GEOGRAPHY RESOURCE & MANAGEMENT 地理與資源管理學系

香港中文大學 The Chinese University of Hong Kong

casa

UCL University College London

Session 3

Data and Technology:
Big Data, Network Data, Measuring Disruption in the 24 Hour City

Michael Batty

http://www.spatialcomplexity.info/

April 20th 2017

m.batty@ucl.ac.uk

@jmichaelbatty

Let me point you to some reading: there is a special issue of **Built Environment** from last year and I will put my two papers up on my blog – you will be able to find these at www.spatialcomplexity.info

I can't put all the issues up on the blog because they are not open access – only my own – and I don't know if the University here gets **Built Environment** but I am sure you can retrieve these from various web sites; and here is the list of papers.

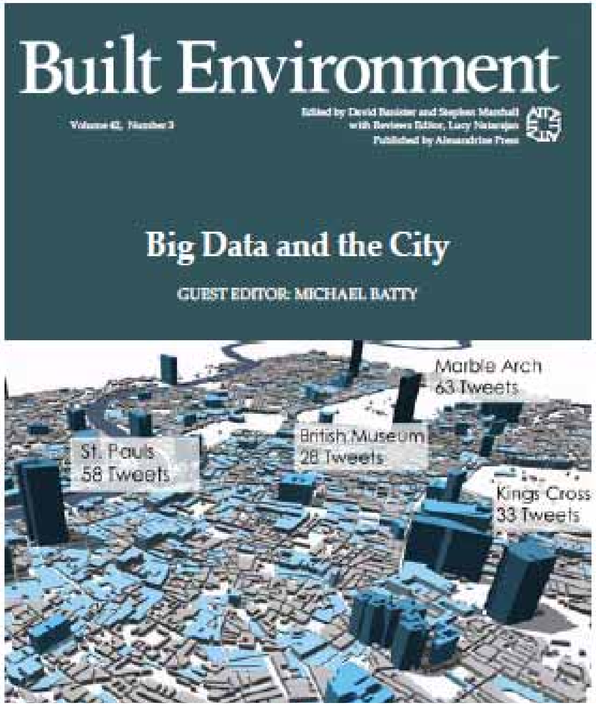
Big Data and the City

Editor: Michael Batty
Centre for Advanced Spatial Analysis, University College London

Built Environment

Volume 42, number 3, September 2016

Big data is everywhere, largely generated by automated systems operating in real time that potentially tell us how cities are performing and changing. A product of the smart city, it is providing us with novel data sets that suggest ways in which we might plan better,anddesignmore sustainable environments. The articles in this issue tell us how scientists and planners are using big data to better understand everything from new forms of mobility in transport systems to new uses of social media. Together, they reveal how visualization is fast becoming an integral part of developing a thorough understanding of our cities.



http://www.spatialcomplexity.info/archives/3026

CONTENTS

Edited by David Banister and Stephen Marshall with Reviews Editor Lucy Natarajan, *Built Environment* is published four times a year, both print and online. For more information or to order a copy of this issue contact:

Alexandrine Press, 1 The Farthings, Marcham, Oxfordshire OX13 6QD

phone : **01865 391518** fax : **01865 391687**

e-mail : **alexandrine@rudkinassociates.co.uk**

visit : **www.alexandrinepress.co.uk**

Outline

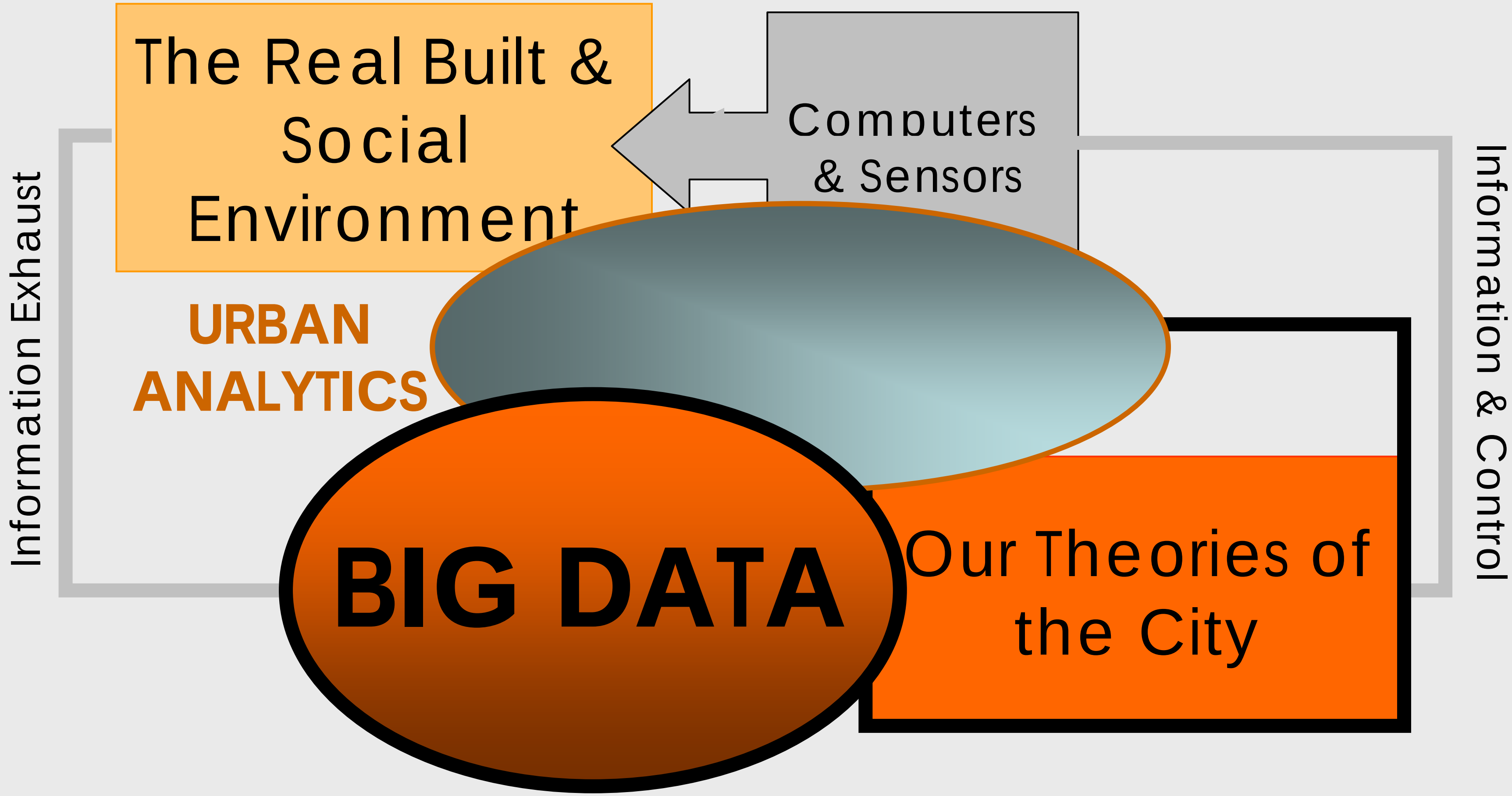
- Back to our Framework of the Smart City
- A Short History of Big Data: How Big is Big?
- Mobility, Transit, & Real-Time Streaming: The Oyster Card Data Set
- Learning about Mobility from the Data

  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

Our Framework once again



Some Basic Points Again

- 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 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**
- This is all part and parcel of increasing complexity; more time scales, more opportunities, more diversity ……

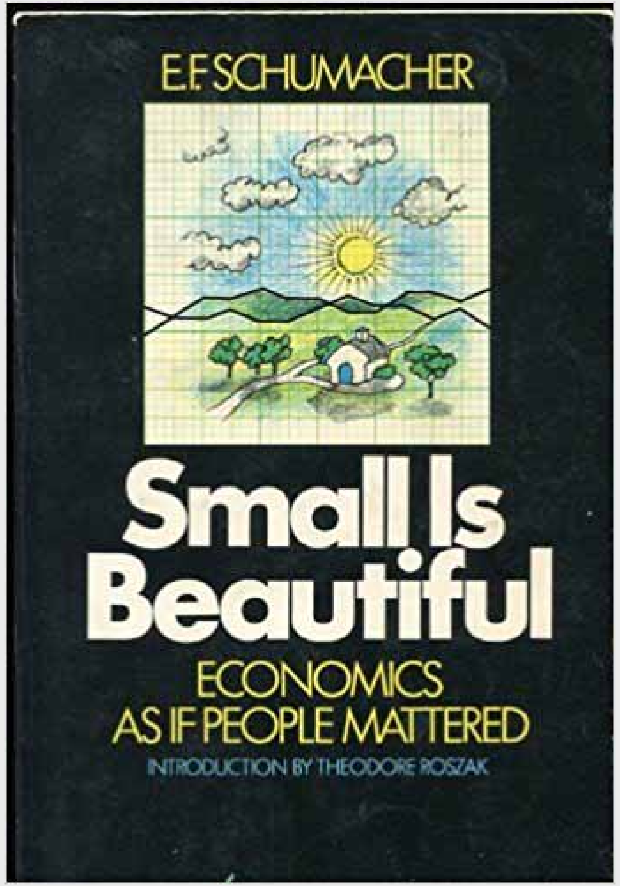
How Big is Data? Big Can Be Small & Small Big

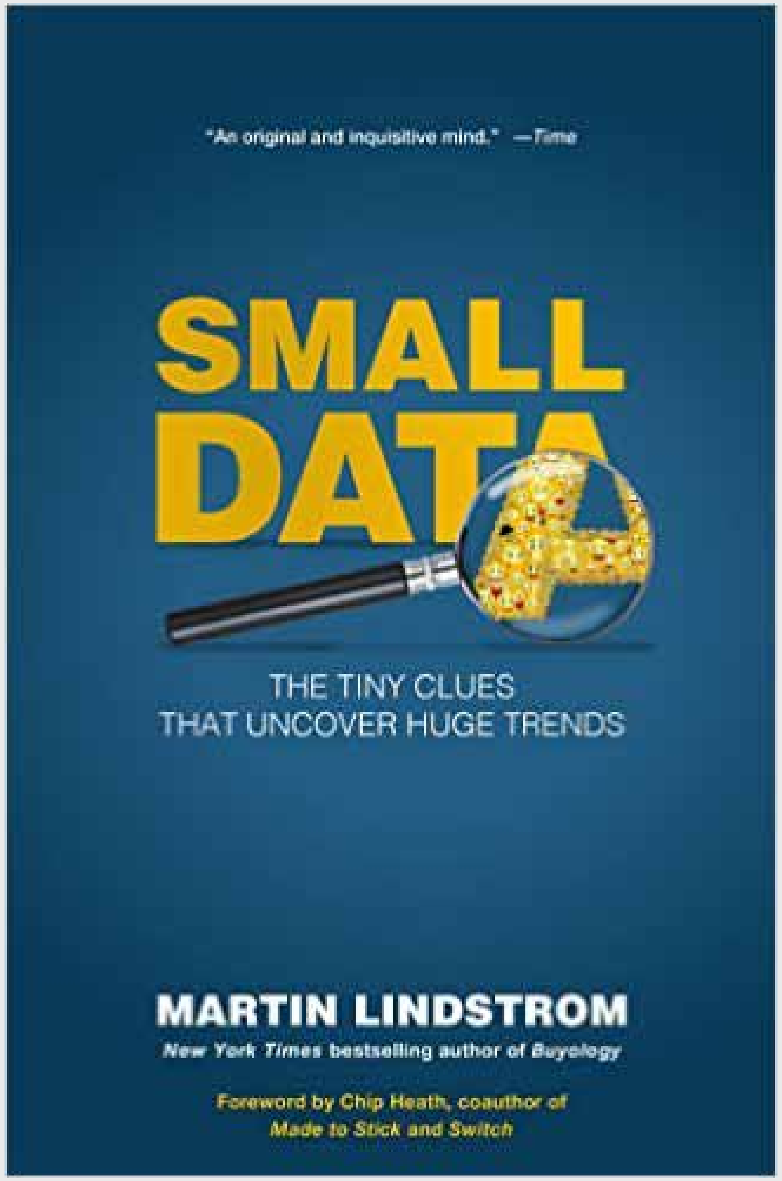
- Data is big with respect to its volume. I know there are other definitions – velocity, variety etc. but to me, data is big if it requires large use of computer memory implying volume.
- The conventional definition in business is the Five V's – volume, velocity, variety, veracity, value
- In cities, 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

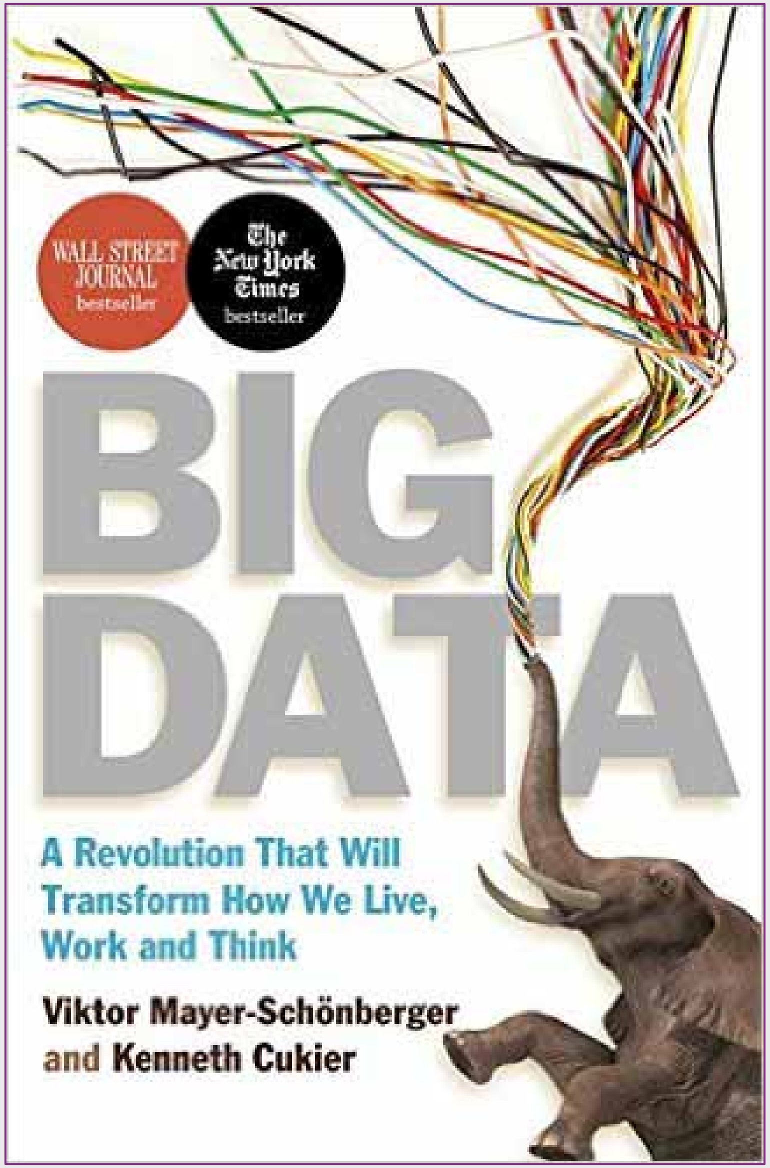
- But there is controversy about how big.
- Big is not necessarily beautiful – small is beautiful – this was the watch word of the 1970s
- So we need to be careful

We can still develop good ideas and good theory with small data.

In fact the idea that the truth or even the path to progress lies in big data is problematic; ***there are limits to machine learning***




"An original and inquisitive mind." —Time
SMALL DATA
THE TINY CLUES THAT UNCOVER HUGE TRENDS
MARTIN LINDSTROM
New York Times bestselling author of Buyology
Foreword by Chip Heath, coauthor of Made to Stick and Switch


WALL STREET JOURNAL bestseller
The New York Times bestseller
BIG DATA
A Revolution That Will Transform How We Live, Work and Think
Viktor Mayer-Schönberger and Kenneth Cukier

- Thus when we have data that contained relations, as we do in spatial analysis, the data can begin to blow up, I explodes – this is Metcalfe's law again
- This is true of flow and network data and arguably this is the most important data that we can get because all locations are really the product of flows
- That is the point I keep making in my ***New Science of Cities*** book; data are relations between locations and in essence if we have ***n locations***, we have ***n^2 interactions***.
- Thus small data can become big as I implied earlier

Examples: Dublin 1837, Ireland 1888, London 1953

Posts About Complexity Fractals Networks Simulations Media Books Articles Editorials

← Visualising Fast Flows

Movies Are Now Online →

The Oldest Flow Map

Posted on June 27, 2011 by Michael Batty

Tweet 3

.... 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?

Be Sociable, Share!

Tweet 3 Like 0 +1 0 Share Submit

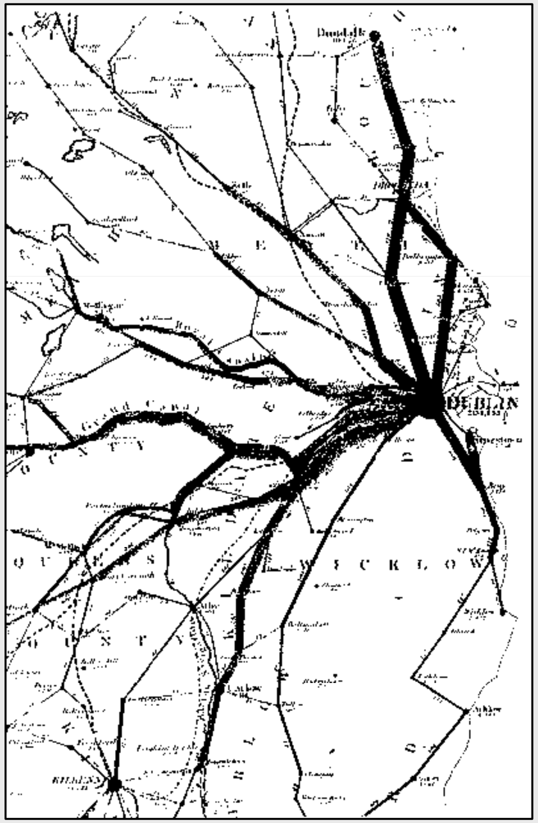
About Michael Batty

$\int \frac{dN(t)}{dt}$

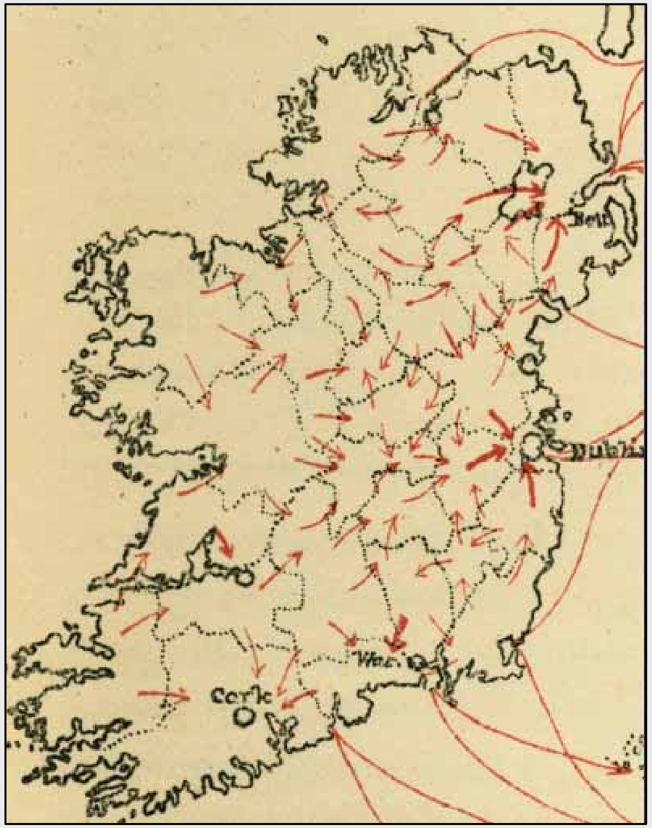
I chair CASA at UCL which I set up in 1995. I am Bartlett Professor In UCL.

View all posts by Michael Batty →

Examples: Dublin 1837, Ireland 1888, London 1955



Harness, 1837



Ravenstein 1888

- So big data can derive from small if we think of it as relations.
- But also big data is relative to our ability to process it – the machines or ‘brains; we have to crunch it
- If I have a big box of punched cards, say several hundred then I can reduce the physical size by putting on another media – punched tape for example, or magnetic tape
- But the size of the machine to process it is a limit – the computer may only be able to process so much in core memory and the problem may be too big for the machine. There is a great example ….

https://www.youtube.com/watch?v=pcBJfkE5UwU

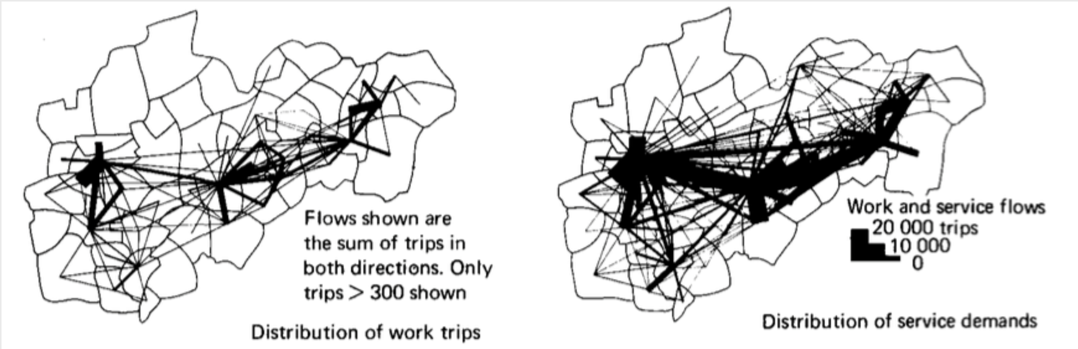
Locations and Interactions: Flow Systems in Cities

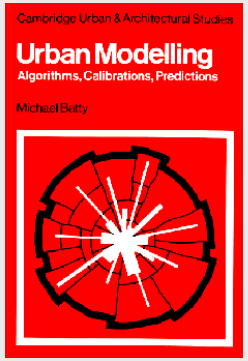
Elsewhere I have argued that we should treat cities as flow systems – as networks. This has been a focus for a long time in transport and land use and we have always been up against the problem of big data.

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

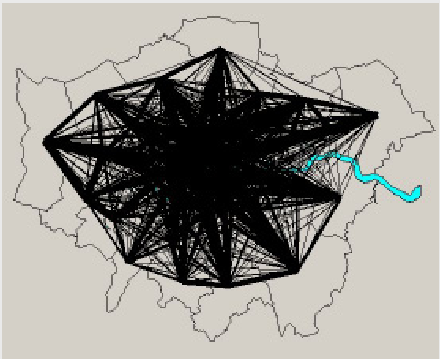
Understanding and Visualising Flows

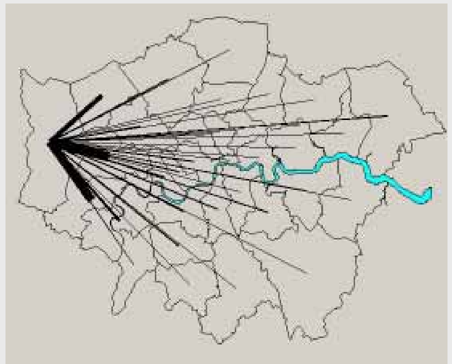
An early model circa 1967-8 Central and NE Lancs



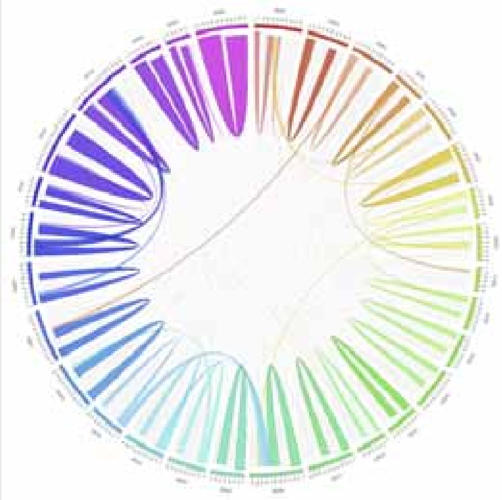


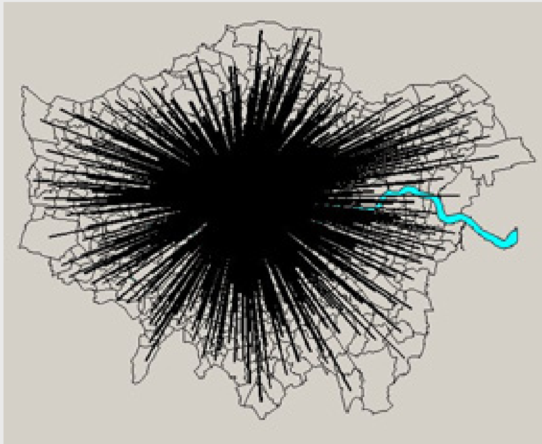
M. Batty (1976)
Urban Modelling
Cambridge UP



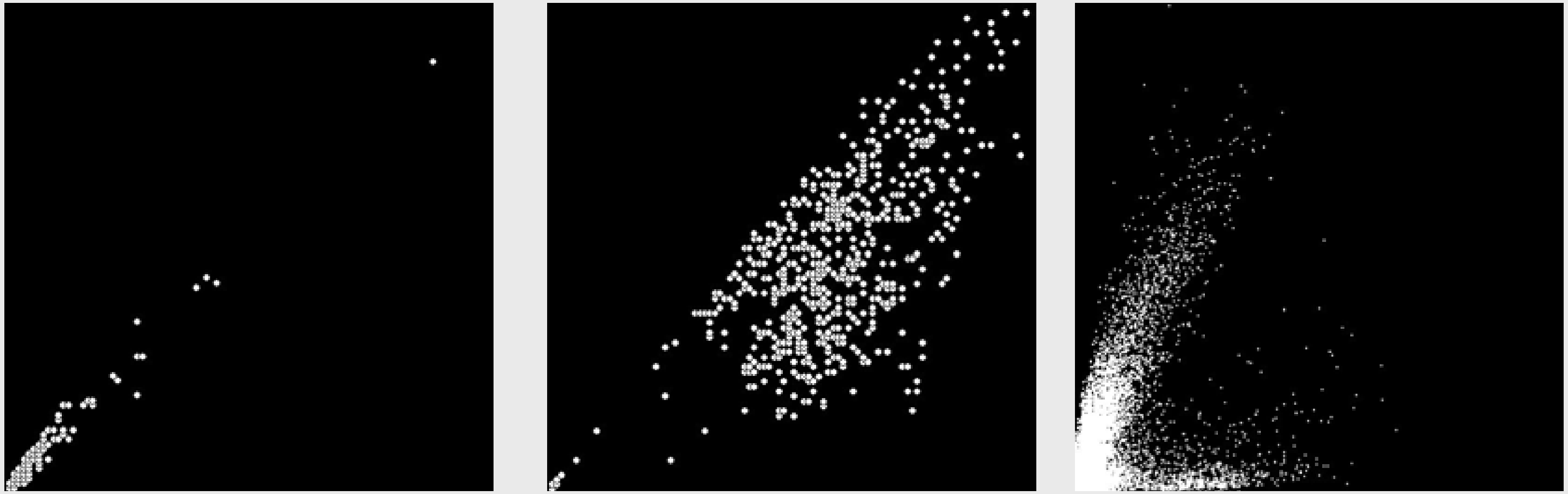


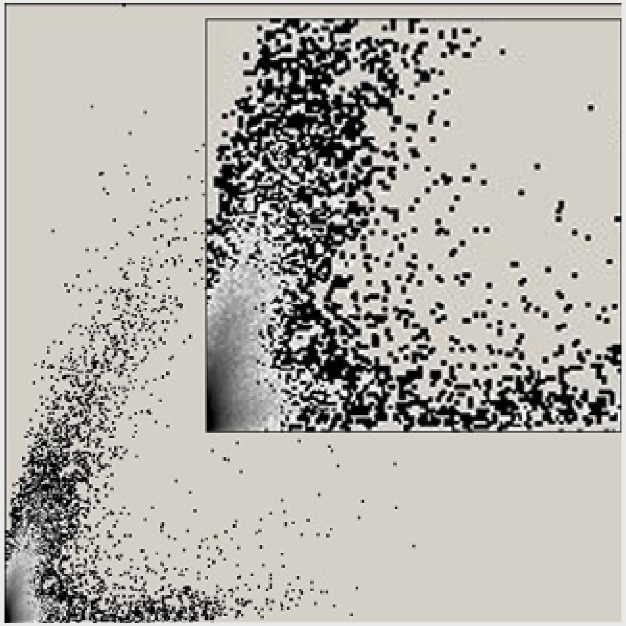
$n^2=33^2=1089$, not so big but hard to visualise





$n^2 = 633^2 = 400,689$, bigger but impossible to visualise

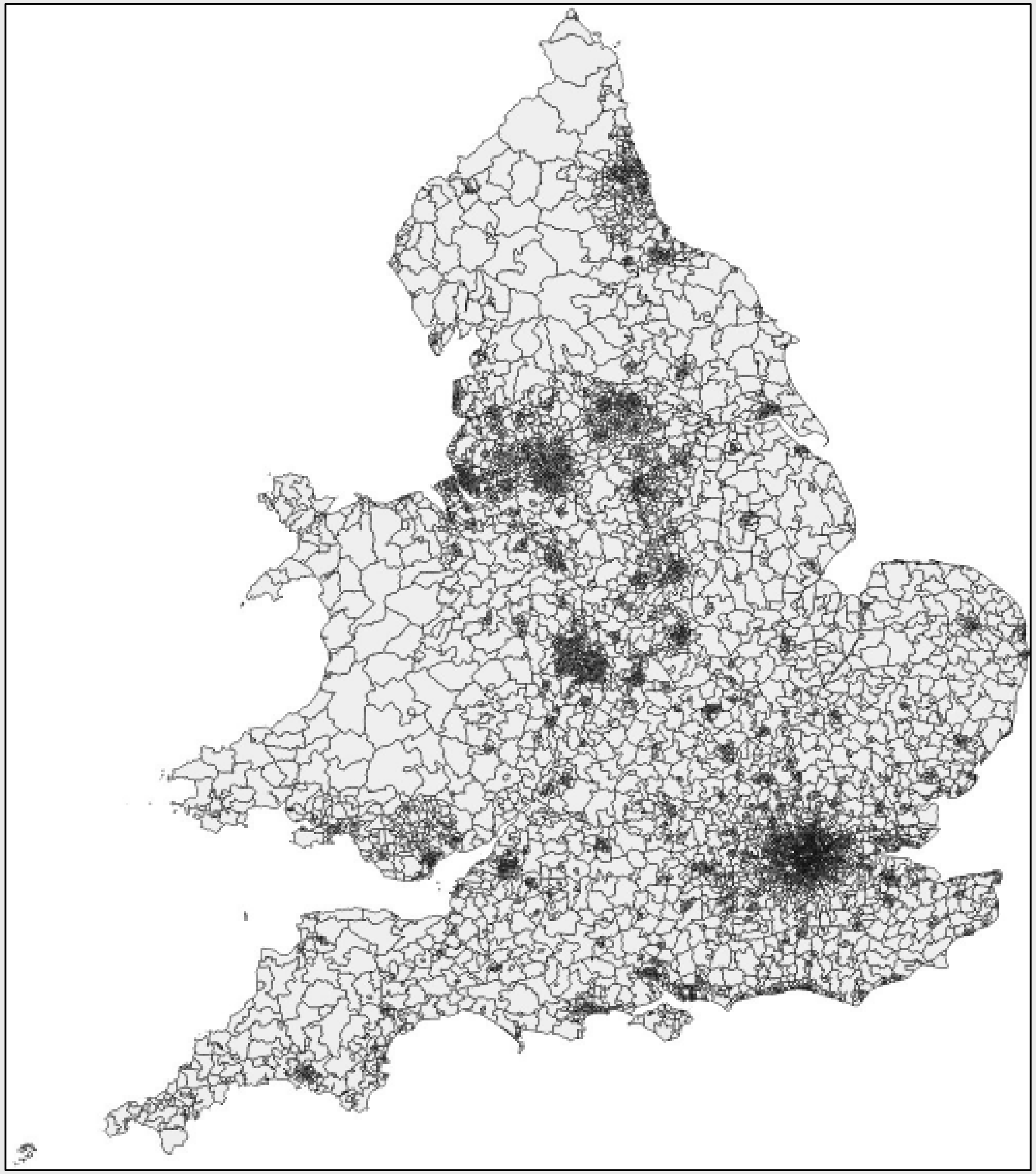




Even our statistics breaks down when we get large numbers like over several thousand as you can see on the left and above right for 400K data points where the pattern is highly convoluted. This is from a gravity model.

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 $n^2=7201^2=51,854,401$.

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 problems scaled up and this is what we are grappling with at present.



(a) MSOA (A=7201) (b) LSOA (A=34753) (c) OA (A=181408)

Figure 8.2: ONS Geographies for MSOA, LSOA and OA levels.

The Web and the Desktop: Users are also Data

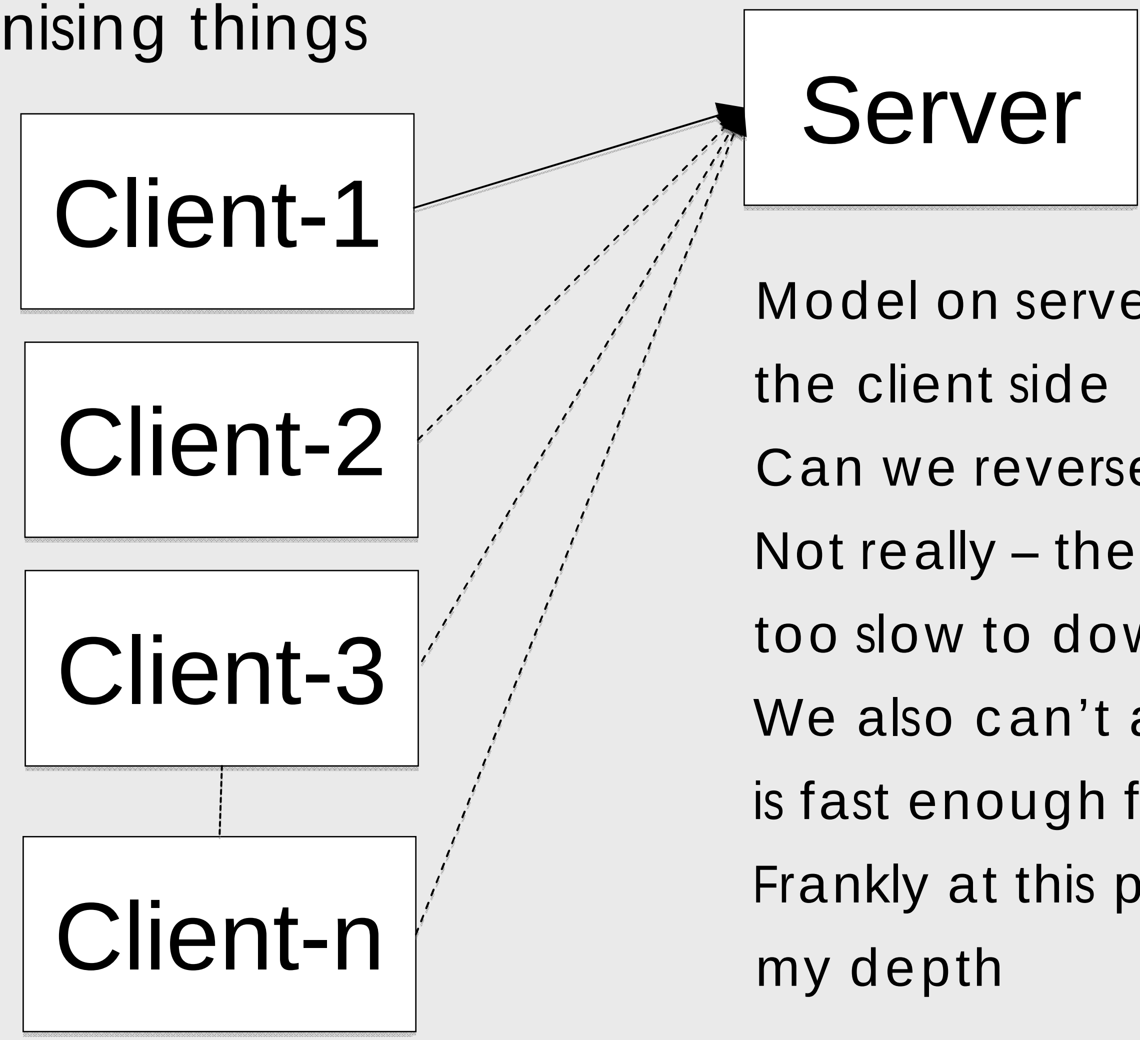
We are building a model of the UK – well E&W at present – we will add Scotland before long – which is of the nature we have been implying – Without going into details, the model takes a few seconds to run – it will take a lot longer when finished as we will add sectors and of course the number of big data we have to hold in RAM might be very large – currently we need to hold 4 such 52 million sized matrices – we may need to go up to 8 in time and that will involve a lot of packing and moving in and out of core, I think

But the real issue is users – if our model is this large, and we have many users, then our data problem is exploded by the users –

Our big data is our original and predicted data from the model, times the number of users. Why are users data ? Well because they are using data differently – they are making their own predictions and thus scaling up the data.

We could have one model for each users but we don't know who the users are? We thus want them to access this on the web. This is where it all hits the fan ..

Here is a block diagram of how we are currently organising things



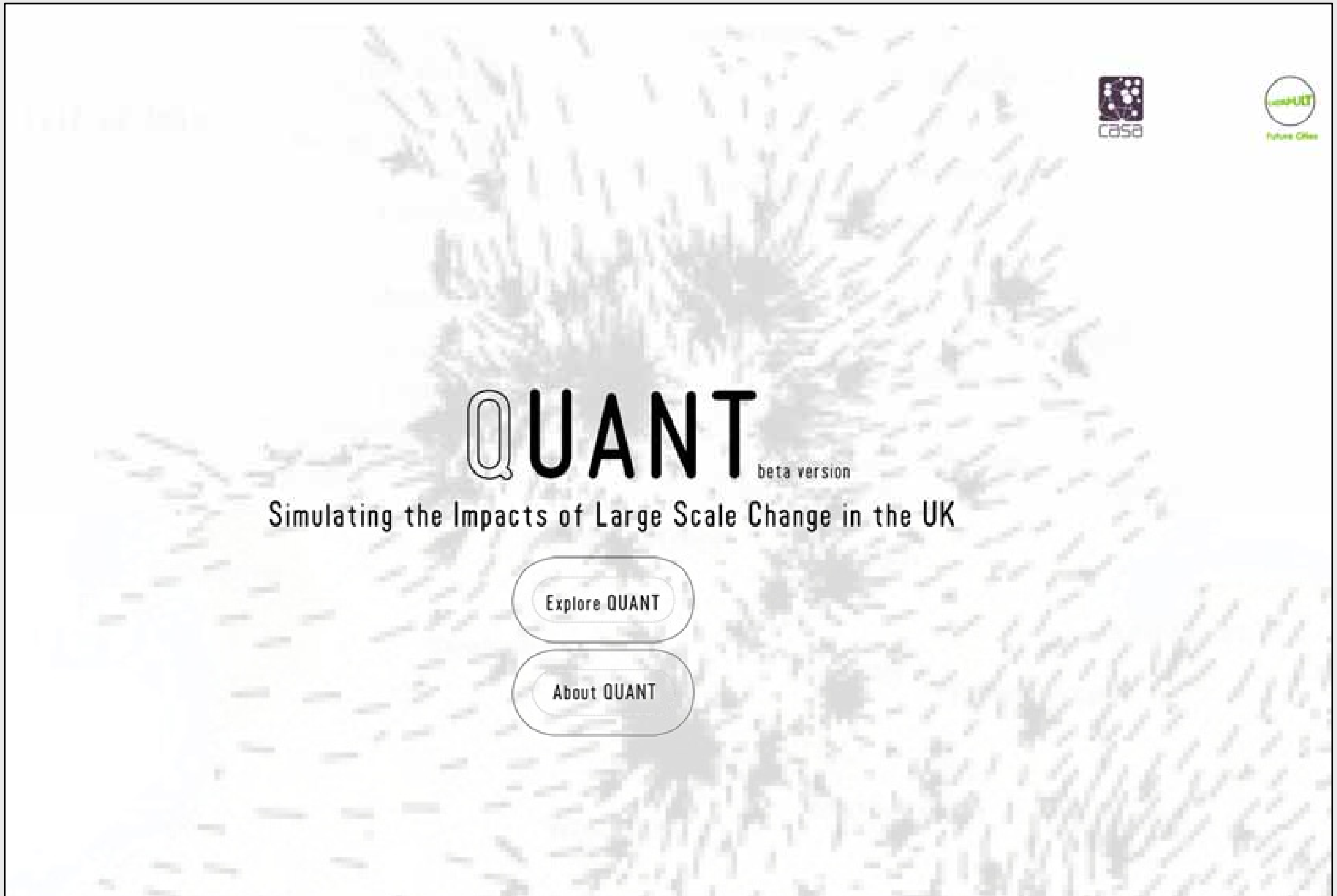
Model on server side; Maps on the client side

Can we reverse this?

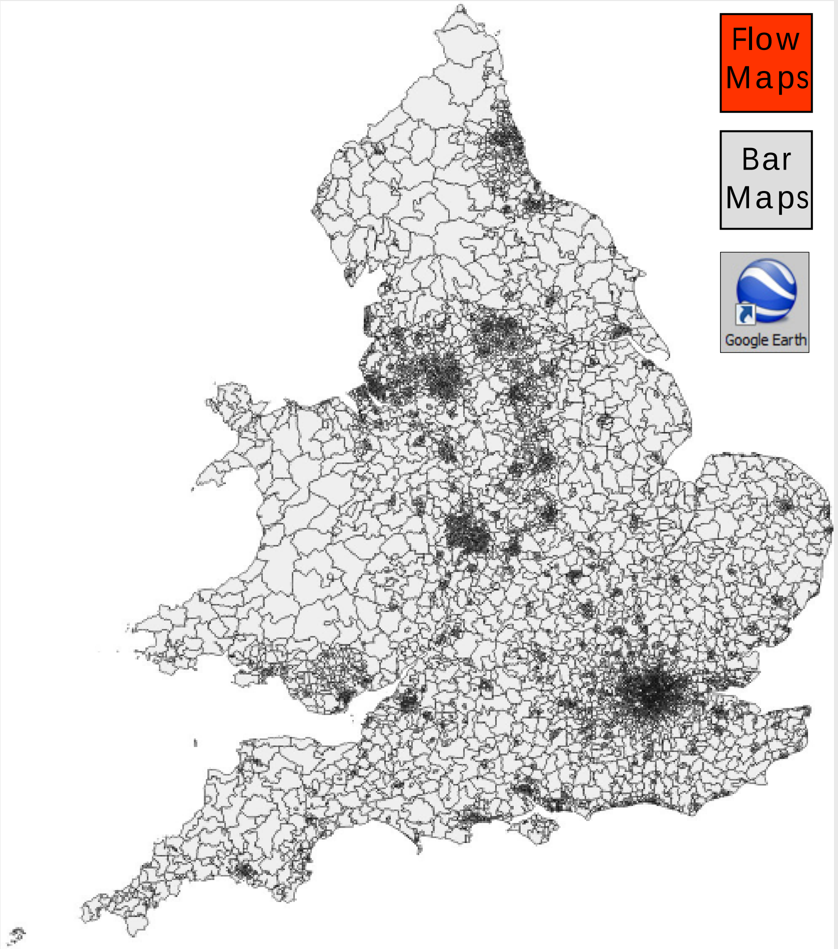
Not really – the matrices are too slow to download to client?

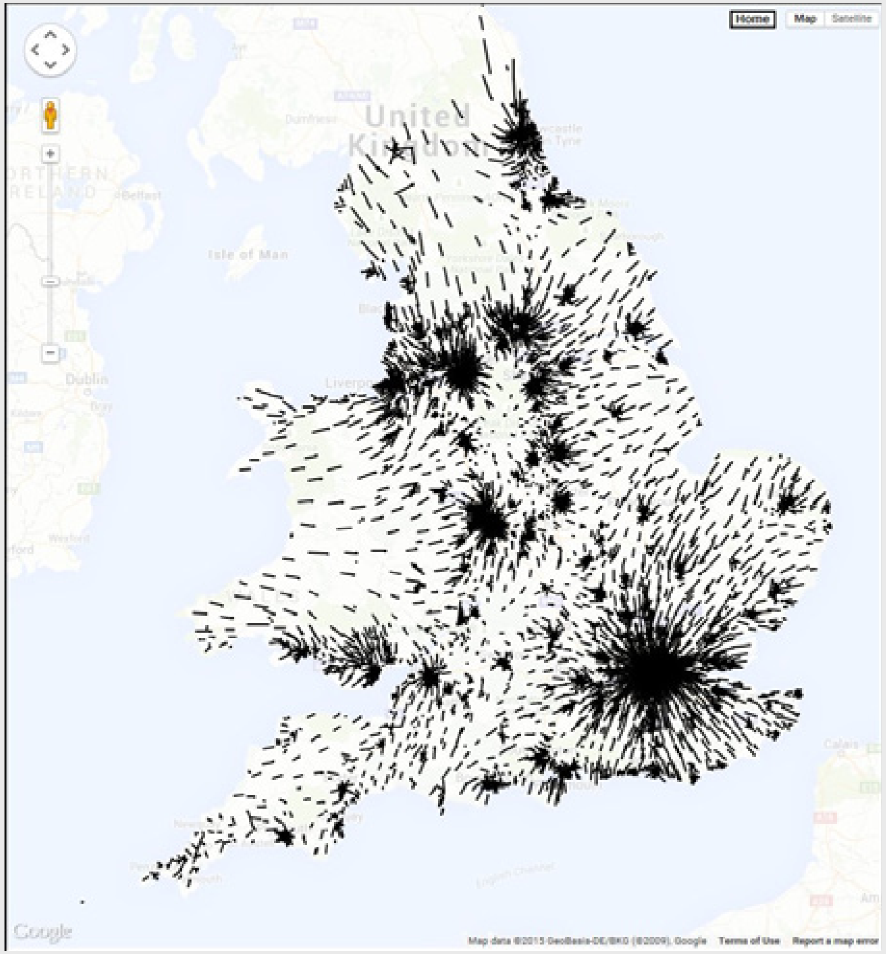
We also can't assume the client is fast enough for computation.

Frankly at this point, I am out of my depth

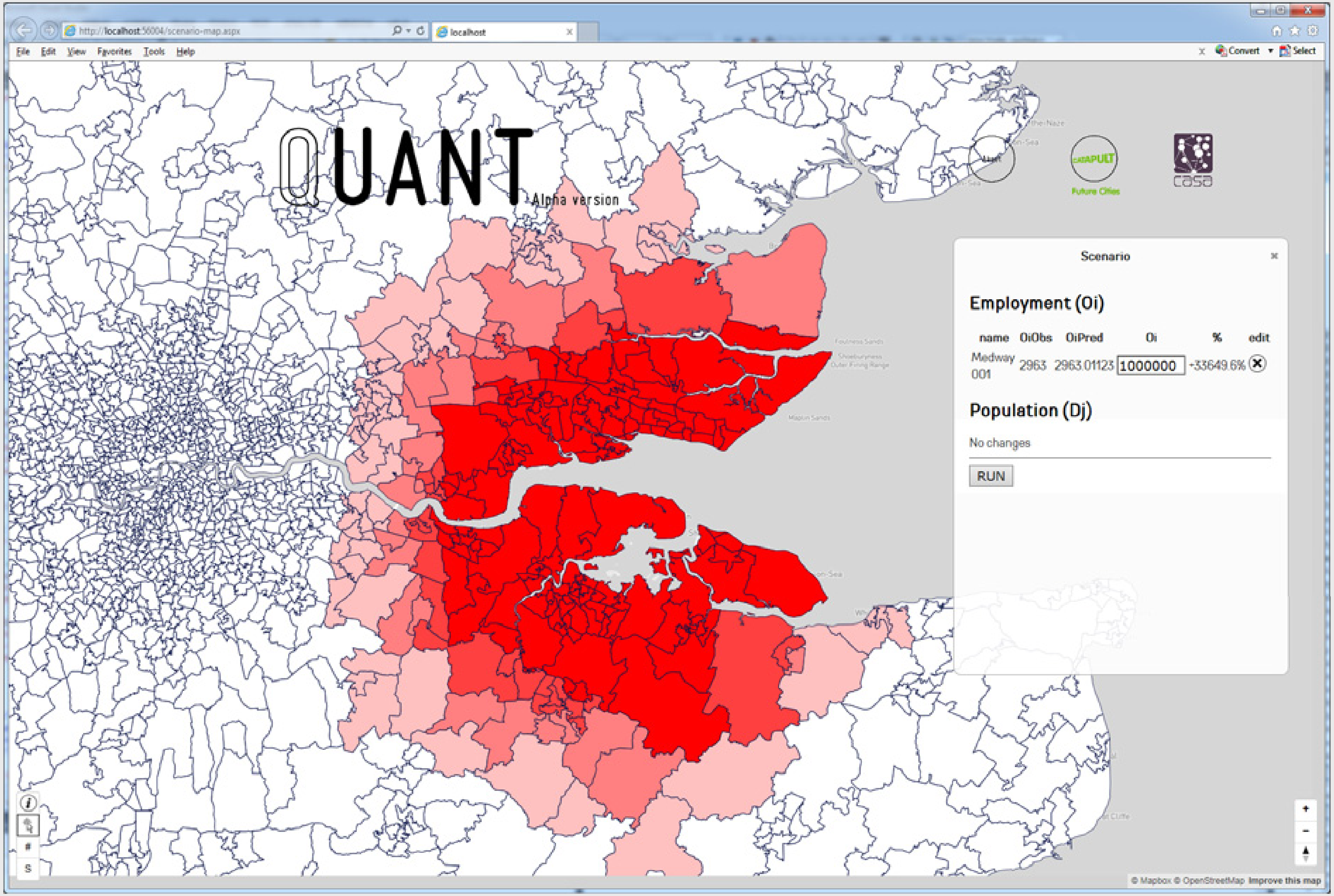


http://quant.casa.ucl.ac.uk/





$$[x_i, y_i] = [[x_i, y_i], \left[\left[x_i + \frac{\sum_j T_{ij}\,[x_i - x_j]}{n}\right], \left[y_i + \frac{\sum_j T_{ij}\,[y_i - yy_j]}{n}\right]\right]]$$


QUANT
Alpha version
Scenario
Employment (Oi)
name OiObs OiPred Oi % edit
Medway 001 2963 2963.01123 1000000 -33649.6%
Population (Dj)
No changes
RUN

Real-Time Streaming: What Sort of Data ?

http://www.citydashboard.org/

London Panopticon

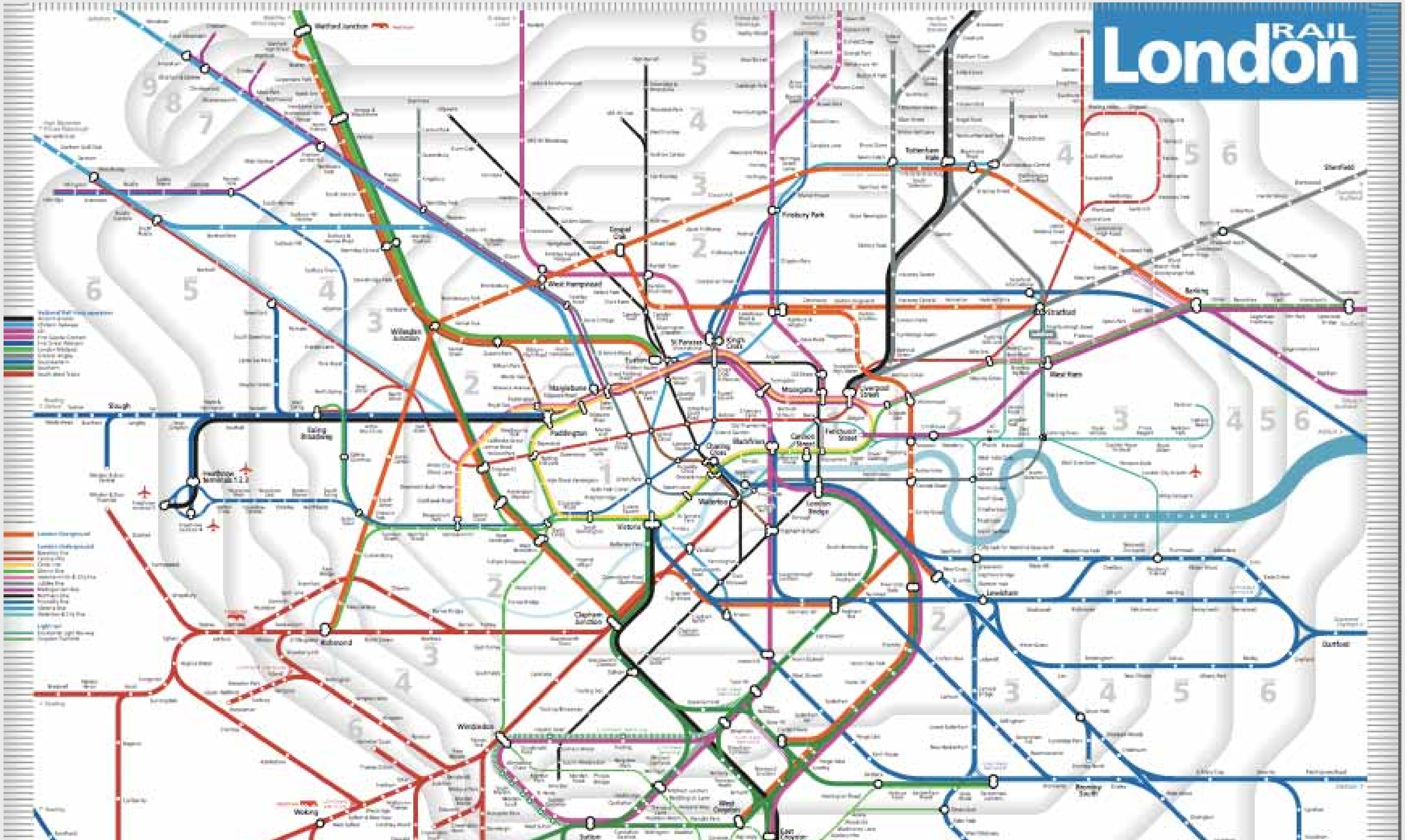
6 April 2016 London

http://vis.oobrien.com/panopticon/

Tweet Like Share 0

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

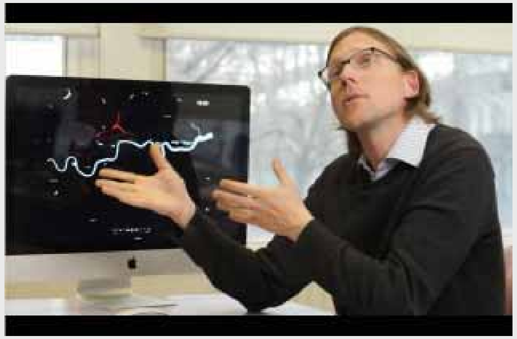


Tube, Overground and National Rail Networks in London where Oyster cards can be used


UCL ENGINEERING
Change the world
UCL
Menu +
Home – Projects – Oyster gives up pearls
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 Soong 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.
UCL Engineering - Oyster Gives up Pearls
0:00 / 3:32
They used Transport for London's database of 11 million records taken over one week from the Oyster Card electronic ticketing system.
Latest news from UCL Engineering
New web privacy system could revolutionise the safety of surfing
UCL host Google Girls Coding Programme with Generating Genius and University of West indies
Professor Polina Bayvel to Give Royal Society Lecture
Twitter feed
RT @markmiodownik: Am giving a ENGins seminar today for @UCLEngineering @UCLENGins all UCL engineers welcome - Roberts G06, 6:30pm. http:/...
8:58am Thu 9th October 2014
RT @Centre4EngEdu: We're hiring! Multi-talented Centre Administrator required to help us launch and expand! bit.ly/ZsERSM
10:54am Wed 8th October 2014
Join our mailing list

And how can we make sense of this

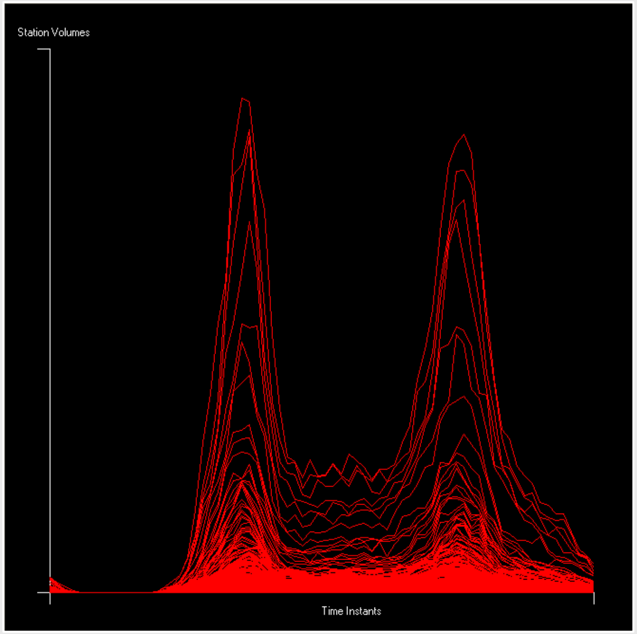
http://www.simulacra.info/

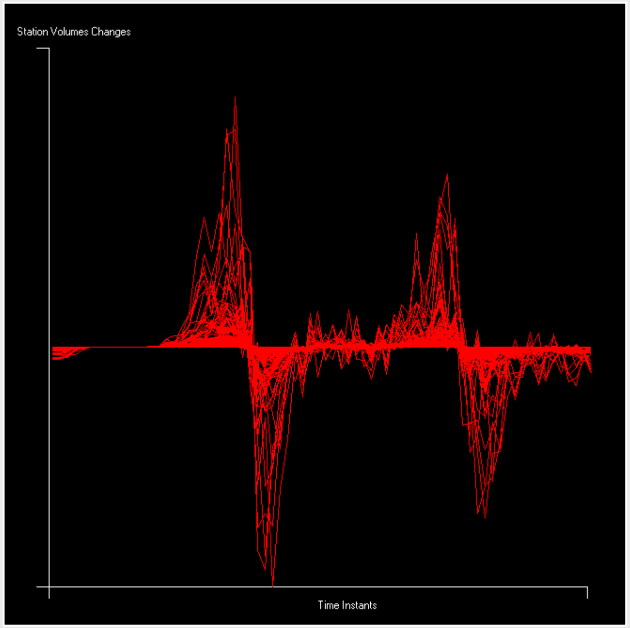



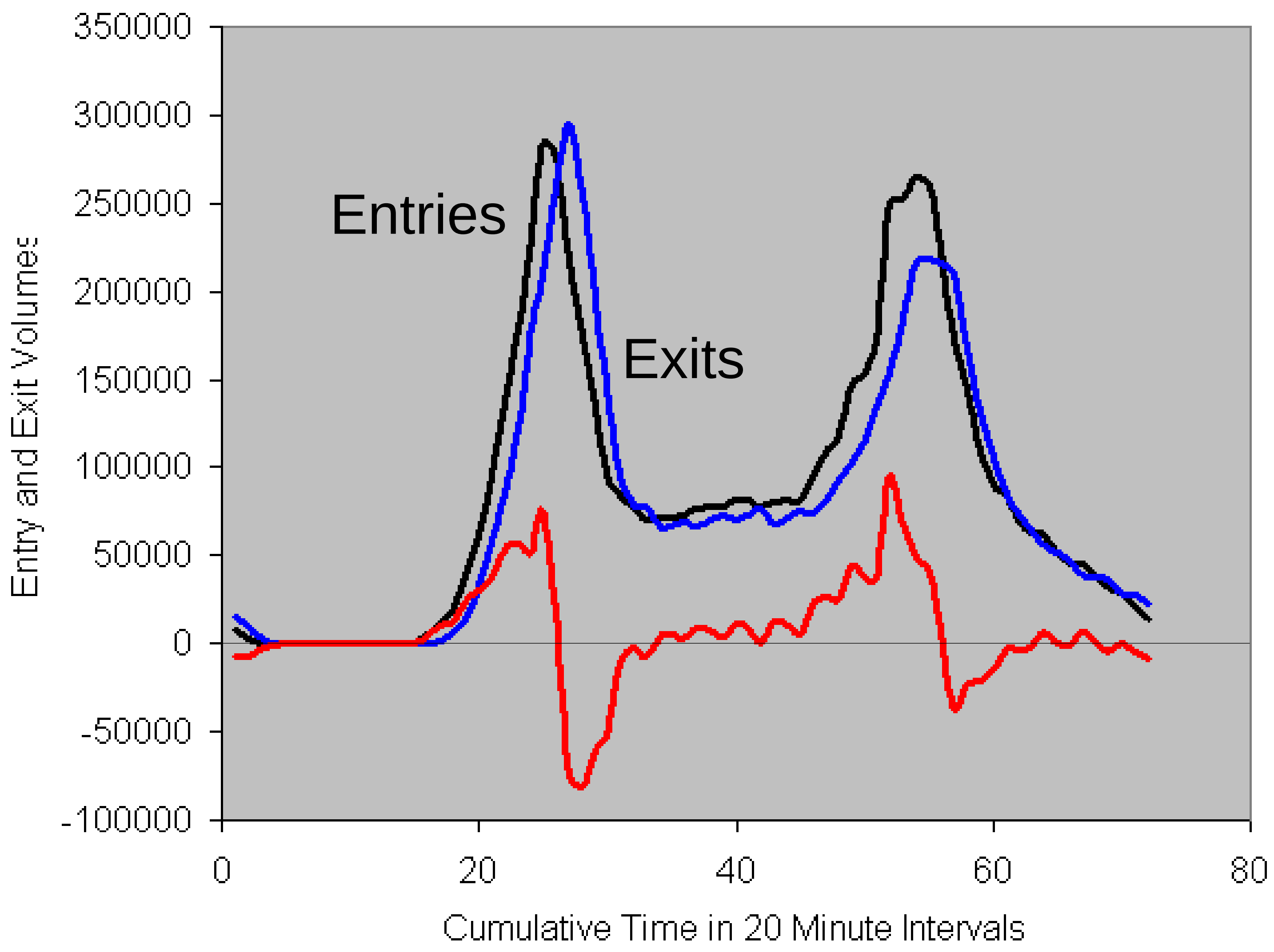

This of course was the thing that Lt Henry Harness did in Dublin in 1837 and what Minard et al. did a little later. In our LUTI models, this is an enormous problem as the scale of this assignment to networks is different

Variabilities – Heterogeneity and Travel Profiles

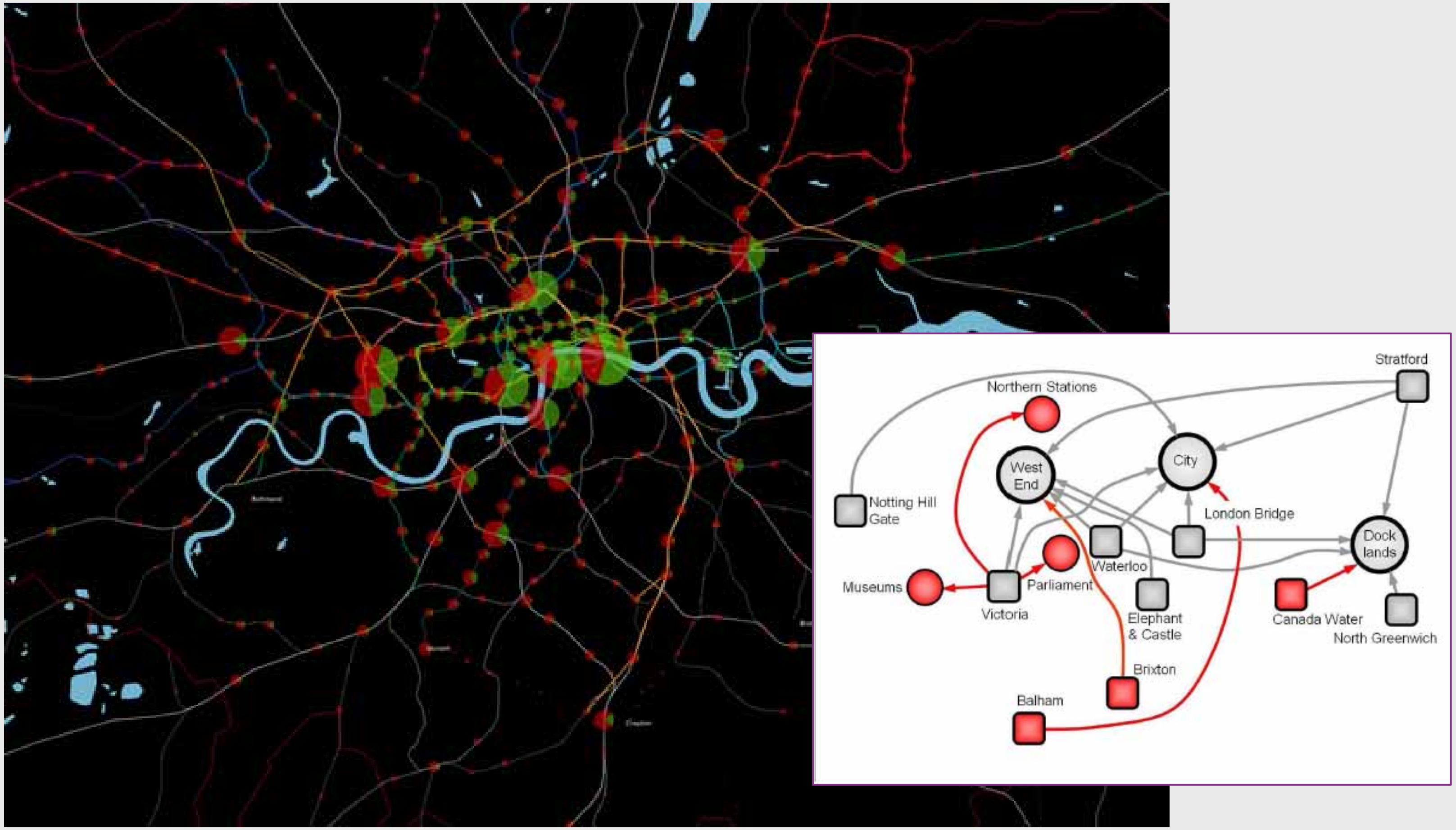
First we will look at some of the data and how it varies in terms of the diurnal flows usually morning and evening peaks, with a small blip (peak) around 10pm at night





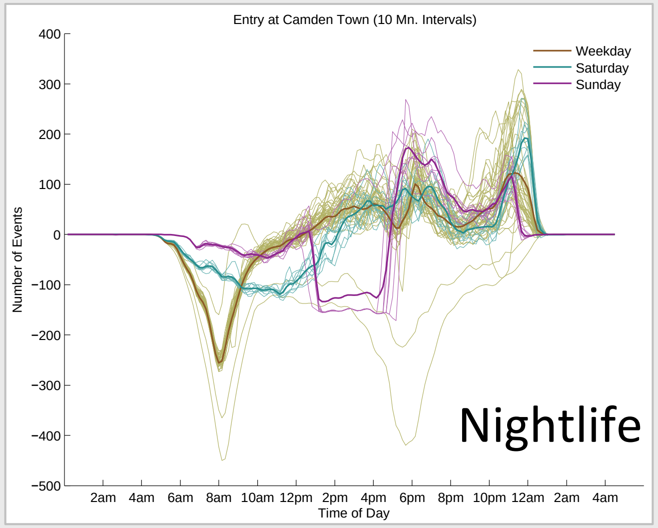

Entries
Exits
Entry and Exit Volumes
Cumulative Time in 20 Minute Intervals
350000
300000
250000
200000
150000
100000
50000
0
-50000
-100000
0
20
40
60
80

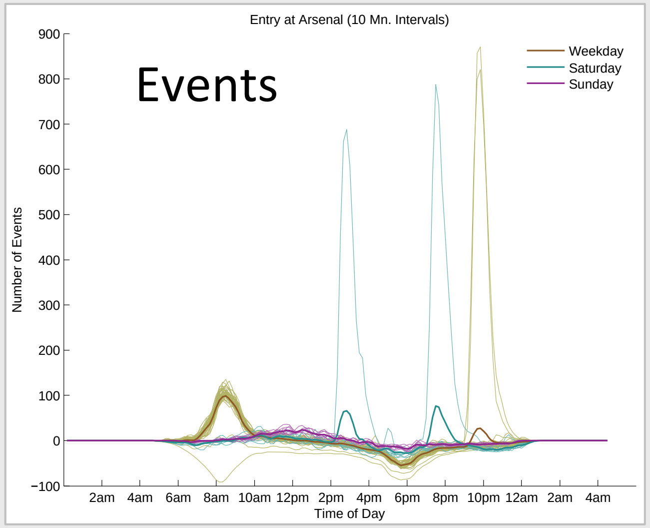
Oyster Card Data – interpreting urban structure, multitrips, etc.

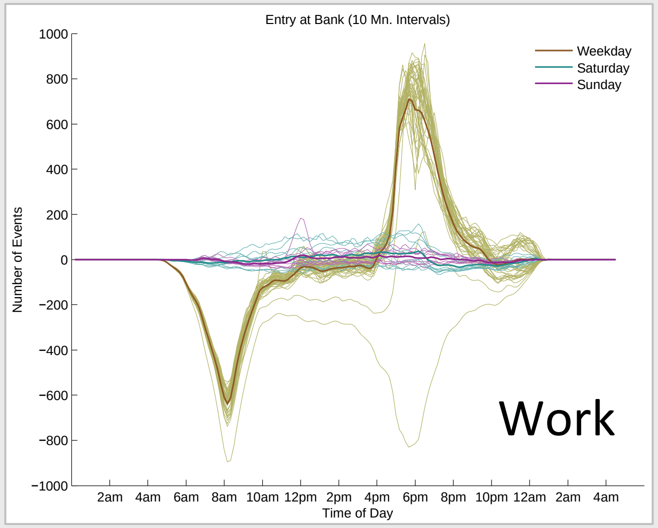


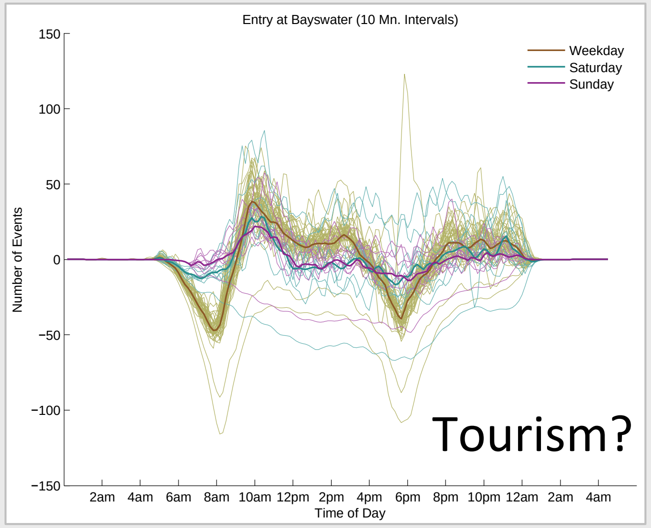
Roth C., Kang S. M., Batty, M., and Barthelemy, M. (2011) Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. **PLoS ONE 6(1):** e15923. doi:10.1371/journal.pone.0015923

Particular Events: Weekdays, Saturdays and Sundays









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)

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

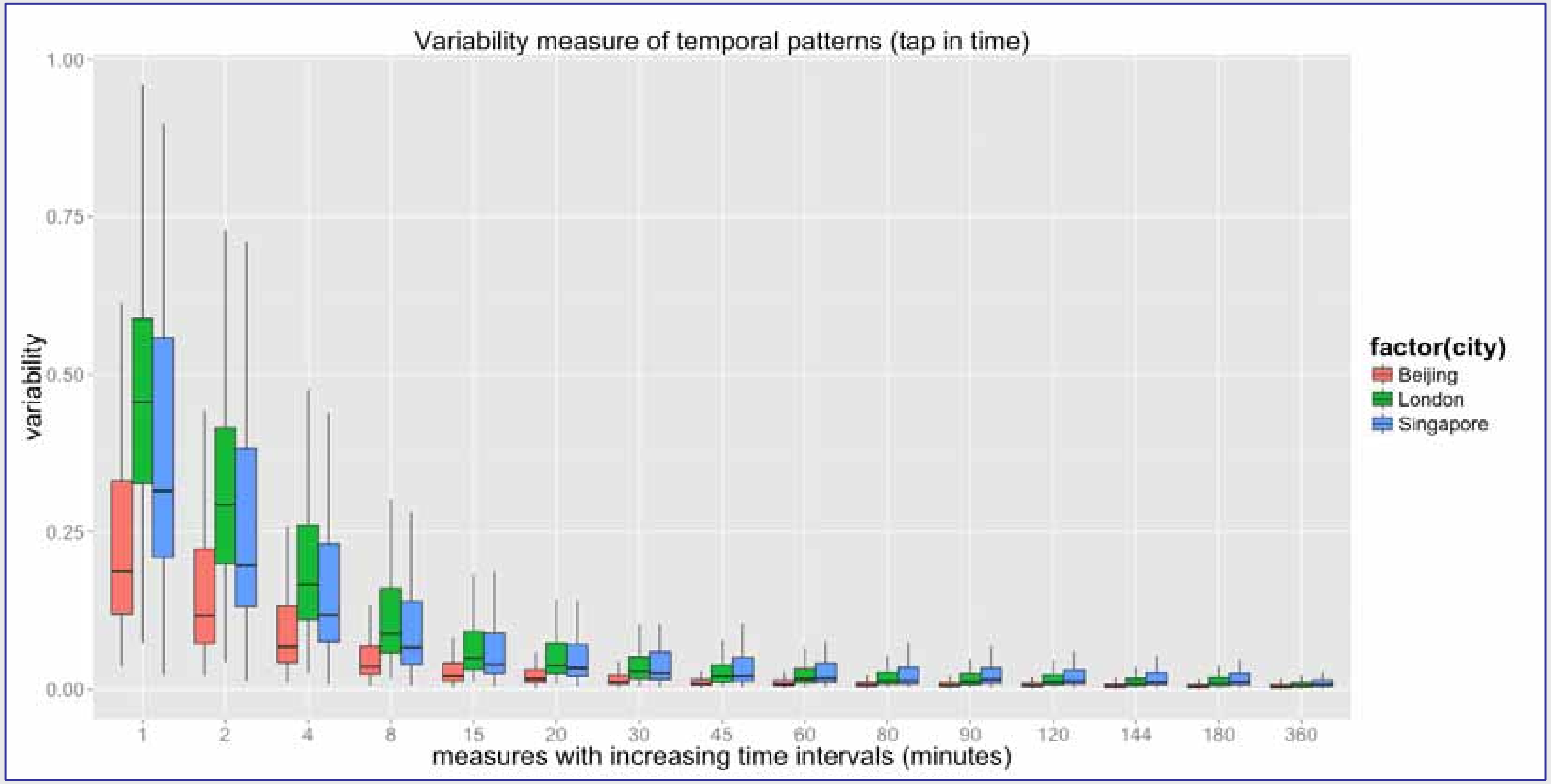
(1) Number of stations is the number of stations with smart-card records generated.

(2) The area of Beijing only counts the area enclosed by the 6th ring road for a fair comparison.

(3) From the World Population Review, http://worldpopulationreview.com/world-cities/ accessed 17 January 2016

Zhong, C., Batty, M., Manley, E., Wan, J., Wang, Z., Che, F., and Schmitt, G. (2016) Variability in Regularity: Mining Temporal Mobility Patterns in London, Singapore and Beijing using Smart-Card Data., **PLOS One**, http://dx.doi.org/10.1371/journal.pone.0149222

From 1 minute intervals to the whole day



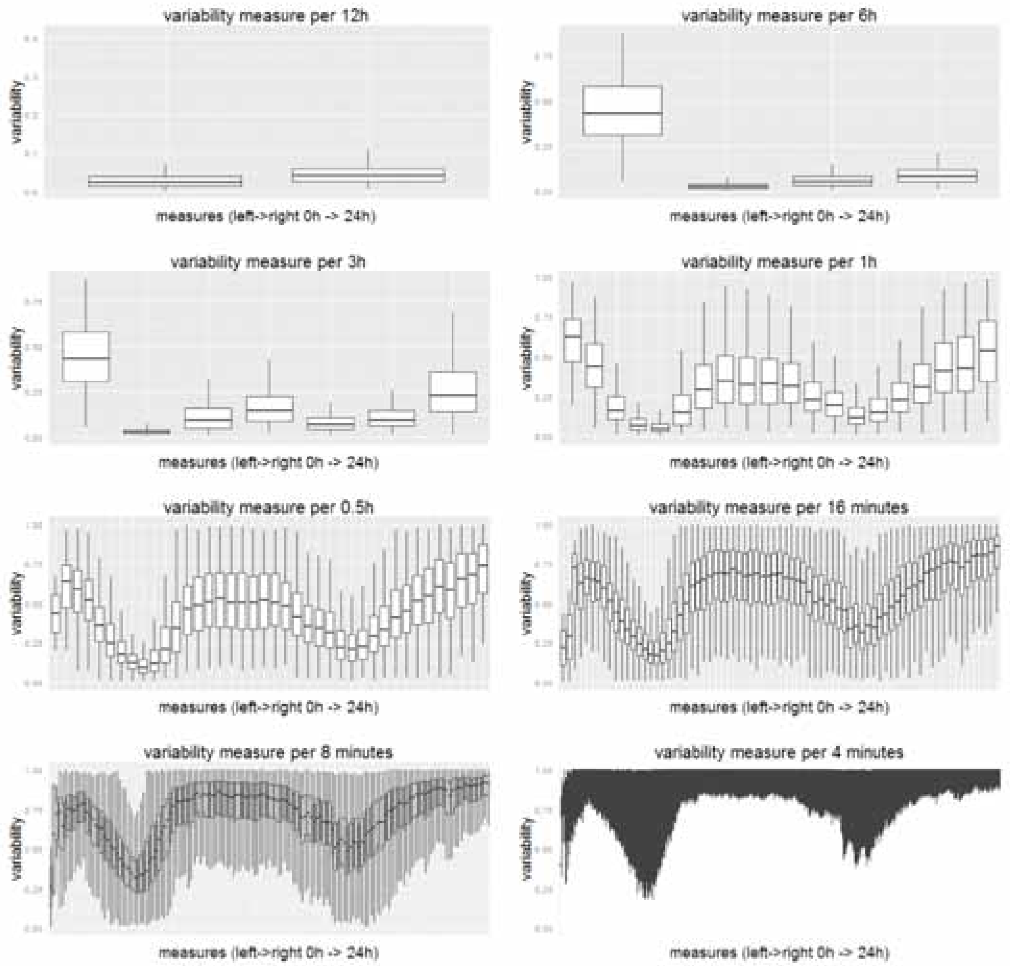
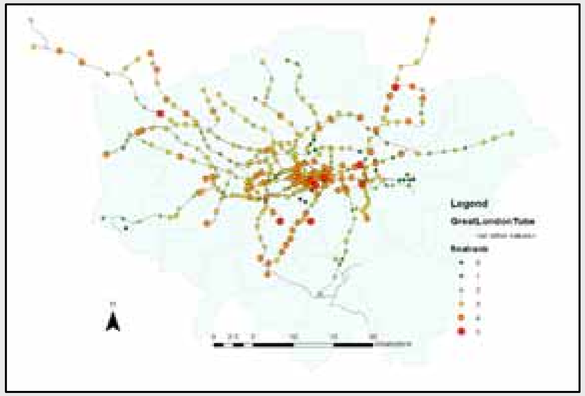


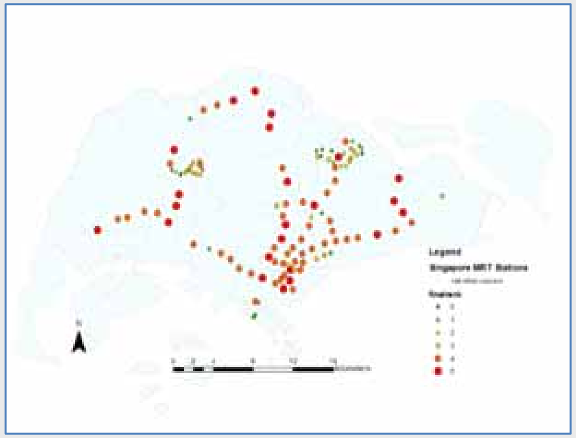
Figure 1. Variability of regularity in the trip matrix over time.

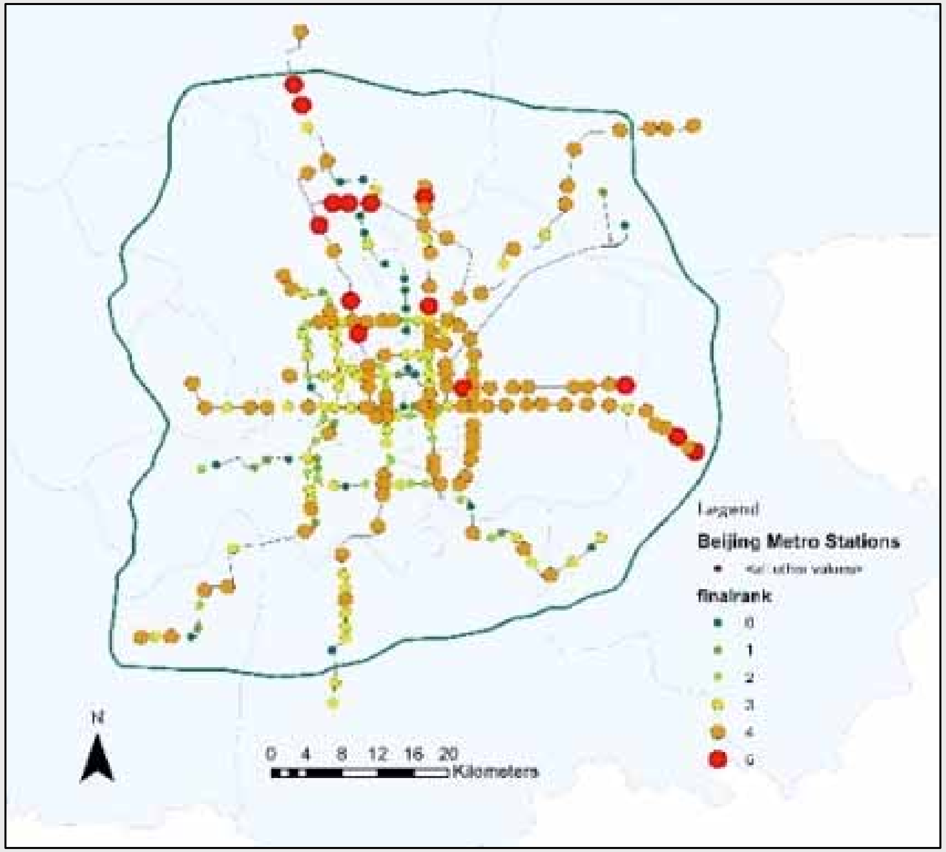
Note: Each box plot shows the variability of 400 stations over time measured at different temporal scales. Overall, eight subplots give a similar trend where lower variability appears during peak hours (around 9 am in the morning and 6pm in the evening). More details can be captured as differences of variability between each time unit are magnified as we decrease the temporal scale from 12h to 4 minutes.

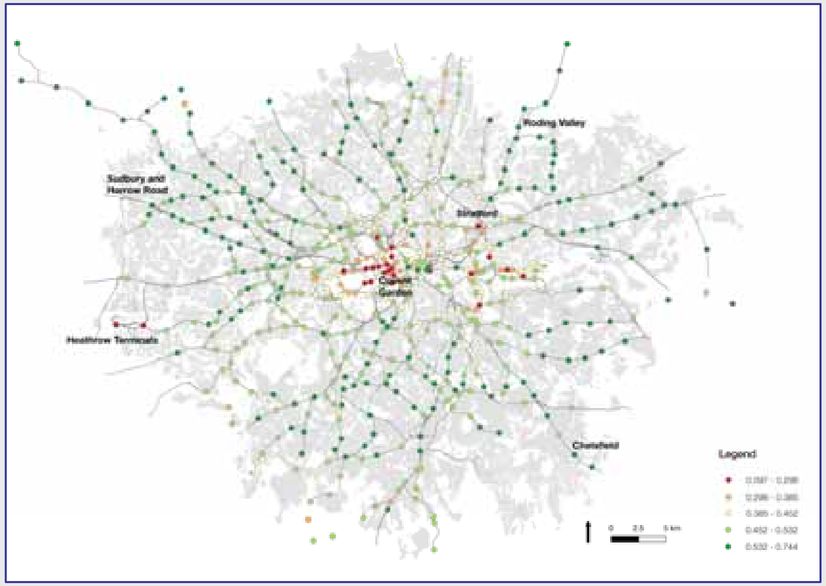
Comparing Variability for different time Intervals over the day

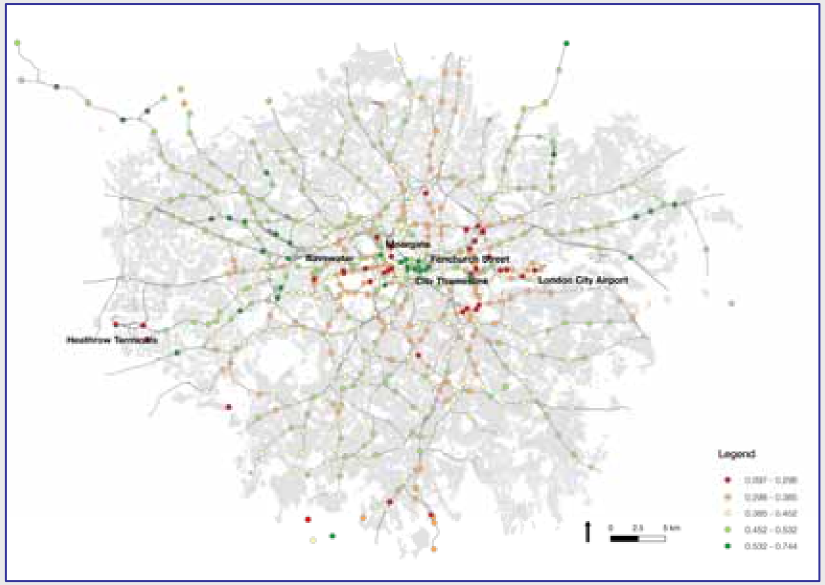
Comparing Variability for different time intervals for Three World Cities: London, Beijing and Singapore

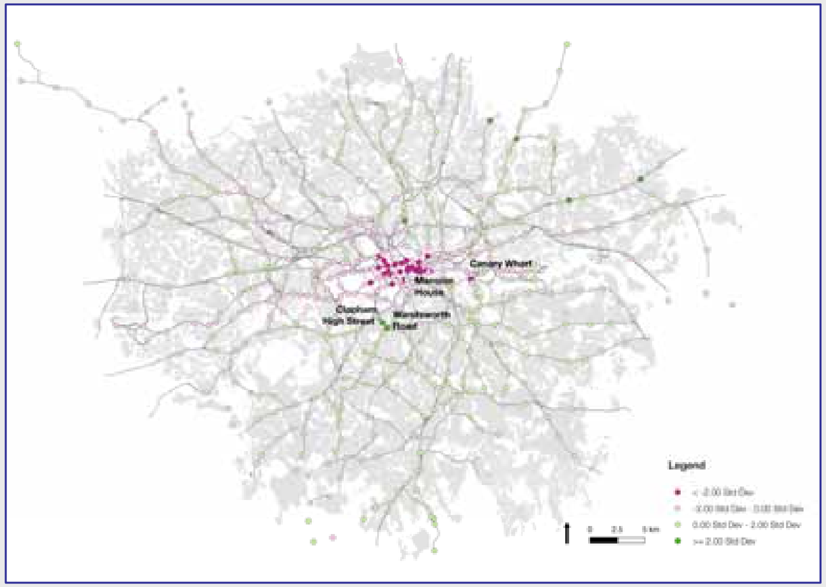












Maps of Underground and Rail stations in London visualised by the proportion of regular trips

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

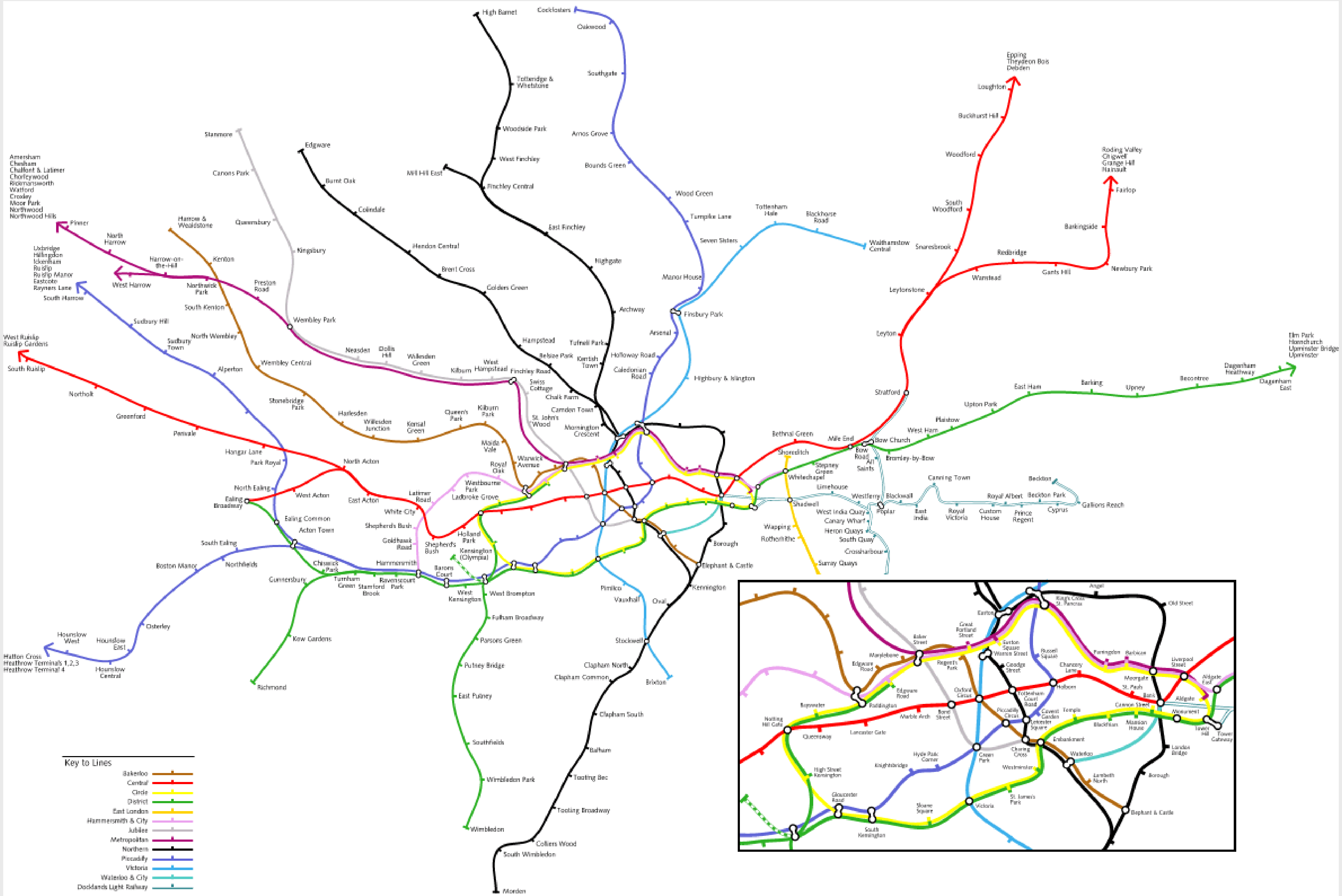
- Then an example of the Circle and District Lines which had a 4 hour stoppage on July 19th 2012

- 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

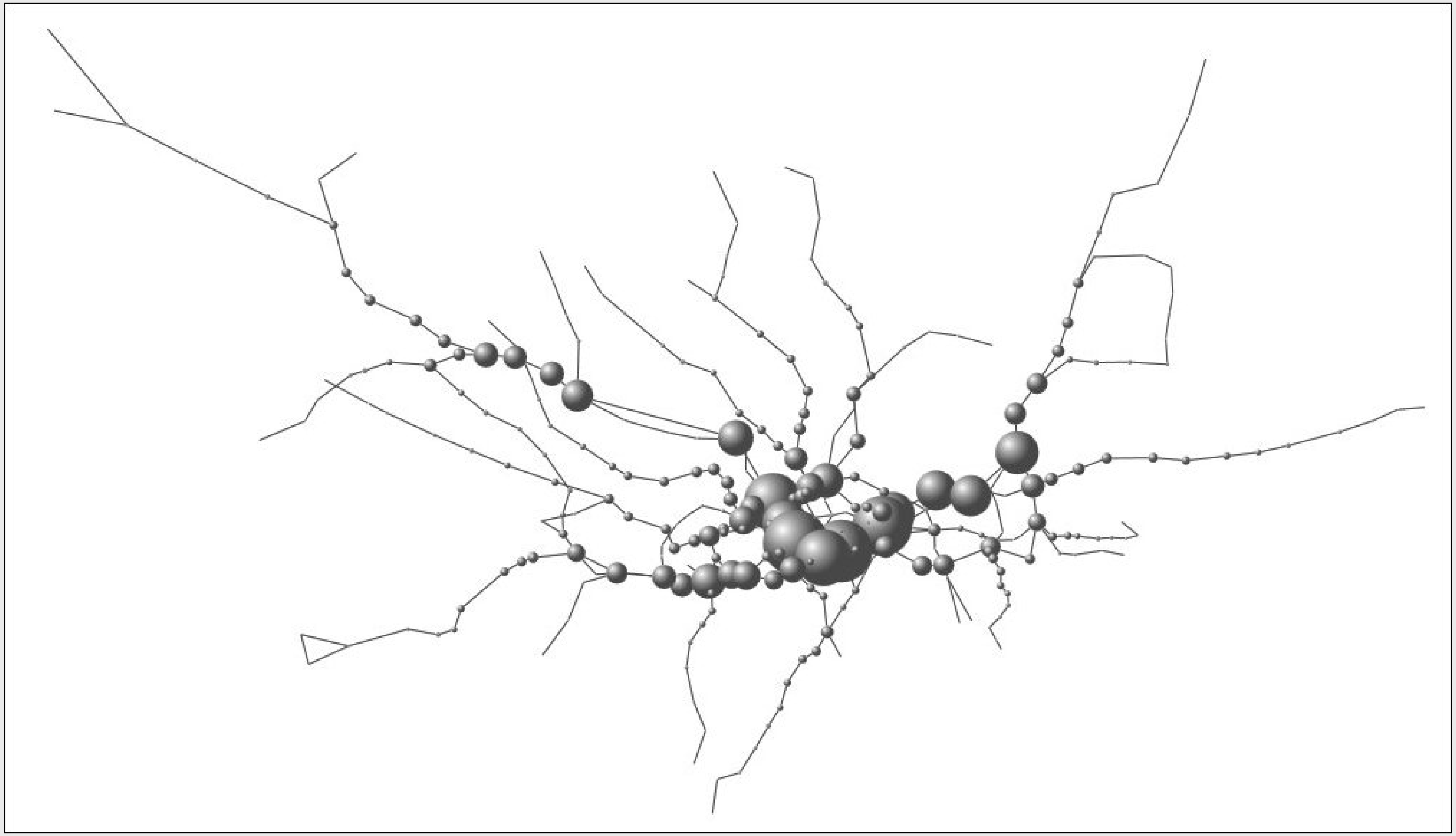

Service information
Date Saturday 19th January
Time Planned Weekend Closures
No Service On The Entire Line
And Kings Cross
Between Baker Street
And Aldgate Also Between Wembley Park
And Uxbridge / Northwood
District Line Between Earls Court
And Edgware Road
Jubilee Line Between Green Park
And Stratford
(Charing Cross Branch) Trains Not Stopping At Euston Station
Holloway Road Station Closed


UNDERGROUND


Key to Lines
Bakerloo
Central
Circle
District
East London
Hammersmith & City
Jubilee
Metropolitan
Northern
Piccadilly
Victoria
Waterloo & City
Docklands Light Railway
High Barnet
Cockfosters
Oakwood
Southgate
Totteridge & Whetstone
Woodside Park
West Finchley
Arnos Grove
Bounds Green
Mill Hill East
Finchley Central
Wood Green
Turnpike Lane
Seven Sisters
Tottenham Hale
Blackhorse Road
Walthamstow Central
Stanmore
Canons Park
Edgware
Burnt Oak
Colindale
Queensbury
Kingsbury
Hendon Central
Brent Cross
Golders Green
East Finchley
Highgate
Manor House
Finsbury Park
Archway
Arsenal
Holloway Road
Caledonian Road
Highbury & Islington
Hampstead
Belsize Park
Tufnell Park
Kentish Town
Chalk Farm
Camden Town
Mornington Crescent
Swiss Cottage
St. John's Wood
Finchley Road
West Hampstead
Kilburn
Willesden Green
Dollis Hill
Neasden
Wembley Park
Preston Road
Northwick Park
Kenton
Harrow & Wealdstone
Harrow-on-the-Hill
North Harrow
Pinner
West Harrow
South Kenton
North Wembley
Wembley Central
Stonebridge Park
Harlesden
Willesden Junction
Kensal Green
Queen's Park
Kilburn Park
Maida Vale
Warwick Avenue
Royal Oak
Westbourne Park
Ladbroke Grove
Latimer Road
White City
Shepherds Bush
Goldhawk Road
Holland Park
Kensington (Olympia)
Hammersmith
Barons Court
West Kensington
West Brompton
Fulham Broadway
Parsons Green
Putney Bridge
East Putney
Southfields
Wimbledon Park
Wimbledon
Amersham
Chesham
Chalfont & Latimer
Chorleywood
Rickmansworth
Watford
Croxley
Moor Park
Northwood
Northwood Hills
Uxbridge
Hillingdon
Ickenham
Ruislip
Ruislip Manor
Eastcote
Rayners Lane
South Harrow
Sudbury Hill
Sudbury Town
Alperton
West Ruislip
Ruislip Gardens
South Ruislip
Northolt
Greenford
Perivale
Hangar Lane
Park Royal
North Acton
North Ealing
West Acton
East Acton
Ealing Broadway
Ealing Common
Acton Town
South Ealing
Northfields
Boston Manor
Chiswick Park
Gunnersbury
Turnham Green
Stamford Brook
Ravenscourt Park
Kew Gardens
Richmond
Osterley
Hounslow West
Hounslow East
Hounslow Central
Hatton Cross
Heathrow Terminals 1,2,3
Heathrow Terminal 4
Pimlico
Vauxhall
Oval
Stockwell
Brixton
Clapham North
Clapham Common
Clapham South
Balham
Tooting Bec
Tooting Broadway
Colliers Wood
South Wimbledon
Morden
Kennington
Elephant & Castle
Borough
Epping
Theydon Bois
Debden
Loughton
Buckhurst Hill
Woodford
South Woodford
Snaresbrook
Roding Valley
Chigwell
Grange Hill
Hainault
Fairlop
Barkingside
Newbury Park
Gants Hill
Redbridge
Wanstead
Leytonstone
Leyton
Stratford
Mile End
Bethnal Green
Shoreditch
Whitechapel
Stepney Green
Bow Road
Bow Church
Bromley-by-Bow
All Saints
Limehouse
Shadwell
Westferry
Poplar
Blackwall
West India Quay
Canary Wharf
Heron Quays
South Quay
Crossharbour
Wapping
Rotherhithe
Surrey Quays
East India
Royal Victoria
Custom House
Prince Regent
Canning Town
Beckton
Royal Albert
Beckton Park
Cyprus
Gallions Reach
West Ham
Plaistow
Upton Park
East Ham
Barking
Upney
Becontree
Dagenham Heathway
Dagenham East
Elm Park
Hornchurch
Upminster Bridge
Upminster
Angel
King's Cross St. Pancras
Old Street
Euston
Great Portland Street
Euston Square
Warren Street
Russell Square
Baker Street
Marylebone
Edgware Road
Regent's Park
Goodge Street
Farringdon
Barbican
Chancery Lane
Liverpool Street
Aldgate East
Aldgate
Moorgate
St. Paul's
Bank
Monument
Cannon Street
Mansion House
Blackfriars
Temple
Tower Hill
Tower Gateway
Bayswater
Paddington
Notting Hill Gate
Queensway
Lancaster Gate
Marble Arch
Bond Street
Oxford Circus
Tottenham Court Road
Holborn
Covent Garden
Leicester Square
Piccadilly Circus
Charing Cross
Embankment
Green Park
Hyde Park Corner
Knightsbridge
High Street Kensington
Gloucester Road
South Kensington
Sloane Square
Victoria
St. James's Park
Westminster
Waterloo
Lambeth North
London Bridge

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

$$\left.\begin{aligned} \sigma_i &= \sum_j a_{ij} \\ \sigma_j &= \sum_i a_{ij} \end{aligned}\right\} \quad \sigma = \sum_i \sigma_i = \sum_j \sigma_j = \sum_i \sum_j a_{ij}$$

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

Entries and Exits

$$T_i = \sum_j T_{ij}$$

$$T_j = \sum_i T_{ij}$$

$$\left.\begin{array}{l} T_i = \sum_j T_{ij} \\ T_j = \sum_i T_{ij} \end{array}\right\} T = \sum_i T_i = \sum_j T_j = \sum_i \sum_j T_{ij}$$

Changes in

Trip Volumes

$$\left.\begin{array}{l} \Delta_i = T_i - T_i' \\ \Delta_j = T_j - T_j' \end{array}\right\} \sum_i \Delta_i = \sum_i \Delta_j = 0$$

Weighted Betweenness

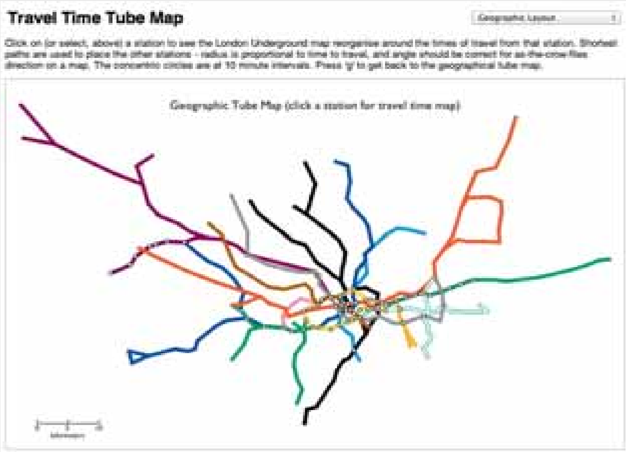
Centrality

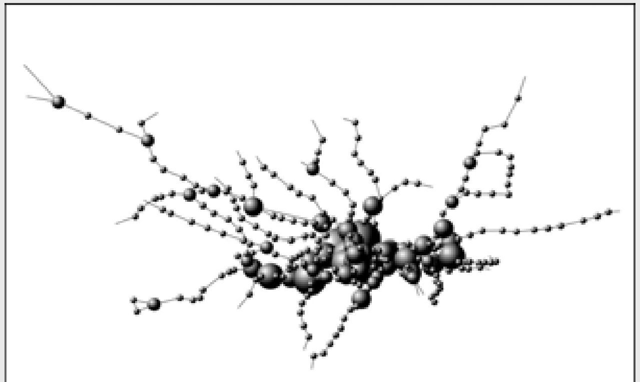
$$p_{ijk} = \frac{\sigma_{ikj}}{\sigma_{ij}} = \frac{\sigma_{ikj}}{\sum_\ell \sigma_{i\ell j}} \quad , \quad \sum_k p_{ikj} = 1$$

$$\tilde{C}_k = \sum_i \sum_j T_{ij} p_{ikj} = \sum_i \sum_j T_{ij} \frac{\sigma_{ikj}}{\sigma_{ij}}$$

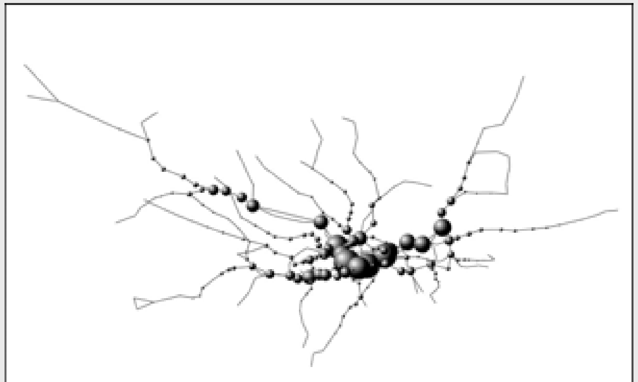
A Preliminary Analysis (1)

The Minimal Tube Network and the Three Centrality Indices

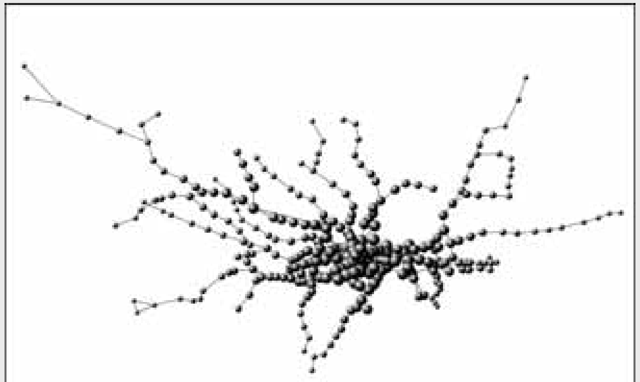




Degree



Betweenness



Closeness

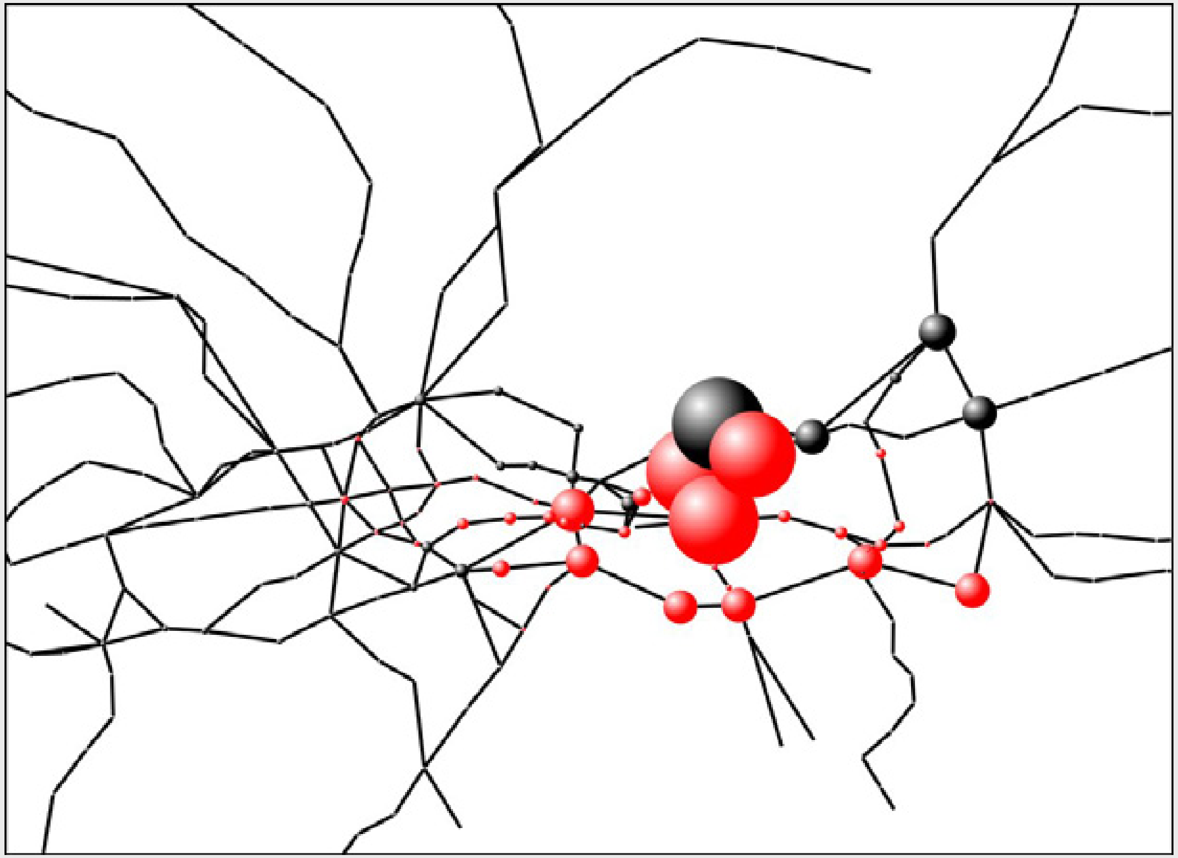
A Preliminary Analysis (2)

- Top Stations
- By Centrality

Station	d_i	Station	$\hat{C}_i$	Station	$\hat{L}_i$
Baker Street	7	*Green Park*	16399	*Green Park*	2.137
King's Cross	7	Waterloo	15644	Westminster	2.107
Bank	6	Bank	15008	Bond Street	2.101
Earl's Court	6	Baker Street	14441	Oxford Circus	2.089
Green Park	6	Westminster	14139	Waterloo	2.089
Oxford Circus	6	Bond Street	11429	Bank	2.074
Waterloo	6	***Liverpool Street***	11186	Baker Street	2.071
Canning Town	5	Stratford	10814	Victoria	2.065
Liverpool Street	5	Mile End	10302	Hyde Pk Corner	2.053
Paddington	5	Bethnal Green	10017	Embankment	2.041
Shadwell	5	Finchley Road	8905	Piccadilly Circus	2.041
Turnham Green	5	Earl's Court	8706	St. James's Park	2.035
Acton Town	4	King's Cross	8679	Regent's Park	2.032
Bond Street	4	Wembley Park	7968	King's Cross	2.029
Camden Town	4	South Ken	7182	***Liverpool Street***	2.026
Canada Water	4	Euston	7156	Marble Arch	2.026
Canary Wharf	4	Gloucester Rd	7042	Tottenham Ct Rd	2.026
Embankment	4	Paddington	7028	Moorgate	2.020
Euston	4	Victoria	6558	Charing Cross	2.017
Finchley Road	4	Harrow-o-t-Hill	6253	Great Portland St	2.017

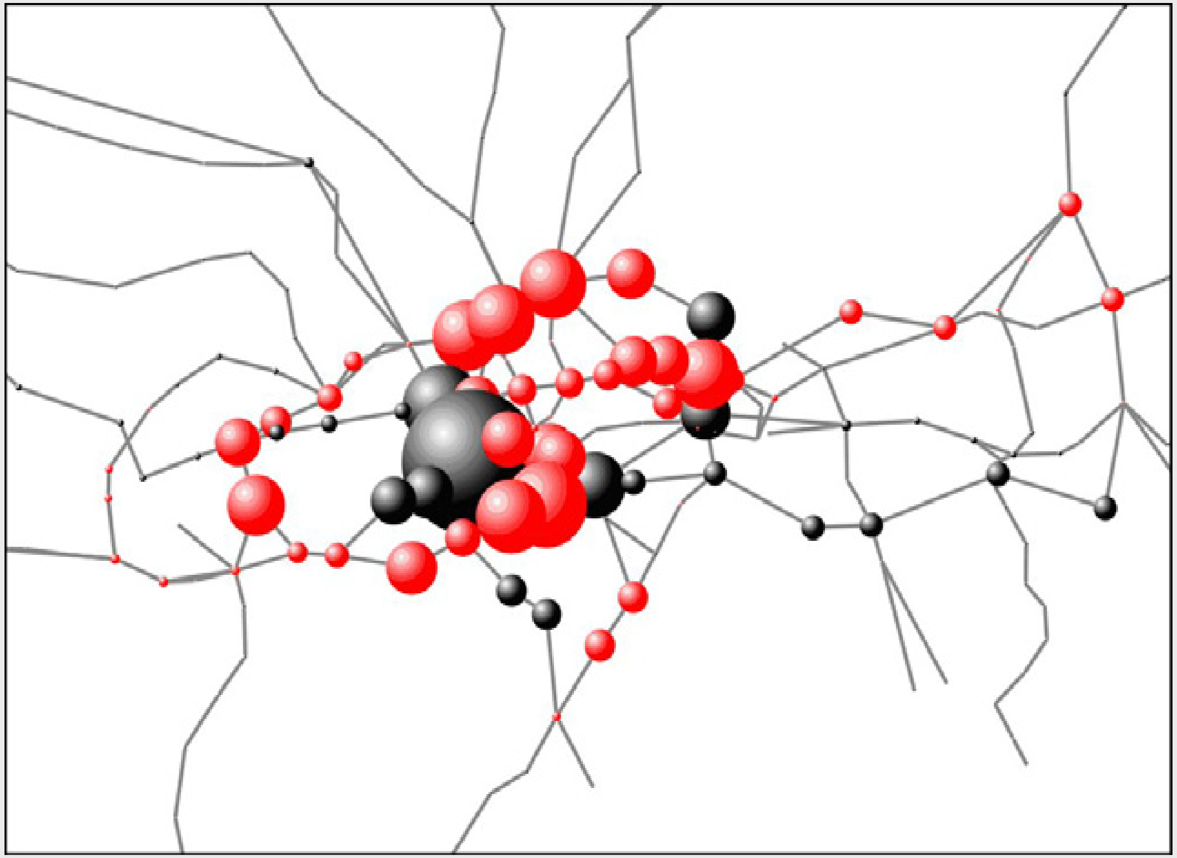
A Preliminary Analysis (3)

Closing Liverpool Street

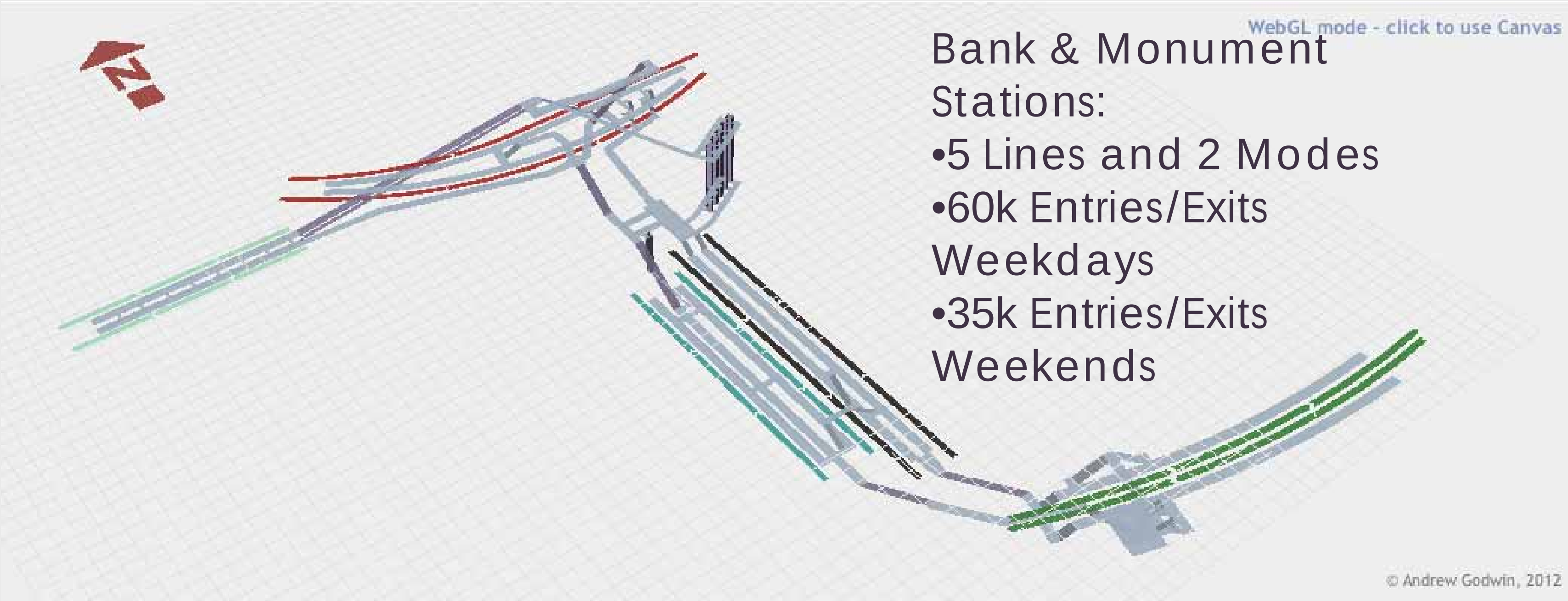


A Preliminary Analysis (3)

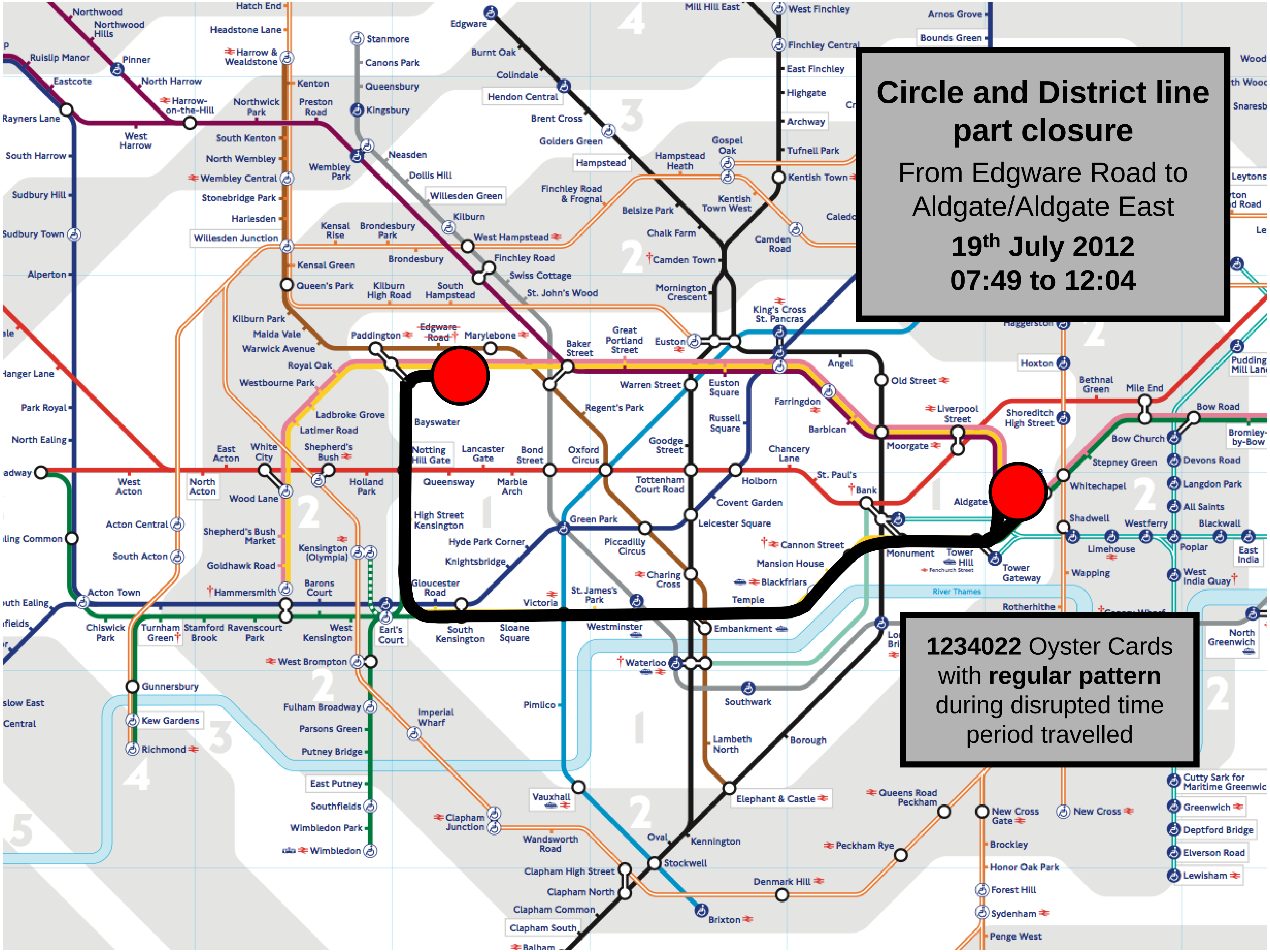
Closing Green Park



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.



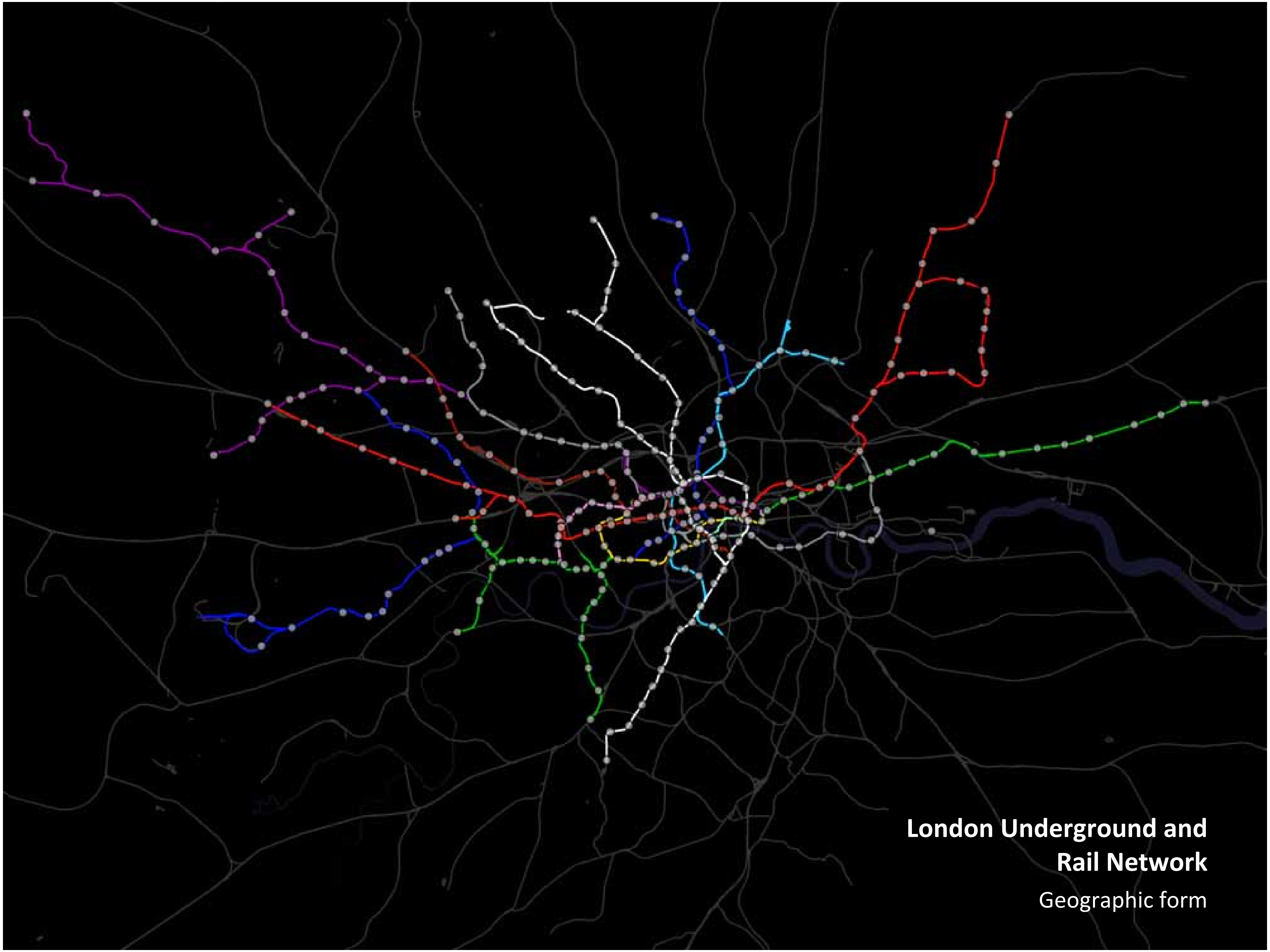
Circle and District line part closure

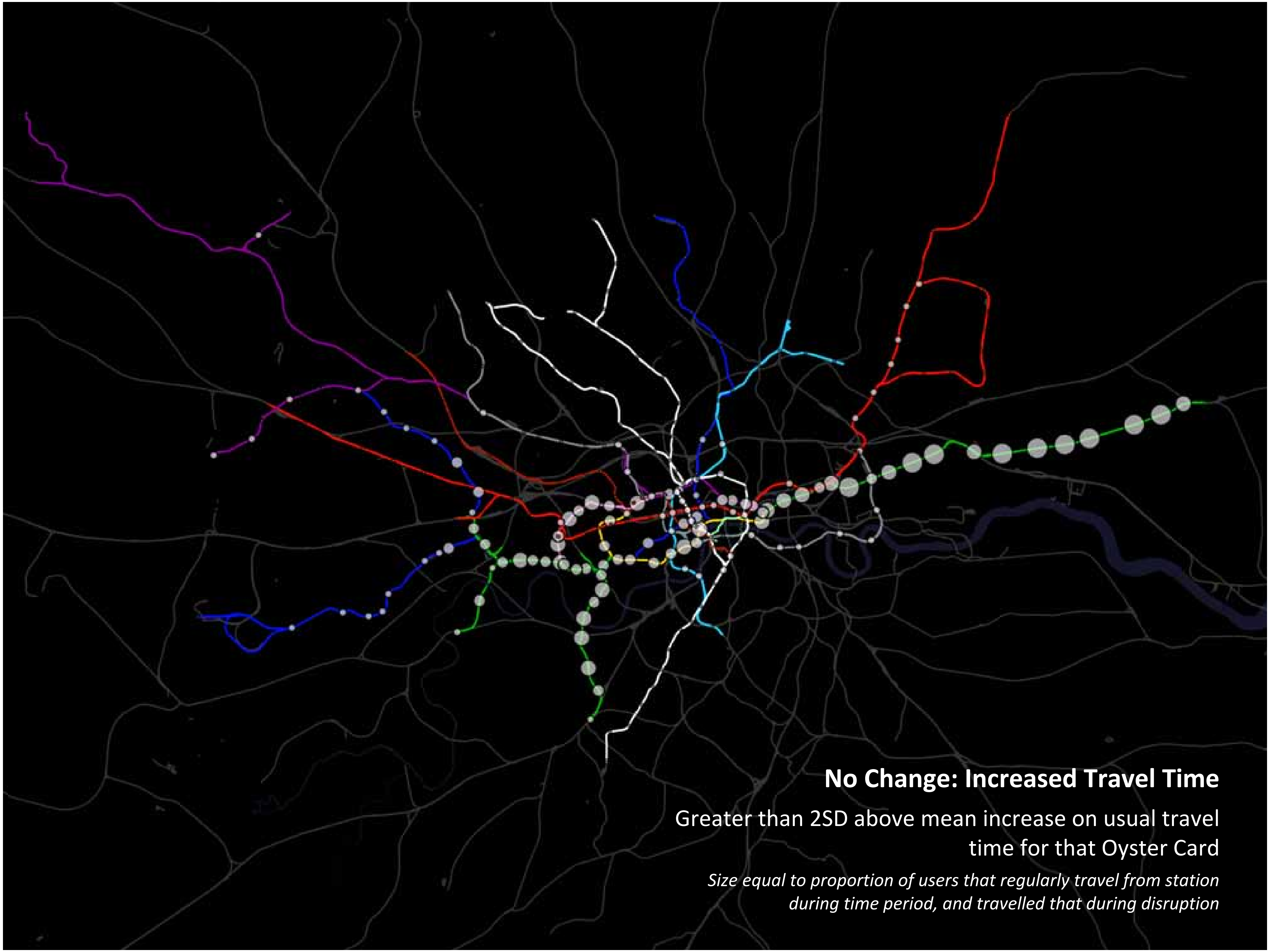
From Edgware Road to Aldgate/Aldgate East

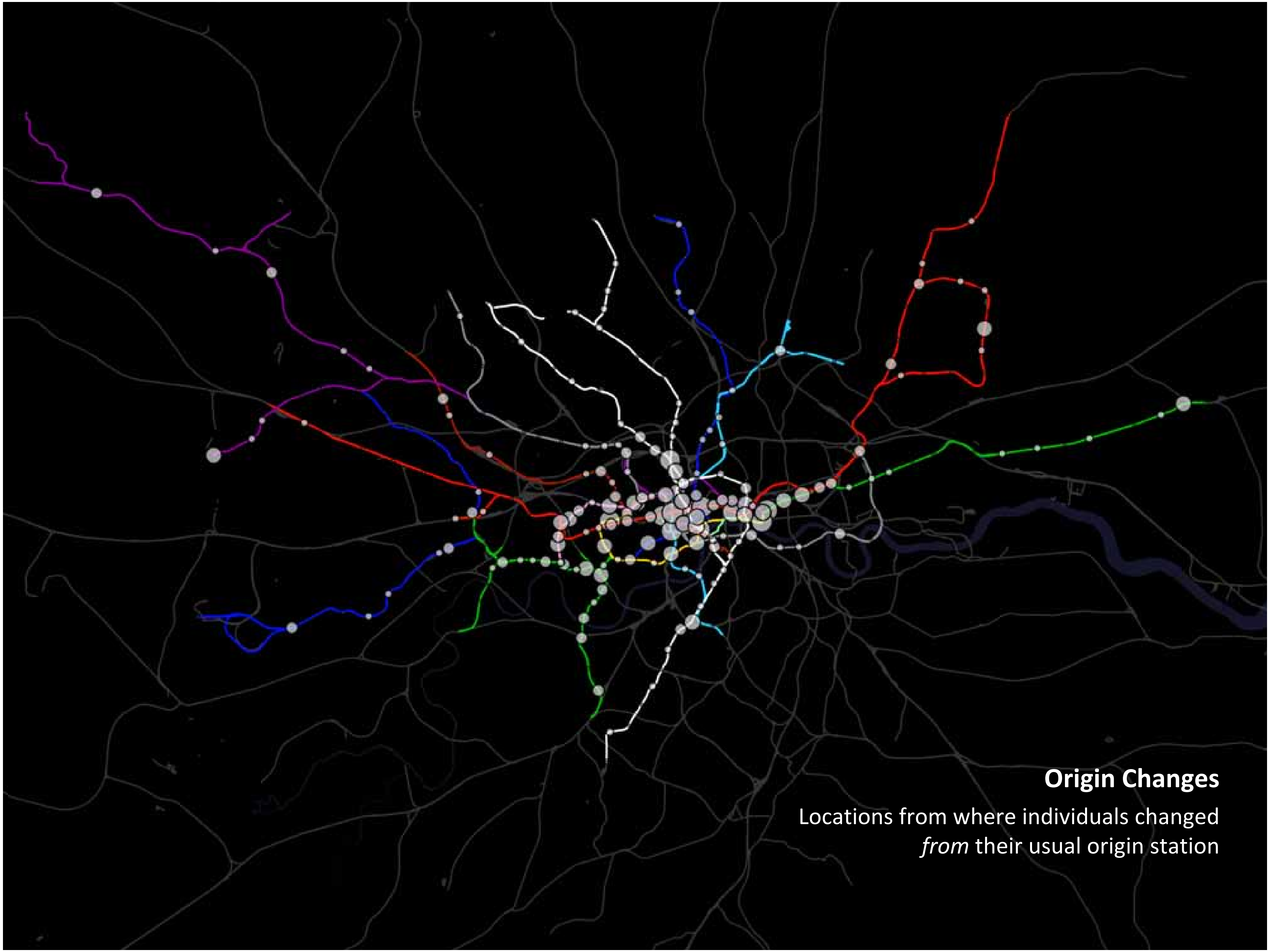
19th July 2012

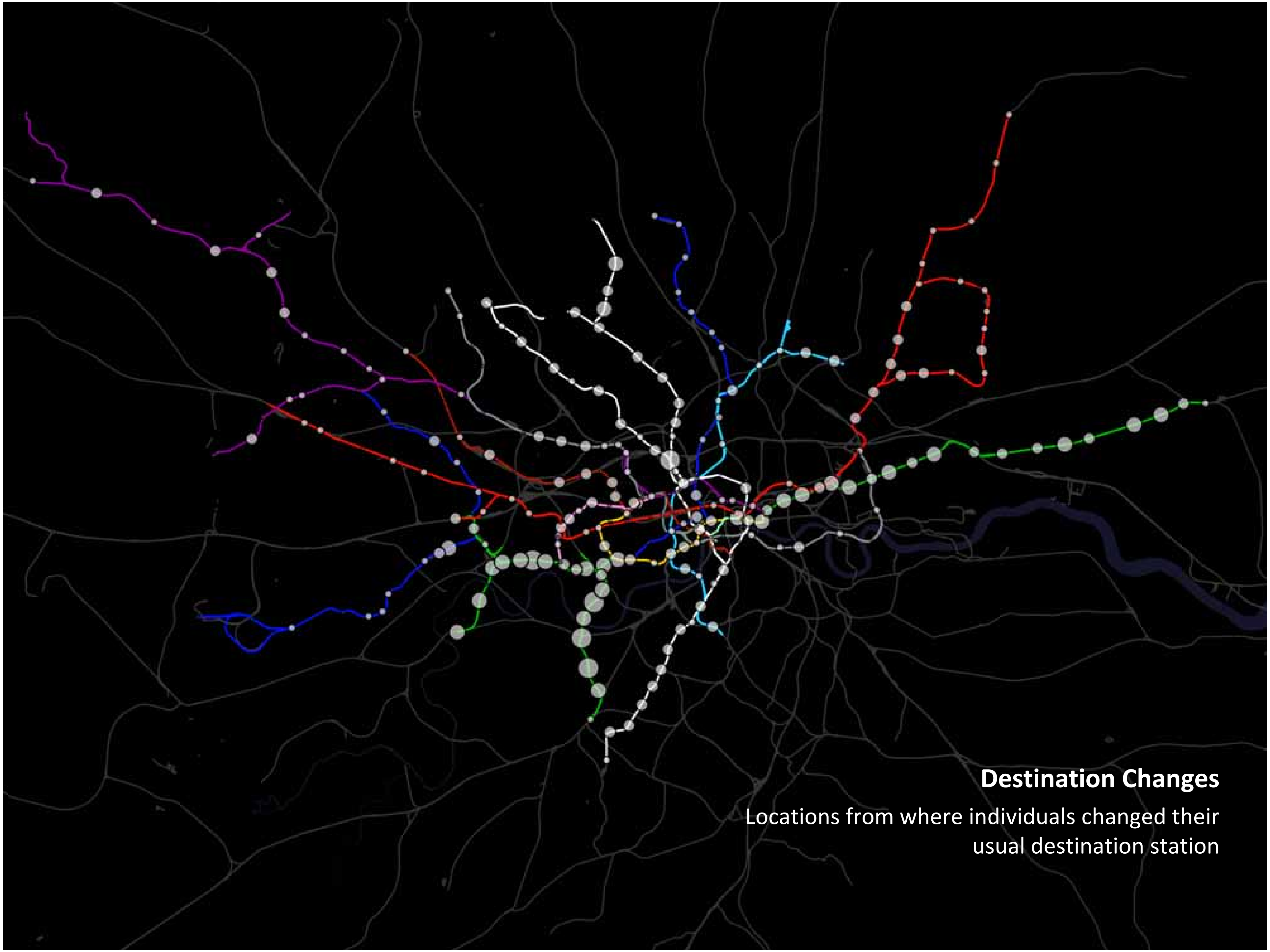
07:49 to 12:04

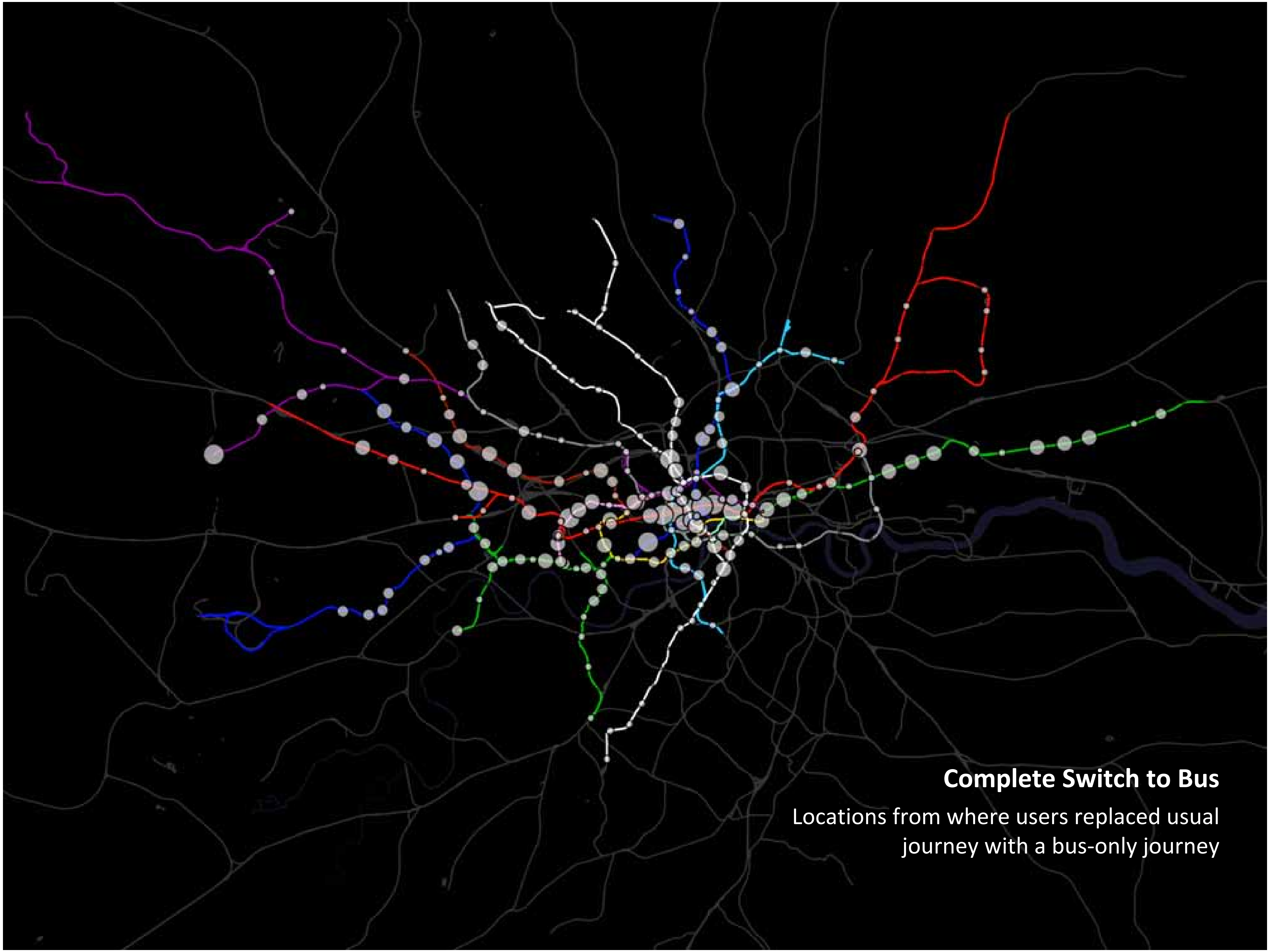
1234022 Oyster Cards with **regular pattern** during disrupted time period travelled

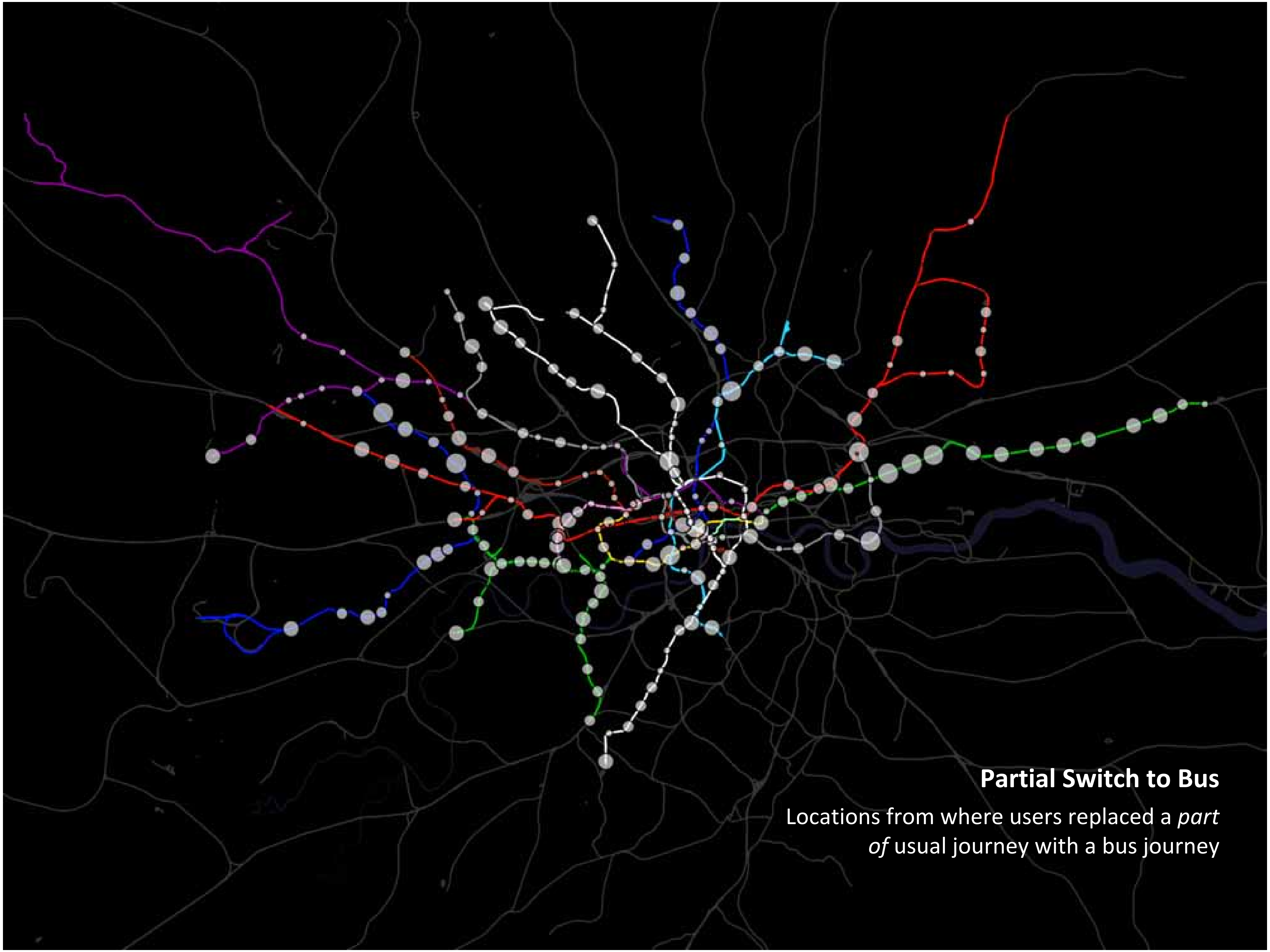

London Underground and
Rail Network
Geographic form

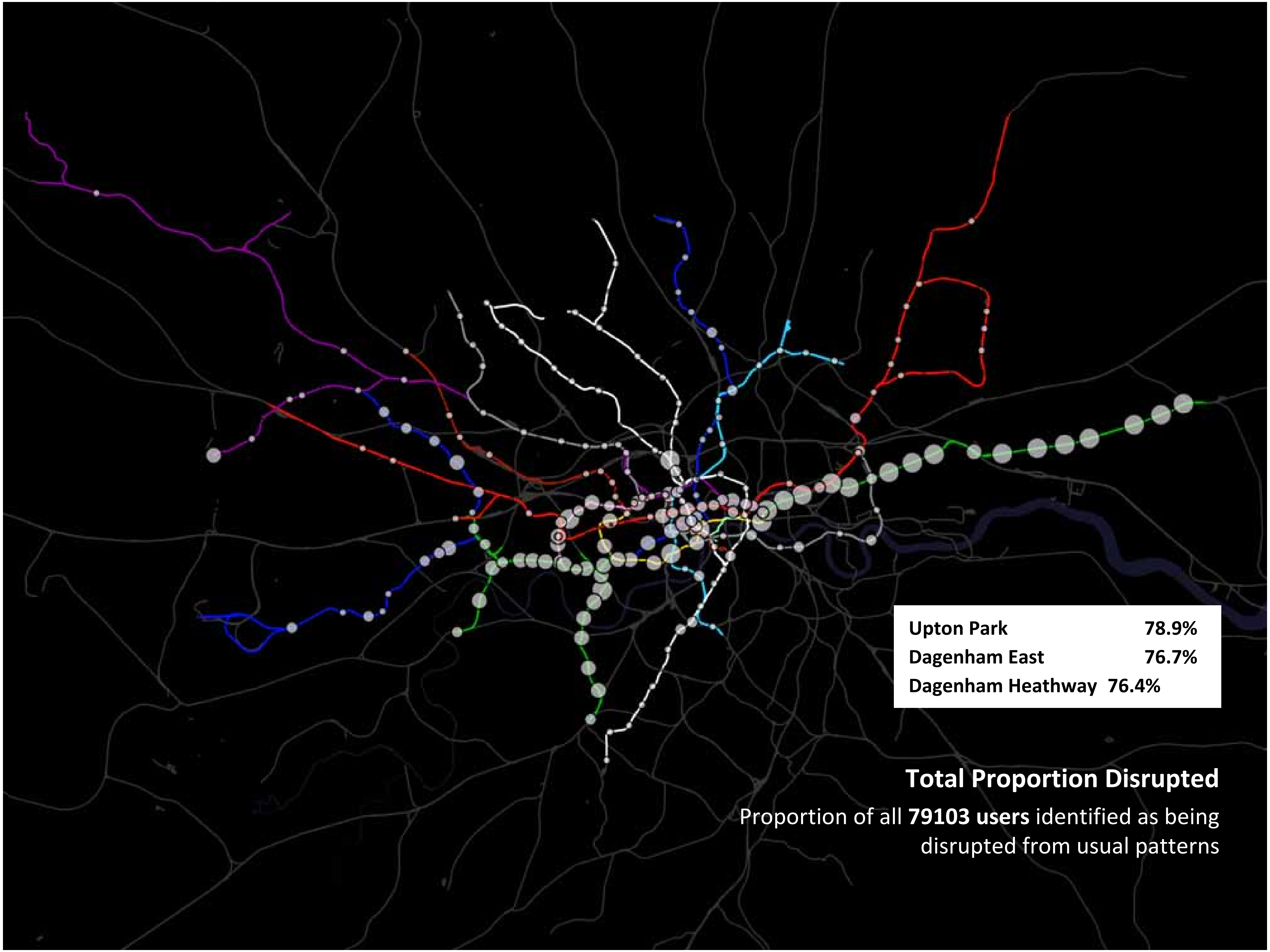

No Change: Increased Travel Time
Greater than 2SD above mean increase on usual travel time for that Oyster Card
Size equal to proportion of users that regularly travel from station during time period, and travelled that during disruption


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

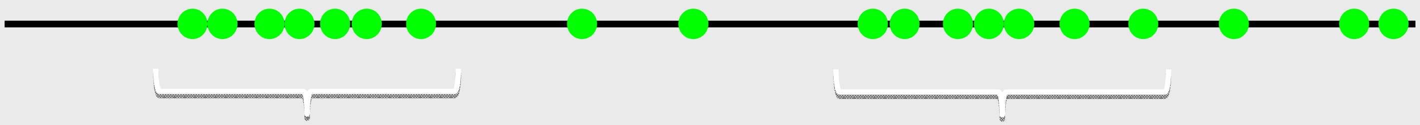

Partial Switch to Bus
Locations from where users replaced a *part of* usual journey with a bus journey


Upton Park 78.9%
Dagenham East 76.7%
Dagenham Heathway 76.4%
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



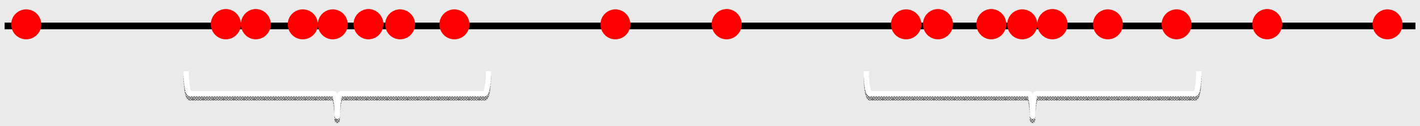
Cluster 1

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

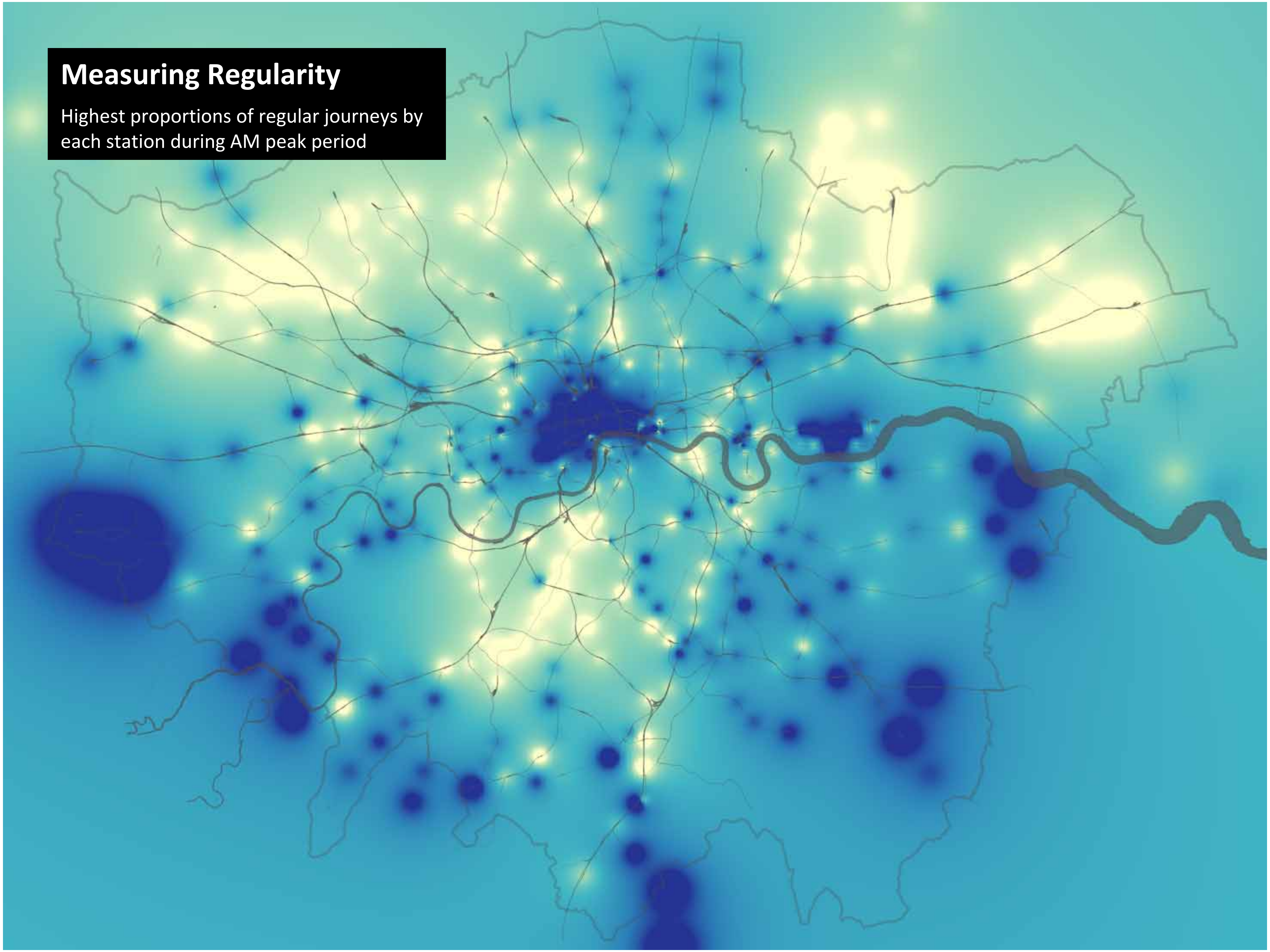


Cluster 3

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

Cluster 4

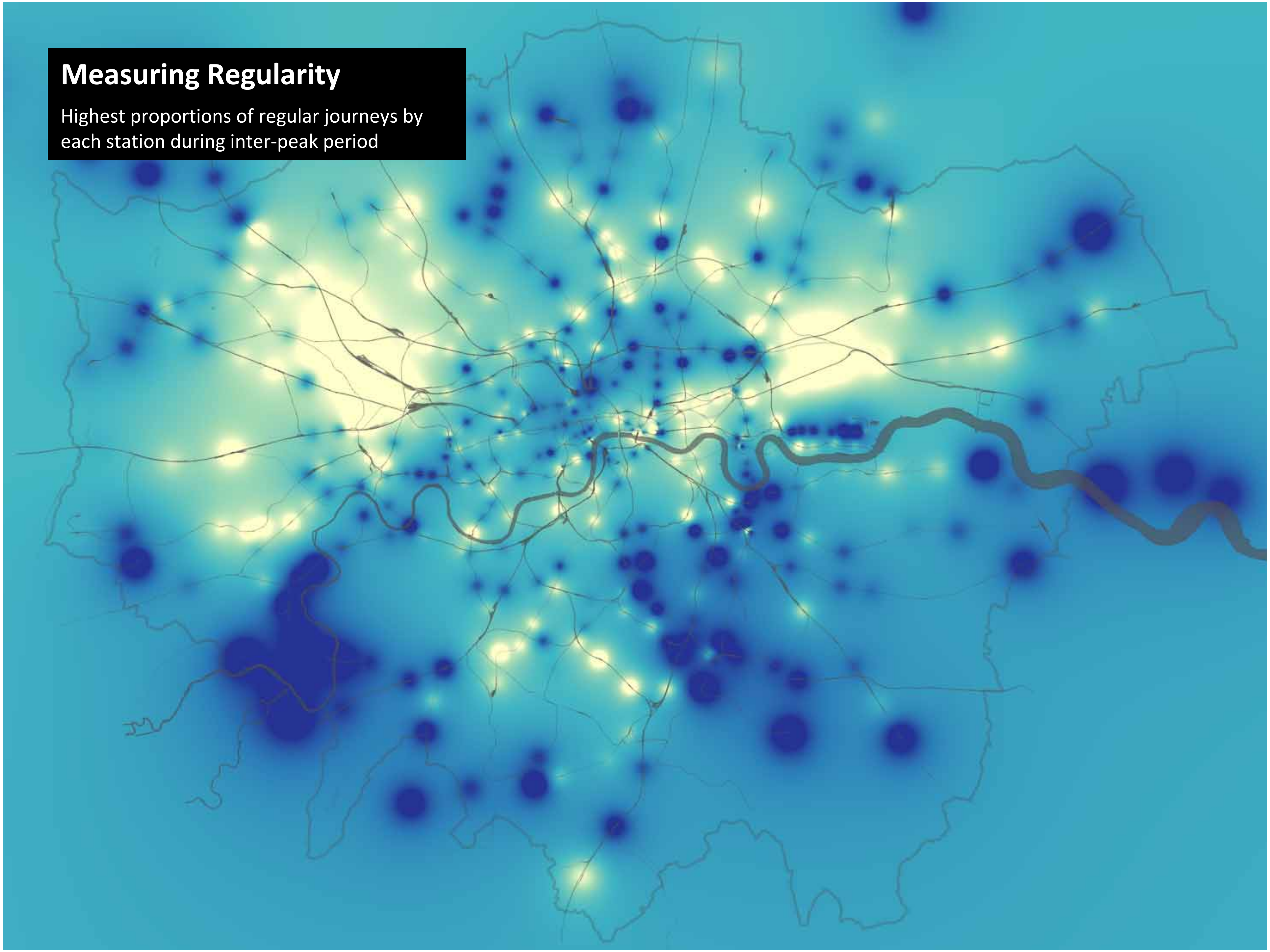
Mean: 18:06; SD = 18.1
Max: 17:26; Min: 19:28
Proportion of Days = 0.6



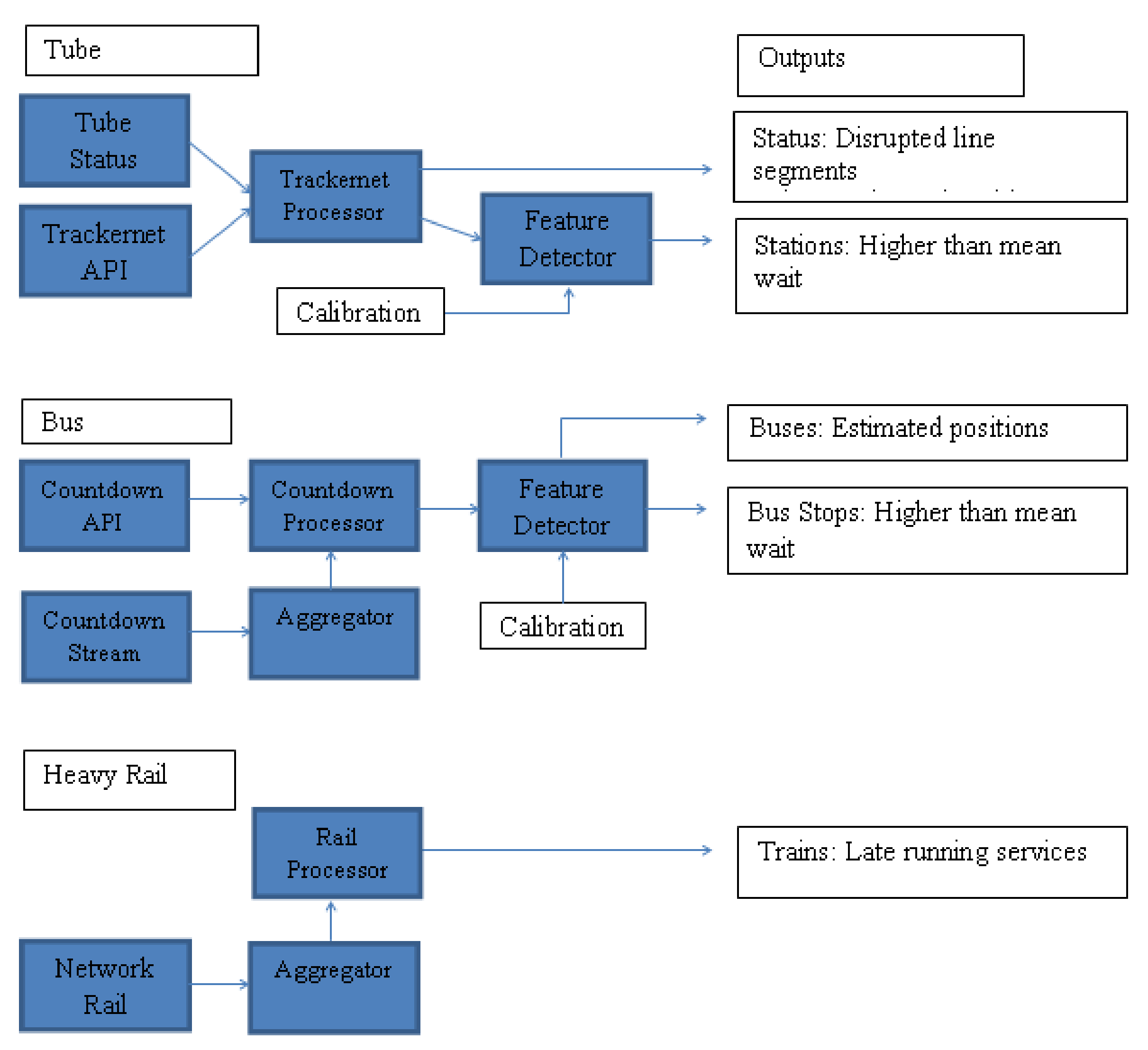
Measuring Regularity
Highest proportions of regular journeys by each station during AM peak period

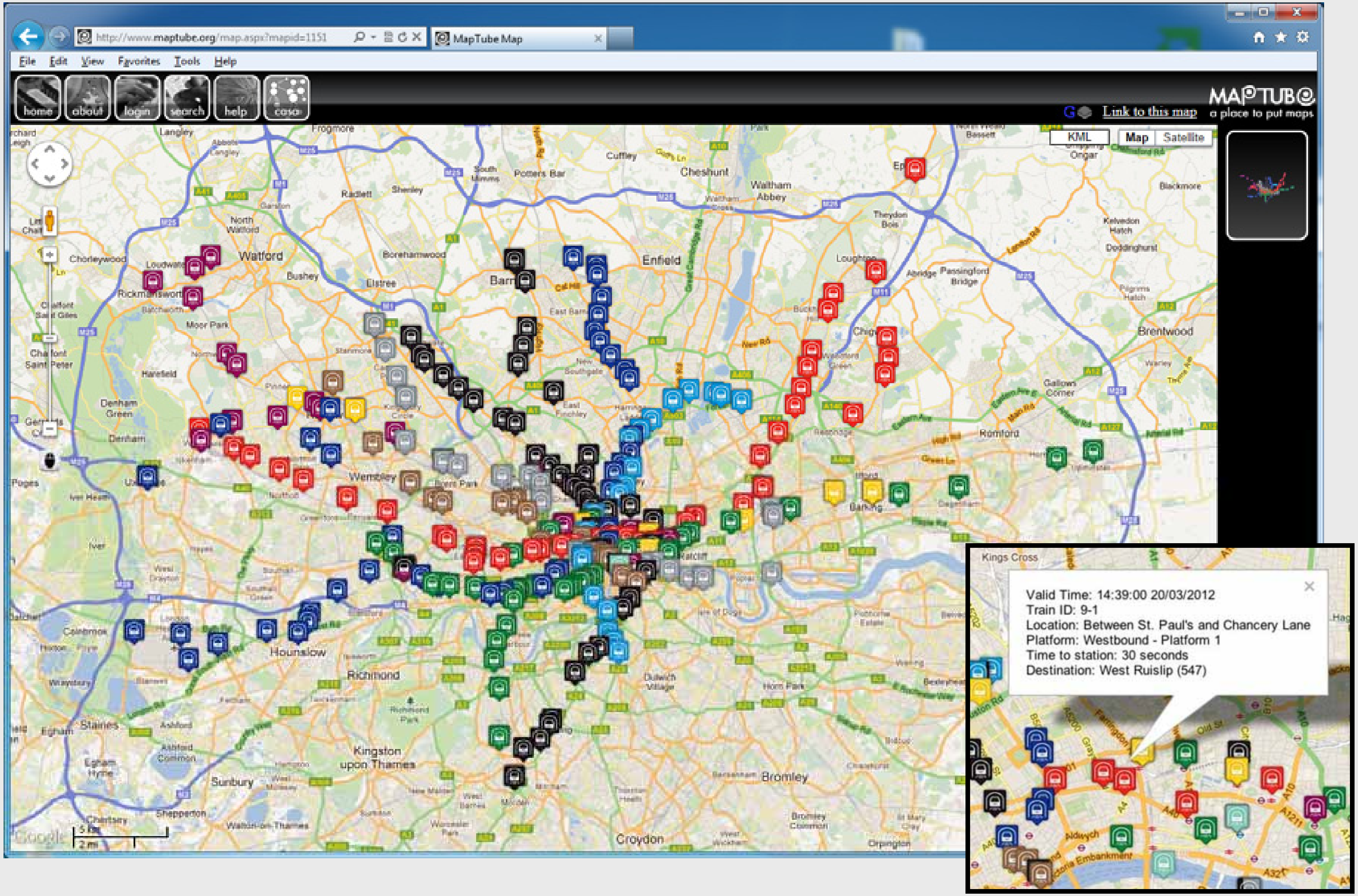
Measuring Regularity

Highest proportions of regular journeys by each station during inter-peak period

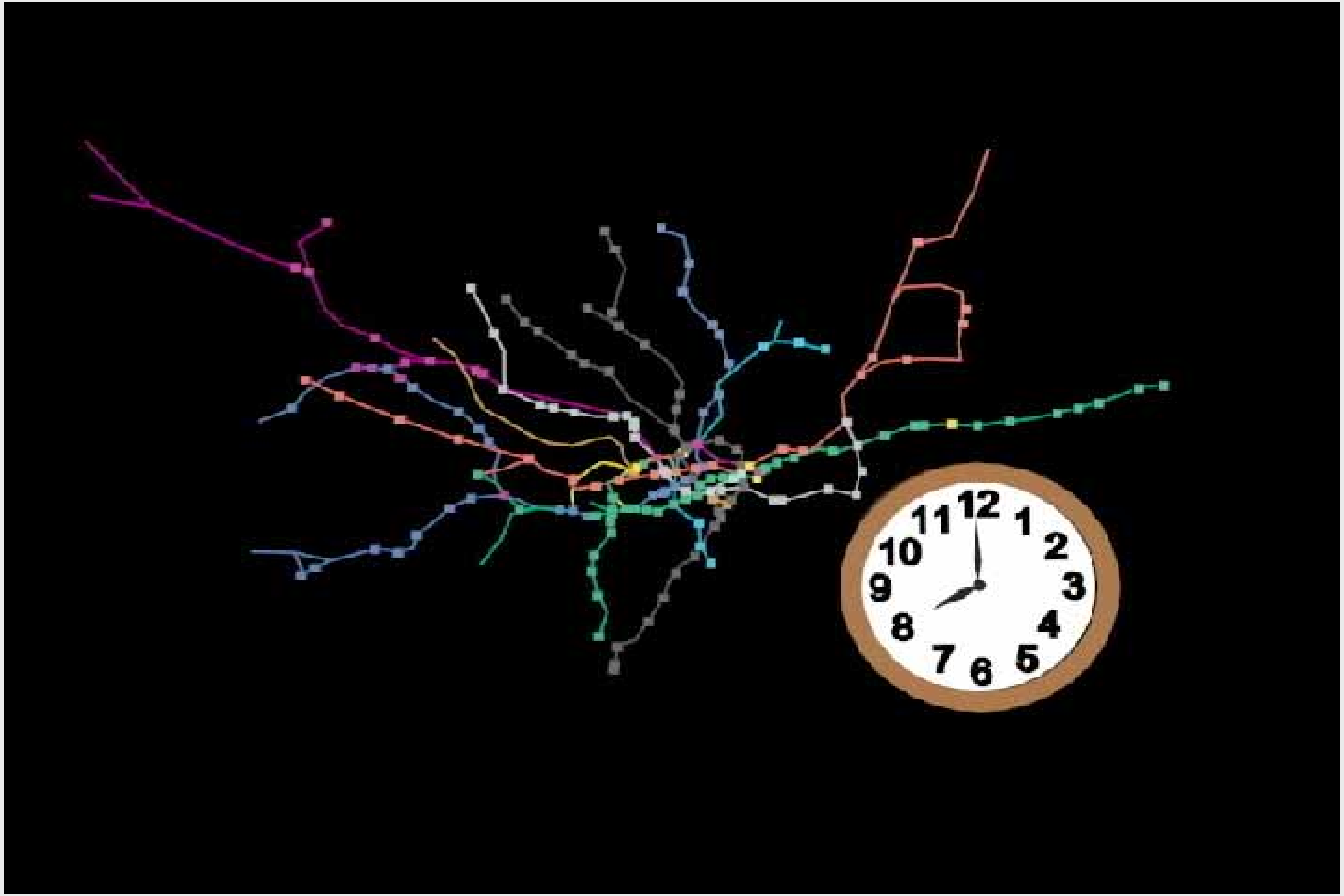


The Public Transport System in Terms of Vehicle Flows

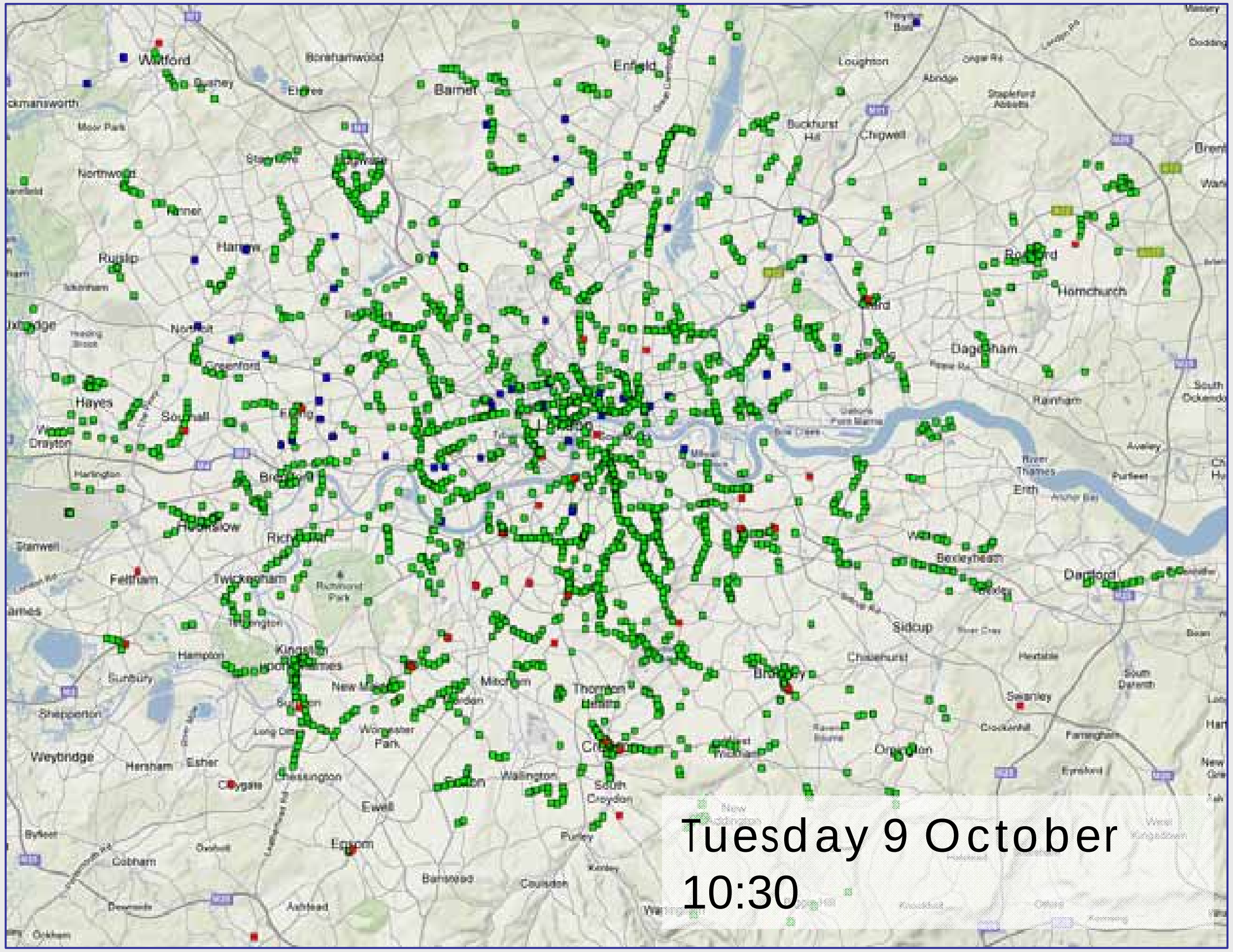




http://www.maptube.org/map.aspx?mapid=1151
MapTube Map
MAPTUBE
a place to put maps
Link to this map
Valid Time: 14:39:00 20/03/2012
Train ID: 9-1
Location: Between St. Paul's and Chancery Lane
Platform: Westbound - Platform 1
Time to station: 30 seconds
Destination: West Ruislip (547)


12
1
2
3
4
5
6
7
8
9
10
11

Delays from Tube, National Rail and Bus Fused



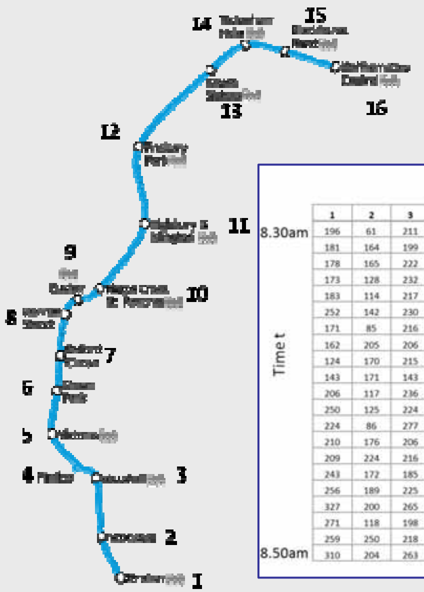
Key

- 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

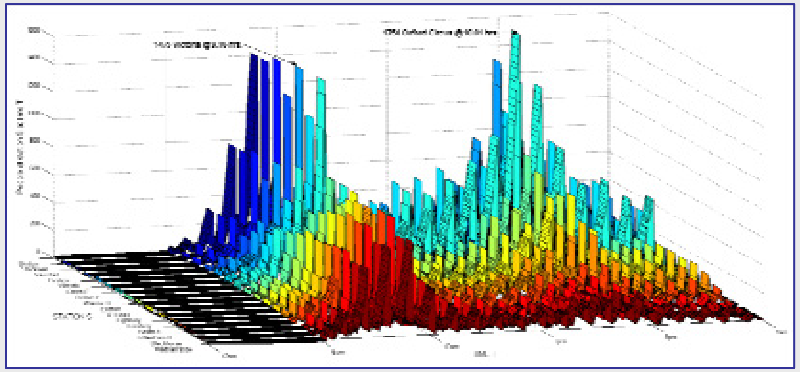
Locational Dynamics of Demand

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 <u>**transfer entropy**</u> to do this. I don't have time to say much about this but here is a picture about this for one line



Stations S — Euston @ 8.29am

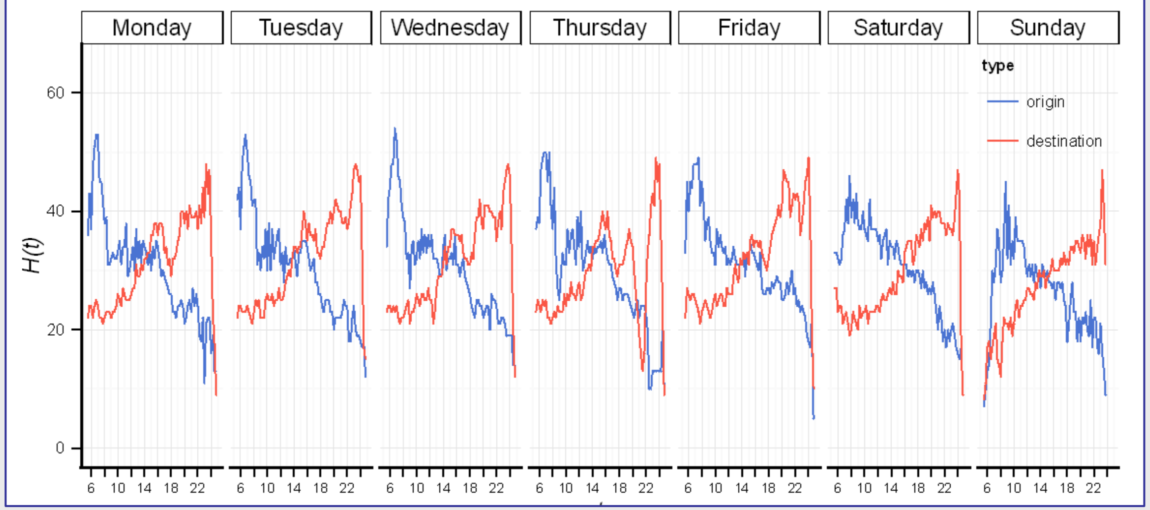
Time t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
8.30am	196	61	211	132	400	152	488	358	252	252	188	170	128	180	144	159
	181	164	199	138	304	144	489	415	210	647	228	157	218	76	55	160
	178	165	222	179	288	188	395	280	604	274	211	137	178	178	111	137
	173	128	232	91	276	167	326	206	299	246	180	243	227	65	72	153
	183	114	217	132	322	279	342	562	244	278	143	153	111	135	148	130
	252	142	230	87	275	212	614	313	214	264	262	218	222	103	72	156
	171	85	216	162	375	203	469	257	251	155	136	183	82	201	140	128
	162	205	206	161	299	372	424	286	206	158	193	248	116	132	60	165
	124	170	215	176	366	265	422	301	[illegible]	312	181	200	106	241	121	158
	143	171	143	138	470	200	427	262	$\boxed{300}$	170	202	254	223	147	46	207
	206	117	236	161	381	315	431	190	178	241	201	133	142	213	120	186
	250	125	224	138	320	253	310	325	223	231	292	144	239	149	42	202
	224	86	277	163	367	221	467	216	202	281	173	123	170	209	112	149
	210	176	206	173	313	210	400	301	242	232	164	203	206	121	56	162
	209	224	216	178	297	273	447	266	192	361	138	177	176	200	164	94
	243	172	185	102	299	143	452	292	313	223	181	257	176	90	73	110
	256	189	225	127	348	216	520	263	220	218	163	211	95	179	162	112
	327	200	265	80	271	220	475	376	191	203	254	194	211	110	62	145
	271	118	198	148	312	239	549	270	198	221	213	179	119	208	119	142
	259	250	218	210	258	188	499	276	235	163	216	183	213	92	61	141
8.50am	310	204	263	163	338	286	508	285	170	288	252	135	162	176	97	171



$$T_{YX} = \sum_{t=1} p(y_{t+1}, y_t, x_t) \log \frac{p(y_{t+1} | y_t, x_t)}{p(y_{t+1} | y_t)}$$

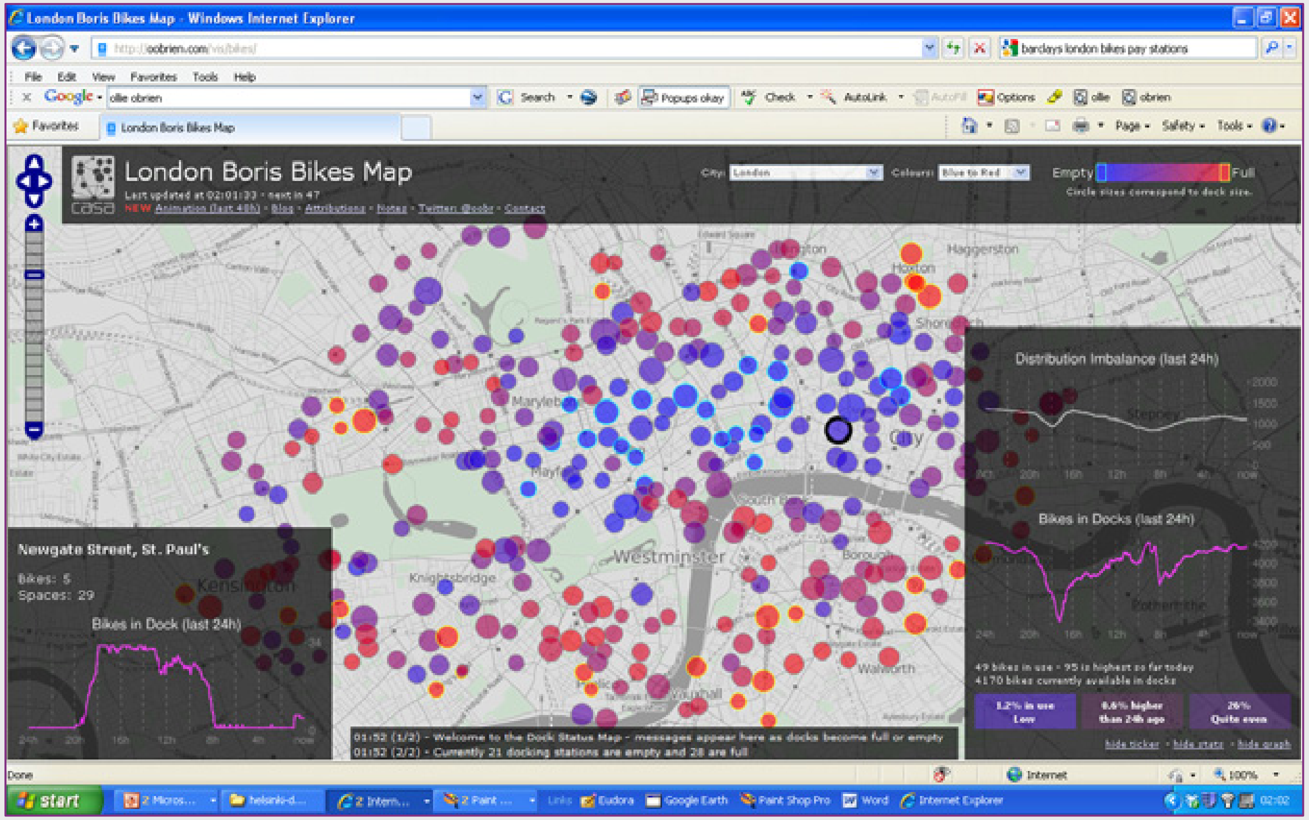
Second we are working with the Oyster data again with Melanie Bosredon in out group and Marc Barthelemy in Paris on extracting clusters from the travel data using a new method of defining intensity. I will show this as a simple movie of origin and destination intensities as they change over time of day.

Number of hotspots (stations) vs. time



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 world wide 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

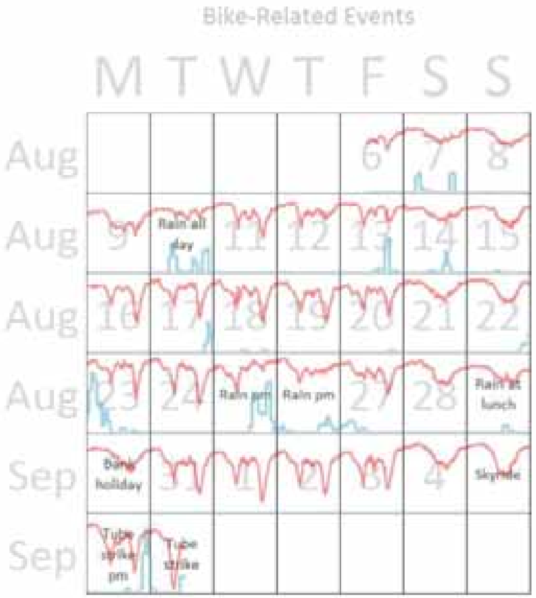


Animations of Public Bike Movements

Animations of Changes in the Bike Nodes: Docking

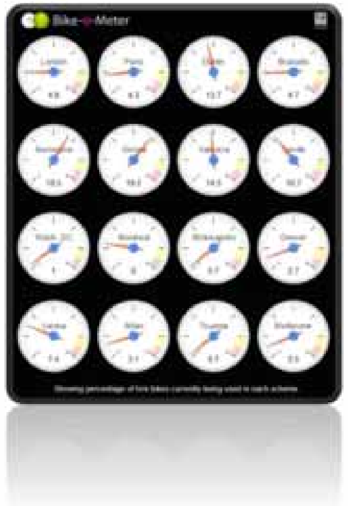
More Analysis

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



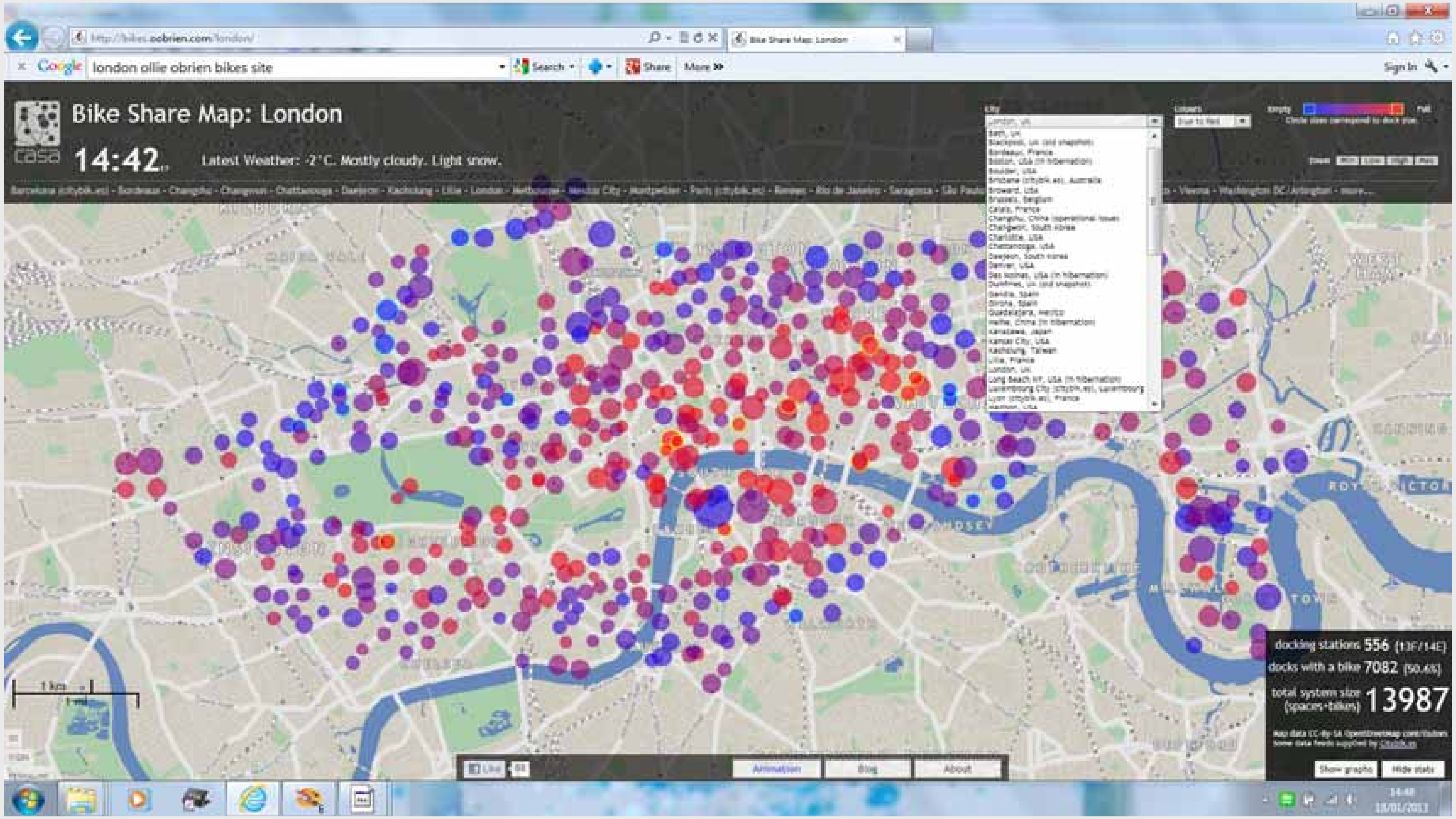
Bike-o-Meter
casa.ucl.ac.uk/bom

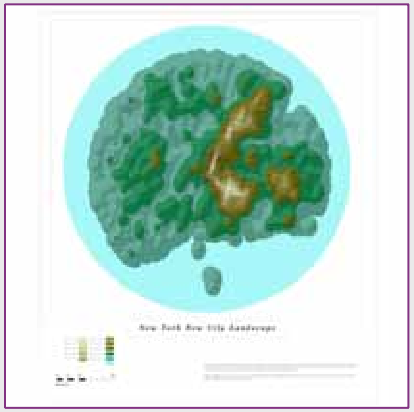
- Tweet-o-Meter for bikes
  - Steven Gray (@frogo)
  - Using Google Gauges
- See the real life Tweet-o-Meters at the new British Library "Growing Knowledge" exhibition
  - Should be easy to hack to show the Bike-o-Meters instead ☺

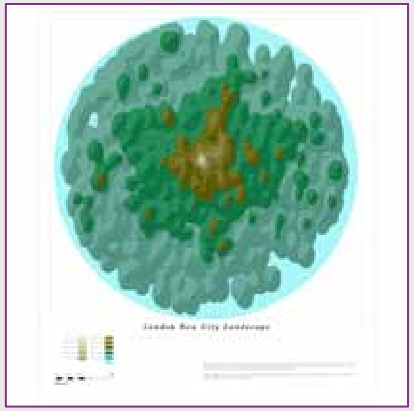


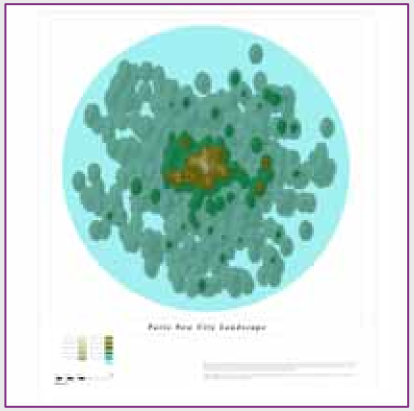
The Website: Real Time Visualisation of Origins and Destinations Activity

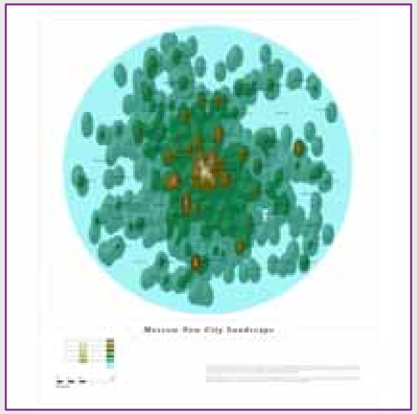
http://bikes.oobrien.com/london/









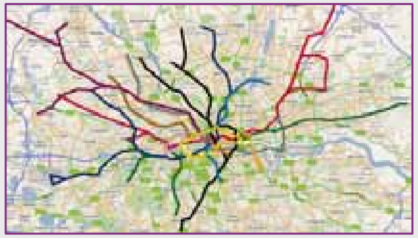


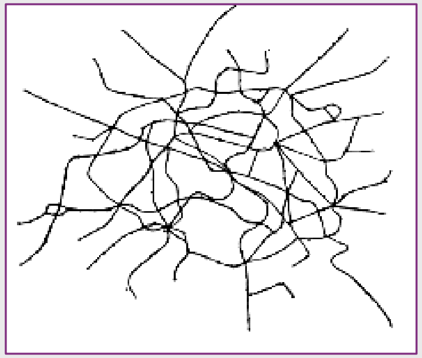
New York

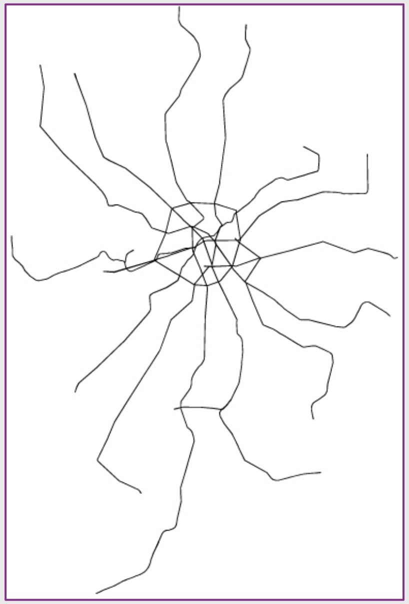
London

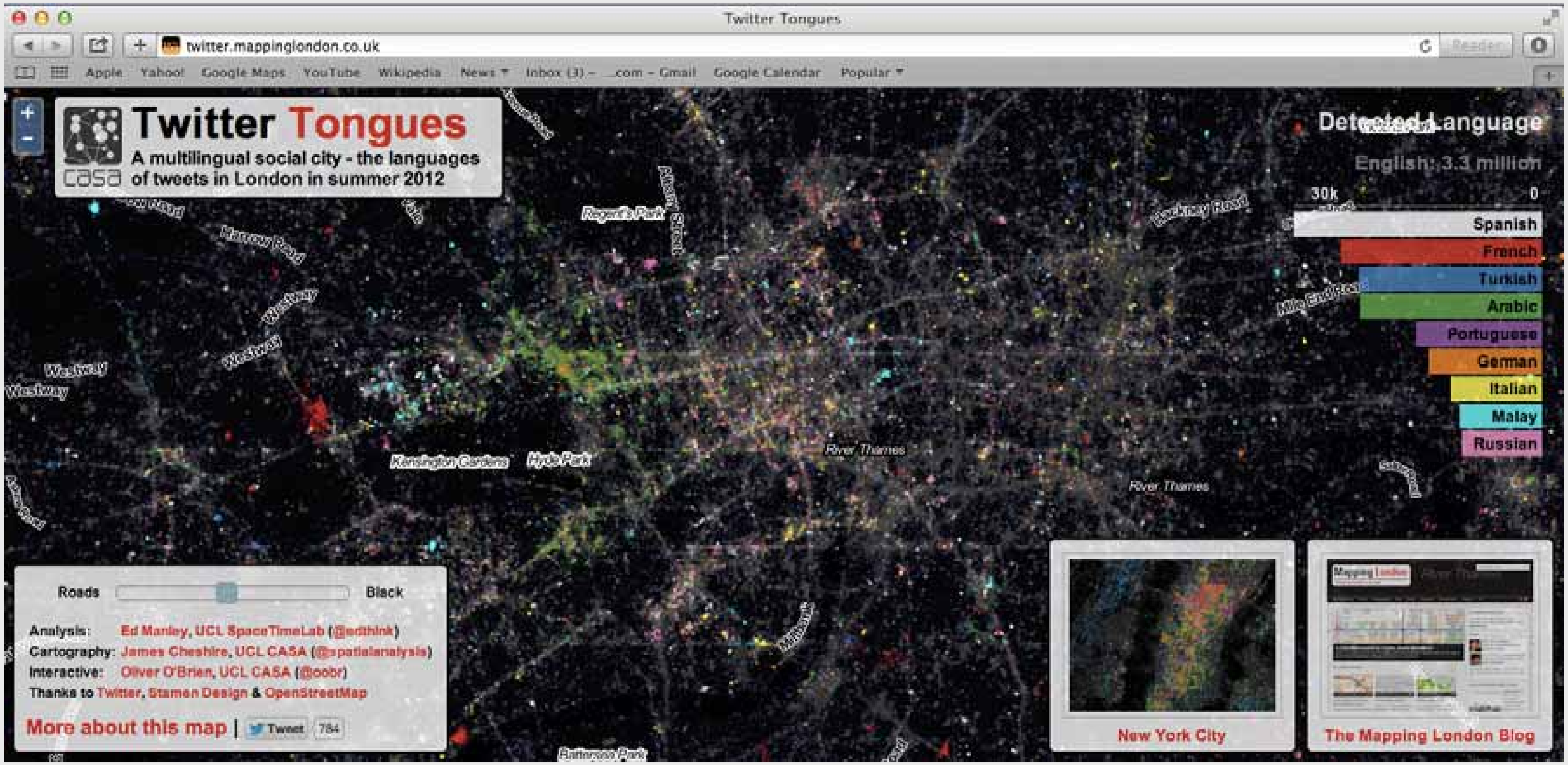
Paris

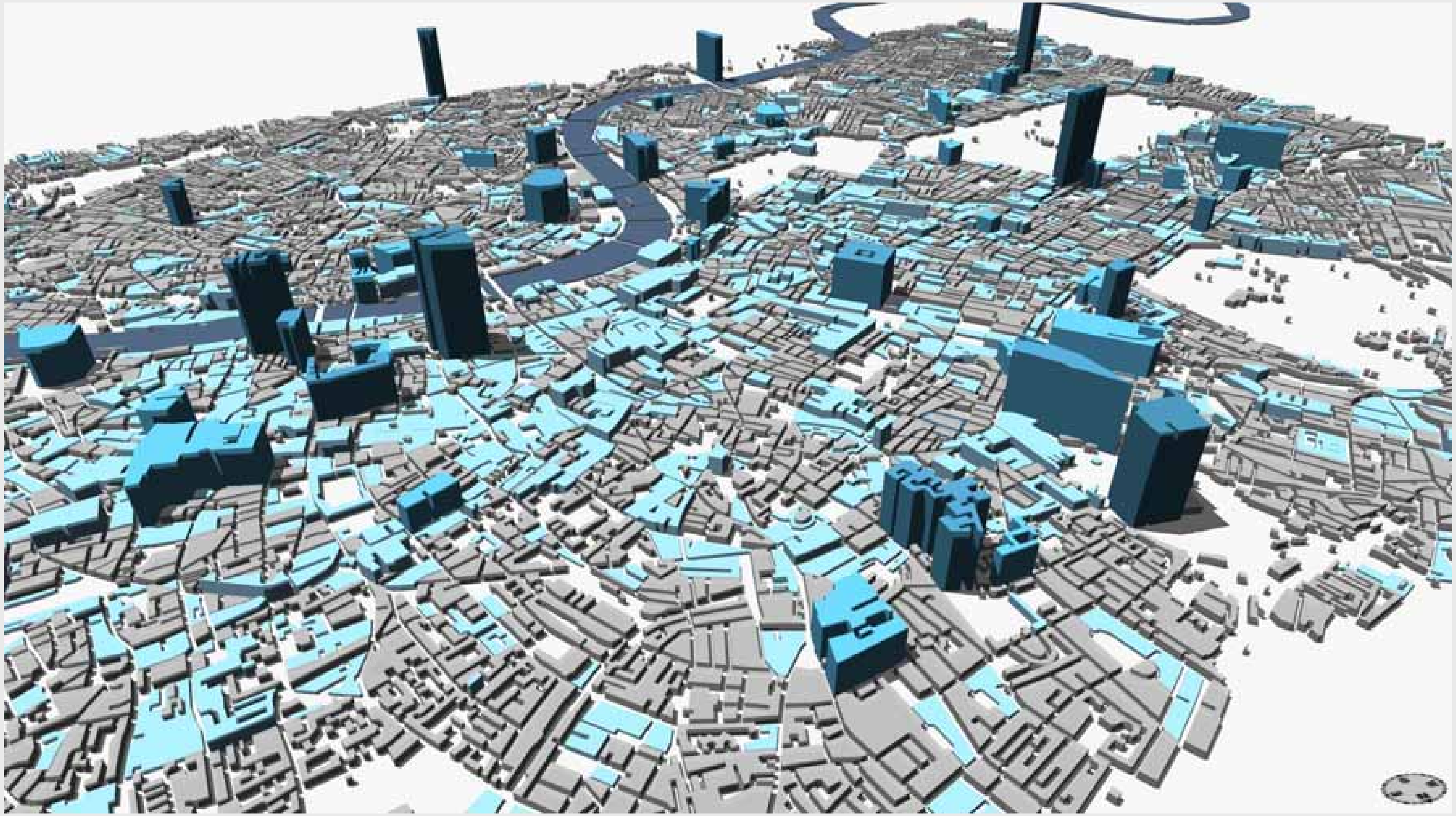
Moscow








Twitter Tongues
twitter.mappinglondon.co.uk
Apple Yahoo! Google Maps YouTube Wikipedia News Inbox (3) – ...com – Gmail Google Calendar Popular
Twitter Tongues
A multilingual social city - the languages of tweets in London in summer 2012
casa
Detected Language
English: 3.3 million
30k
0
Spanish
French
Turkish
Arabic
Portuguese
German
Italian
Malay
Russian
Regent's Park
Harrow Road
Westway
Kensington Gardens
Hyde Park
River Thames
Battersea Park
Roads
Black
Analysis: Ed Manley, UCL SpaceTimeLab (@edthink)
Cartography: James Cheshire, UCL CASA (@spatialanalysis)
Interactive: Oliver O'Brien, UCL CASA (@oobr)
Thanks to Twitter, Stamen Design & OpenStreetMap
More about this map | Tweet 784
New York City
The Mapping London Blog



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 IJGIS

International Journal of Geographical Information Science, 2014
http://dx.doi.org/10.1080/13658816.2014.914521

Taylor & Francis
Taylor & Francis Group

Detecting the dynamics of urban structure through spatial network analysis

Chen Zhong[a*], Stefan Müller Arisona[a,b], Xianfeng Huang[c], Michael Batty[d] and Gerhard Schmitt[a]

References

Manley, E., Chen, Z., and Batty, M. (2016) Spatiotemporal Variation in Travel Regularity through Transit User Profiling, to be submitted.

O'Brien, O, Cheshire, J. and Batty (2014) Mining Bicycle Sharing Data for Generating Insights in Sustainable Transport Systems, **Journal of Transport Geography**, **34**, 262–273

Roth C., Kang S. M., Batty, M., and Barthelemy, M. (2011) Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. **PLoS ONE 6(1):** e15923. doi:10.1371/journal.pone.0015923

Zhong, C., Arisona, S. M., Huang, X., Schmitt, G. and Batty, M. (2014)) Detecting the Dynamics of Urban Structure through Spatial Network Analysis, **International Journal of Geographical Information Science**, http://dx.doi.org/10.1080/13658816.2014.914521

Zhong, C., Batty, M., Manley, E., Wan, J., Wang, Z., Che, F., and Schmitt, G. (2016) Variability in Regularity: Mining Temporal Mobility Patterns in London, Singapore and Beijing using Smart-Card Data., **PLOS One**, http://dx.doi.org/10.1371/journal.pone.0149222

Zhong, C., Huang, X., Arisona, S. M., Schmitt, G., and Batty, M. (2014) Inferring building functions from a probabilistic model using public transportation data, **Computers, Environment and Urban Systems, 48**, 124–137

Zhong, C., Manley, E., Stefan Muller Arisona, S., Batty, M.,and Schmitt, G. (2015) Measuring Variability of Mobility Patterns from Multiday Smart-card Data, **Journal of Computational Science**, doi.org/doi:10.1016/j.jocs.2015.04.021

Finding Pearls in London's Oysters

JONATHAN READES, 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 into 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 fine temporal intervals. Here 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 13 years. We explore the generic properties of the Tube and Overground rail system focusing first on the scale and distribution of the flow volumes at stations, then engaging in an analysis of temporal flows that can be decomposed into various patterns using principal components analysis (PCA) which smoothes out normal fluctuations and leaves a residual in which significant deviations can be tracked and explained. We then explore the heterogeneity in the data set with respect to how travel behaviour varies over different time 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; they 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 their 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 Seoul with the UPass card, and then in Hong Kong where they introduced the Octopus card, which was then extended to other purchases 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 which 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

RESEARCH ARTICLE

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

Chen Zhong[1]*, Michael Batty[1], Ed Manley[1], Jiaqiu Wang[1], Zijia Wang[2,3], Feng Chen[2,3], Gerhard Schmitt[4]

1 Centre for Advanced Spatial Analysis, University College London, London, United Kingdom, 2 School of Civil and Architectural Engineering, Beijing Jiaotong University, No.3 Shangyuancun, Haidian District, Beijing, P. R. China, 3 Beijing Engineering and Technology Research Centre of Rail Transit Line Safety and Disaster Prevention, No.3 Shangyuancun, Haidian District, Beijing, P. R. China, 4 Future Cities Laboratory, Department of Architecture, ETH Zurich, Zurich, Switzerland

* c.zhong@ucl.ac.uk

OPEN ACCESS

Citation: Zhong C, Batty M, Manley E, Wang J, Wang Z, Chen F, et al. (2016) Variability in Regularity: Mining Temporal Mobility Patterns in London, Singapore and Beijing Using Smart-Card Data. PLoS ONE 11(2): e0149222. doi:10.1371/journal.pone.0149222

Editor: Tobias Preis, University of Warwick, UNITED KINGDOM

Received: November 24, 2015

Accepted: January 28, 2016

Published: February 12, 2016

Data Availability Statement: Data are available from the Transport for London (TFL) in UK, Land Transport Authority (LTA) in Singapore and Beijing Transport Committee in China for researchers who meet the criteria for access to confidential data.

Funding: This work was co-funded by the European Research Council (https://erc.europa.eu/) under 249393-ERC-2009-AdG (PI: Michael Batty) and the National Natural Science Foundation of China(http://www.nsfc.gov.cn/) under grant number: 51408029 (PI: Feng Chen). The funders had no role in study

Abstract

To discover regularities in human mobility is of fundamental importance to our understanding of urban dynamics, and essential to city and transport planning, urban management and policymaking. Previous research has revealed universal regularities at mainly aggregated spatio-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 at what scales are the regularities we detect stable, explicable, and sustainable. 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 one-week of smart-card data. The results show that variations in regularity scale as non-linear functions of the temporal resolution, which we measure over a scale from 1 minute to 24 hours thus reflecting the diurnal cycle of human mobility. A particularly dramatic increase in variability occurs up to the temporal scale of about 15 minutes in all three cities and this implies that limits exist when we look forward or backward with respect to making short-term predictions. The degree of regularity varies in fact from city to city with Beijing and Singapore showing higher regularity in comparison to London across all temporal scales. A detailed discussion is provided, which relates the analysis to various characteristics of the three cities. In summary, this work contributes to a deeper understanding of regularities in patterns of transit use from variations in volumes of travellers entering subway stations, it establishes a generic analytical framework for comparative studies using urban mobility data, and it provides key points for the management of variability by policy-makers intent on for making the travel experience more amenable.