Improving Urban Mobility
Transit Systems, New Technologies & Smart Cities

Michael Batty

m.batty@ucl.ac.uk
@jmichaelbatty

http://www.complexcity.info/
http://www.spatialcomplexity.info/
Outline

• Some Ideas about Smart Cities and Big Data
• Real-Time Streaming: The Oyster Card Data Set
• Learning about Mobility from the Data
  Variabilities – Heterogeneity and Travel Profiles
  Disruptions – Signal Failures, Stalled Trains
  Variable Locational Dynamics of Demand
• Related Real-Time Data: Bikes, Social Media
• What Can We Learn: The Limits to Big Data
Some Ideas about Smart Cities and Big Data

The spreading out of computers into public places & the built environment and all their consequences

The Real Built & Social Environment

Computers & Sensors

routine models

real-time streamed data

BIG DATA

Our Theories of the City

Information Exhaust

Information & Control

Centre for Advanced Spatial Analysis
• The way we access the smart city is through technologies that let us generate and use data and its useful equivalent – information (data) is key.

• Access through mobile and fixed devices like phones, smart cards, through fixed sensors.

• These usually complement rather than substitute for data which we collected and used in the past. This data still essential and highly relevant.

• This has introduced time into our thinking – in the past most urban planning for future cities was timeless – garden cities, new towns, master plans.

• This is all part and parcel of increasing complexity; more time scales, more opportunities, more diversity.
Real-Time Streaming: The Oyster Card Data Set

Tap at **start** and **end** of train journeys
Tap at **start only** on buses

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

**Many Variants of the Data Sets**

**991 million** Oyster Card taps over Summer 2012 – this is big data
Tube, Overground and National Rail Networks in London where Oyster cards can be used
OYSTER GIVES UP PEARLS

How studying millions of Oyster Card journeys reveals London’s ‘polycentres’

Researchers from UCL have analysed millions of Oyster Card journeys in a bid to understand how, why and where we travel in London.

Professor Michael Batty (UCL Centre for Advanced Spatial Analysis) and Dr Song Hang (UCL Management Science and Innovation) applied the techniques of statistical physics to their mountain of raw data.

The pair joined forces with a computational social scientist and a physicist, both based in Paris, to explore patterns of commuting by tube into central London.

They used Transport for London’s database of 11 million records taken over one week from the Oyster Card electronic ticketing system.
And how can we make sense of this

http://www.simulacra.info/
Variabilities - Heterogeneity and Travel Profiles

First we will look at some of the data and how it varies in terms of the diurnal flows usually morning and evening peaks, with a small blip (peak) around 10pm at night.
Oyster Card Data – interpreting urban structure, multitrips, etc.
Particular Events: Weekdays, Saturdays and Sundays

Nightlife

Entry at Camden Town (10 Mn. Intervals)

Weekday
Saturday
Sunday

Events

Entry at Arsenal (10 Mn. Intervals)

Weekday
Saturday
Sunday

Work

Entry at Bank (10 Mn. Intervals)

Weekday
Saturday
Sunday

Tourism?

Entry at Bayswater (10 Mn. Intervals)

Weekday
Saturday
Sunday
Comparing Variability for different time intervals for Three World Cities: London, Beijing and Singapore

<table>
<thead>
<tr>
<th>Day</th>
<th>London</th>
<th>Singapore</th>
<th>Beijing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>3,457,234</td>
<td>2,208,173</td>
<td>4,577,500</td>
</tr>
<tr>
<td>Tuesday</td>
<td>3,621,983</td>
<td>2,250,597</td>
<td>4,421,737</td>
</tr>
<tr>
<td>Wednesday</td>
<td>3,677,807</td>
<td>2,277,850</td>
<td>4,564,335</td>
</tr>
<tr>
<td>Thursday</td>
<td>3,667,126</td>
<td>2,276,408</td>
<td>4,582,144</td>
</tr>
<tr>
<td>Friday</td>
<td>3,762,336</td>
<td>2,409,600</td>
<td>4,880,267</td>
</tr>
<tr>
<td>Number of stations (1)</td>
<td>400</td>
<td>130</td>
<td>233</td>
</tr>
<tr>
<td>Number of tube line</td>
<td>13</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>Area (2)</td>
<td>1,572 km²</td>
<td>718.3 km²</td>
<td>2267 km²</td>
</tr>
<tr>
<td>Total population (3)</td>
<td>8.63 million</td>
<td>5.3 million</td>
<td>21.15 million</td>
</tr>
<tr>
<td>Ridership of Metro</td>
<td>20%</td>
<td>35%</td>
<td>21%</td>
</tr>
<tr>
<td>Length of metro lines</td>
<td>402km</td>
<td>182km (MRT+LRT)</td>
<td>465 km</td>
</tr>
</tbody>
</table>

(1) Number of stations is the number of stations with smart-card records generated.
(2) The area of Beijing only counts the area enclosed by the 6th ring road for a fair comparison.
From 1 minute intervals to the whole day

Variability measure of temporal patterns (tap in time)
Comparing Variability for different time Intervals over the day

Figure 1. Variability of regularity in the trip matrix over time.

Note: Each box plot shows the variability of 400 stations over time measured at different temporal scales. Overall, eight subplots give a similar trend where lower variability appears during peak hours (around 9 am in the morning and 6pm in the evening). More details can be captured as differences of variability between each time unit are magnified as we decrease the temporal scale from 12h to 4 minutes.
Comparing Variability for different time intervals for Three World Cities: London, Beijing and Singapore
Maps of Underground and Rail stations in London visualised by the proportion of regular trips originating at each location ending at each location starting and ending at each location.
Disruptions - Signal Failures, Stalled Trains

• We will look at three disruptions - the Circle and District Lines which had a 4 hour stoppage on July 19th 2012

• And a Bus Strike in East London and how this shows up in the data

• And typical pattern of delay on all modes visualised for Greater London
Circle and District line part closure
From Edgware Road to Aldgate/Aldgate East
19th July 2012
07:49 to 12:04

1234022 Oyster Cards with regular pattern during disrupted time period travelled
Increased Travel Time

Greater than 2SD above mean increase on usual travel time for that Oyster Card

Size equal to proportion of users that regularly travel from station during time period, and travelled that during disruption
The Public Transport System in Terms of Vehicle Flows

- **Tube**
  - Trackmet Processor
  - Feature Detector
  - Outputs
    - Status: Disrupted line segments
    - Stations: Higher than mean wait

- **Bus**
  - Countdown API
  - Countdown Processor
  - Feature Detector
  - Outputs
    - Buses: Estimated positions
    - Bus Stops: Higher than mean wait

- **Heavy Rail**
  - Rail Processor
  - Outputs
    - Trains: Late running services

- **Network Rail**
  - Aggregator
  - Calibration

- **Trackmet API**
  - Tube Status
  - Calibration
Delays from Tube, National Rail and Bus Fused

Key
- National Rail more than 5 minutes late
- Tube stations showing a wait time 15% above expected
- Bus stops showing a wait time 20% above expected
- Tube delays from the TfL status feed are also plotted as lines

Tuesday 9 October 10:30
We are currently using information theory to figure out how much information from trips is transmitted from station to station through time by working out how many passengers are in stations or on trains in stations over time. We are using the concept of **transfer entropy** to do this. I don’t have time to say much about this but here is a picture about this for one line:

\[
T_{yx} = \sum_{t=1} p(y_t+1, y_t, x_t) \log \frac{p(y_{t+1}|y_t, x_t)}{p(y_{t+1}|y_t)}
\]
Second we are working with the Oyster data again with Melanie Bosredon in our group and Marc Barthelemy in Paris on extracting clusters from the travel data using a new method of defining intensity. I will show this as a simple movie of origin and destination intensities as they change over time of day.
Related Real-Time Data: Bikes, Social Media

A lot of data is now coming online for travel and one of our group Oliver O’Brien has some 97 bike schemes worldwide for which he has online data in real time - Bikes Data – 4200 bikes, started Nov 2010, all the data – everything – all trips, all times, all stations/docks
More Analysis

- London
- Graph shows number of bikes available to hire
- Effect of rain
  - Using the CASA weather station
- Effect of the tube strikes

Bike-o-Meter

casa.ucl.ac.uk/bom

- Tweet-o-Meter for bikes
  - Steven Gray (@frogo)
  - Using Google Gauges
- See the real life Tweet-o-Meters at the new British Library “Growing Knowledge” exhibition
  - Should be easy to hack to show the Bike-o-Meters instead 😊
The Website: Real Time Visualisation of Origins and Destinations Activity

http://bikes.oobrien.com/london/
What Can We Learn: The Limits to Big Data

We need to add geo-demographics to this data – how – we barely have any possibility of doing this because of confidentiality.

We only have a difference between young and old in terms of the card data.

Chen Zhong my post doc has done a lot of work on this relating to extracting such data from related data sets producing synthetic results – our paper in IJGIS.

http://dx.doi.org/10.1080/13658816.2014.914521

Detecting the dynamics of urban structure through spatial network analysis

Chen Zhong\textsuperscript{a*}, Stefan Müller Arisona\textsuperscript{a,b}, Xianfeng Huang\textsuperscript{c}, Michael Batty\textsuperscript{d} and Gerhard Schmitt\textsuperscript{a}
References


Thanks

http://www.spatialcomplexity.info/
http://www.complexcity.info/
http://blogs.casa.ucl.ac.uk/

Acknowledgements