Uncovering the spatial structure of urban mobility networks
Methods and applications to Spanish cities

T. LOUAIL

M. LENORMAND, JJ. RAMASCO, M. BARTHELEMY
M. PICORNELL, O. GARCIA-CANTU, R. HERRANZ, E. FRIAS-MARTINEZ
J. MURILLO ARIAS, M. SAN MIGUEL
• Mobility Data
• Types of mobility networks (regular daily flows & individual trajectories)
• Extracting an expressive signature of the structure of large mobility networks
  • Method
  • Application to the comparison of Spanish cities
• Discussion
Why?

Mobility is the vector of other processes (epidemiology, city growth, etc.)

Many phenomena we don't properly understand (mobility and social networks, individual strategies of space exploration, invisible networks of familiar networks, etc.)

Mobility is interesting by itself!
Why?

Mobility is the vector of other processes (epidemiology, city growth, etc.)

Many phenomena we don't properly understand
- (mobility and social networks, individual strategies of space exploration, invisible networks of familiar networks, etc.)

Mobility is interesting by itself!

What changed?

Data data data data data data data data data data data...

... suggested new questions/topics

... bring new actors

Complex systems/networks approach
New sources of transport and mobility data

Collective, voluntary geographic information, public data, e.g. infrastructure networks (OSM), transport timetables, locations of public infrastructures, etc.
Collective, voluntary geographic information, public data, e.g. infrastructure networks (OSM), transport timetables, locations of public infrastructures, etc.

Geotagged (meta)data passively generated by individuals when using their mobile devices (mobile phones, social networking apps, transport smart cards, car GPS, credit cards, etc.)
1st type of mobility network
Origin-Destination (OD) matrices

Classic object to study commuting patterns

F(i,j) is the number of individuals that live in i and work in j

Provides the complete information on the commuting flows
Simple hypothesis used to determine the home & work locations of each individual

**Home**

is assimilated to
the most frequent location between 8 p.m. and 8 a.m.

**Work/School/Main activity**

is assimilated to
the most frequent location between 8 a.m. and 8 p.m.
Extracting O-D from CDR data

Sorted by frequency

<table>
<thead>
<tr>
<th>Location</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.13</td>
</tr>
<tr>
<td>4</td>
<td>0.04</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.73</td>
</tr>
<tr>
<td>2</td>
<td>0.15</td>
</tr>
<tr>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Hp > Prop  &  Wp > Prop

(a) Number of users (x 10^4)
(b) Proportion of intra-zonal flows

http://ifisc.uib-csic.es
Cross-checking different sources of mobility information

→ (Lenormand et al. Plos One 2014)
How to extract a **simple and expressive footprint** of a large mobility network?

Can we build a typology of cities based on the spatial structure of their commuting patterns?

**Inspiration:** schematic/simple figures of commuting flows in idealized city forms

(from Bertaud & Malpezzi 2003)
1. Determine **residential hotspots** and **employment hotspots**

2. Separate **4 categories of flows**
   - **Integrated**: from residential hotspots to work hotspots
   - **Convergent**: from elsewhere to work hotspots
   - **Divergent**: from residential hotspots to elsewhere
   - **Random**: from elsewhere to elsewhere
Method

1. \[ OD = \begin{bmatrix} C_{1,1} & \cdots & C_{1,n} \\ \vdots & \ddots & \vdots \\ C_{n,1} & \cdots & C_{n,n} \end{bmatrix} \sum_{j=1..n} C_{i,j} = C_{i}^{out} \]
\[ \sum_{i=1..n} C_{i,j} = C_{j}^{in} \]

2. \[ C_{i}^{out} = (C_{1}^{out}, \ldots, C_{n}^{out}) \]
\[ C_{i}^{in} = (C_{1}^{in}, \ldots, C_{n}^{in}) \]

Determine hotspots

3. \[ OD = \begin{bmatrix} \frac{\sum_{i=1..m} C_{i,j} / \sum C_{i,j}}{\sum_{j=1..p} C_{i,j} / \sum C_{i,j}} & \frac{\sum_{i=1..m} C_{i,j} / \sum C_{i,j}}{\sum_{j=p+1..n} C_{i,j} / \sum C_{i,j}} \\ \frac{\sum_{i=m+1..n} C_{i,j} / \sum C_{i,j}}{\sum_{j=1..p} C_{i,j} / \sum C_{i,j}} & \frac{\sum_{i=m+1..n} C_{i,j} / \sum C_{i,j}}{\sum_{j=p+1..n} C_{i,j} / \sum C_{i,j}} \end{bmatrix} \]

\( \rho \) residential (out) hotspots

\[ OD = \begin{bmatrix} I & D \\ C & R \end{bmatrix} \]

with \( I + D + C + R = 1 \)
Comparing the commuting structure of 31 Spanish cities

Harmonized functional definition (same for all cities)
Weight of Integrated flows decreases when population size $P$ increases, in favor of Random flows.

Weights of Divergent and Convergent flows almost constant whatever the city size.
ICDR values vs. P

Weight of Integrated flows decreases when population size $P$ increases, in favor of Random flows.

Weights of Divergent and Convergent flows almost constant whatever the city size.

Same values ranked by decreasing I

I and R alone seem sufficient to classify cities.
Null model: ICDR values vs. P

ICDR values: null model vs. data

Z-scores
- For all types of flows distance increases with population size
- Convergent Flows (C) are the longest and the most penalized when $P$ increases
- As cities grow, the spatial organisation of commuting flows is more and more rationale (i.e. advantageous when compared to random flows)
- Some small cities display a value less than 1, indicating the lesser importance of space at shorter scales
Biggest cities are clustered together

Robust with different sizes of the aggregation grid
<table>
<thead>
<tr>
<th>Cluster</th>
<th>Average population of cities in cluster</th>
<th>I</th>
<th>R</th>
<th>D</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cordoba, Gijon, Vitoria, etc.</td>
<td>255,330</td>
<td>0.43</td>
<td>0.27</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td>Zaragoza, Malaga, etc.</td>
<td>392,970</td>
<td>0.37</td>
<td>0.36</td>
<td>0.15</td>
<td>0.13</td>
</tr>
<tr>
<td>Valencia, Sevilla, etc.</td>
<td>732,992</td>
<td>0.31</td>
<td>0.41</td>
<td>0.16</td>
<td>0.13</td>
</tr>
<tr>
<td>Madrid</td>
<td>2,463,551</td>
<td>0.25</td>
<td>0.46</td>
<td>0.17</td>
<td>0.12</td>
</tr>
</tbody>
</table>

4 clusters of cities
Tap at **start** and **end** of train journeys

Accepted at 695 Underground and rail stations, and on thousands of buses

A week of data  (week starting 02/07/2012)

For each 15 minutes time-slice, from Monday to Sunday, An O-D matrix containing the numbers of commuters:
• 597 stations
• 22,938,721 taps
• 18,858,038 Mon-Fri
• 4,080,683 Sat-Sun
Aggregated stations used in our analysis (597 stations)
Structure of commuting in the London train+subway network

ICDR values are almost constant during the week
They constitute a signature of the mobility structure in the city
ICDR is a versatile method to extract a simple footprint of the structure of large, weighted and directed networks.

The method is independent from any particular definition of hotspots and from the spatial resolution at which data are available.

Application on 31 Spanish cities:
- As the population size increases, I decreases while R increases.
- The comparison with a reasonable null model revealed that commuting flows are more organized in the largest cities.
- Convergent flows are the more penalized when the population size increases.

Application to London proved that ICDR values are a signature of the mobility structure in the city.