Big Data, Little Data, Real-Time Streaming & Smart Cities

Michael Batty
Centre for Advanced Spatial Analysis
CASA-UCL

m.batty@ucl.ac.uk
@jmichaelbatty
Themes in My Lecture

• Future Cities, Smart Cities, What Are They?
• Cities are Getting More Complex. We Have to Run to Stand Still
• New Developments in IT Are Changing the Focus onto the Routine and the Short Term, More So than Ever in the Past
• New Forms of Data from the Smart City are emerging very fast: Big Data (and Little Data too)
• Three Sets of Examples: All from London but I Hope They Resonate with Urban Issues Here in Australia
What Are Future Cities? Smart Cities?

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

The Real Built & Social Environment

Computers & Sensors

routine models

strategic models

real-time streamed data

BIG DATA

Our Theories of the City
• The way we access the smart city is through technologies that let us generate and use data and its useful equivalent – information which is key
• Access through mobile and fixed devices like phones, smart cards, through fixed sensors which record transactions and so on
• These usually complement rather than substitute for data which we collected and used in the past
• This has introduced time into our thinking – in the past most urban planning for future cities was timeless – think of garden cities, new towns, master plans
• This is all part and parcel of increasing complexity; more time scales, more opportunities, more diversity ……
How Big is Data? Big Can Be Small & Small Big

- Data is big with respect to its volume. I know there are other definitions – velocity, variety etc. but to me, data is big if it requires large use of computer memory implying volume.
- In cities, data usually implies numbers of locations and their attributes but locations imply interactions.
- Thus data are relations between locations and in essence if we have \( n \) locations, we have \( n^2 \) interactions. Thus small data can become big. EG:
Examples: Dublin 1837, Ireland 1888, London 1953

The Oldest Flow Map
Posted on June 27, 2011 by Michael Batty

,... according to the great cartographer Arthur Robinson, the two maps of traffic between Dublin and the rest of Ireland by Lt. Harness of the British Army in 1837, are the oldest. Mapped for the Irish Railway Commissioners prior to construction of the railway. Are these actually the first?
Examples: Dublin 1837, Ireland 1888, London 1955

Harness, 1837
Ravenstein 1888
Big Data Problems have been around longer than you think

The Strata Conference is in town and one presentation that caught my eye was titled The Great Railway Caper: Big Data in 1955

https://www.youtube.com/watch?v=pcBJfkE5UwU
Locations and Interactions: Flow Systems in Cities

Elsewhere I have argued, in my recent book, that we should treat cities as flow systems – as networks as this is the real focus – it has been for a long time in my world of transport and land use and we have always been up against the problem of big data. So let me begin my illustration of this dilemma and how we are thinking about it with some problems that have very small data. Problems of spatial interaction where our numbers of locations is small < 100, ~ 50
Understanding and Visualising Flows

An early model circa 1967-8 Central and NE Lancs

\[ n^2 = 33^2 = 1089, \text{ not so big but hard to visualise} \]
\[ n^2 = 633^2 = 400,689, \text{ bigger but impossible to visualise} \]
Even our statistics breaks down when we get large numbers like over several thousand as you can see on the left and above right for 400K data points where the pattern is highly convoluted. This is from a gravity model.
Now what happens when we really do scale up to the level of MSOAs of which there are 7201 in the UK – do we partition and argue we don’t need to scale up to $n^2 = 7201^2 = 51,854,401$. Do we simplify the spatial scale?

Circa 52 million points is an issue but our models run in a matter of seconds but that is a lot of data to store – ok it is sparse but sparsity isn’t structured so we can’t easily partition and in any case we want to compute any possible flows between central London say and Newcastle. Here is the problem scaled up and this is what we are grappling with at present.
QUANT
Simulating the Impacts of Large Scale Change in the UK

http://quant.casa.ucl.ac.uk/
My Three Examples

• New Ways of Managing Transportation: Public Transport in London – from real time data
• Climate Change in London: The Major Issue is Sea Level Rise – using the old science in a public context
• New Forms of Data: Crowd-Sourcing, Social Media – the potential of new data

There is good and bad in all these examples and they suggest to us that we need a concerted effort to explore how we can best respond to using these new technologies to reinforce the good and squeeze out the bad.

It will not be easy, it never is.
Example 1: New Ways of Managing Transportation
Smart Card Data

Oyster Card Taps

- 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. **1.053 billion** Oyster Card taps over July to September 201
- 762 million OD trips, 291 million UG and rail trips, 11.5m Oyster Cards. No routing data
Researchers from UCL have analyzed 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 Seong Kang (UCL Management Science and Innovation) applied the techniques of statistical physics to the 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/
Particular Events: Weekdays, Saturdays and Sundays

- **Nightlife**: Entry at Camden Town (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

- **Events**: Entry at Arsenal (10 Mn. Intervals)
  - Weekday
  - Saturday
  - Sunday

Time of Day: 2am, 4am, 6am, 8am, 10am, 12pm, 2pm, 4pm, 6pm, 8pm, 10pm, 12am, 2am, 4am

Number of Events: 0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000
Oyster Card Data – interpreting urban structure, multitrips
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.
Delays from Tube, National Rail and Bus Fused

Key

- Red: National Rail more than 5 minutes late
- Blue: Tube stations showing a wait time 15% above expected
- Green: Bus stops showing a wait time 20% above expected

Tube delays from the TfL status feed are also plotted as lines.
WE are working with the Oyster data 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.

Number of hotspots (stations) vs. time
Example 2: Climate Change in London: Flooding
The Thames Barrier built from 1978 to 1984 in operation, likely to be ineffective now by 2030 due to new predictions of sea level rise.
There are many stages to figuring out what the future might be in this region – sea level rise we think (from IPCC) will be about 1 metre by 2100

Enough to cause a major problem – but of course on the way to 2100 many things will be put in place

Central London will not flood but to give you an idea of the potential for flooding let me show a movie very quickly which I will come back to later.
Flooding from our 3D Virtual London Model
Shifts in Traffic Accessibility if all Bridges across the Thames are Inoperable as far West as Hammersmith
We have developed in fact an integrated assessment which is built around various models and groups of modellers. Let me show some of this.
Example 2: Exploring the Impact of Large Infrastructure Projects: New Airports/Jobs
Interactive Input of Changes to Origin-Destination Crow-Fly Distances

Point your mouse at the two zones whose link you wish to change and click.

Use slider to input percentage change for Zone 5 to 219.

Old Distance from 5 to 219 is 35
New Distance is 7

Updated Distances So Far
Infrastructure in London: Olympics, 3rd Runway, Cross-rail,
Example 3: Better Data: A New Understanding
A new credit crunch survey started in October and currently has 3,802 responses.
http://www.maptube.org/creditcrunch/
Manchester Congestion Charge

15,902 responses
October to December 2008

I am not affected by these changes, 9
- Drive and pay the charge, 13
- Drive at different times, 13
- Use public transport/motorbike/bicycle, 1
- Work or shop elsewhere, 45

9.25% 13% 13% 6.93%
Conclusions, Reflections

I would like to stress increasing complexity in cities and contemporary society is being driven by new information technologies.

Data from the real time city is changing our focus. This data is BIG data but Big data can also be conventional and not generated in real time. Big can actually be little and vice versa.

Technology is never neutral – it is always value laden and we do have choices and the biggest problems tend to be tangential to new IT.
Thank You

Michael Batty
Centre for Advanced Spatial Analysis
CASA-UCL

m.batty@ucl.ac.uk
@jmichaelbatty