Big Data and data-driven modeling for transport systems

José Javier Ramasco
jramasco@ifisc.uib-csic.es
Motivation

Bibliography

Roads

Boats & Cargo

Airtraffic

Schedules and delay propagation

City mobility

Conclusions
- 40,000 years
- 5,000 years
- 2,000 years
- 700 years
- 100 years

Spanish flu 1918
These are some references to works in the area:

- The science of cities and regions, Alan Wilson, Springer (2012).
Modeling urban street patterns, M. Barthelemy & A. Flamini, PRL 100, 138702 (2008).
Modeling urban street patterns, M. Barthelemy & A. Flamini, PRL 100, 138702 (2008).
Modeling urban street patterns, M. Barthelemy & A. Flamini, PRL 100, 138702 (2008).
- Total cost of flight delay in US in 2007 was **41B** dollars.
- Rich transport dynamics.
- Cascading failure.

(Air transportation)

(\(http://www.transtats.bts.gov/\))

(\(http://www.eurocontrol.int\))
Database:

- Airline On-Time Performance Data (www.bts.gov)
  - Schedule & actual departure (arrival) times
  - Origin & destination airports
  - Airline id
  - Tail number

- 2010 flights:
  - 6,450,129 flights (74%)
  - 18 carriers
  - 305 airports

Network:

- Nodes: airports
- Edges: direct flights between airports
- Node attributes: average delay per flight
Clusters:

- Formed by airports in problem
  - average delay per flight > 29 min

- Must be connected (flight route between them)

- A group of airports connected by flights that their average delay is higher than 29 minutes

<table>
<thead>
<tr>
<th>Date</th>
<th>Daily departure delay [min]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-04-19</td>
<td>16.9</td>
</tr>
<tr>
<td>2010-10-09</td>
<td>17.2</td>
</tr>
<tr>
<td>2010-11-11</td>
<td>17.3</td>
</tr>
<tr>
<td>2010-04-14</td>
<td>17.6</td>
</tr>
<tr>
<td>2010-10-08</td>
<td>18.0</td>
</tr>
<tr>
<td>2010-09-11</td>
<td>18.4</td>
</tr>
<tr>
<td>2010-04-15</td>
<td>18.4</td>
</tr>
<tr>
<td>2010-10-13</td>
<td>18.5</td>
</tr>
<tr>
<td>2010-04-17</td>
<td>18.5</td>
</tr>
<tr>
<td>2010-11-10</td>
<td>18.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Daily departure delay [min]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-10-27</td>
<td>54.3</td>
</tr>
<tr>
<td>2010-03-12</td>
<td>53.0</td>
</tr>
<tr>
<td>2010-12-12</td>
<td>51.9</td>
</tr>
<tr>
<td>2010-01-24</td>
<td>49.8</td>
</tr>
<tr>
<td>2010-02-24</td>
<td>49.1</td>
</tr>
<tr>
<td>2010-05-31</td>
<td>46.8</td>
</tr>
<tr>
<td>2010-05-21</td>
<td>45.5</td>
</tr>
<tr>
<td>2010-05-14</td>
<td>44.6</td>
</tr>
<tr>
<td>2010-06-23</td>
<td>44.6</td>
</tr>
<tr>
<td>2010-07-13</td>
<td>44.3</td>
</tr>
</tbody>
</table>
Clusters:

- Formed by airports in problem
  - average delay per flight > 29 min
- Must be connected (flight route between them)
- A group of airports connected by flights that their average delay is higher than 29 minutes

- April 19, 2010
- Average delay per delayed flight:
  - 16.9 min
Clusters:

* Formed by airports in problem
  - average delay per flight > 29 min

* Must be connected (flight route between them)

* A group of airports connected by flights that their average delay is higher than 29 minutes

- March 9, 2010
- Average delay per delayed flight:
  - 25.7 min
Clusters:

• Formed by airports in problem
  > average delay per flight > 29 min

• Must be connected (flight route between them)

• A group of airports connected by flights that their average delay is higher than 29 minutes

March 12, 2010
• Average delay per delayed flight:
  > 53.2 min
Great variety
Consecutive days are very different each other.
**Flight rotation (same tail number)**

\[ T^{\text{act.d}}_{ij}(p_{ij}) = \max[T^{\text{sch.d}}_{ij}(p_{ij}); T^{\text{act.a}}_{ij}(p_{ij}) + T_s] \]

**Flight connectivity (different tail number)**

\[ T^{\text{act.d}}_{ij}(p_{ij}) = \max[T^{\text{sch.d}}_{ij}(p_{ij}); T^{\text{act.a}}_{ij}(p_{ij}) + T_s; \max[T^{\text{act.a}}_{ij}(p_{i'j})], \forall i' \neq i] \]

**Initial Conditions**

- **From the data...**
  - Known → when, where and the departure delay for the first flight of the sequence.

- **Random initial conditions...**
  - Fixed initial delay (min)
  - % of initially delayed planes

**Airport Congestion**

![Graph showing SAAR over time](image)
Delay propagation dynamics
Data and model comparison for March 12 and April 19
• With random initial conditions…

- Each day is potentially a bad day, if some initial conditions are met.
- Flight connectivity is a key factor for the rise of congestion in the network.
- Sensitivity to initial conditions.
From mobile phone data to the spatial structure of cities, T. Louail et al, arXiv: 1401.4540
31 urban areas with more than 200,000 inhabitants
55 days in fall 2009
On average 2% population (max 5%).

From mobile phone data to the spatial structure of cities, T. Louail et al, arXiv: 1401.4540
From mobile phone data to the spatial structure of cities, T. Louail et al, arXiv: 1401.4540
From mobile phone data to the spatial structure of cities, T. Louail et al, arXiv: 1401.4540
From mobile phone data to the spatial structure of cities, T. Louail et al, arXiv: 1401.4540
From mobile phone data to the spatial structure of cities, T. Louail et al, arXiv: 1401.4540
From mobile phone data to the spatial structure of cities, T. Louail et al, arXiv: 1401.4540
Concentration areas

MATSIM + Phone users agenda

1. Initial conditions
2. Execution
3. Score
4. Analysis
5. Replan
Calibration + Validation
● First statistics

**Results**

![Graph: Inner ring](image)

![Graph: Peripheral ring](image)

- Baseline scenario
- All day 5 € toll
- All day 2 € toll
- All day 10 € toll
Large scale

Modal split

Baseline scenario

Inner ring 10€ toll

Peripheral ring 10€ toll

24.5%  
23.9%  
51.6%  

23.0%  
24.8%  
52.2%  

22.5%  
25.2%  
52.3%  

Private transport
Not motorised transport
Public transport
Large scale

Results

- **Start and end inside the ring**
  - Base scenario: Inner ring:
    - 9.2% (21.4% Private transport, 69.4% Not motorised transport)
    - 8.3% (21.7% Private transport, 70.0% Not motorised transport)
    - 11.0% (26.4% Private transport, 62.6% Not motorised transport)
    - 10.2% (26.8% Private transport, 63.0% Not motorised transport)

- **Entering or exiting the ring**
  - Base scenario: Inner ring:
    - 23.4% (44.7% Private transport, 31.9% Not motorised transport)
    - 19.2% (47.8% Private transport, 33.0% Not motorised transport)
    - 37.6% (37.9% Private transport, 24.5% Not motorised transport)
    - 31.0% (42.8% Private transport, 26.2% Not motorised transport)

- **Start and end outside the ring: Peripheral ring**
  - Base scenario: Inner ring:
    - 33.1% (15.8% Private transport, 51.1% Not motorised transport)
    - 32.5% (16.1% Private transport, 51.4% Not motorised transport)
    - 42.0% (7.8% Private transport, 50.2% Not motorised transport)
    - 41.2% (8.2% Private transport, 50.6% Not motorised transport)
- Large scale
Trips

a)
b)
c)
d)

- 0 - 3649
- 3649 - 7298
- 7298 - 10946
- 10946 - 14595
- 14595 - 18244
- 18244 - 29190

Results
Results

● Travel times

Average travel time per resident:
- 19.00 - 21.25 min
- 21.25 - 23.50 min
- 23.50 - 25.75 min
- 25.75 - 28.00 min
- 28.00 - 30.25 min
- 30.25 - 33 min

a) [Map]
b) [Map]
c) [Map]
d) [Map]
e) [Map]
- Public transport
• Public transport
● Public transport
At the level of neighborhood
1) The availability of data has opened a new era in mobility research

2) To extract the maximum potential, one needs imagination for the analysis but mainly **models**

3) The applications are ready to explore, from technical and the deployment of new technologies to the socioeconomic implications and to health and pollution.
THANK YOU
for your attention