Session 3: Big Data, Network Data, Measuring Disruption in the 24 Hour City

Michael Batty

http://www.spatialcomplexity.info/

July 7, 2018
Let me point you at www.spatialcomplexity.info where I have now added the material under the menu item Shanghai.

There you will find

- The Journal Publication: Big Data and the City
- Shanghai Smart Cities Lecture 1, 2, 3, and 4
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
- Quality of Data
- What Can We Use It For
- Missing Data and Noise
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 Kang (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/
This of course was the thing that Lt Henry Hamess did in Dublin in 1837 and what Minard et al. did a little later. In our LUTI models, this is an enormous problem as the scale of this assignment to networks is different.
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

Events

Work

Tourism?
Comparing Variability for different time intervals for Three World Cities: London, Beijing and Singapore

Table 1. Summary statistics of one-week of smart-card data (metro trips only)

<table>
<thead>
<tr>
<th></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 (MRT+LRT)</td>
<td>402 km</td>
<td>182 km</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)

factor(city)
- Beijing
- London
- Singapore

measures with increasing time intervals (minutes)
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 - Routine Analysis of Daily Events

- Behaviours vary across network
- Different areas of network more resilient to disruption, due to available infrastructure and individual ability to change
- But areas of network are naturally closely tied through established usage patterns
- Individual-based analyses provide insight into behaviours underlying macroscopic flows

We will look at several kinds of disruption
• First hypothetical disruptions simply by examining breaks in the network

• Then an example of 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
Resilience and Disruption

Examining Network Disruption: Station and Line Closures
Representing the Tube Network

We use standard graph algebra to represent the network where we define three indices of centrality:

Degrees of the graph:
\[
\sigma_i = \sum_j a_{ij} \quad \quad \sigma = \sum_i \sigma_i = \sum_j \sigma_j = \sum_i \sum_j a_{ij}
\]

Betweenness Centrality:
\[
C_k = \sum_i \sum_j \frac{\sigma_{ikj}}{\sigma_{ij}}
\]

Closeness Centrality:
\[
L_i = KD_i^{-1} = K \left( \sum_j \, d_{ij} \right)^{-1}
\]
Representing Flows

Trip Volume
\[ T_i = \sum_j T_{ij} \]

Entries and Exits
\[ T_j = \sum_i T_{ij} \]

Changes in Trip Volumes
\[ \Delta_i = T_i - T_i' \]
\[ \Delta_j = T_j - T_j' \]

Weighted Betweenness
\[ p_{ijk} = \frac{\sigma_{ikj}}{\sigma_{ij}} = \sum_\ell \frac{\sigma_{ikj}}{\sigma_{i\ell j}} \quad , \quad \sum_k p_{ikj} = 1 \]

Centrality
\[ \tilde{C}_k = \sum_i \sum_j T_{ij} p_{ikj} = \sum_i \sum_j T_{ij} \frac{\sigma_{ikj}}{\sigma_{ij}} \]
A Preliminary Analysis (1)
The Minimal Tube Network and the Three Centrality Indices

Degree

Betweenness

Closeness
A Preliminary Analysis (2)

- Top Stations
- By Centrality

<table>
<thead>
<tr>
<th>Station</th>
<th>$d_i$</th>
<th>Station</th>
<th>$\hat{C}_i$</th>
<th>Station</th>
<th>$\hat{L}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baker Street</td>
<td>7</td>
<td>Green Park</td>
<td>16399</td>
<td>Green Park</td>
<td>2.137</td>
</tr>
<tr>
<td>King's Cross</td>
<td>7</td>
<td>Waterloo</td>
<td>15544</td>
<td>Westminster</td>
<td>2.107</td>
</tr>
<tr>
<td>Bank</td>
<td>6</td>
<td>Bank</td>
<td>15008</td>
<td>Bond Street</td>
<td>2.101</td>
</tr>
<tr>
<td>Earl's Court</td>
<td>6</td>
<td>Baker Street</td>
<td>14441</td>
<td>Oxford Circus</td>
<td>2.089</td>
</tr>
<tr>
<td>Green Park</td>
<td>6</td>
<td>Westminster</td>
<td>14139</td>
<td>Waterloo</td>
<td>2.089</td>
</tr>
<tr>
<td>Oxford Circus</td>
<td>6</td>
<td>Bond Street</td>
<td>11429</td>
<td>Bank</td>
<td>2.074</td>
</tr>
<tr>
<td>Waterloo</td>
<td>6</td>
<td>Liverpool Street</td>
<td>11186</td>
<td>Baker Street</td>
<td>2.071</td>
</tr>
<tr>
<td>Canning Town</td>
<td>5</td>
<td>Stratford</td>
<td>10814</td>
<td>Victoria</td>
<td>2.065</td>
</tr>
<tr>
<td>Liverpool Street</td>
<td>5</td>
<td>Mile End</td>
<td>10302</td>
<td>Hyde Pk Corner</td>
<td>2.053</td>
</tr>
<tr>
<td>Paddington</td>
<td>5</td>
<td>Bethnal Green</td>
<td>10017</td>
<td>Embankment</td>
<td>2.041</td>
</tr>
<tr>
<td>Shadwell</td>
<td>5</td>
<td>Finchley Road</td>
<td>8905</td>
<td>Piccadilly Circus</td>
<td>2.043</td>
</tr>
<tr>
<td>Turnham Green</td>
<td>5</td>
<td>Earl's Court</td>
<td>8706</td>
<td>St. James's Park</td>
<td>2.035</td>
</tr>
<tr>
<td>Acton Town</td>
<td>4</td>
<td>King's Cross</td>
<td>8679</td>
<td>Regents Park</td>
<td>2.032</td>
</tr>
<tr>
<td>Bond Street</td>
<td>4</td>
<td>Wembley Park</td>
<td>7968</td>
<td>King's Cross</td>
<td>2.029</td>
</tr>
<tr>
<td>Camden Town</td>
<td>4</td>
<td>South Ken</td>
<td>7182</td>
<td>Liverpool Street</td>
<td>2.026</td>
</tr>
<tr>
<td>Canada Water</td>
<td>4</td>
<td>Euston</td>
<td>7155</td>
<td>Marble Arch</td>
<td>2.026</td>
</tr>
<tr>
<td>Canary Wharf</td>
<td>4</td>
<td>Gloucester Rd</td>
<td>7042</td>
<td>Tottenham Ct. Rd</td>
<td>2.026</td>
</tr>
<tr>
<td>Embankment</td>
<td>4</td>
<td>Paddington</td>
<td>7028</td>
<td>Moorgate</td>
<td>2.020</td>
</tr>
<tr>
<td>Euston</td>
<td>4</td>
<td>Victoria</td>
<td>6558</td>
<td>Charing Cross</td>
<td>2.017</td>
</tr>
<tr>
<td>Finchley Road</td>
<td>4</td>
<td>Harrow-on-the-Hill</td>
<td>6253</td>
<td>Great Portland St</td>
<td>2.017</td>
</tr>
</tbody>
</table>
A Preliminary Analysis (3)

Closing Liverpool Street
A Preliminary Analysis (3)

Closing Green Park
Looking at Station Closures and Shifts of Travellers Using The ‘Shortest’-Paths

Although a simple station/line network may be sufficient for small cities, for ‘Mega-Cities’ such as London, New York, or Tokyo a much more detailed network is needed with interchanges measured down to the platform level. The ‘penalties’ for changing lines (and permitted Out-of-Station Interchanges) can be severe and should be included in a schematic network representation.
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
No Change: 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
Origin Changes

Locations from where individuals changed from their usual origin station
Destination Changes
Locations from where individuals changed their usual destination station
Complete Switch to Bus

Locations from where users replaced usual journey with a bus-only journey
Partial Switch to Bus

Locations from where users replaced a \textit{part} of usual journey with a bus journey.
Total Proportion Disrupted

Proportion of all 79103 users identified as being disrupted from usual patterns
Measuring Regularity
Version 2: DBSCAN Method

Oyster Card A – Origin 747

Cluster 1
Mean: 08:07; SD = 6.2
Max: 07:45; Min: 08:16
Proportion of Days = 0.8

Cluster 2
Mean: 17:34; SD = 17.1
Max: 16:58; Min: 18:57
Proportion of Days = 0.6

Oyster Card A – Destination 647

Cluster 3
Mean: 08:37; SD = 7.4
Max: 08:13; Min: 08:48
Proportion of Days = 0.8

Cluster 4
Mean: 18:06; SD = 18.1
Max: 17:26; Min: 19:28
Proportion of Days = 0.6
Measuring Regularity

Highest proportions of regular journeys by each station during AM peak period
**Measuring Regularity**

Highest proportions of regular journeys by each station during inter-peak period
The Public Transport System in Terms of Vehicle Flows

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

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

- Heavy Rail
  - Network Rail
  - Rail Processor
  - Outputs
    - Trains: Late running services
  - Aggregator
Smart Cities Lectures: The Shanghai University of Finance and Economics SUFE
Delays from Tube, National Rail and Bus Fused

Tuesday 9 October
10:30

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
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.

\[ T_{yx} = \sum_{t=1}^{\infty} p(y_{t+1}, y_t, x_t) \log \frac{p(y_{t+1}|y_t, x_t)}{p(y_{t+1}|y_t)} \]
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, now a lecturer at Kings (KCL) has done a lot of work on this relating to extracting such data from related data sets producing synthetic results – a paper in IJ GIS.

Detected the dynamics of urban structure through spatial network analysis

Chen Zhong, Stefan Müller Arisona, Xianfeng Huang, Michael Batty, and Gerhard Schmitt

http://dx.doi.org/10.1080/1365881x.2014.914521
References


Finding Pearls in London’s Oysters

JONATHAN READS, CHEN ZHONG, ED MANLEY, RICHARD MILTON and MICHAEL BATTY

Public transport is perhaps the most significant component of the contemporary smart city currently being automated using sensor technologies that generate data about human behaviour. This is largely due to the fact that the travel associated with such transport is highly ordered. Travellers move collectively in closed vehicles between fixed stops and their entry and from the system is unambiguous and easy to automate using smart cards. Flows can thus be easily calculated at specific station locations and bus stops and within the temporal intervals. How we outline work we have been doing using a remarkable big data set for public transport in Greater London generated from the Oyster Card, the smart card which has been in use for over 15 years. We explore the generic properties of the Tube and Overground rail system focusing on the scale and distribution of the flows volumes at stations, then engaging in an analysis of temporal flows that can be decomposed into various patterns using principal components analysis (PCA) which smooths out normal fluctuations and leaves a residual in which significant deviations can be tracked and explained. We then explore the heterogeneity in the data with respect to how travel behaviour varies over different times intervals and suggest how we can use these ideas to detect and manage disruptions in the system.

Big Data, Automation and Smart Transit

Automation in transit systems is the most visible sign of how the city is being transformed to enhance the travel experience and efficiency of movement (Batty et al., 2012). There are many ways of achieving this but one of the most significant is the use of smart cards for fully automatic fare collection. These smart cards usually contain the value that the consumer has agreed to load onto the card, meet stringent requirements for anonymity and security, and their use is such that by tapping in and out of an automated system, correct payments are ensured. Smart cards like this, in fact, go back to the late 1960s and rapid progress in their development was achieved in the 1970s and 1980s when they first made these appearance as phone cards in France. Different varieties of credit card were then emerging too, and by 1984 in places like Hong Kong, stored value cards for use on their new Mass Transit Railway (MTR) had been introduced. By the mid-1990s, contactless cards came onto the scene, first in Japan with the UPass card, and then in Hong Kong where they introduced the Octopus card, which was then extended to other purchasers in the local retail system.

Several other cities followed, but one of the most comprehensive rollouts was in London where, in 2003, the first cards were introduced on the underground (Tube) system. These are called ‘Oyster’ cards — partly in tribute, it would seem, to Hong Kong’s Octopus card — but the official reason is that the Oyster Card protects its ‘pearl’ — the stored value — in a ‘hard shell,’ hence the name we have used in the title to this paper. Our particular interest in these ‘pearls’ is not in their value but in the raw data that can be extracted which covers ‘where’ the owner of

PLOS ONE

Variability in Regularity: Mining Temporal Mobility Patterns in London, Singapore and Beijing Using Smart-Card Data

Chen Zhong1, Michael Batty1, Ed Manley2, Jinyi Wang3, Jing Wang4, Feng Chen1,*, Gerard Salter1,2

1Center for Advanced Spatial Analysis, University College London, London, United Kingdom. 2School of Civil and Architectural Engineering, Beijing University of Technology, Beijing, China. 3Rohto Engineering and Technology Research Center of Hatani Trans Line Group and Disaster Prevention, HATANI Shingyuu; Beijing, China. 4Future Cities Laboratory, Department of Architecture, ETH Zurich, Zurich, Switzerland.

*corresponding author

Abstract

To discover regularities in human mobility is of fundamental importance for understanding urban dynamics, and essential to city and transport planning, urban management and policymaking. Previous research has revealed universal regularities at mainly aggregated space-temporal scales but when we zoom into finer scales, considerable heterogeneity and diversity is observed instead. The fundamental question we address in this paper is what are the regularities we detect spatially, temporally, and socioeconomically. This paper thus proposes a basic measure of variability to assess the stability of such regularities focusing mainly on changes over a range of temporal scales. We demonstrate this by comparing regularities in the urban mobility patterns in three world cities, namely, London, Singapore and Beijing using data from smart-card data. The results show that variations in regularly scale as non-linear functions of the temporal resolution, which we observe over a scale from 1 minute to 24 hours thus reflecting the daily cycle of human mobility. A particularly strong increase in variability occurs up to the temporal scale of about 16 minutes in all three cities and this limits what we look for forward with respect to making short-term predictions. The degree of regularity varies in fact from city to city with walking and cycling showing higher regularity in comparison to London across all temporal scales. A detailed discussion is provided, which reveals the analysis to various characteristics of the three cities. In summary, the work contributes to a deeper understanding of regularities in patterns of travel card use variations in volumes of travels entering railway stations. It establishes a general analytical framework for comparative studies using urban mobility data, and it provides key points for the management of variability by policy-makers intent on making the travel experience more amenable.