3 June 2013

Smart Cities

Session 2: Lecture 2:

Material and Electronic Networks: Transport

Michael Batty

m.batty@ucl.ac.uk

@jmichaelbatty

http://www.spatialcomplexcity.info/

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

Outline of the Lecture

1. Context: New Data for and from the Smart City
2. Multimodal Flows: An Analytical Gap
3. Why London?
4. A Partial View of London's Network
5. Three Approaches, Three Problems –
6. Problem 1: Flows on the tube lines – by trains
7. Problem 2: Representing Networks
8. Problem 3: The 'Shortest'-Path Problem

Context: New Data for and from the Smart City

Here we are going to move directly to networks in the literal sense, Remember Metcalfe's Law, so we will look also at flows on networks

Our understanding of urban flows tends to be based almost exclusively on single-mode systems and each mode is handled separately.

Reflects a widespread focus on auto-mobility, Also a function of analytical tractability. But the increasing availability of large behavioural data sets from Public Transport Networks (PTNs), combined with the increasing power and sophistication of computational approaches, creates new ways of exploring travel demand.

Greater spatial resolution – down to station or, soon, bus stop, and greater temporal resolution – down to the minute, and greater coverage – centralised collection of data means a single data store for an entire city

PTN-derived data also avoids many of the privacy issues associated with mobile network data because the user becomes invisible as soon as they leave the system.

Multimodal Flows: An Analytical Gap

	Pros	**Cons**	**Representative Research/Resources**
Graph Theoretical	Quick Full abstraction	Static Load-less Too simplistic	Von Ferber et al. (2007, 2010) Derrible (2010, 2012) Erath et al. (2009)
Agent-Based & Other Dynamic	Realistic Dynamic Capacity constrained	Slow Limited abstraction Can be too complex	Cats & Jenelius (2012) Manley et al. (2012) MATSim

Is there is an opportunity here to leverage digital data in order to drive more realistic models that combine the performance of graph theoretical approaches with the added realism of ABMs?

Why London?

Transport for London's (TfL) RFID-based 'Oyster Card' is particularly attractive because users typically need to use their card at both ends of a trip, providing us with detailed origin and destination data for more than 3 million daily users.

The system is particularly large and complex:

Approximately 640 stations across all modes

340 stations with Oyster Card readers served by National Rail trains

80 stations served by Overground trains

270 stations served by Underground trains

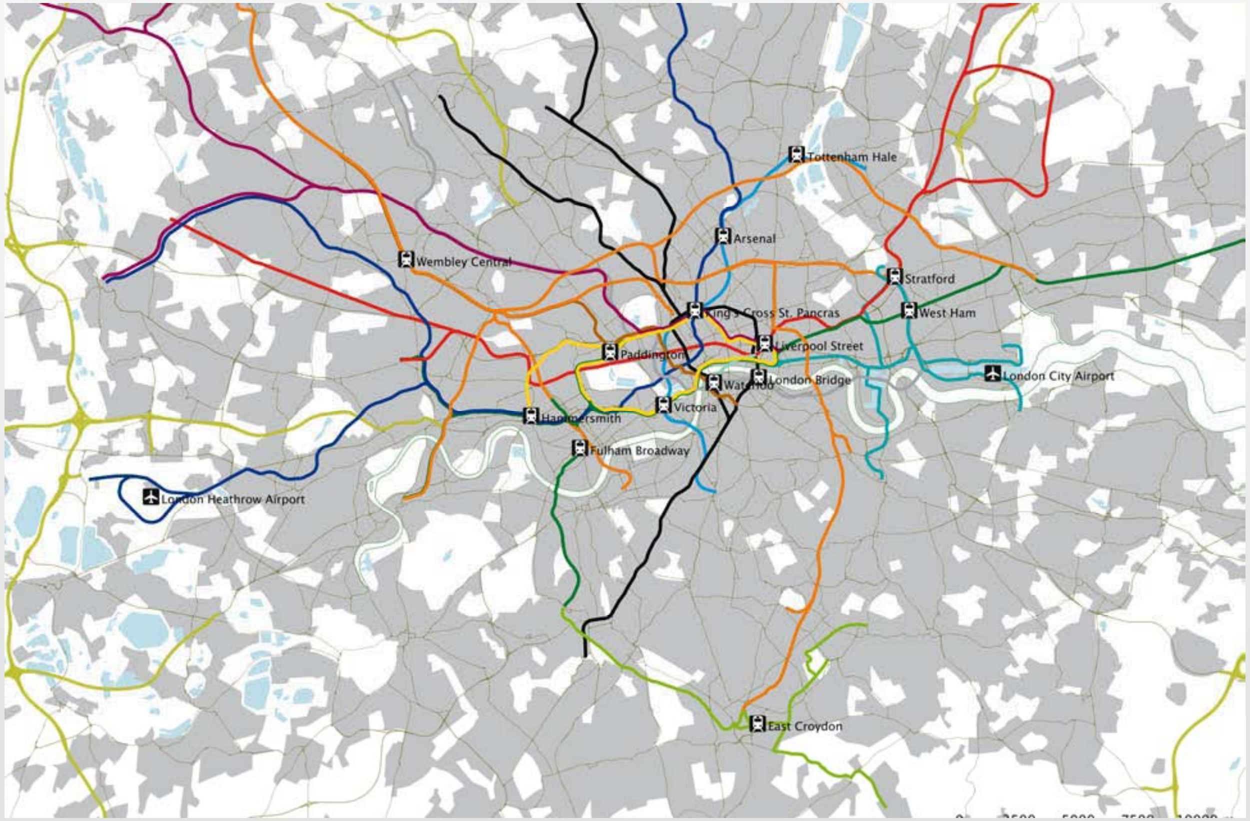
45 stations served by Docklands Light Rail

39 stations served by Tram

147 stations with some kind of interchange (between line or mode)

Aboveground coverage by Open Street Map (OSM) is also particularly good, allowing us to model walking behaviour using open source tools that respect pedestrian preferences for balancing directness with quieter streets.

A Partial View of London's Network



Although many users – especially visitors – are used to thinking about London in terms of the Beck schema, the combination of an online Journey Planner and regular travel on the network enable many to identify the quickest route *even if* it doesn't appear to be the most direct.

Three Approaches, Three Problems

1. To explore what is happening to actual movements of trains – to compute delays – currently Transport for London TfL have an API for tubes and buses that enables the user to query the location and of all trains/buses on the network and to examine how they move – from this we can calculate delay and compute a variety of measures. This is what Richard Milton, one of my colleagues has done and I will show some of this – it is very preliminary
2. Use of classic graph measures to show how the network can be disrupted – this is largely a topological graph/network approach that shows how betweenness centrality or accessibilities is disturbed – I am responsible for this and will show some work
3. Fully fledged flow and graph measures in a multimodal context – tube, overground, walk so far – where we are computing changes in flow – this work that Jon Reades is doing
4. All very preliminary – I show all this in the spirit of work in progress

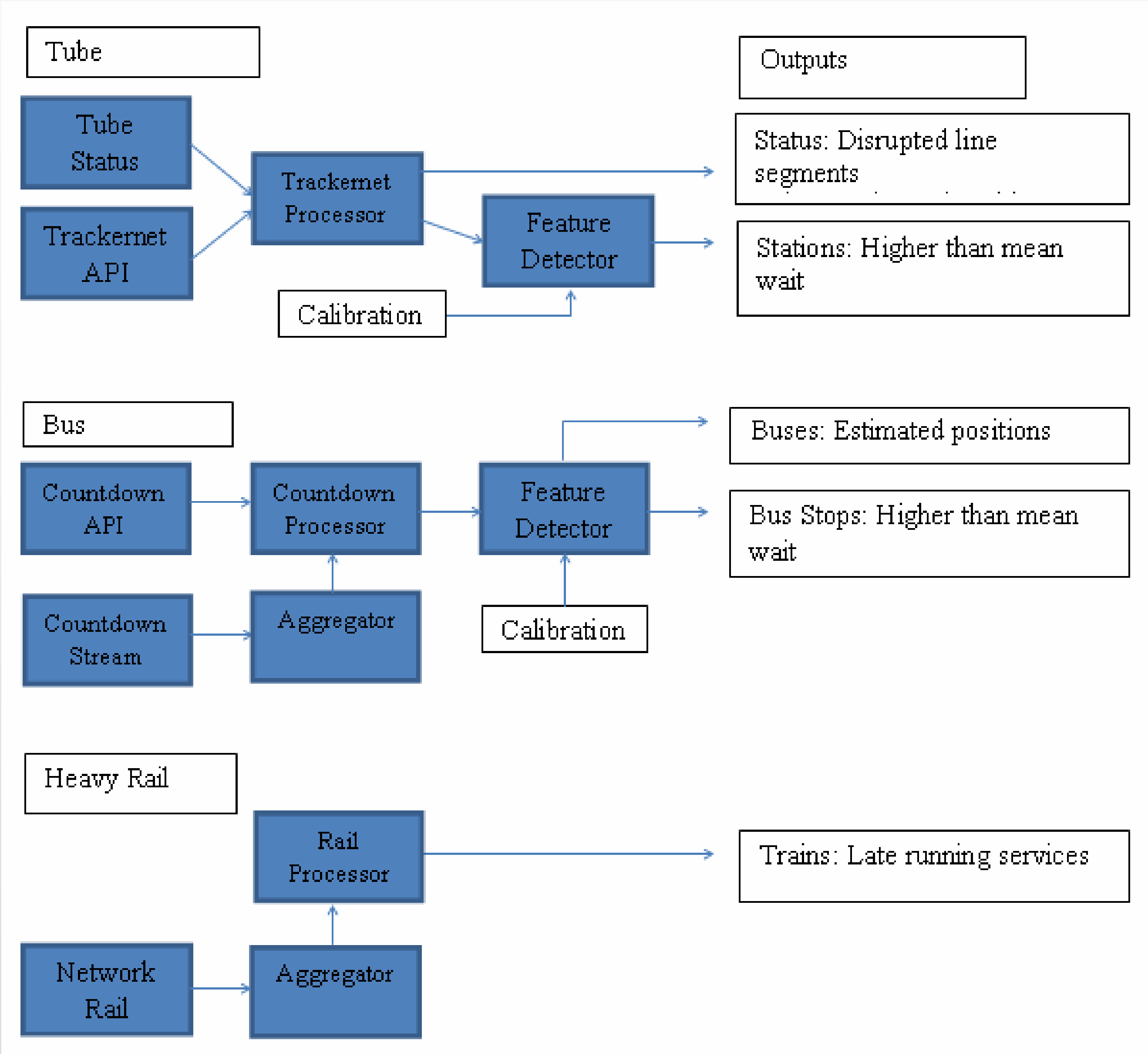
Problem 1: Flows on the tube lines – by trains

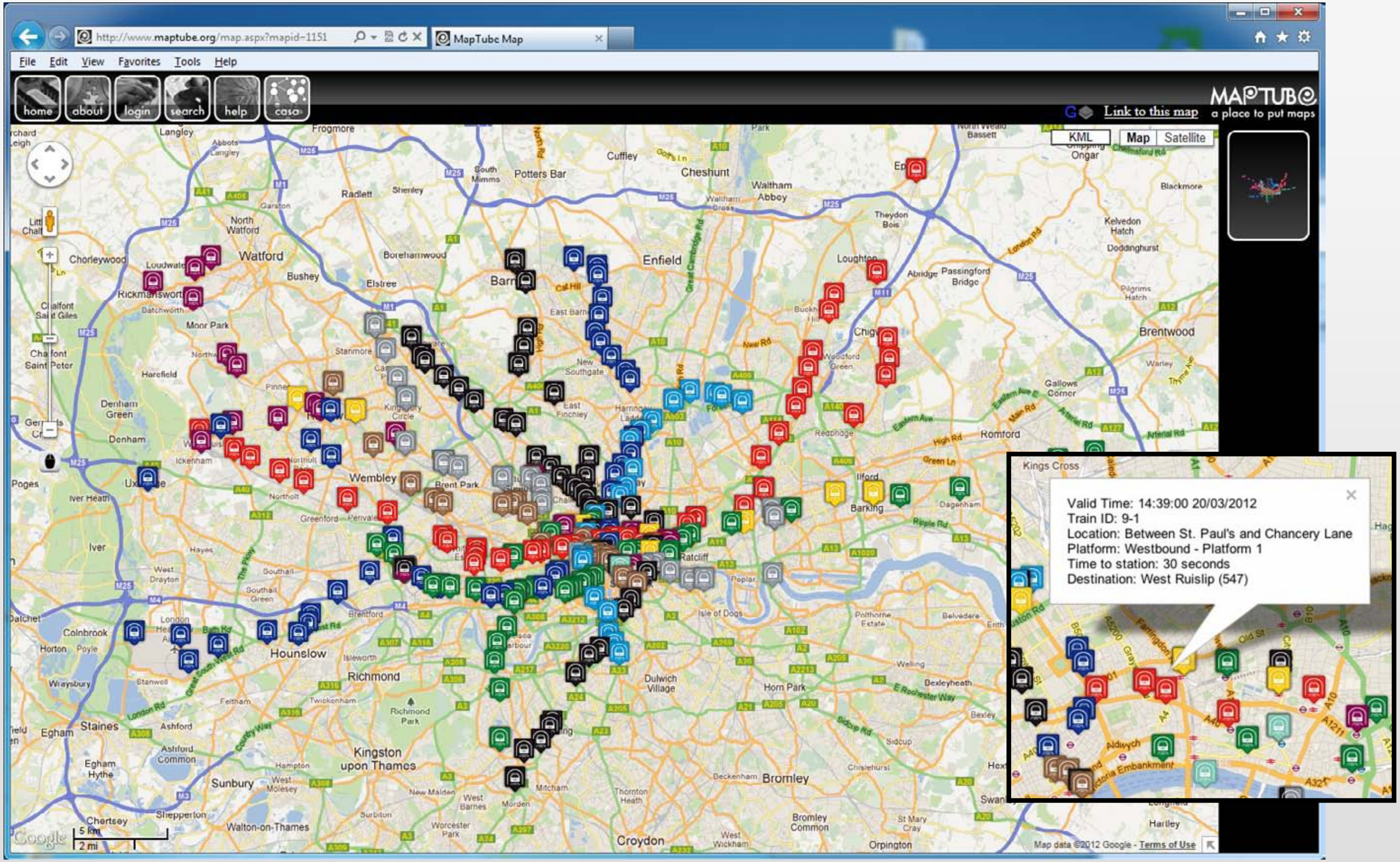
As we will demonstrate, through the "Trackernet" system for London Underground and the "Countdown" system for buses, it is now possible to collect and visualise the positions of vehicles in real-time.

At peak periods there can be 7000 buses, 900 trains and 450 tubes running on the system

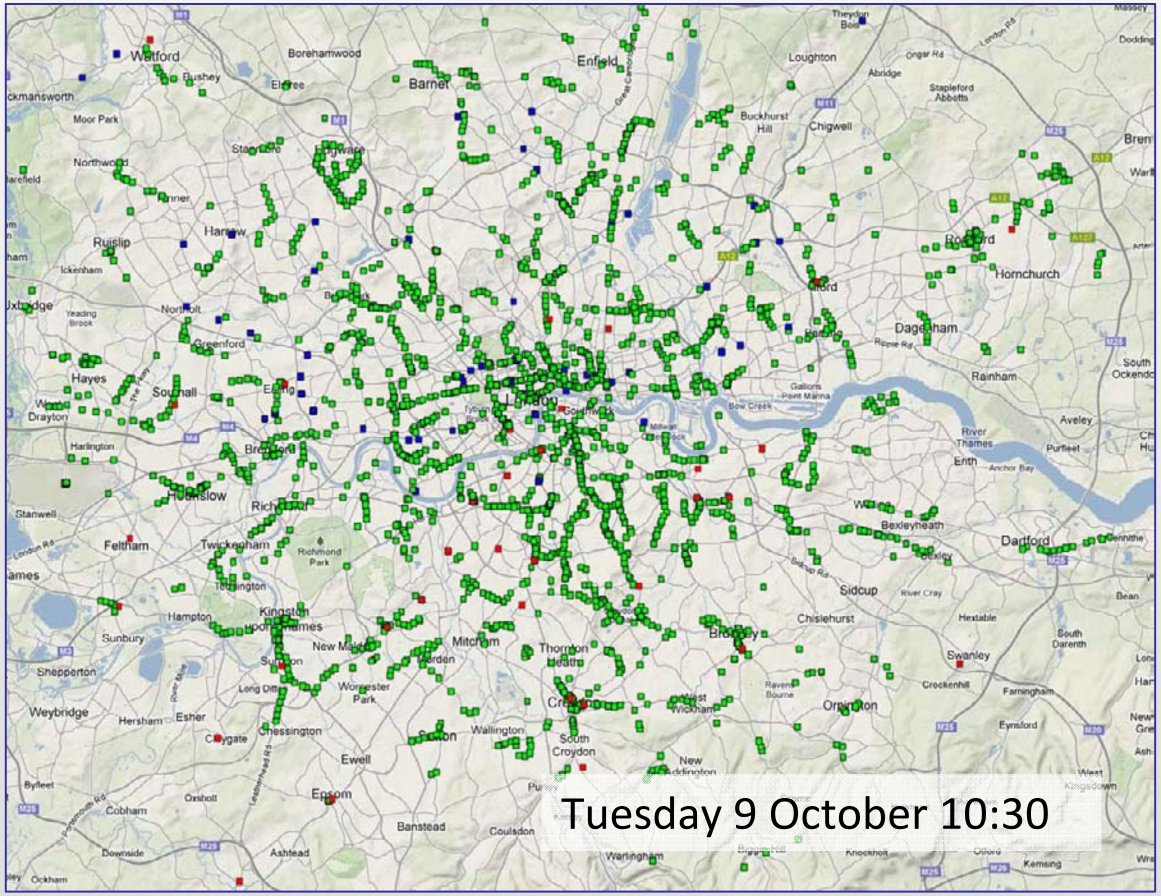
Delays for these transport systems were calculated by using an archive of historic data to find the mean wait time for every hour and every station or bus stop.

This can then be visualised in real time or after the event for further analysis. We show a mix of these visualisations in the figures that follow – as yet we have not developed an integrated analysis but all the ideas are there. We show the analysis first for the tube but here is the block diagram showing how we are assembling the data.


Tube
Tube Status
Trackernet API
Trackernet Processor
Feature Detector
Calibration
Outputs
Status: Disrupted line segments
Stations: Higher than mean wait
Bus
Countdown API
Countdown Processor
Feature Detector
Countdown Stream
Aggregator
Calibration
Buses: Estimated positions
Bus Stops: Higher than mean wait
Heavy Rail
Rail Processor
Network Rail
Aggregator
Trains: Late running services


http://www.maptube.org/map.aspx?mapid=1151
MapTube Map
File Edit View Favorites Tools Help
home about login search help casa
Link to this map
MAPTUBE
a place to put maps
KML Map Satellite
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)
Kings Cross
Map data ©2012 Google - Terms of Use

Delays from Tube, National Rail and Bus Fused

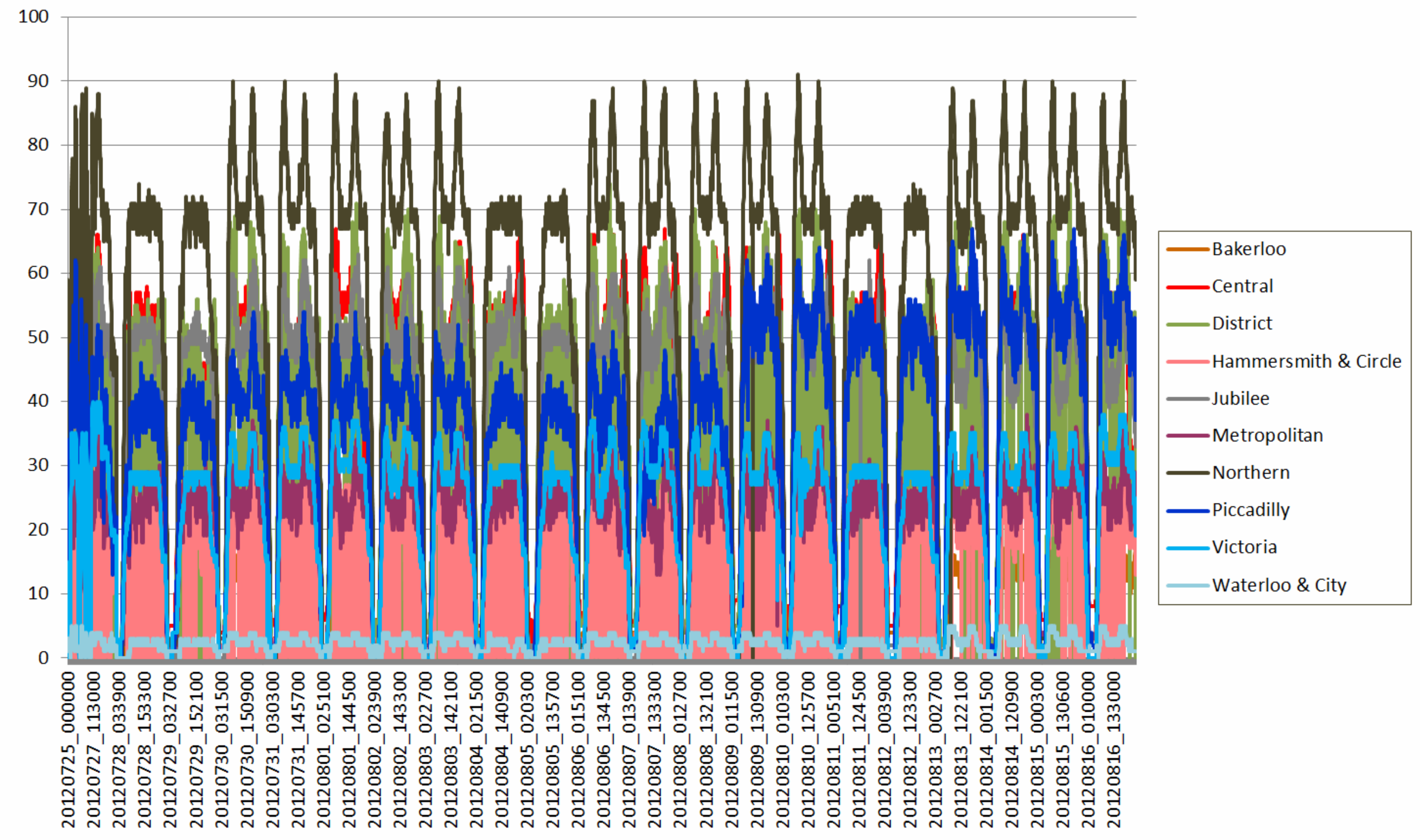


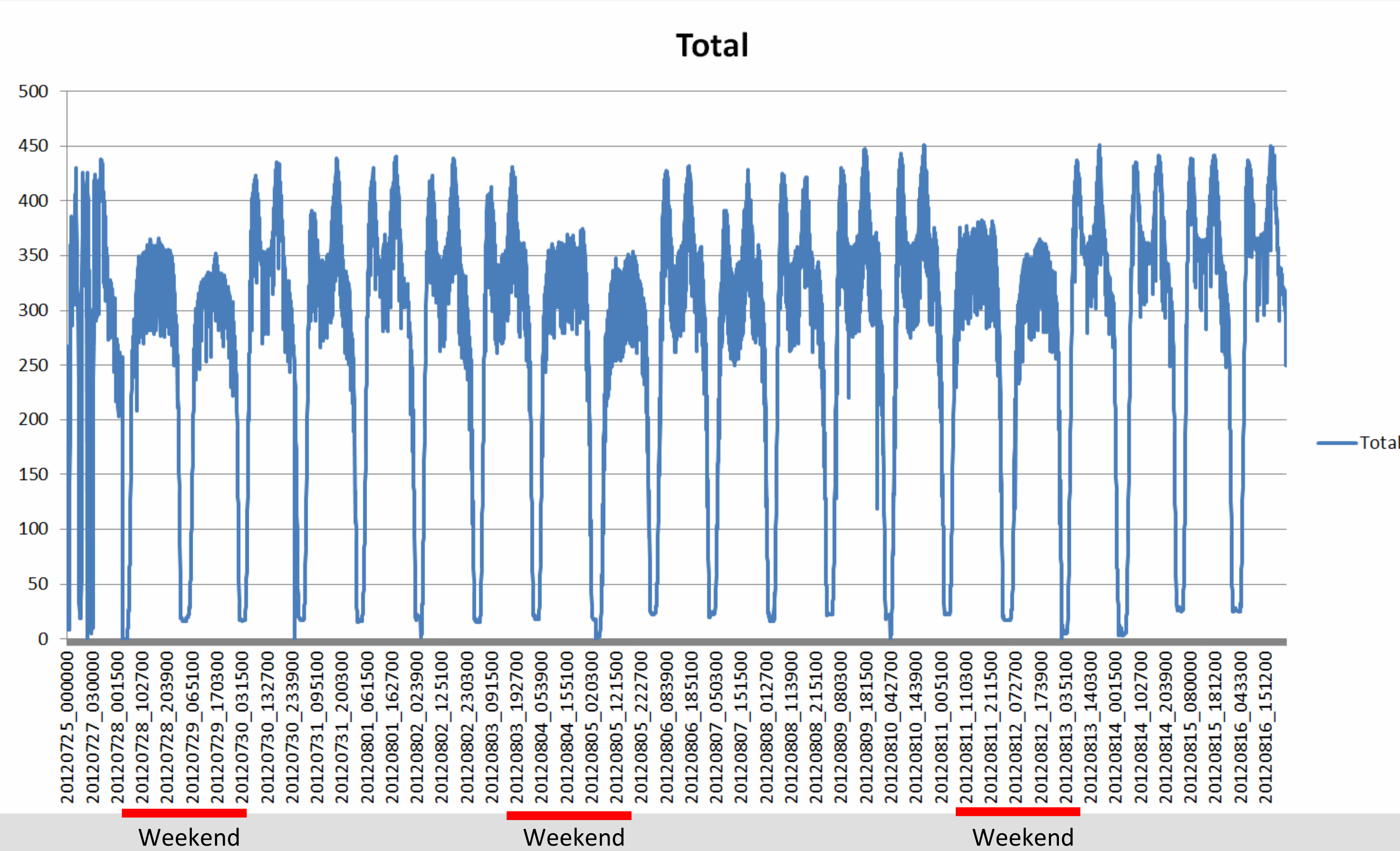
Key

- 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

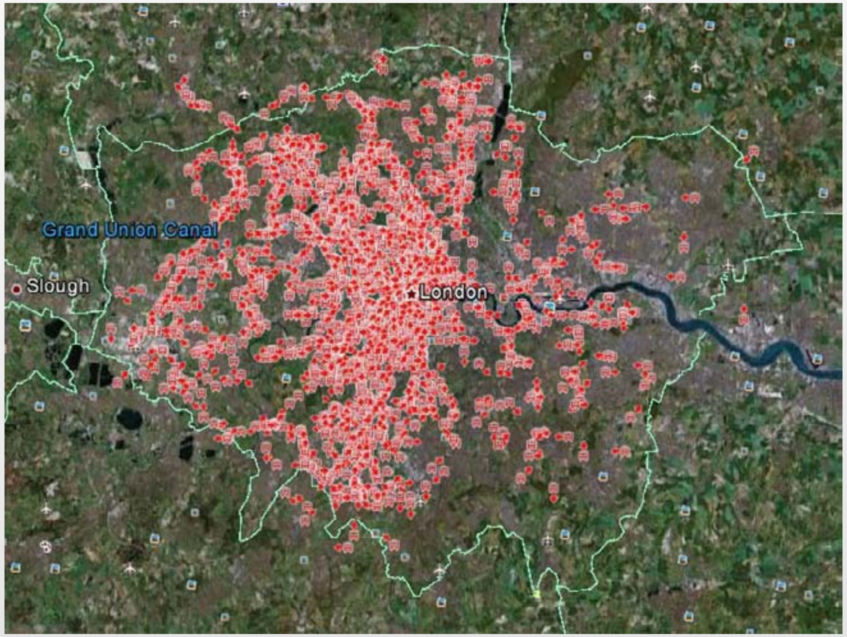
Tube delays from the TfL status feed are also plotted as lines

Flows During the Olympics – we are looking at this as a case study

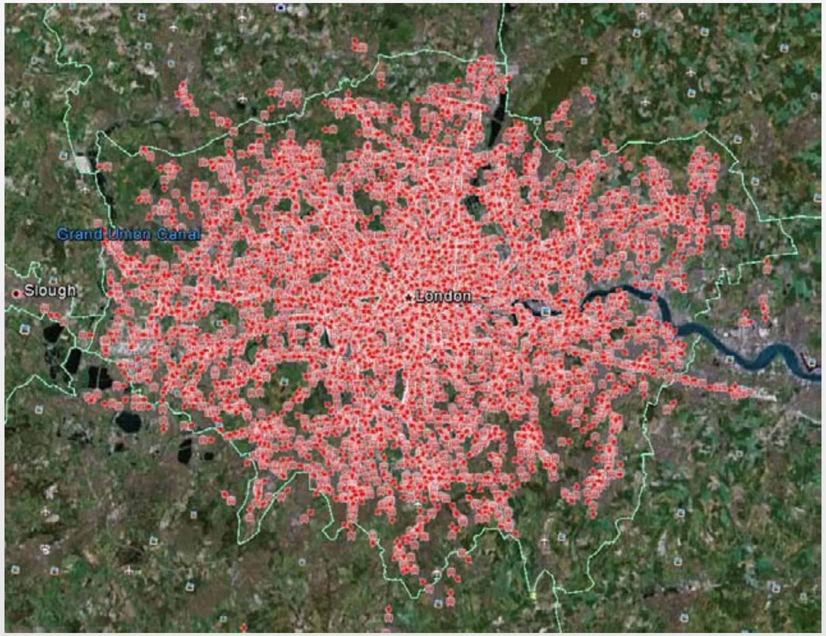



Total
500
450
400
350
300
250
200
150
100
50
0
20120725_000000
20120727_030000
20120728_001500
20120728_102700
20120728_203900
20120729_065100
20120729_170300
20120730_031500
20120730_132700
20120730_233900
20120731_095100
20120731_200300
20120801_061500
20120801_162700
20120802_023900
20120802_125100
20120802_230300
20120803_091500
20120803_192700
20120804_053900
20120804_155100
20120805_020300
20120805_121500
20120805_222700
20120806_083900
20120806_185100
20120807_050300
20120807_151500
20120808_012700
20120808_113900
20120808_215100
20120809_080300
20120809_181500
20120810_042700
20120810_143900
20120811_005100
20120811_110300
20120811_211500
20120812_072700
20120812_173900
20120813_035100
20120813_140300
20120814_001500
20120814_102700
20120814_203900
20120815_080000
20120815_181200
20120816_043300
20120816_151200
Total
Weekend
Weekend
Weekend

The Effect of Bus Strike



Tuesday 22nd May 2012, 09:00



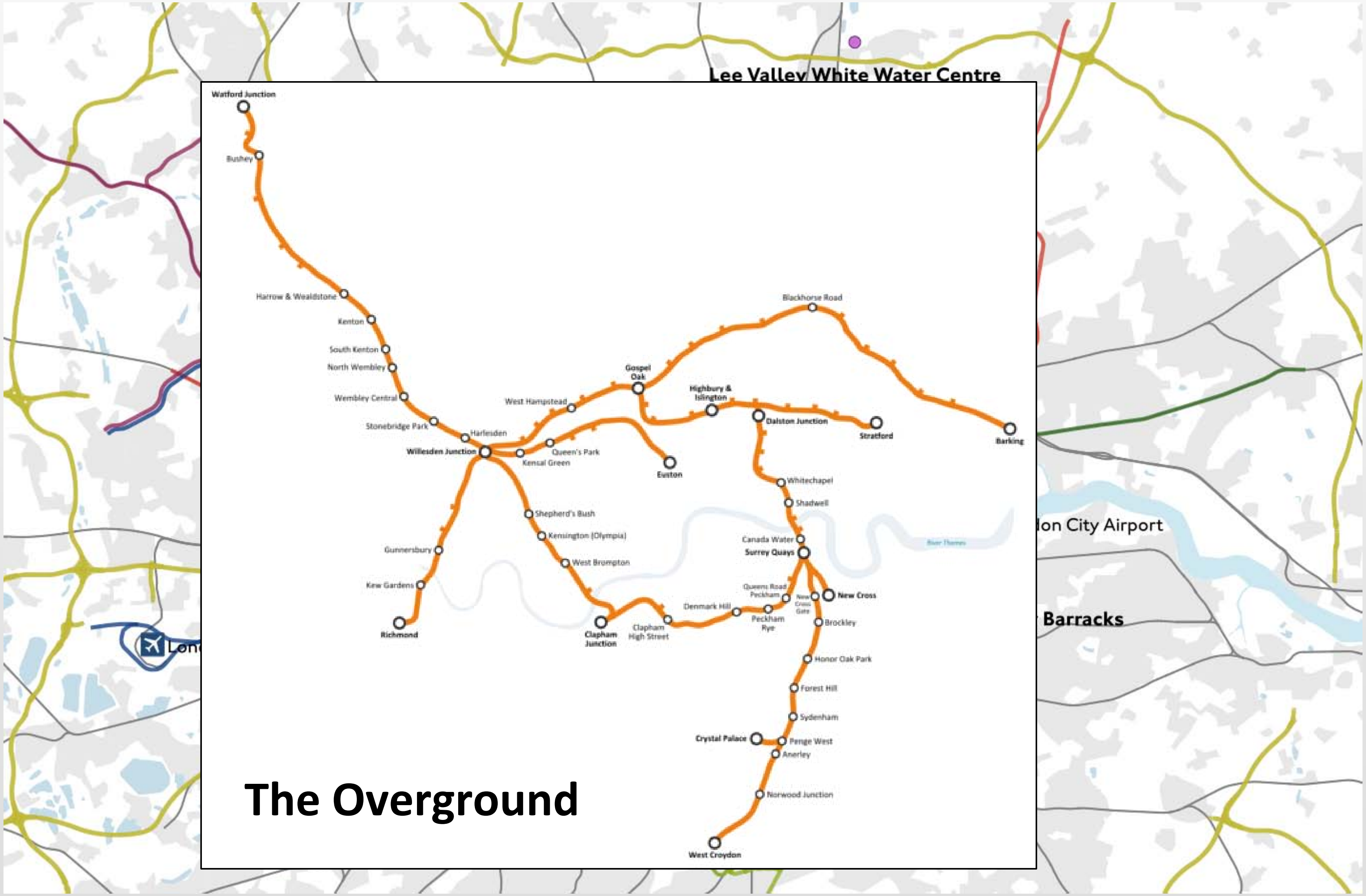
Wednesday 23rd May 2012, 09:00

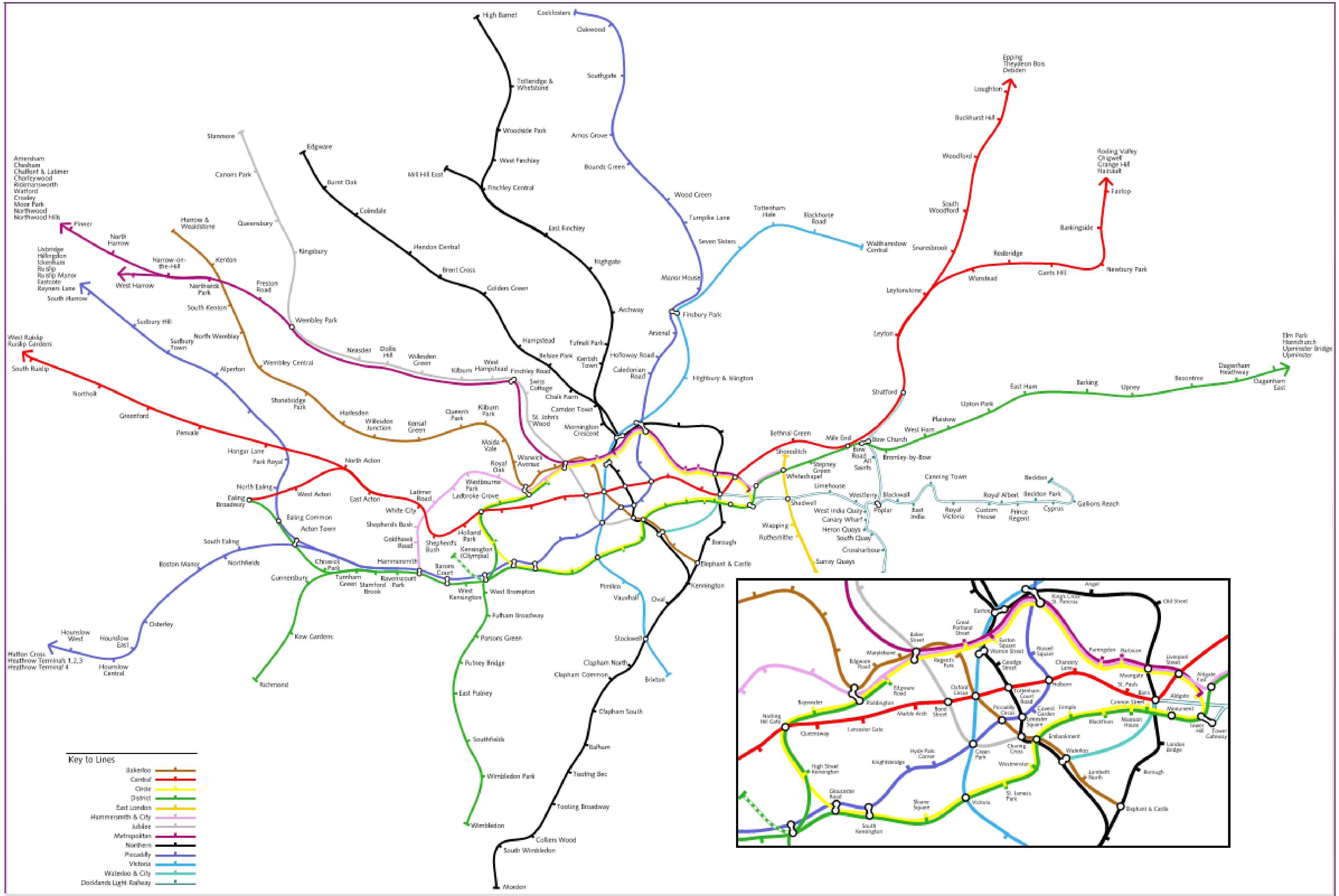
The left image shows the effect of the bus strike on 22nd May 2012, while the image on the right shows a normal day.

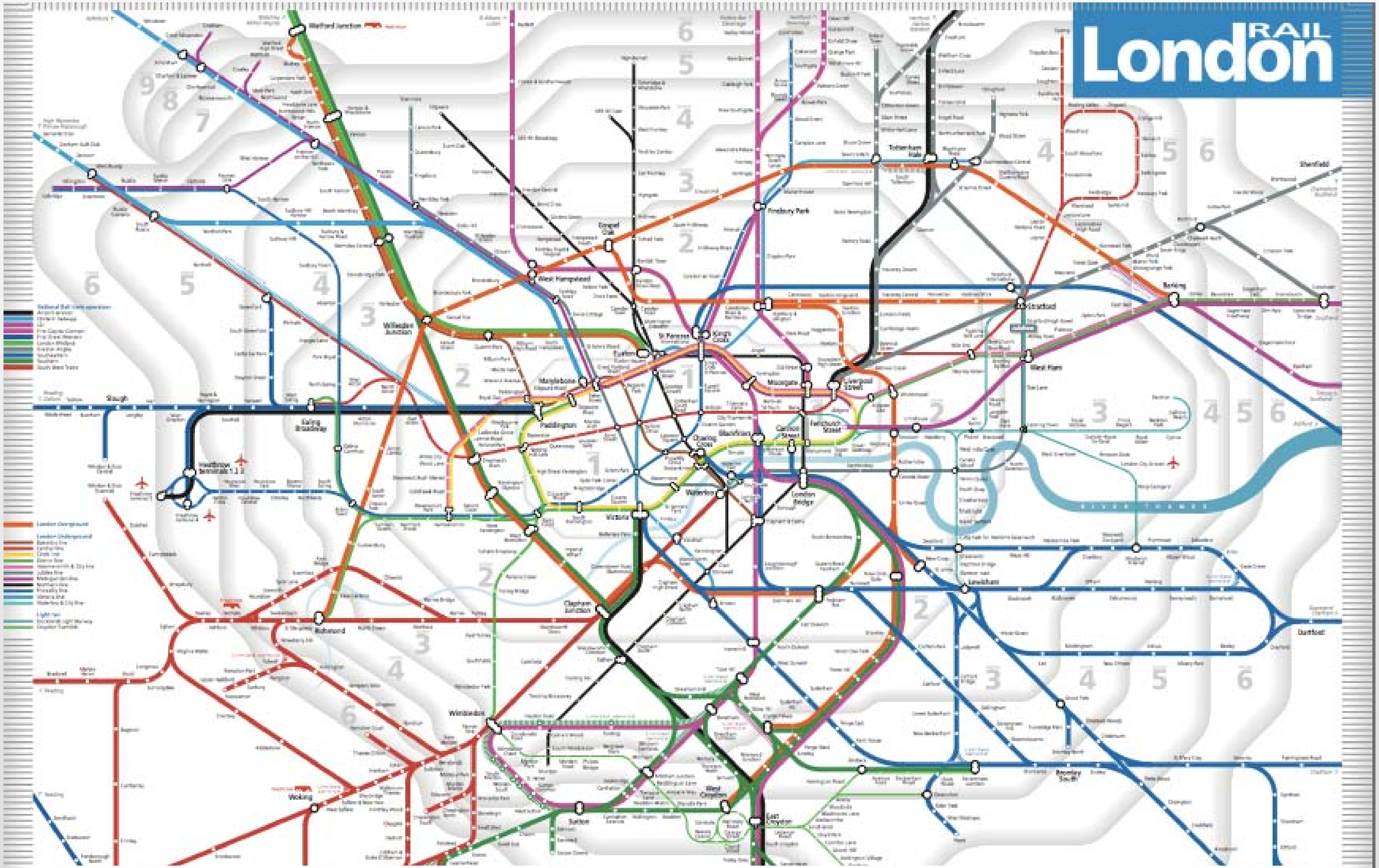
I want to introduce a little more background on the automated scheme in London that we are discussing – some history first and then a map of the network

It is quite confusing because so many networks overlap so we should be clear about the core network – the underground which has 270 stations; the Docklands Light railway has 45; the overground has 83 but this extends out side of Greater London: then national rail is more complicated because it dovetails with overground and underground stations also coincide but are not the same as national rail stations.

We are mainly going to deal here with the greater London networks but the touch card data is for a much bigger area and also for the bus system and it can also be related to national rail.

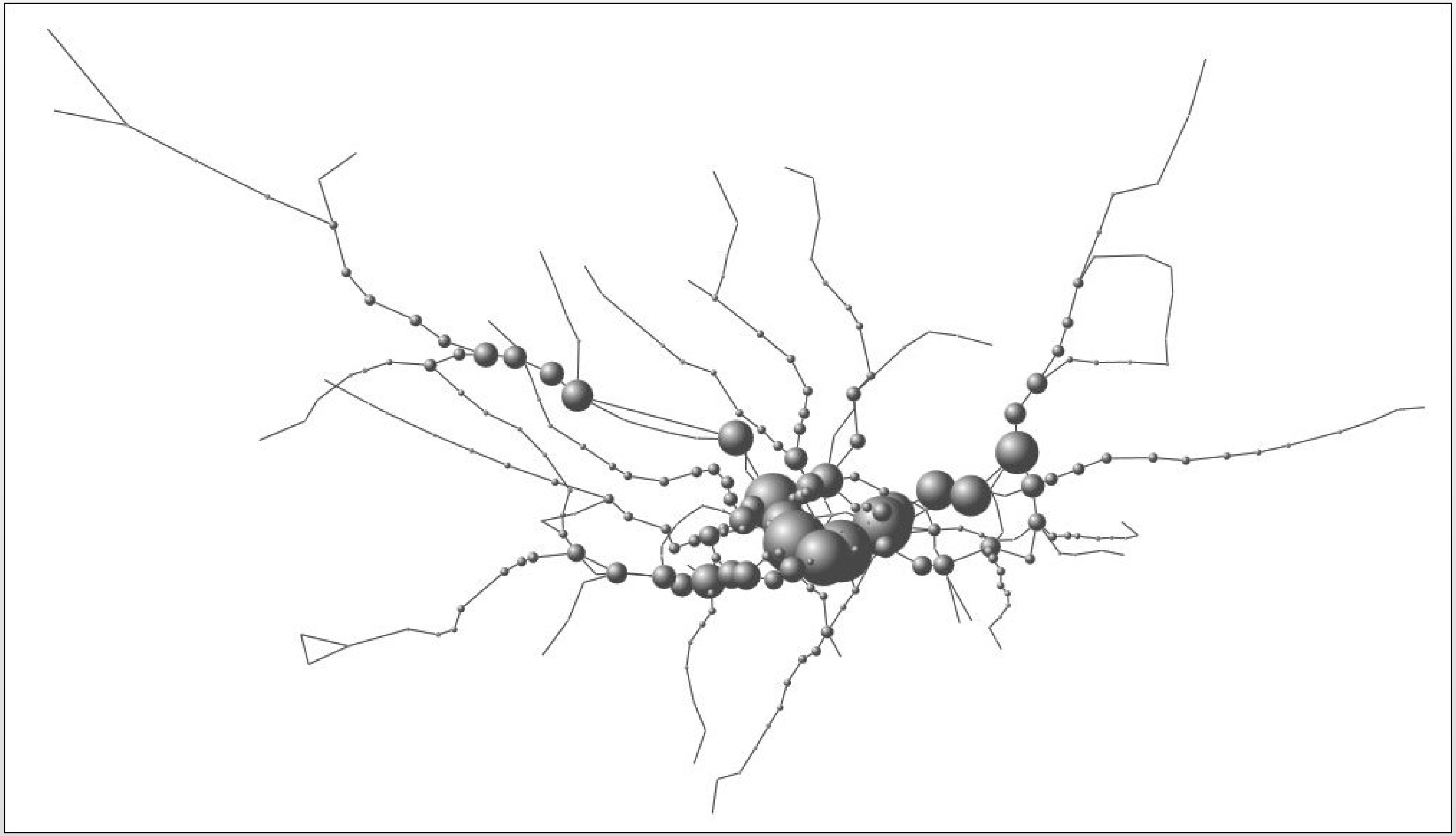

Lee Valley White Water Centre
lon City Airport
Barracks
Watford Junction
Bushey
Harrow & Wealdstone
Kenton
South Kenton
North Wembley
Wembley Central
Stonebridge Park
Harlesden
Willesden Junction
Kensal Green
Queen's Park
West Hampstead
Gospel Oak
Highbury & Islington
Dalston Junction
Stratford
Blackhorse Road
Barking
Euston
Whitechapel
Shadwell
Canada Water
Surrey Quays
New Cross
Brockley
Honor Oak Park
Forest Hill
Sydenham
Crystal Palace
Penge West
Anerley
Norwood Junction
West Croydon
Shepherd's Bush
Kensington (Olympia)
West Brompton
Clapham Junction
Clapham High Street
Denmark Hill
Peckham Rye
Queens Road Peckham
Gunnersbury
Kew Gardens
Richmond
The Overground


Key to Lines
Bakerloo
Central
Circle
District
East London
Hammersmith & City
Jubilee
Metropolitan
Northern
Piccadilly
Victoria
Waterloo & City
Docklands Light Railway



Tube, DLR, Overground and National Rail Networks in London

The simplest network for the first problem, based on an analysis of the network, not the flows: shown below are the degrees



Problem 2: Representing Networks

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

$$L_i = K D_i^{-1} = K\left(\sum_j d_{ij}\right)^{-1}$$

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

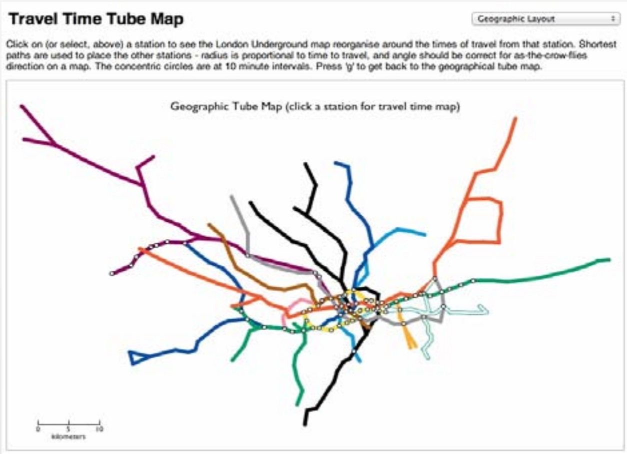
Centrality

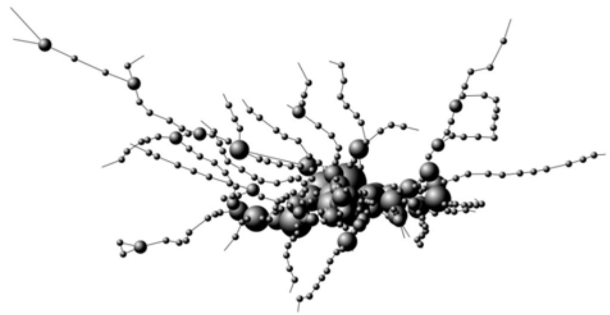
$$p_{ijk} = \frac{\sigma_{ikj}}{\sigma_{ij}} = \frac{\sigma_{ikj}}{\sum_{\ell} \sigma_{i\ell j}} \quad , \quad \sum_k p_{ikj} = 1$$

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

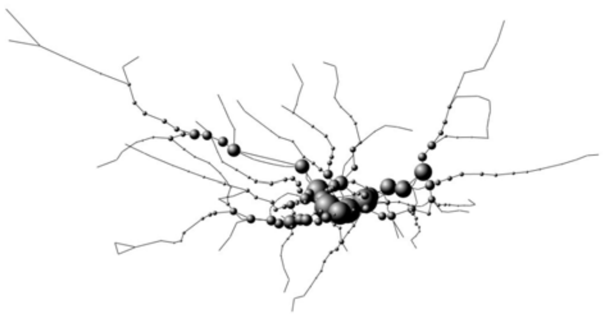
A Preliminary Analysis (1)

The Minimal Tube Network and the Three Centrality Indices

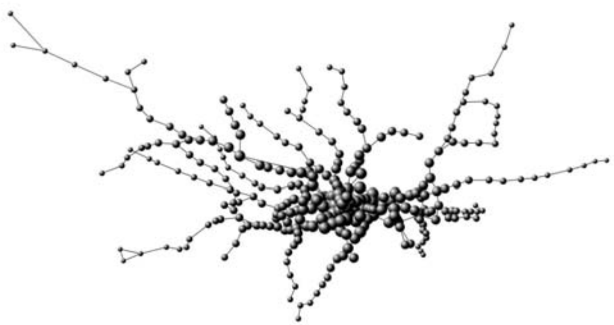




Degree



Betweenness



Closeness

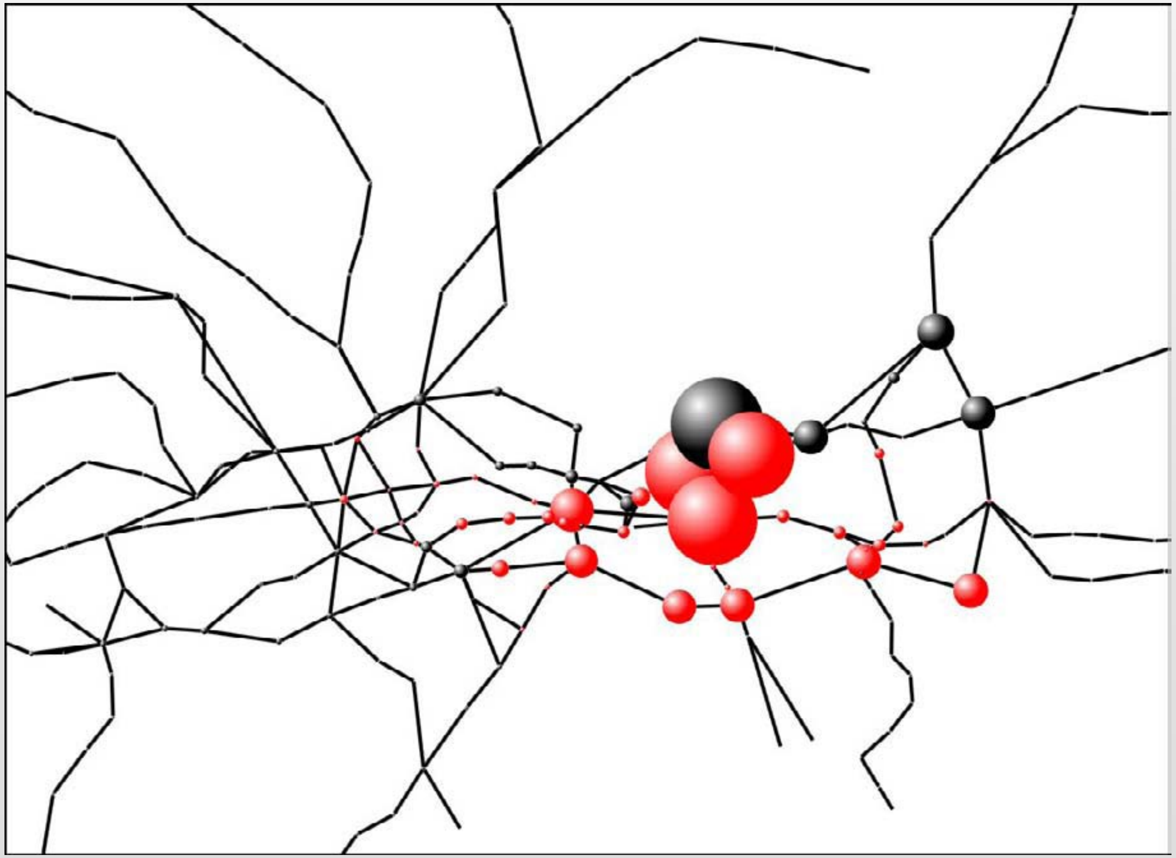
A Preliminary Analysis (2)

- 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

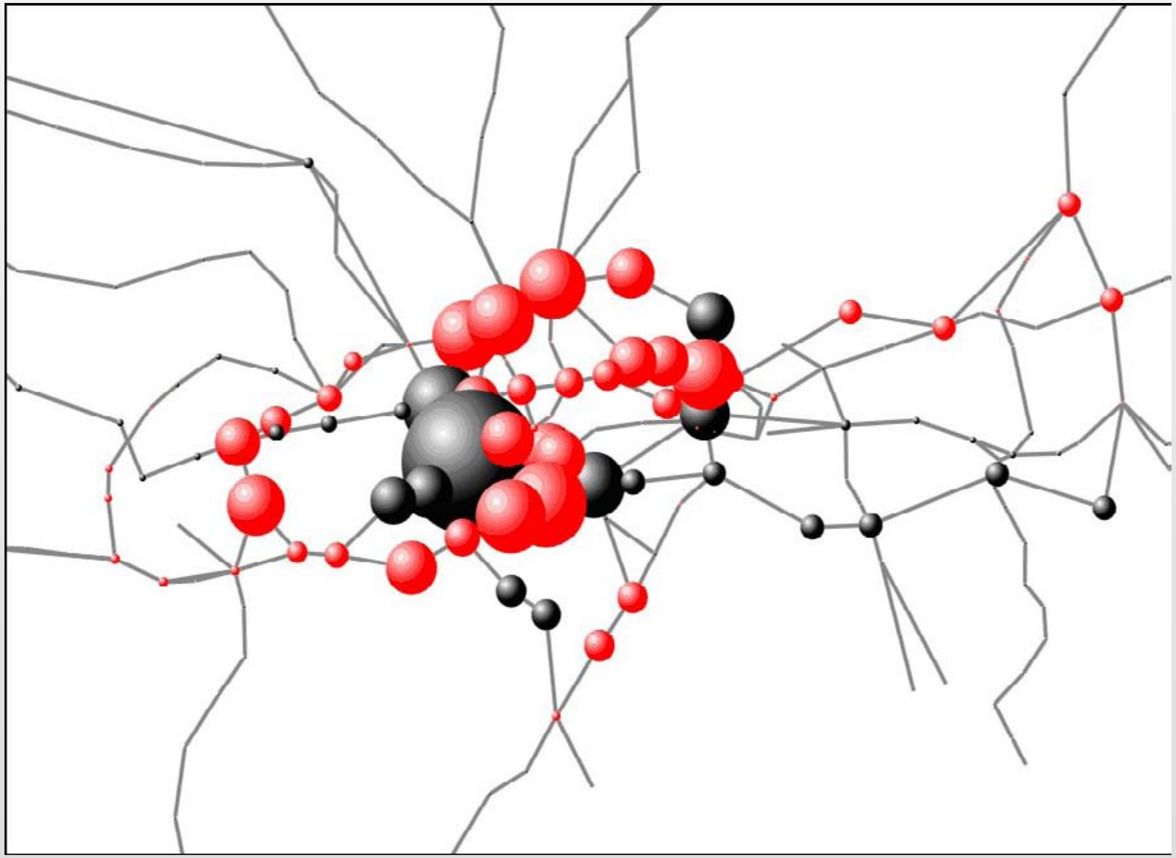
A Preliminary Analysis (3)

Closing Liverpool Street



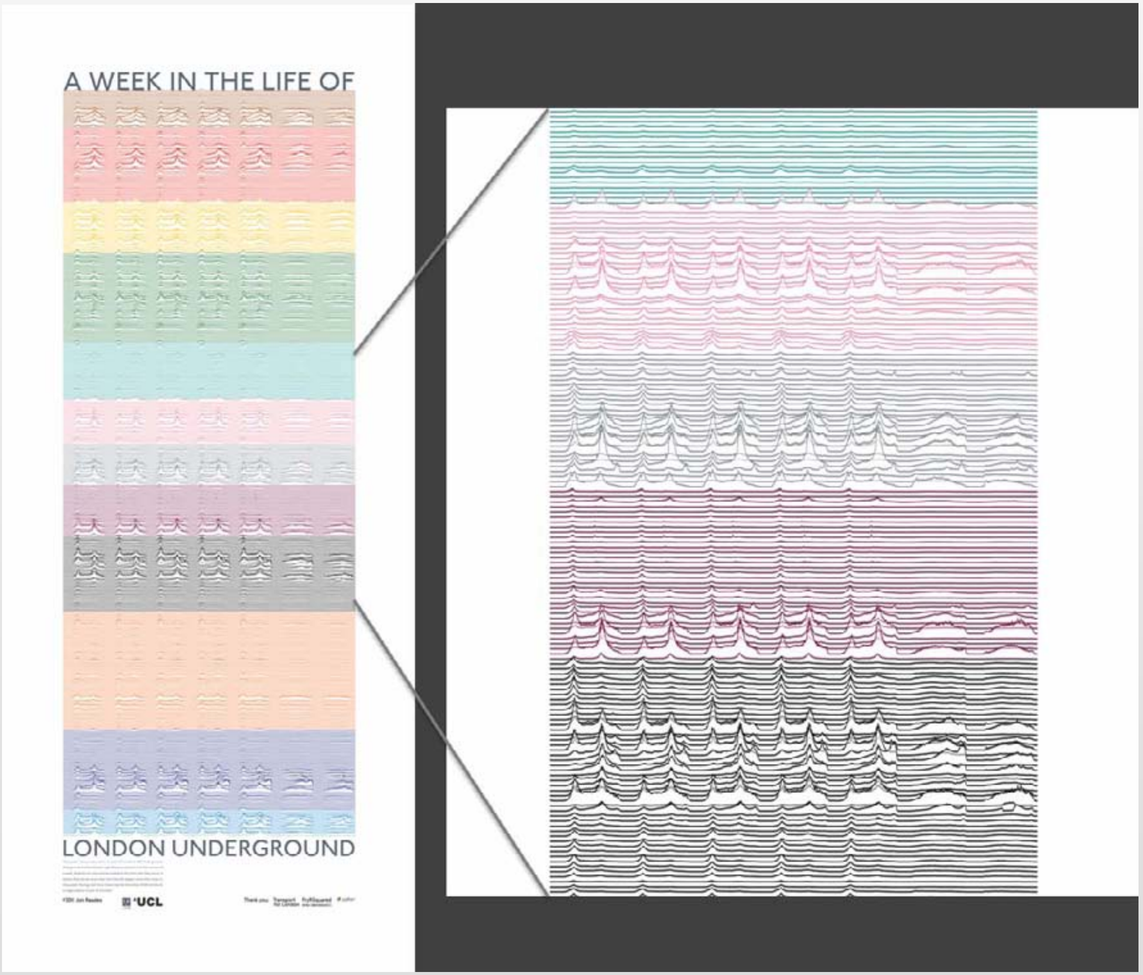
A Preliminary Analysis (3)

Closing Green Park

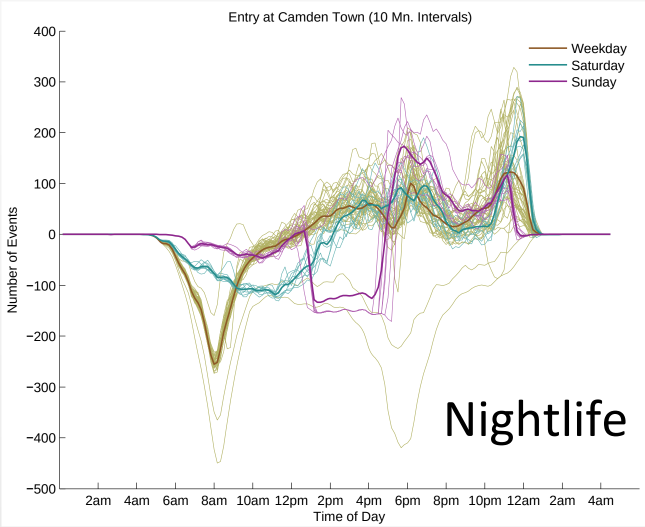


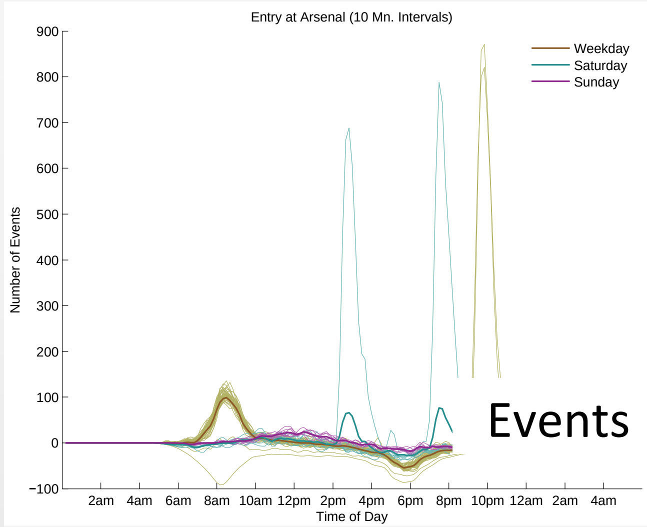
Problem 3: The Shortest Path

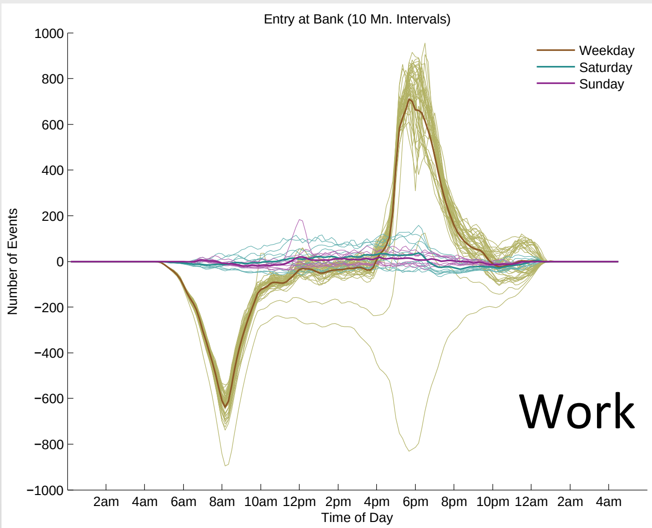
but to remind you, here we are dealing with flow data with the usual rush hour peaks and troughs

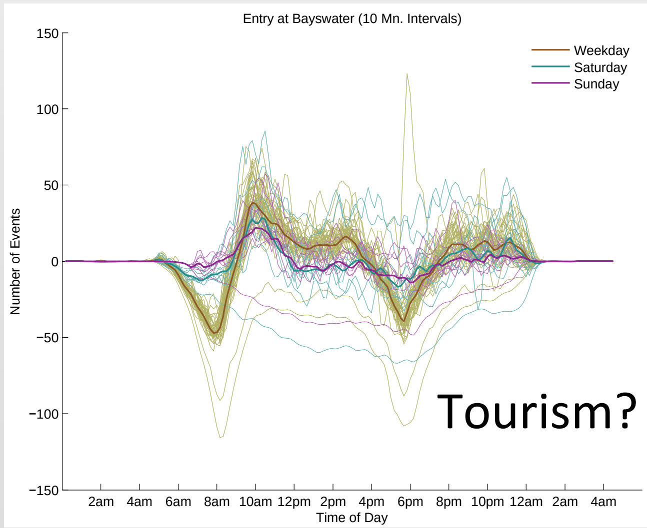


Station Time: Weekdays, Saturdays and Sundays

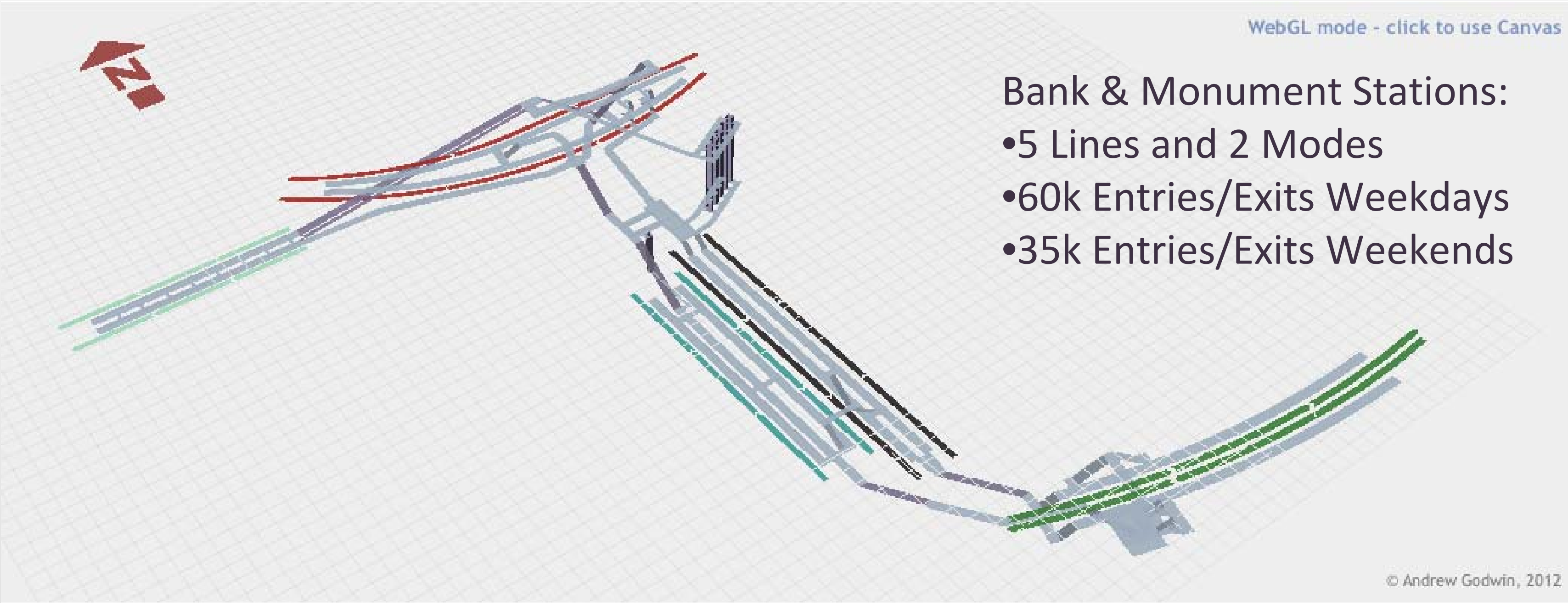




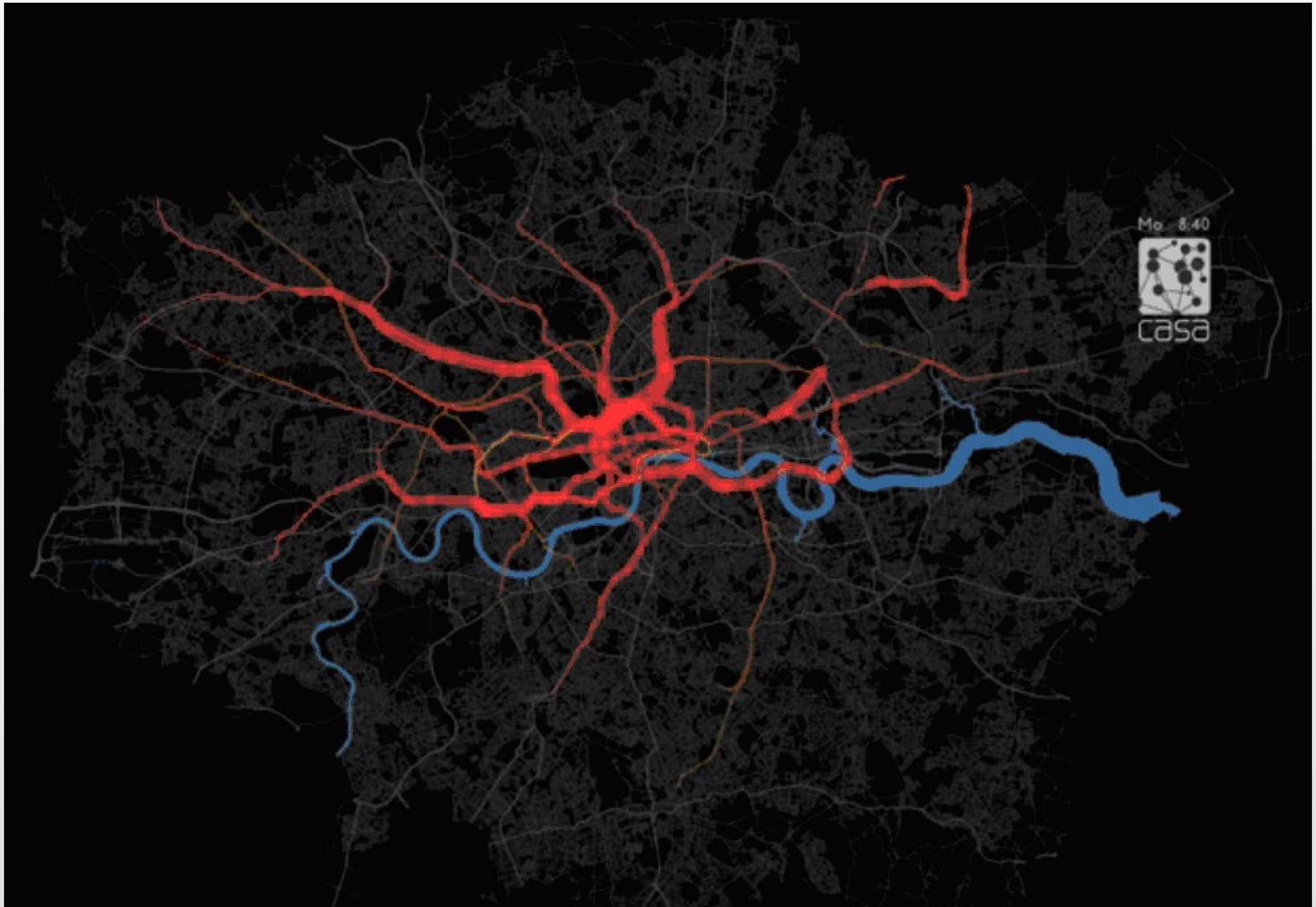




Problem 3: The ‘Shortest’-Path Problem



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.

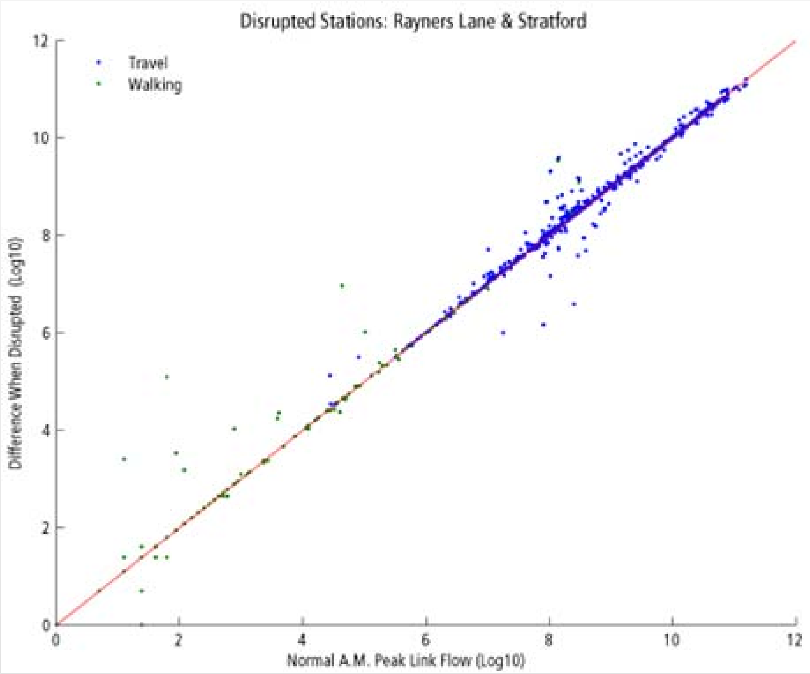


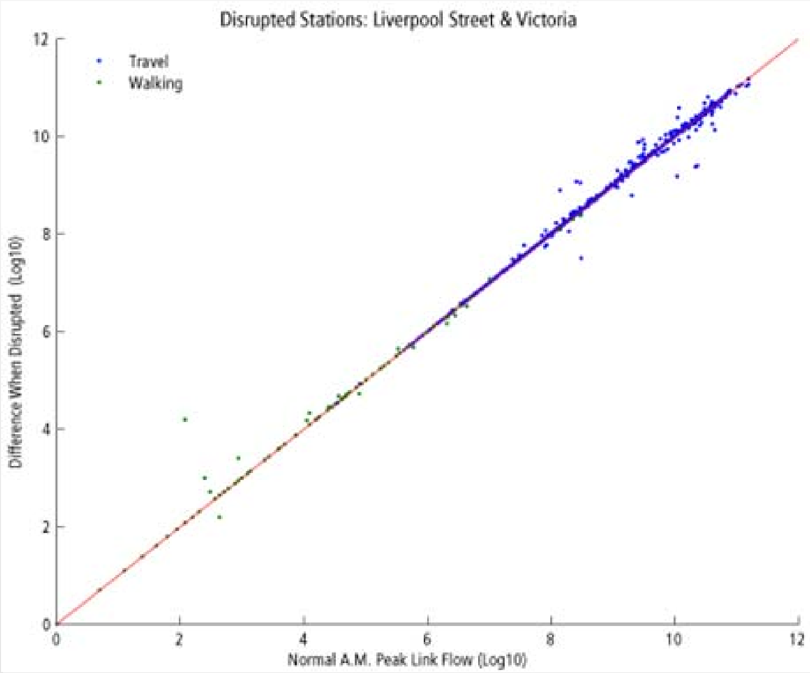
If you go to Vimeo or to http://simulacra.info/ you will get this movie over a 7 day typical work second by second flows

Methodology

1. Build average O/D flows between all Under- and Over-ground stations
   – 33 days of activity with 100% coverage of pseudonymous Oyster cards
   – More than 300 million unique trip segments (of which 120 million by some form of rail)
2. Build walking network between all stations within 5km of each other
   – Routing on OSM network provided by routino using realistic preferences for walk speed and intensity of road usage
3. Build integrated travel-time network representation of both modalities
   – End-to-end travel time extracted from routino routing
   – Physical layout of stations inferred from real-time platform data
4. Simulate simultaneous disruption for 1 or 2 stations
   – Use real O/D matrix and remap disrupted trip segments
   – Realistic disruption on basis of entry/exit/interchange breakdown
   – Measure changes in volumes and 'lost' travel times across segments

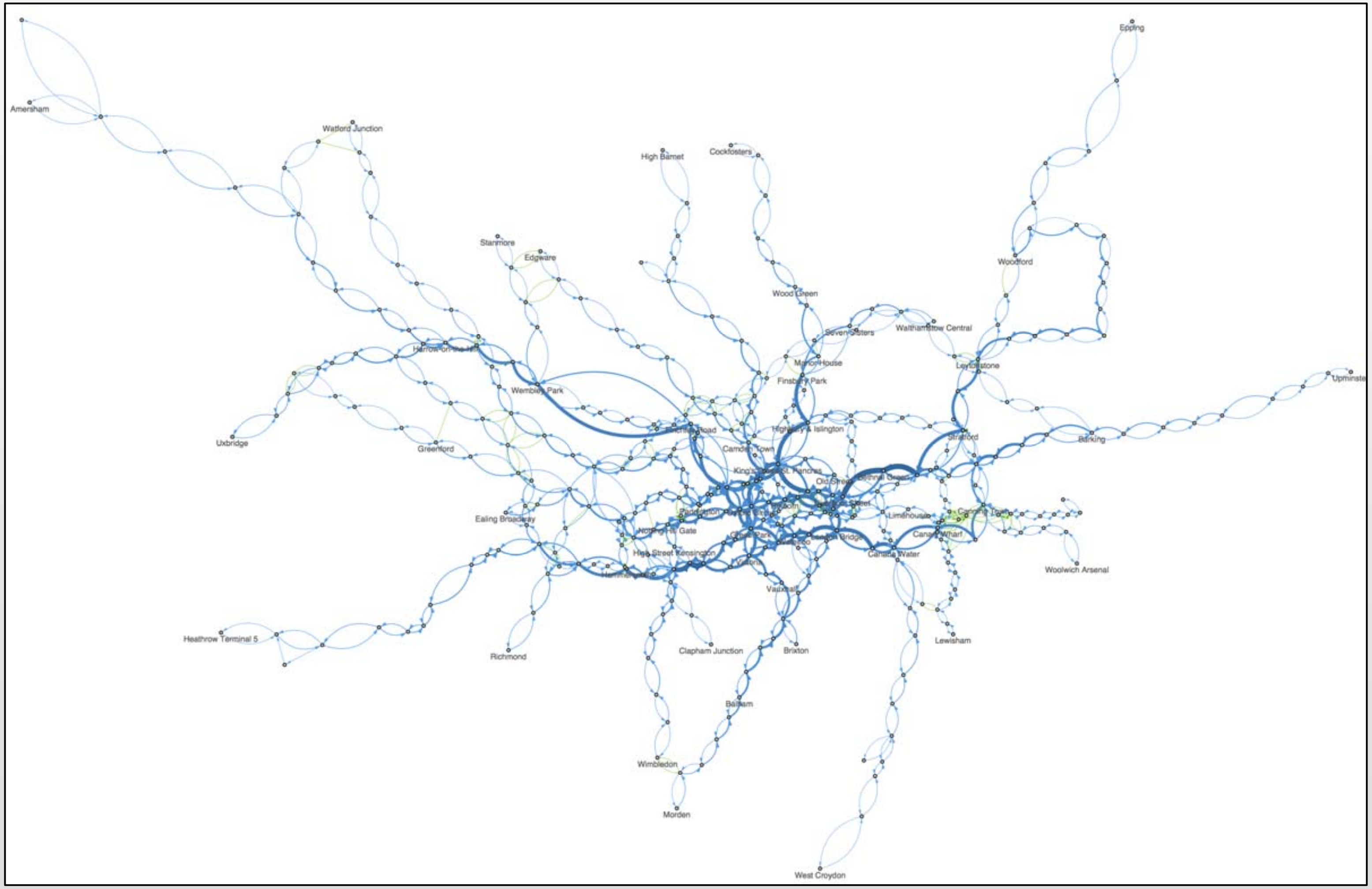
Link-Level Disruption



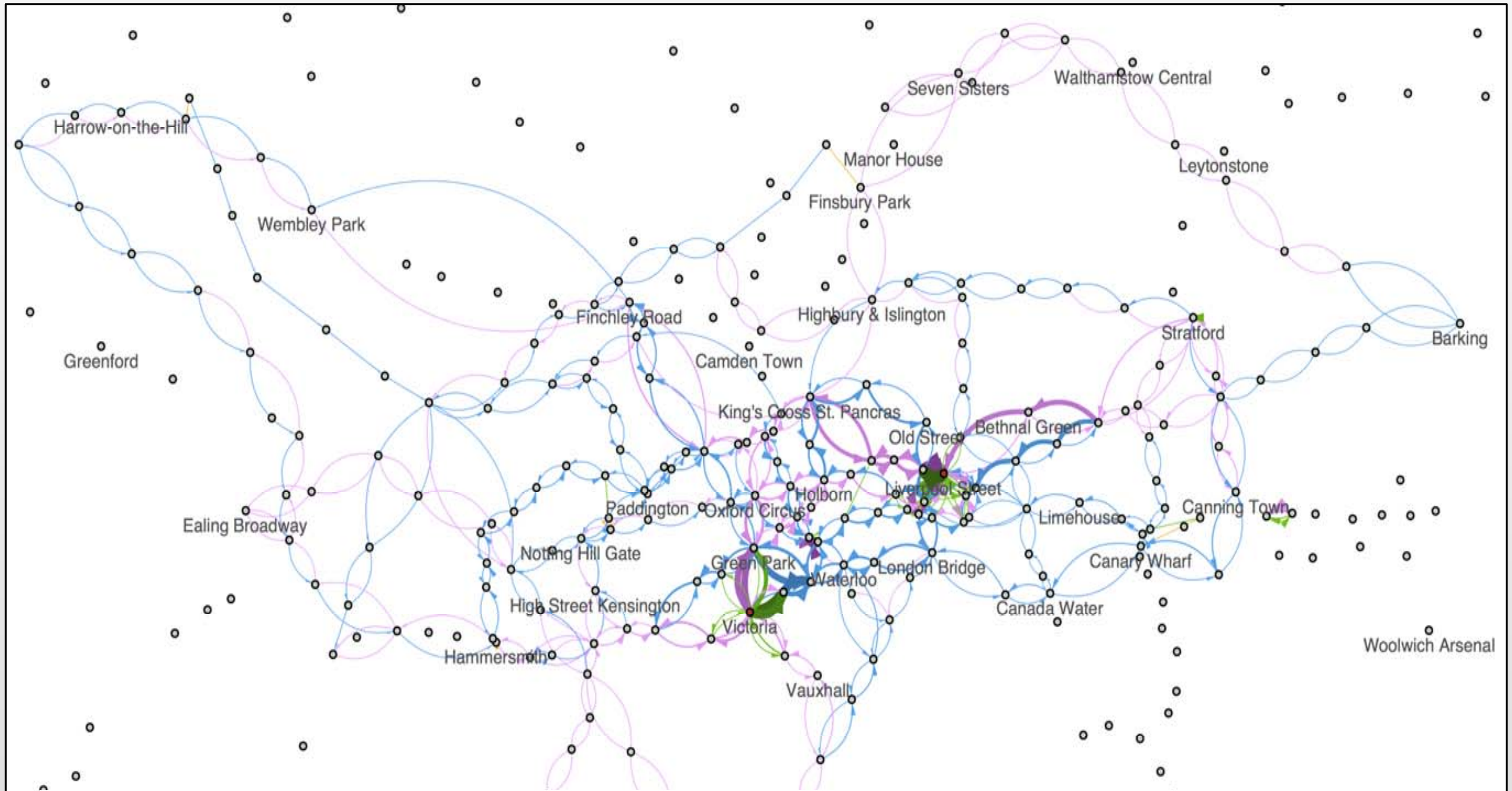


Single- and dual-station disruptions produce unexpected link-level interactions: changes in shortest-path typically cause some links to lose passengers, and gains are often less than expected. Moreover, it is not the biggest and most central stations that cause the largest shifts!

The Undisturbed Network

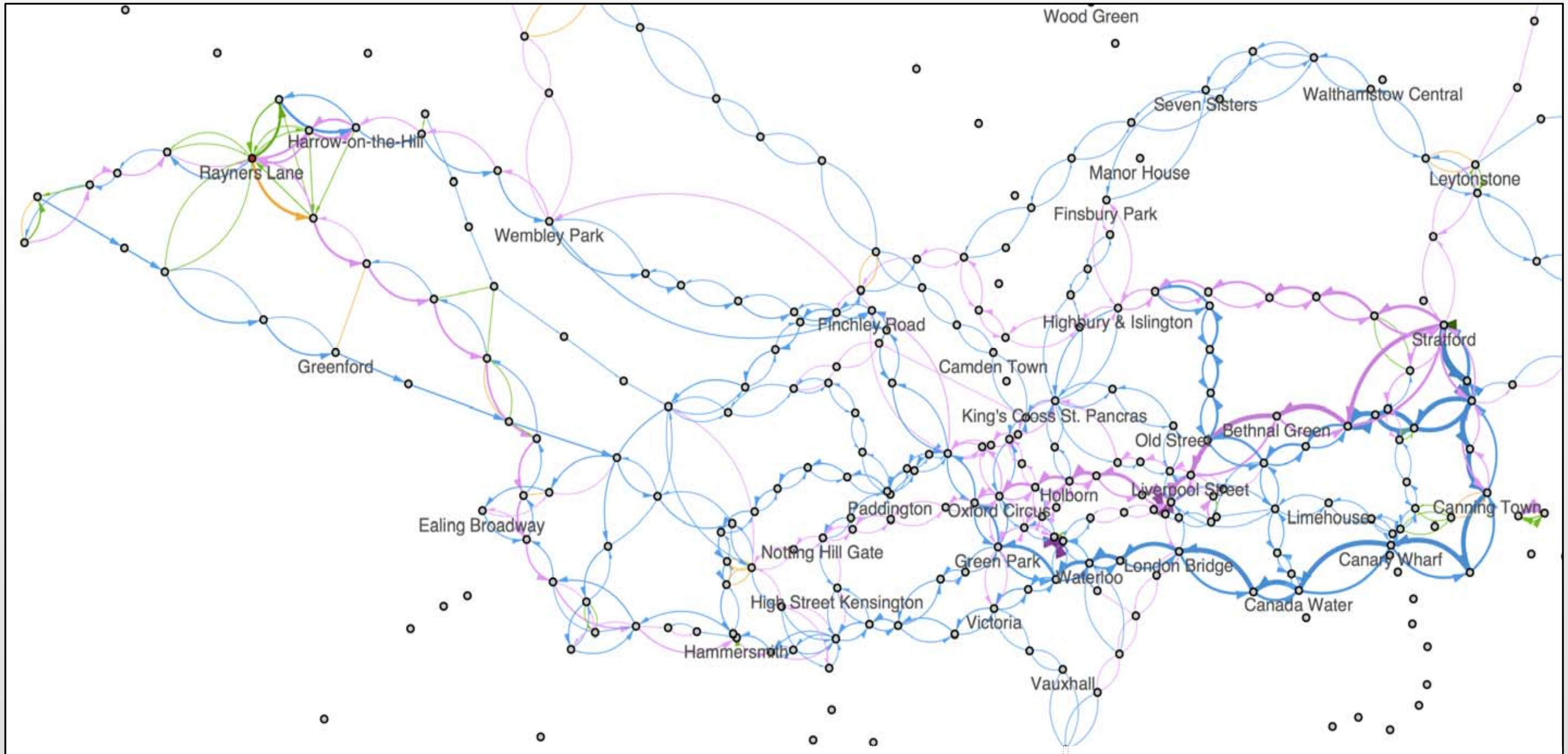


Liverpool Street & Victoria



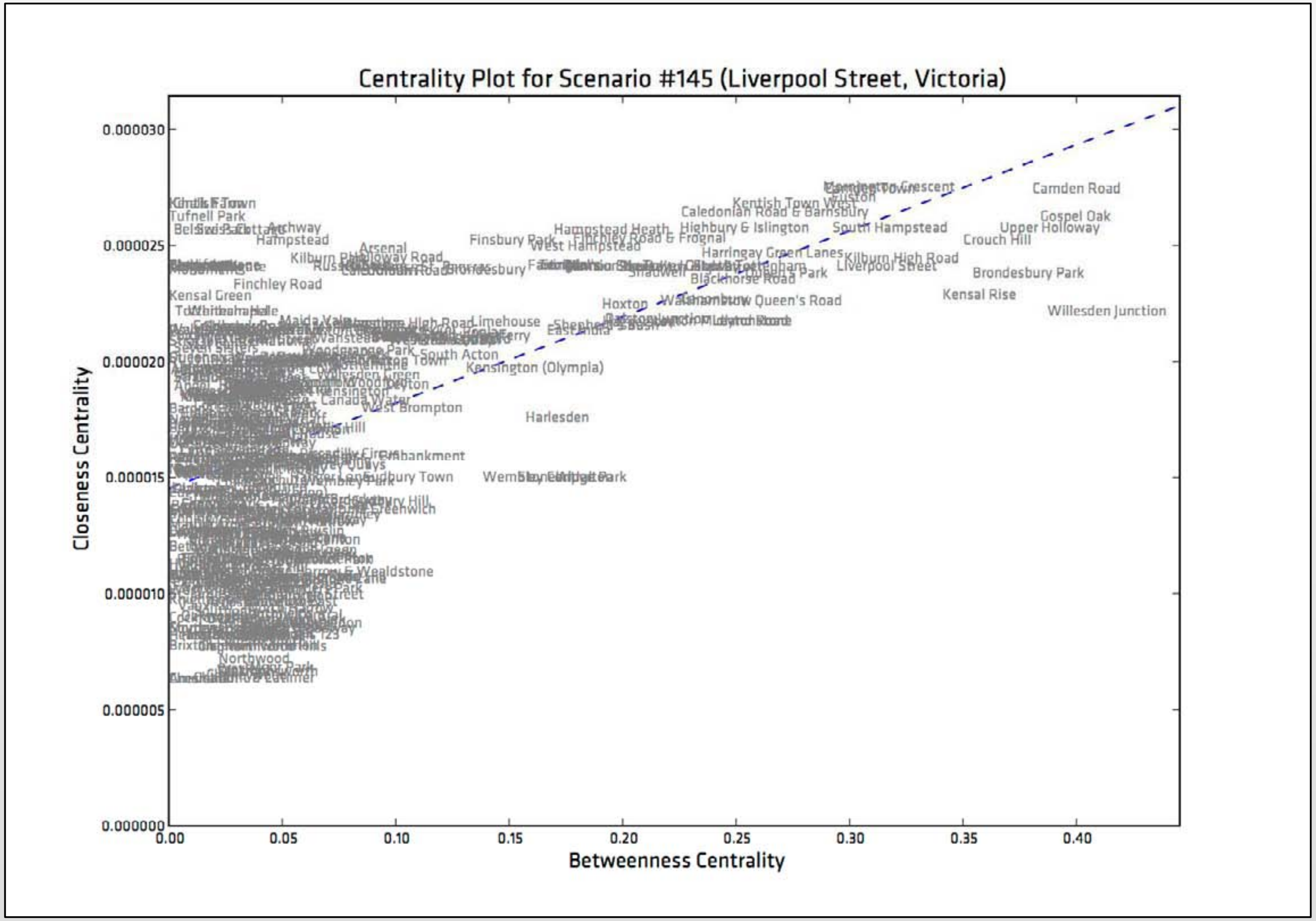
Two of London's busiest stations – because of connections to mainline rail – but if disruptions are localised to the Tube *alone* then there are many more local substitutes.

Rayners Lane & Stratford

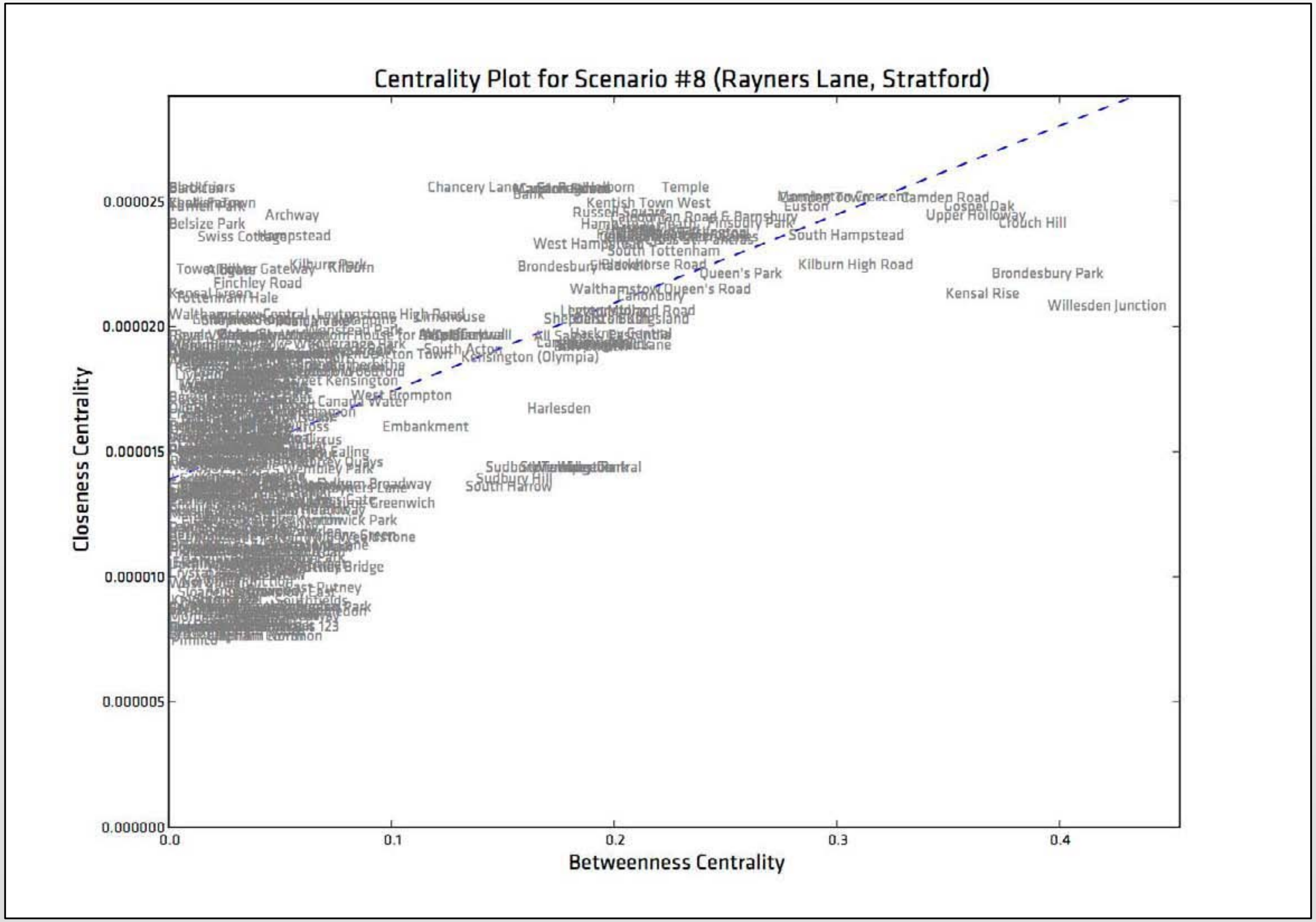


Secondary interchanges outside the core seem to cause greater disruption. Major re-routing required to complete journey, and time lost to walking long distances or travelling via more circuitous routes is much greater.

Betweenness & Closeness: Liverpool Street & Victoria



Betweenness & Closeness: Rayners Lane & Stratford



The Impact of the London Olympics

Strategy #1: Demand Distribution

The distribution of Olympic venues helped to create spectator contra-flows to much of the normal rush hour traffic; however, commuters also adjusted their behaviour in small, but important ways: increasing their use of the routes least affected by Games (e.g. the Victoria line, some NR) and moderating their use of heavily trafficked routes (e.g. the Central line).

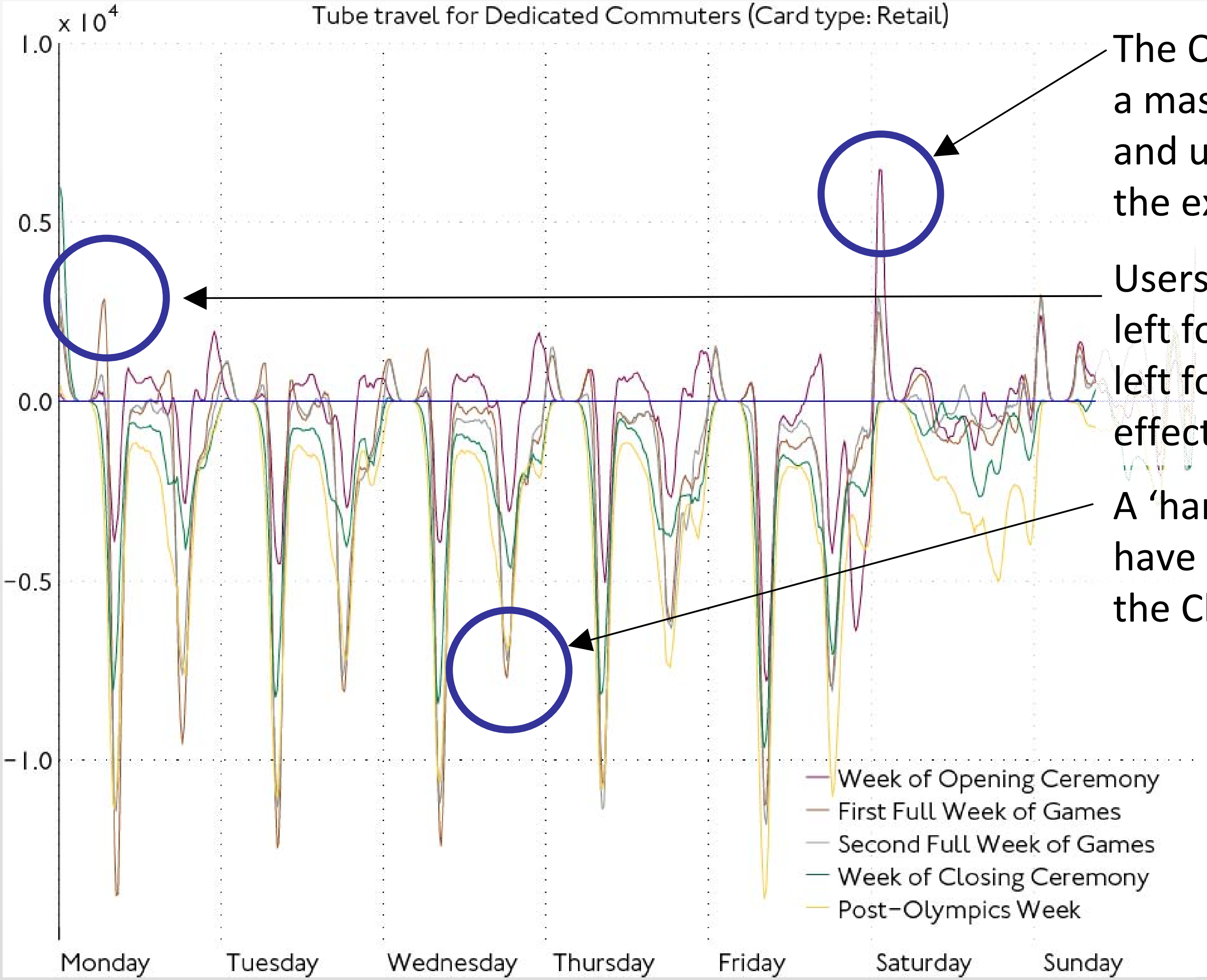
Strategy #1: Demand Distribution

Londoners with greater discretion over their travel clearly responded more actively to suggestions that they avoid intensive travel during the Games. However, the single largest group of users – 'Commuters' who make use of bus, rail and Tube network on a daily basis – made proportionally fewer adjustments to their average weekly activity.

Before the Games	Games Week #1	Games Week #2	Count	Pct.
Commuter	Commuter	Commuter	2,250,000	87%
Commuter	Commuter	Infrequent	122,000	5%
Commuter	Infrequent	Commuter	146,000	5%
1-3 Times/Week	1-3 Times/Week	1-3 Times/Week	782,000	48%
1-3 Times/Week	Infrequent	Infrequent	274,000	17%
1-3 Times/Week	Commuter	Commuter	75,000	5%

For Commuters, n ≅2,580,000; For 1-3 Times/Week, n ≅1,640,000

Strategy #3: Demand Shifting

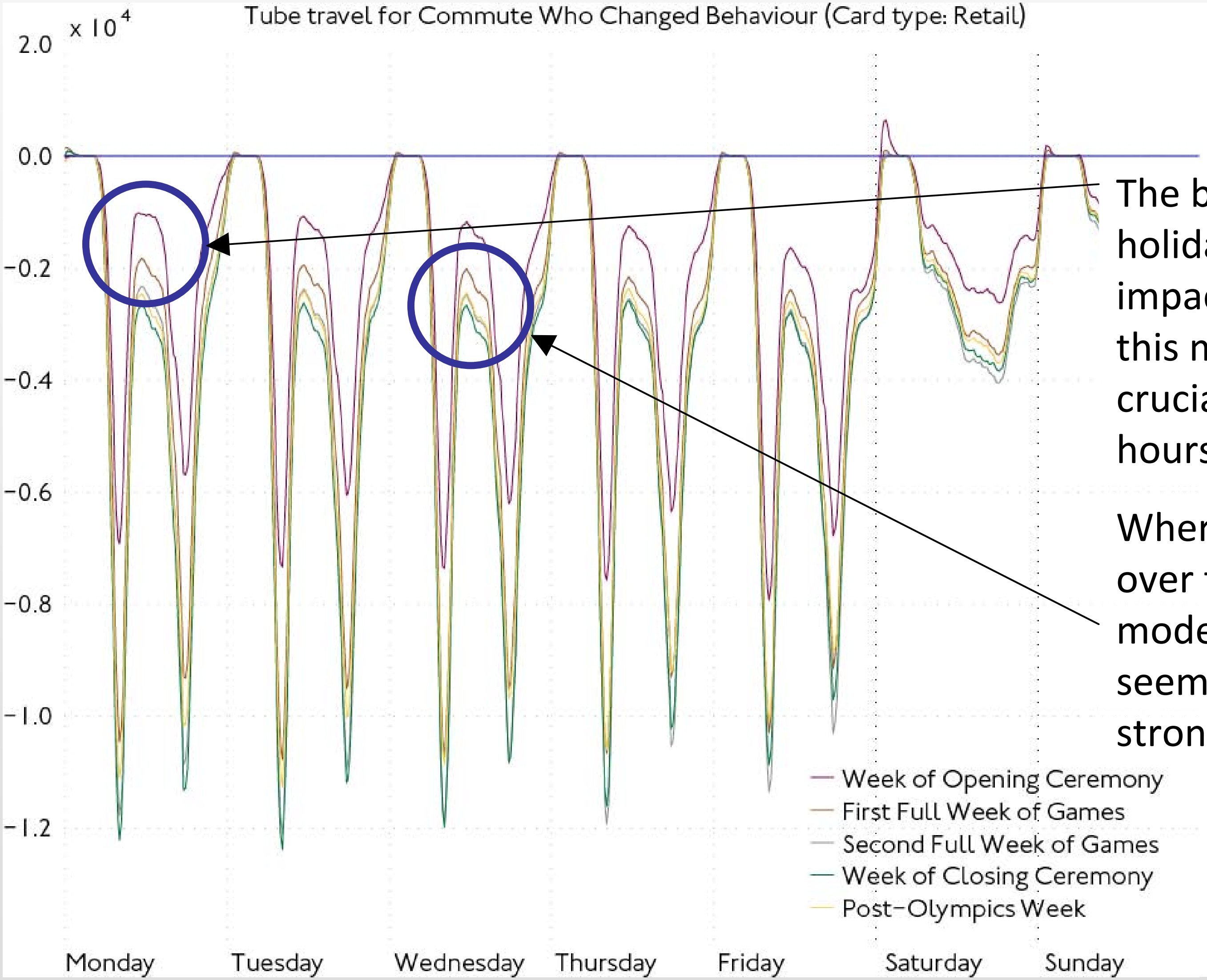


The Opening Ceremonies had a massive impact on activity, and users adjusted quickly to the extra hour.

Users left for work earlier, and left for home later. And some left for home earlier! The effect fell off rapidly.

A 'hangover' effect seems to have come into play following the Closing Ceremonies.

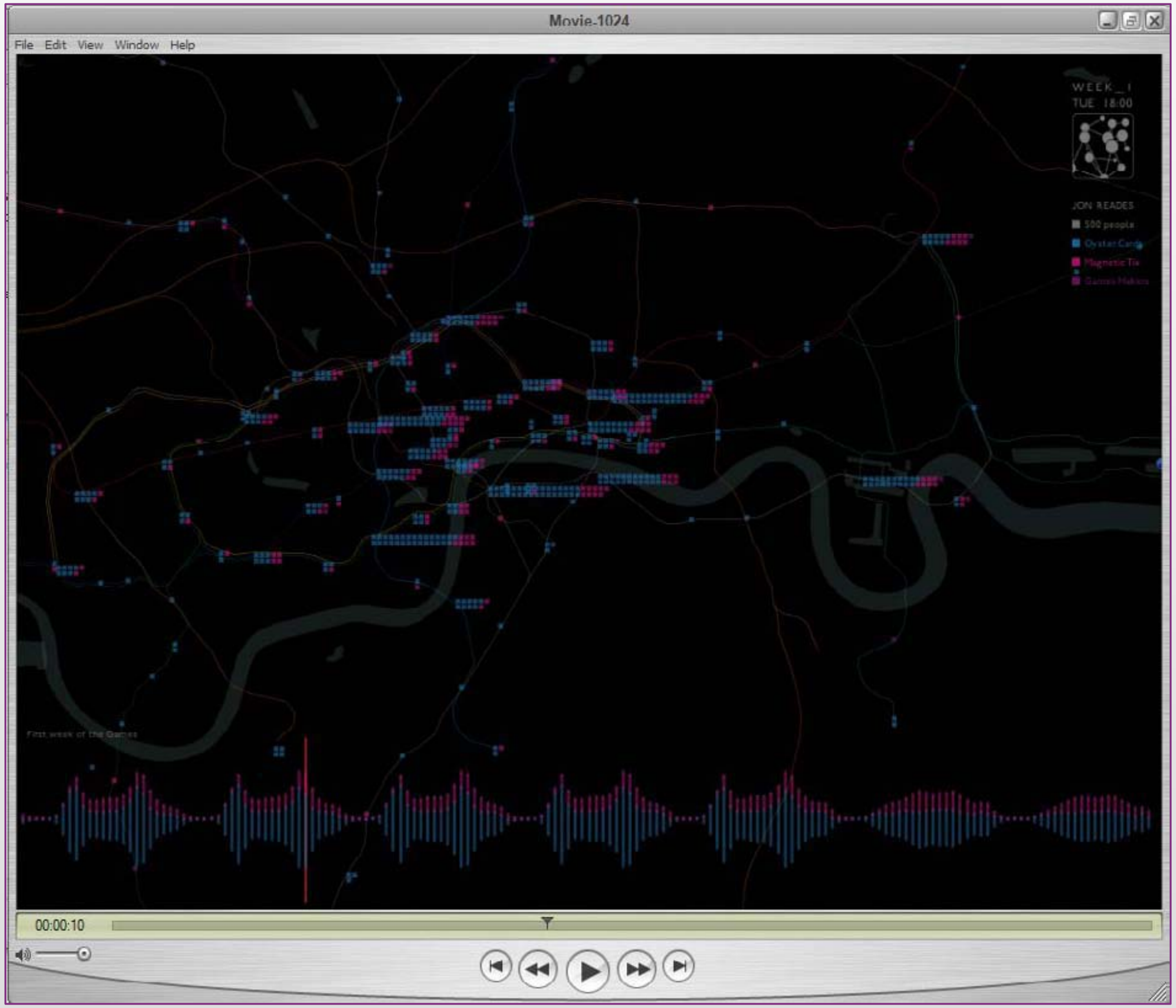
Strategy #3: Demand Shifting



The beginning of the summer holidays had an enormous impact on travel demand – this may well have been crucial to the easier rush hours.

Where users had discretion over travel frequency and mode the demand impact seems to have been much stronger.

To end, an interesting movie of the flows during the Olympics



Next Steps

1. Identify meaningful measures enabling comparison *between* scenarios:
2. Need to capture both individual station and cumulative network impacts
3. As well, the network model could be improved in several ways:
4. Better-respect known route-choice preferences using RODS survey data and, potentially, Space Syntax-like ‘cognitive complexity’ (e.g. compare difference between route time and map complexity)
5. Improve modelling of interchange penalties by taking re-entry and ticket cost into account for different passenger groups
6. Improve modelling of interchange times at particularly large and complex stations (e.g. King’s Cross St. Pancras)
7. Incorporate National Rail system flows since this will change many network measures substantially (although non-Oyster users would make analysis of potential hotspots much, much more complicated).