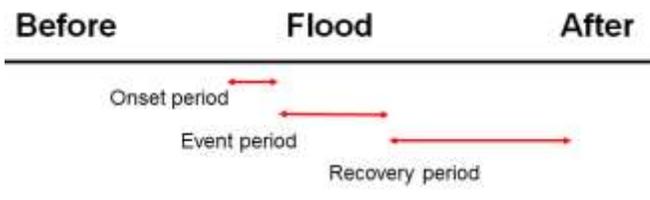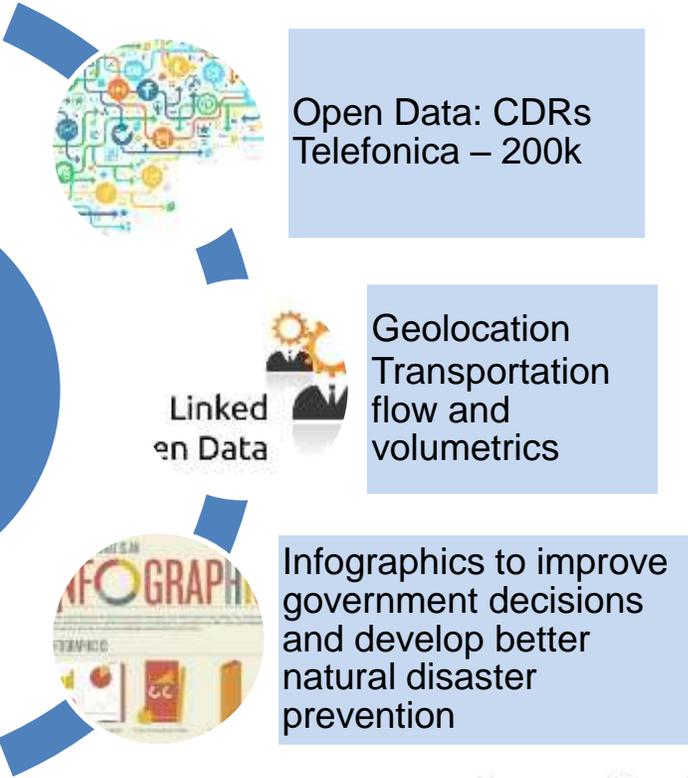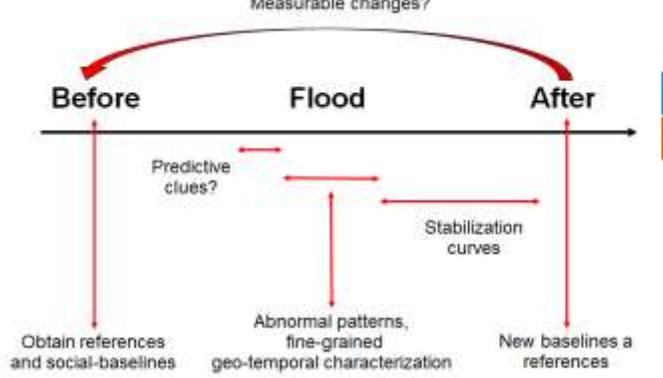
Case Study: Natural Disaster Prevention

Tabasco

The importance of an Open Data Policy

Big Data & Open Data

Open Data: CDRs Telefonica – 200k

Geolocation Transportation flow and volumetrics

Linked en Data

Infographics to improve government decisions and develop better natural disaster prevention

Before Flood After

Onset period

Event period

Recovery period

Two different perspectives of data analysis:

• Geolocation-centered perspective: Data is aggregated at the level of the towers or sets of towers defining regions
  - Displacement networks
  - Activity source and sinks
  - Settlements
  - Calls traffic networks

• User-centered perspective: Data is built around active agents (phone users) or target populations
  - Trajectories
  - Social networks
  - Home locations
  - Migration flows

Can we see if people were affected by changing their home-locations during the flood?

How long did the people stayed out? Did they all come back after the general population came back to normal?

Can we estimate changes in the patterns of mobility inside Tabasco and Tabasco with the rest of Mexico? Were they concentrated in specific areas or links of the mobility network?

Can we estimate same changes in calls traffic network? Were some communication links specially affected? Did everything come back to normal after a recovery period?

Measurable changes?

Before Flood After

Predictive clues?

Stabilization curves

Obtain references and social-baselines

Abnormal patterns, fine-grained geo-temporal characterization

New baselines a references
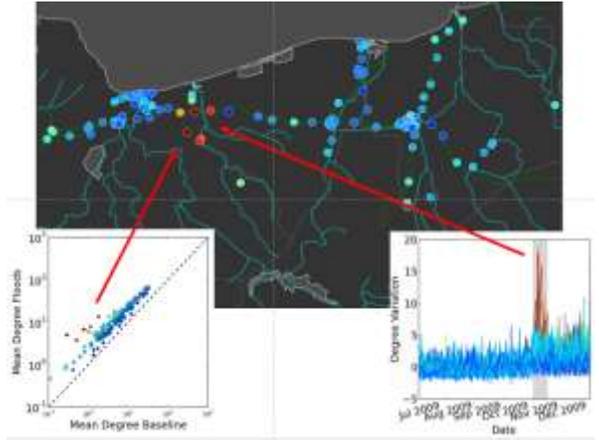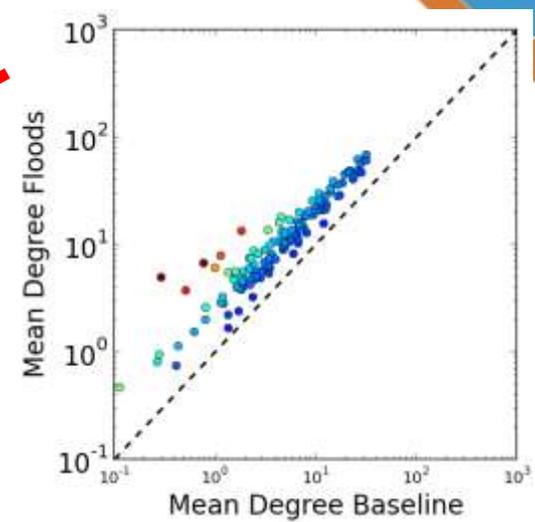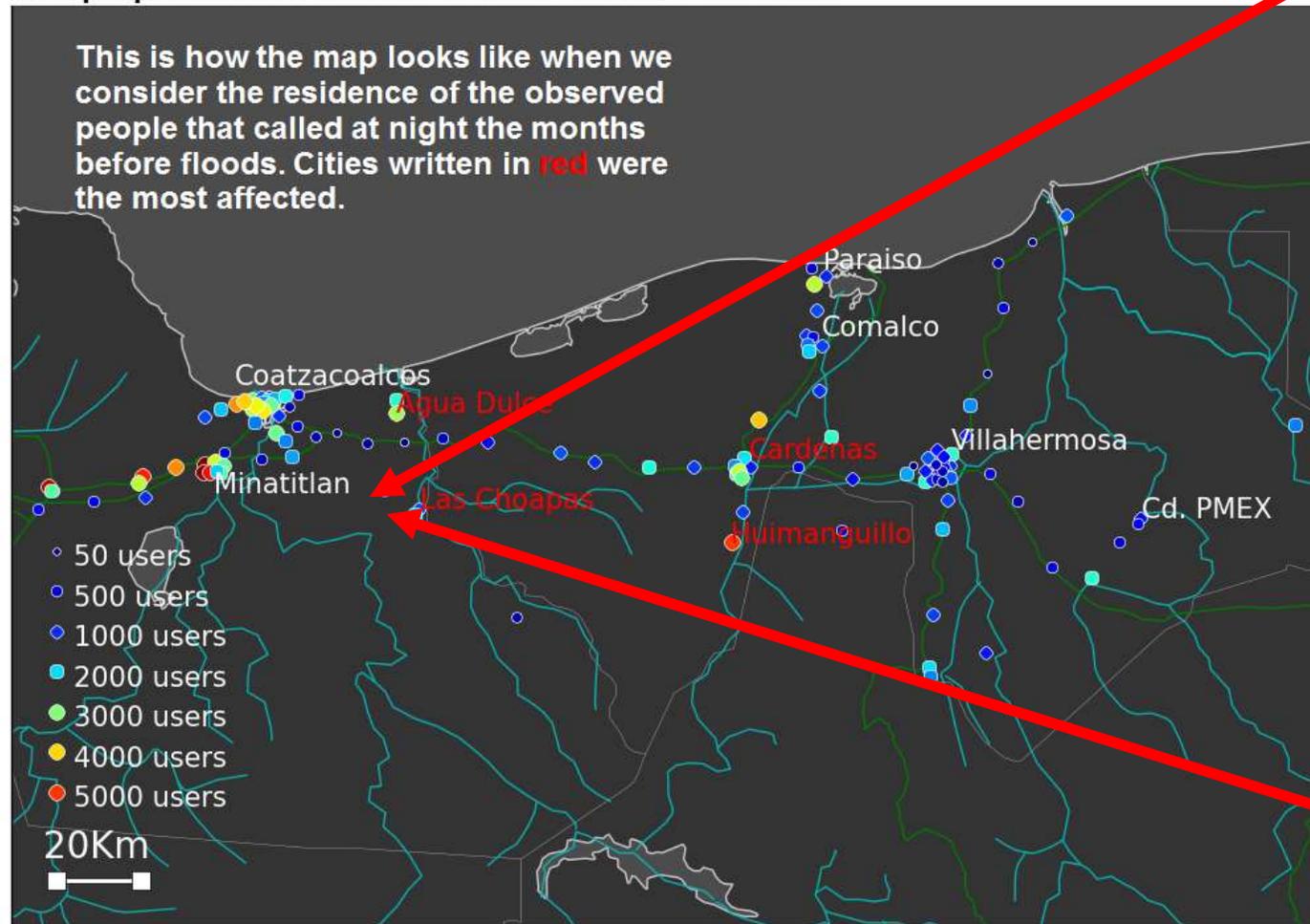
Baselines characterization

- Home locations: users' patterns after 8pm
- Who lives where: history of home locations
- Social baseline: who lives where + user's activity density + geolocated areas
- Settlements: clusters of locations where the social base-line lives.
- Average mobility and social networks
- Average trajectories and in-out patterns

Hurricane IDA – Tabasco 2009

Who lives where
Size proportional to number of users served after 20hrs

This is how the map looks like when we consider the residence of the observed people that called at night the months before floods. Cities written in red were the most affected.

Paraiso
Comalco
Coatzacoalcos
Agua Dulce
Villahermosa
Cardenas
Minatitlan
Las Choapas
Cd. PMEX
Huimanguillo

○ 50 users
● 500 users
● 1000 users
● 2000 users
● 3000 users
● 4000 users
● 5000 users

20Km

Here we present how the number of connections of towers changed in time.

In the lower right figure, we present the towers mean degree during the previous month versus their mean degree during floods.

We see that all of them are above their expected value, and some are way above. These are located around the most affected areas.
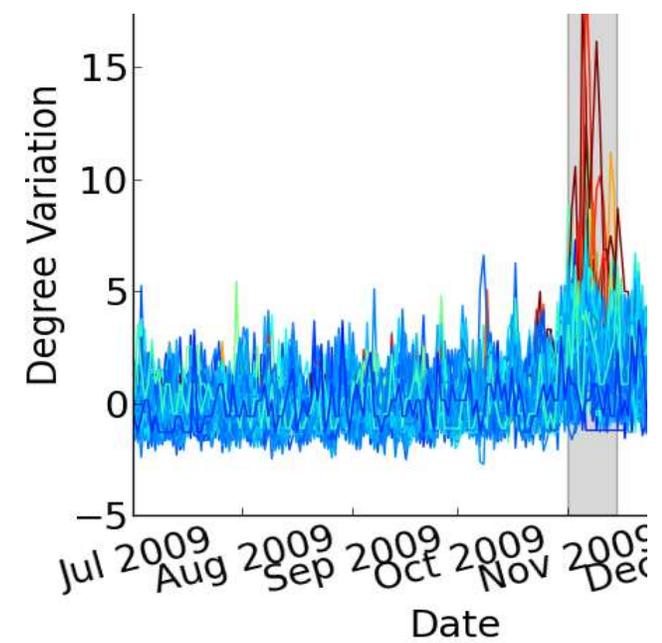
Open Data: People and transportation flow before and after Hurricane IDA 2009

Who changed residence

Network of night displacements during floods

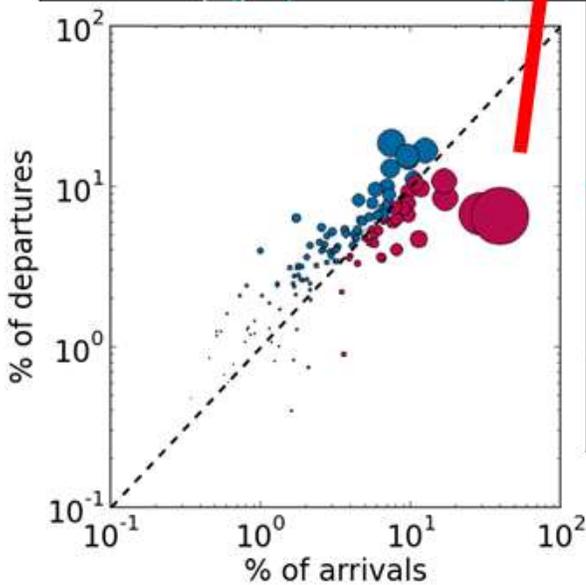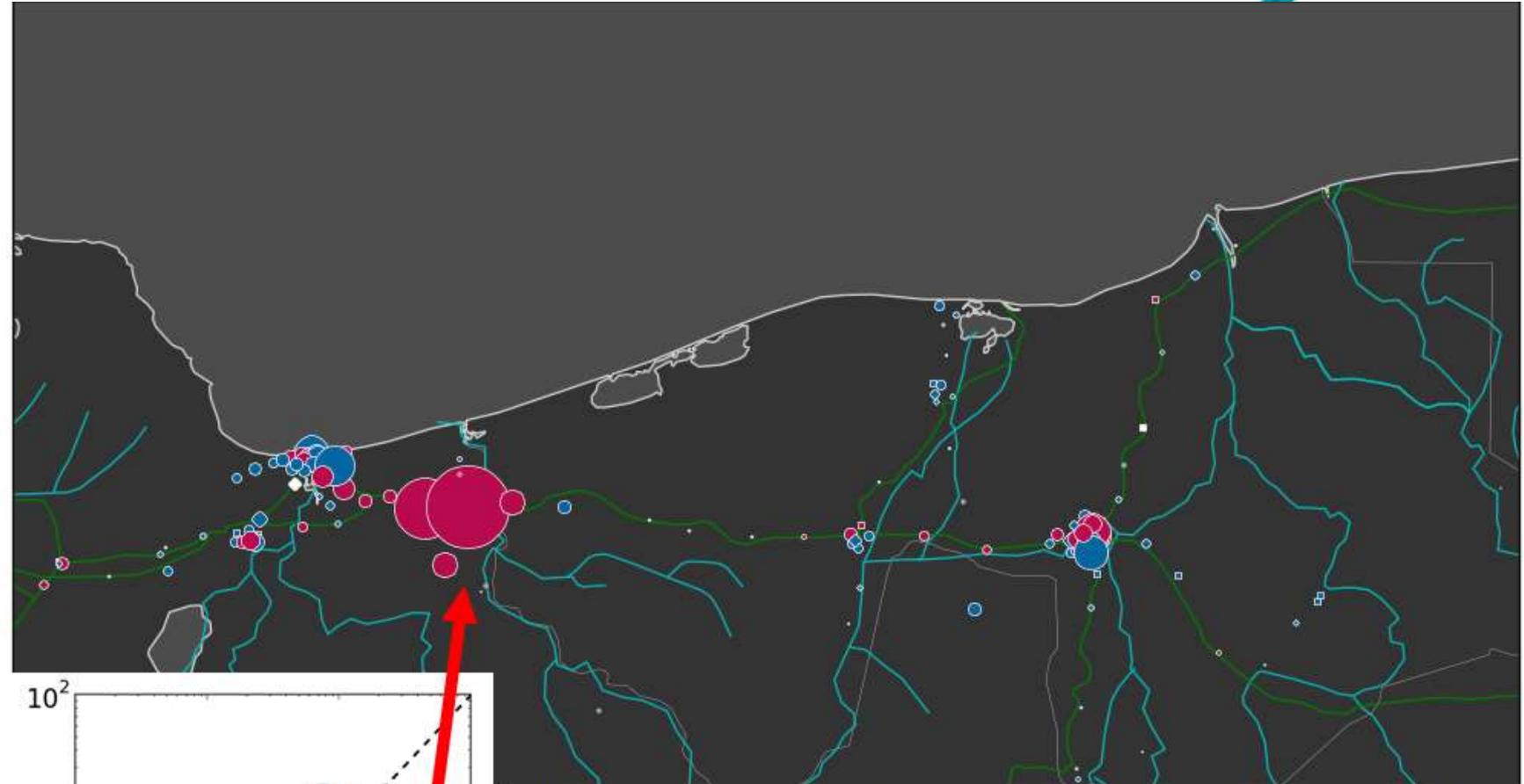Evacuation centers???

Random period
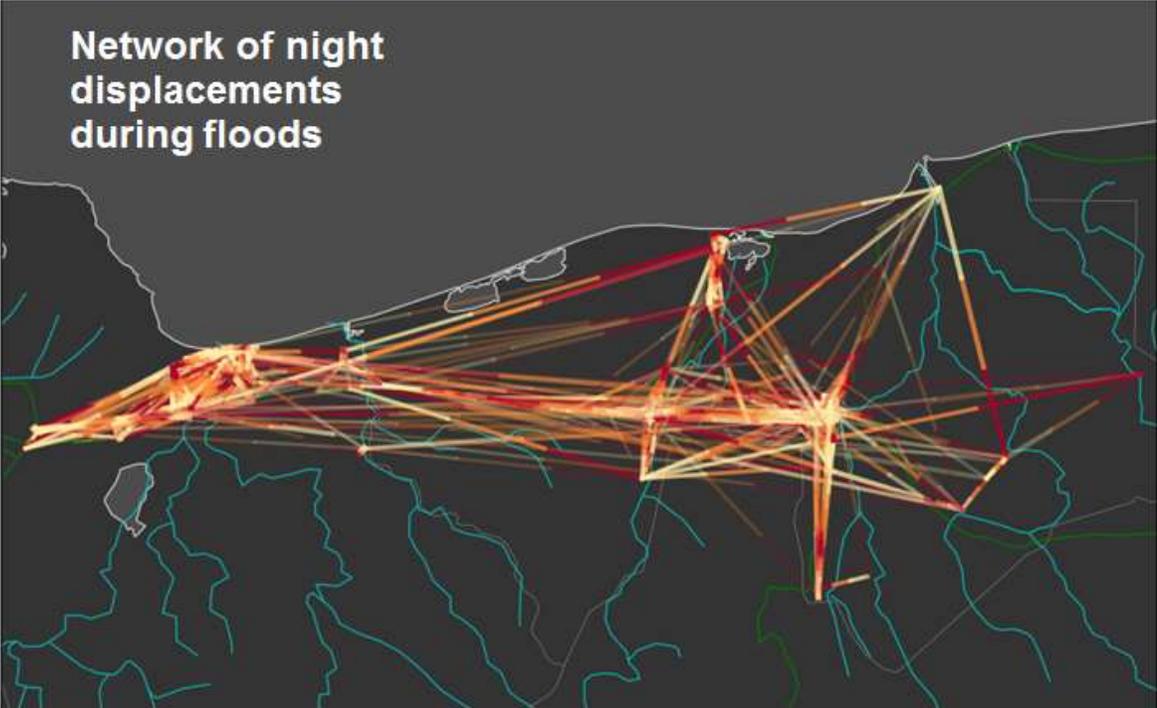
Here we present a snapshot of the night displacements during floods and the same days one month before. The edge goes from red to pale yellow.

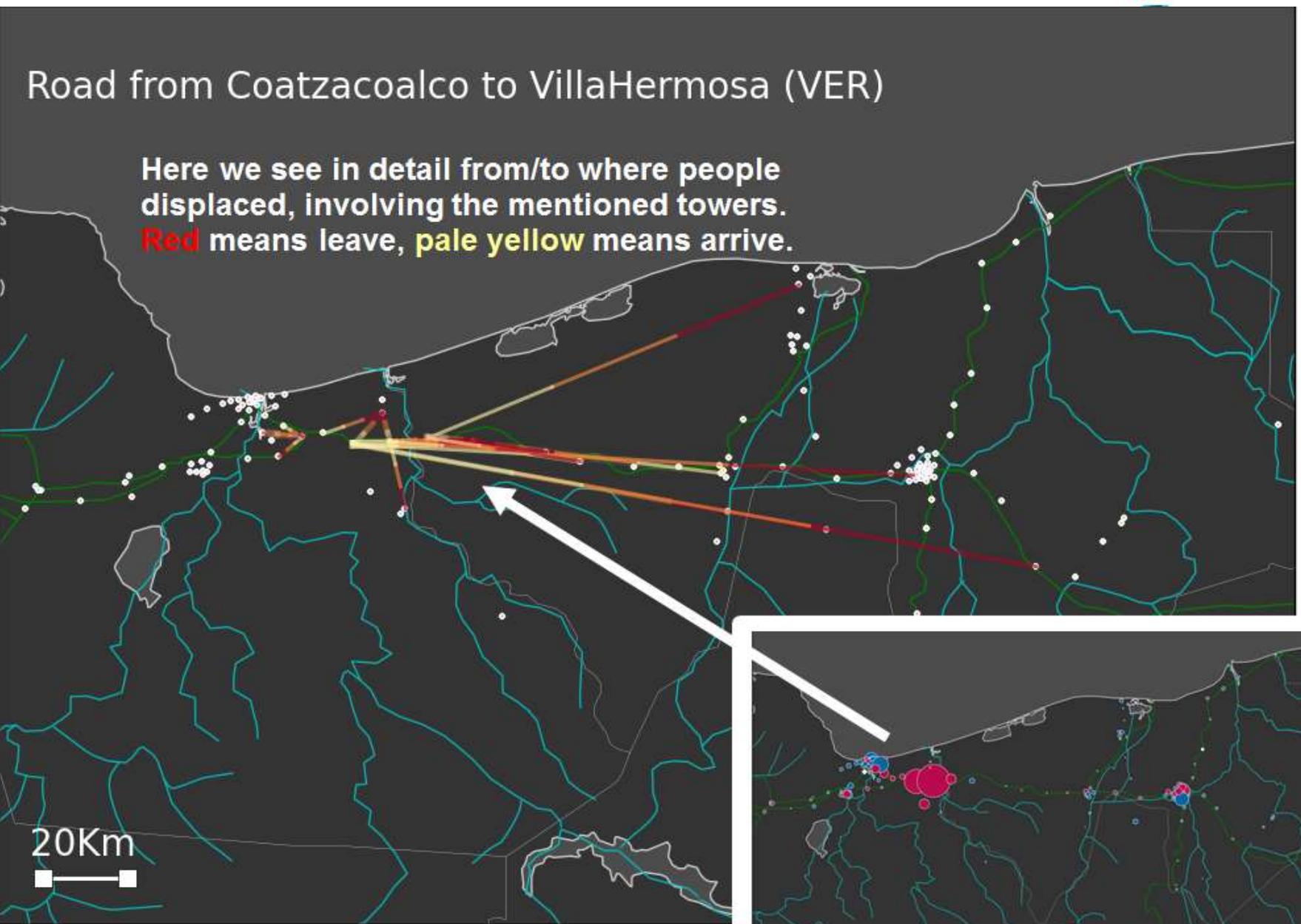We see that the resulting network is much denser and presents more edges than the previous period.
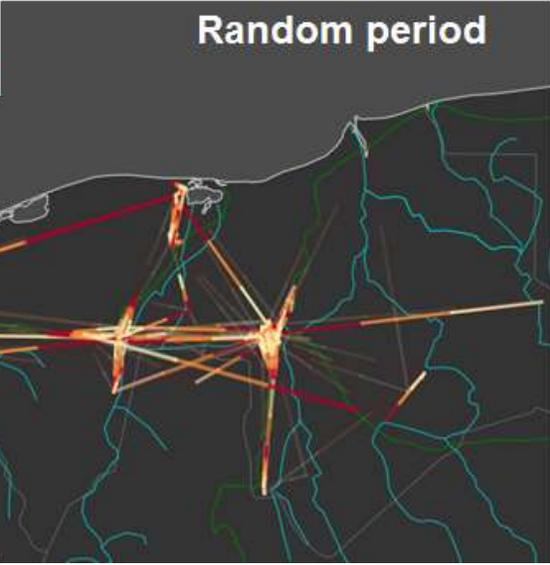
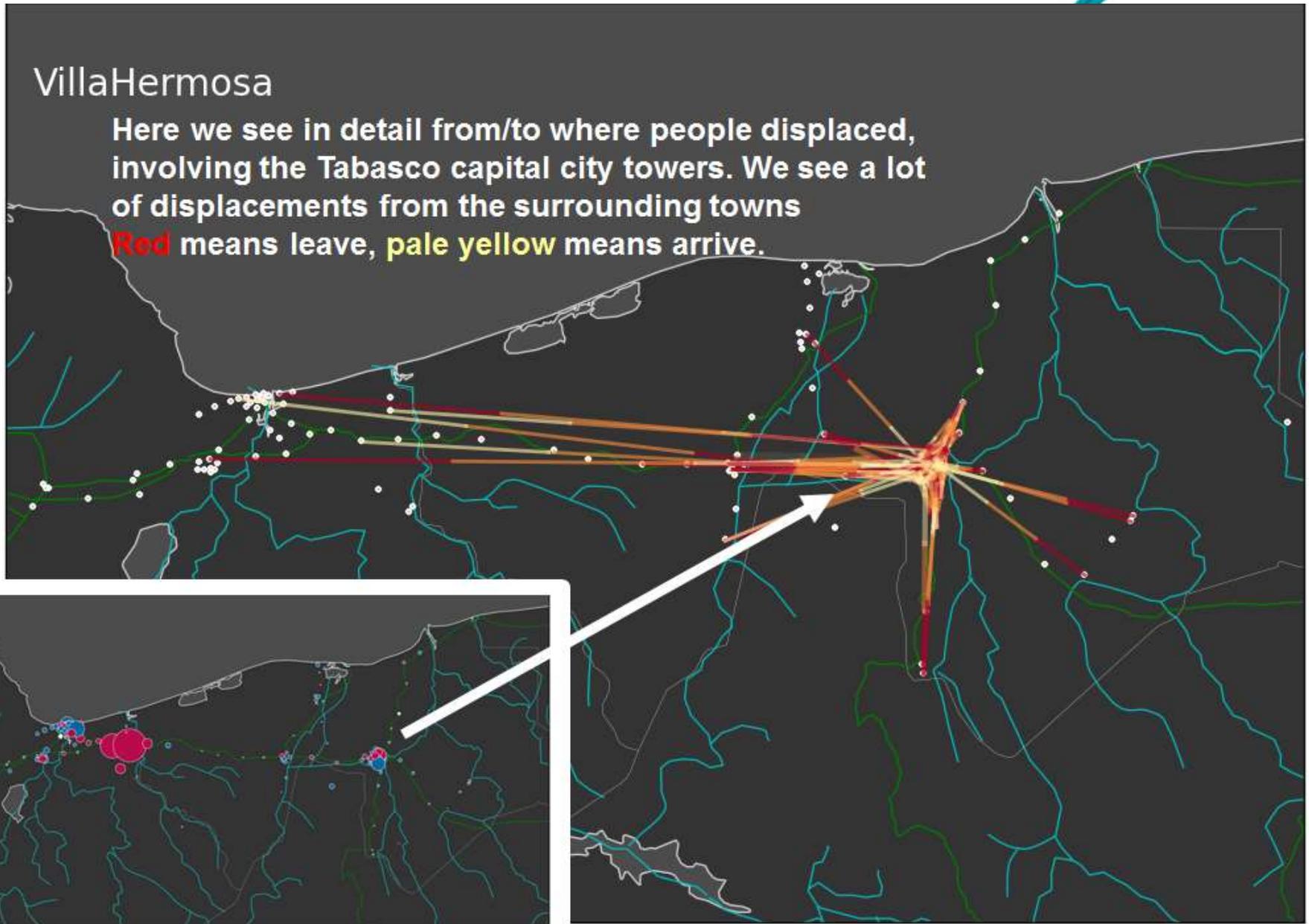# Open Data: Family movement during Hurricane IDA 2009

All displacements imply leaving and arriving to a place. In this map red towers received more people than those who left, and blue towers lost more people than they gained. The size is proportional to the % loss or gain, according to the number of people that we determined as residents of those towers.

We see again that the tower at the road (gren line) between Tabasco and Veracruz received a much larger amount of people than usual.
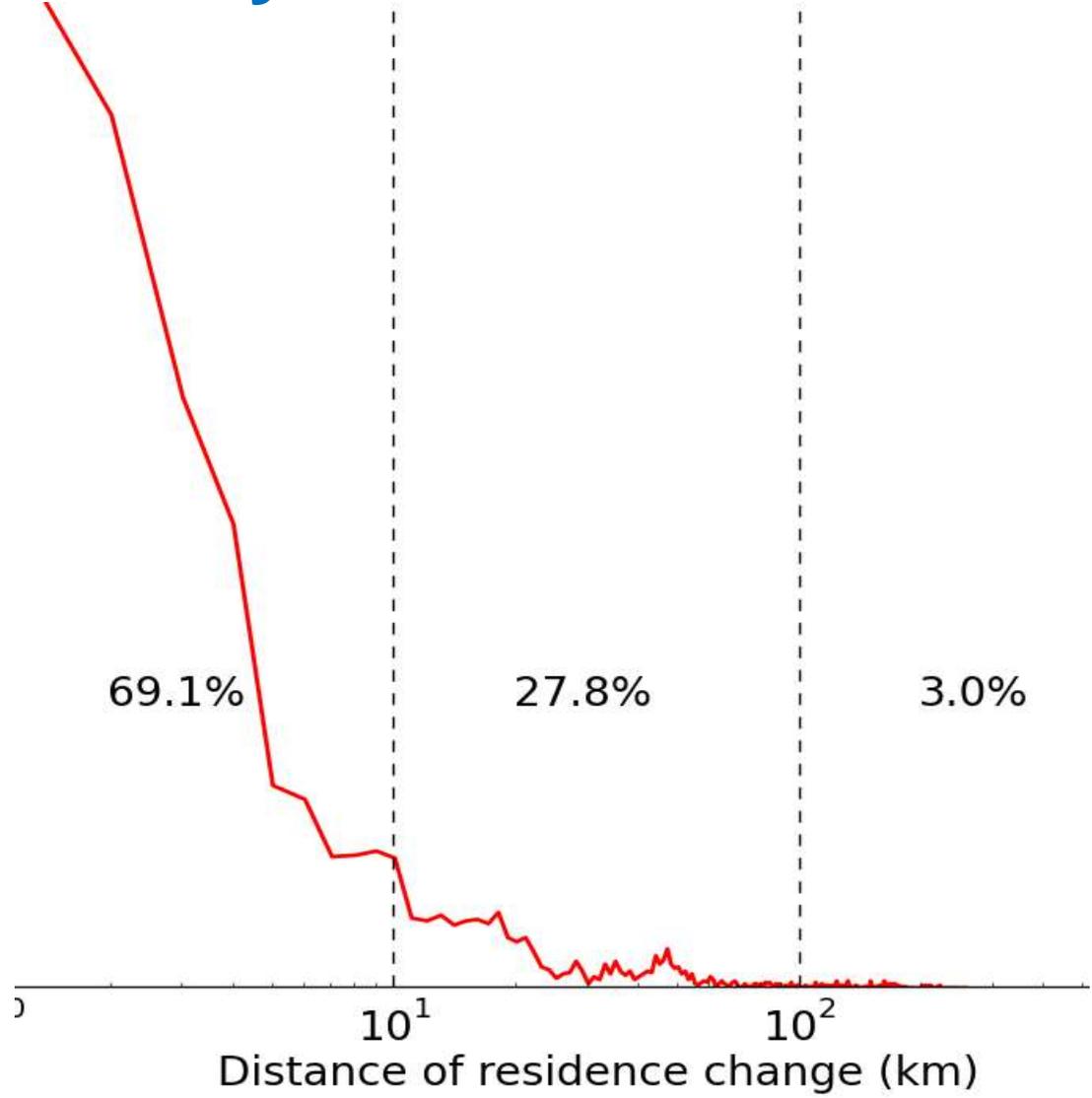
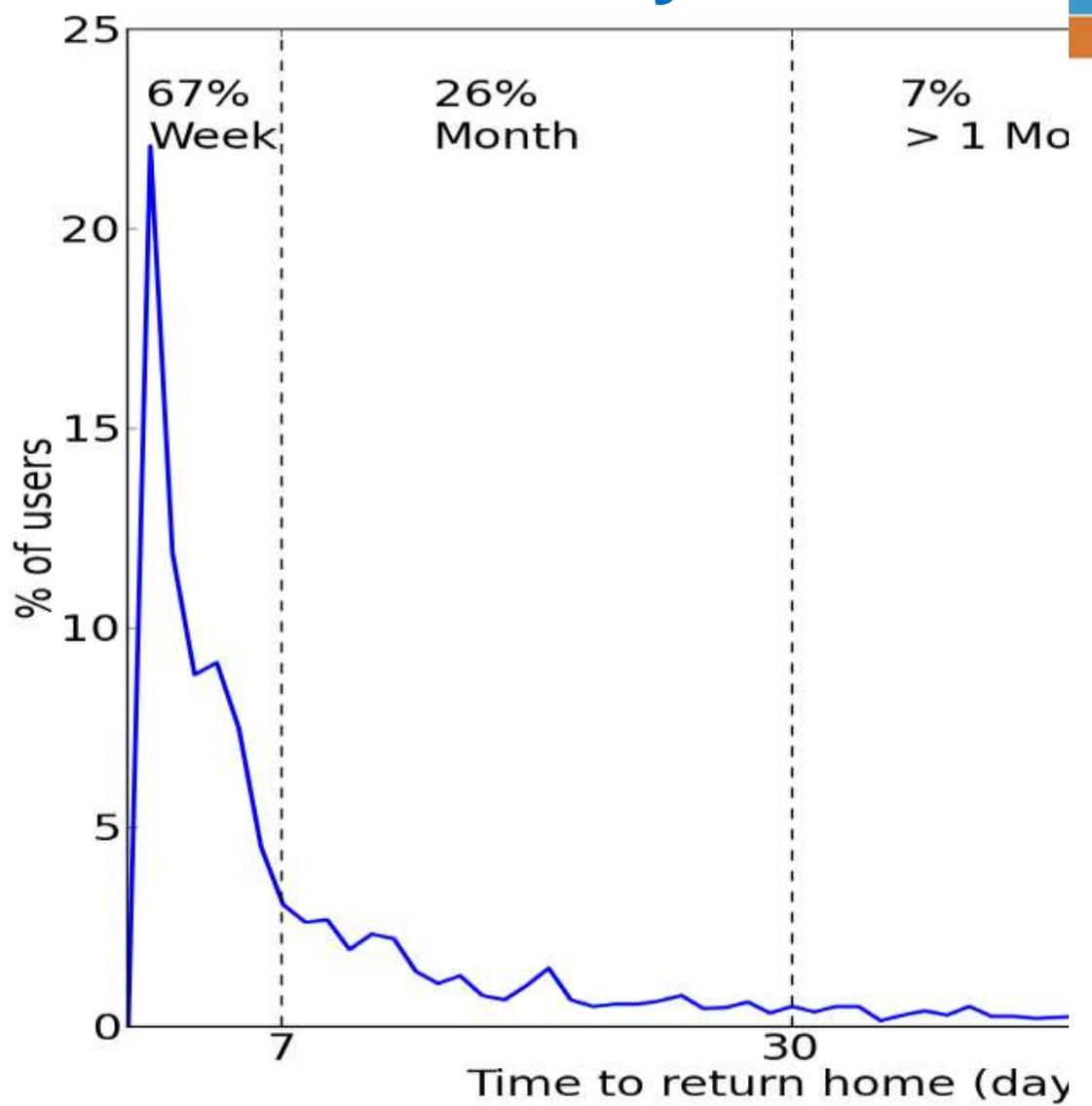# Open Data: Movement between cities during Hurricane IDA 2009

Road from Coatzacoalco to VillaHermosa (VER)

Here we see in detail from/to where people displaced, involving the mentioned towers. Red means leave, pale yellow means arrive.

20Km

Movement in neighbor city (VillaHermosa) during Hurricane IDA 2009

VillaHermosa

Here we see in detail from/to where people displaced, involving the Tabasco capital city towers. We see a lot of displacements from the surrounding towns Red means leave, pale yellow means arrive.

How far they move?

When do they return?

retos.datos.gob.mx

Links:

- Open Data Policy :
  http://www.dof.gob.mx/nota_detalle.php?codigo=5382838&fecha=20/02/2015

- Open Data Public Website:
  http://datos.gob.mx

- Open Data Platform:
  http://adela.datos.gob.mx/

- Open Data and Big Data for a Better Mexico (Retos Mexico)
  http://retos.datos.gob.mx/retos