Smart Cities

SESSION III: Lecture 1: Transport and Transit:
Smart Systems and Big Data

Michael Batty
m.batty@ucl.ac.uk
@jmichaelbatty

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

This Lecture

Note that Lecture 2 in Session 2 has been incorporated into Lecture 1 in Session 3. If you follow the PDFs on the blog, then you find this material repeated for some of the slides in this lecture

Outline of the Lecture

1. The Context Once Again: Smart Cities, Real Time Streaming, Big Data and Movement Data
2. The Example of Transit in London
3. The Supply of Vehicles: Trains and Buses
4. The Demand for Travel: Smart Cards, Oyster
5. Resilience and Disruption
   - *Examining Network Disruption: Station and Line Closures*
   - *Examining Traveller Disruption: Stalled Trains*
6. Other Measures: Daily Polycentricity: Information Flow

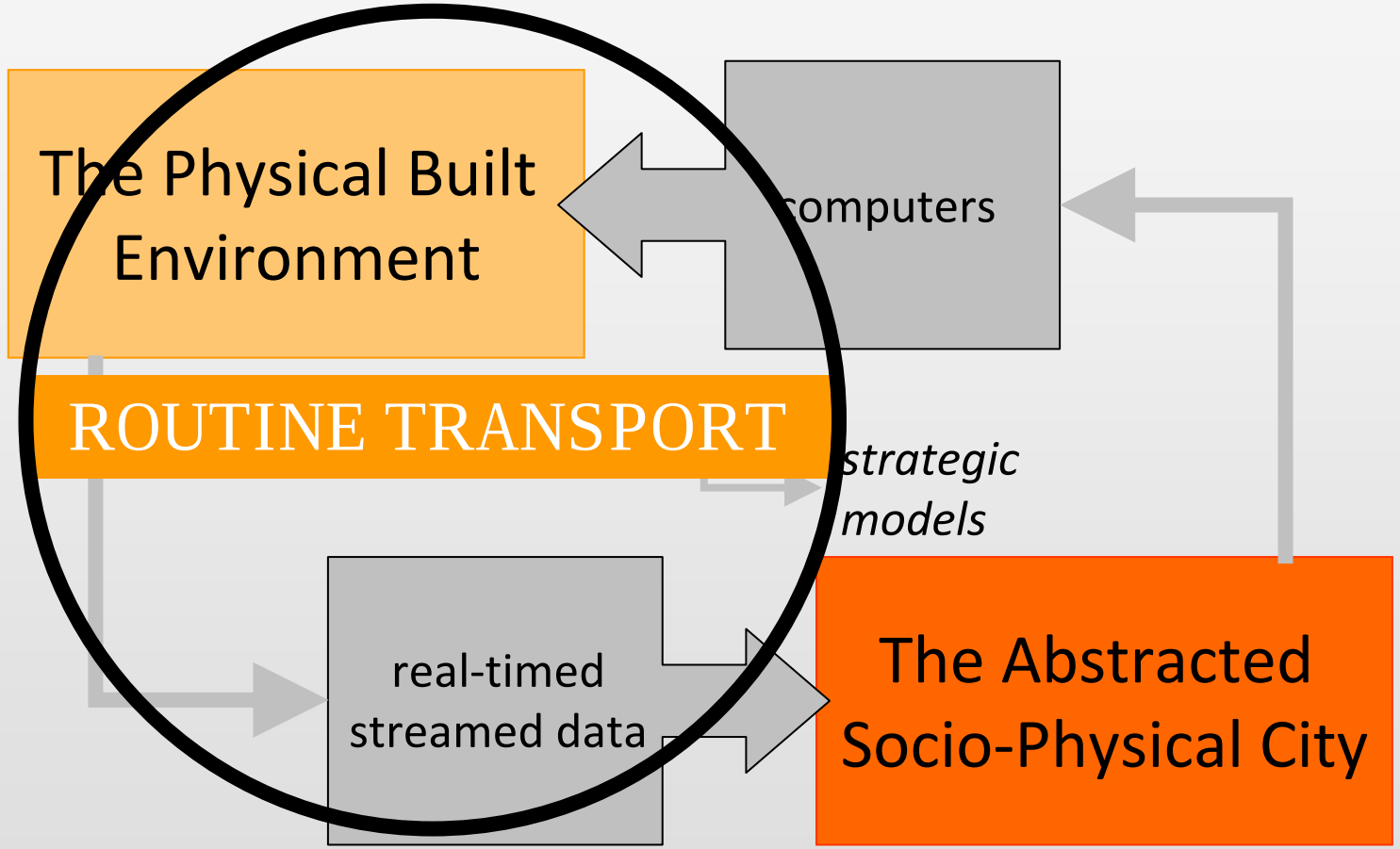
The Context Once Again: Smart Cities, Real Time Streaming, Big Data and Movement Data

Now last time we looked at a simple model of smart cities where we said that data flow and the analysis and prediction of routine functions where computers of various sorts were embedded into the built environment were the main subject of our smart cities analytics.

The very best example of this is *transit*. For in enabling users and vehicles to be computerised, we can collect large volumes of data automatically – and in real time as users and vehicles operate the system – and begin to examine problems of management and control as well as prediction.

Urban analytics in this sense is about analysing the data, visualising it, and building models for short term prediction

Our focus on transport and communications mainly affect the routine aspects of smart cities over short time periods



The Example of Transit in London

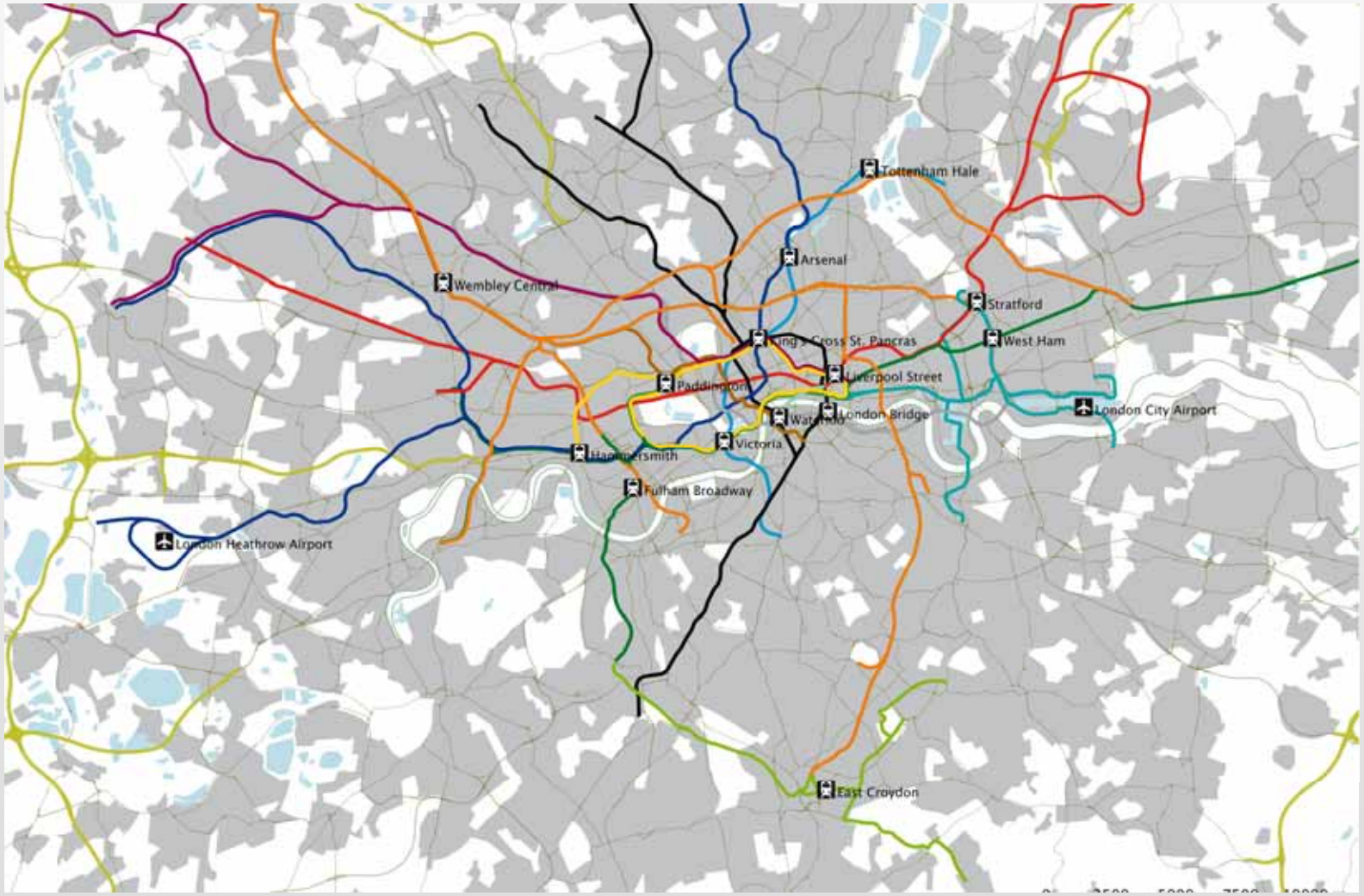
London is a good place because out of 8 million people in the inner metro area, some 50% or more are riding public transport – bus, tube (subway and light rail), heavy rail

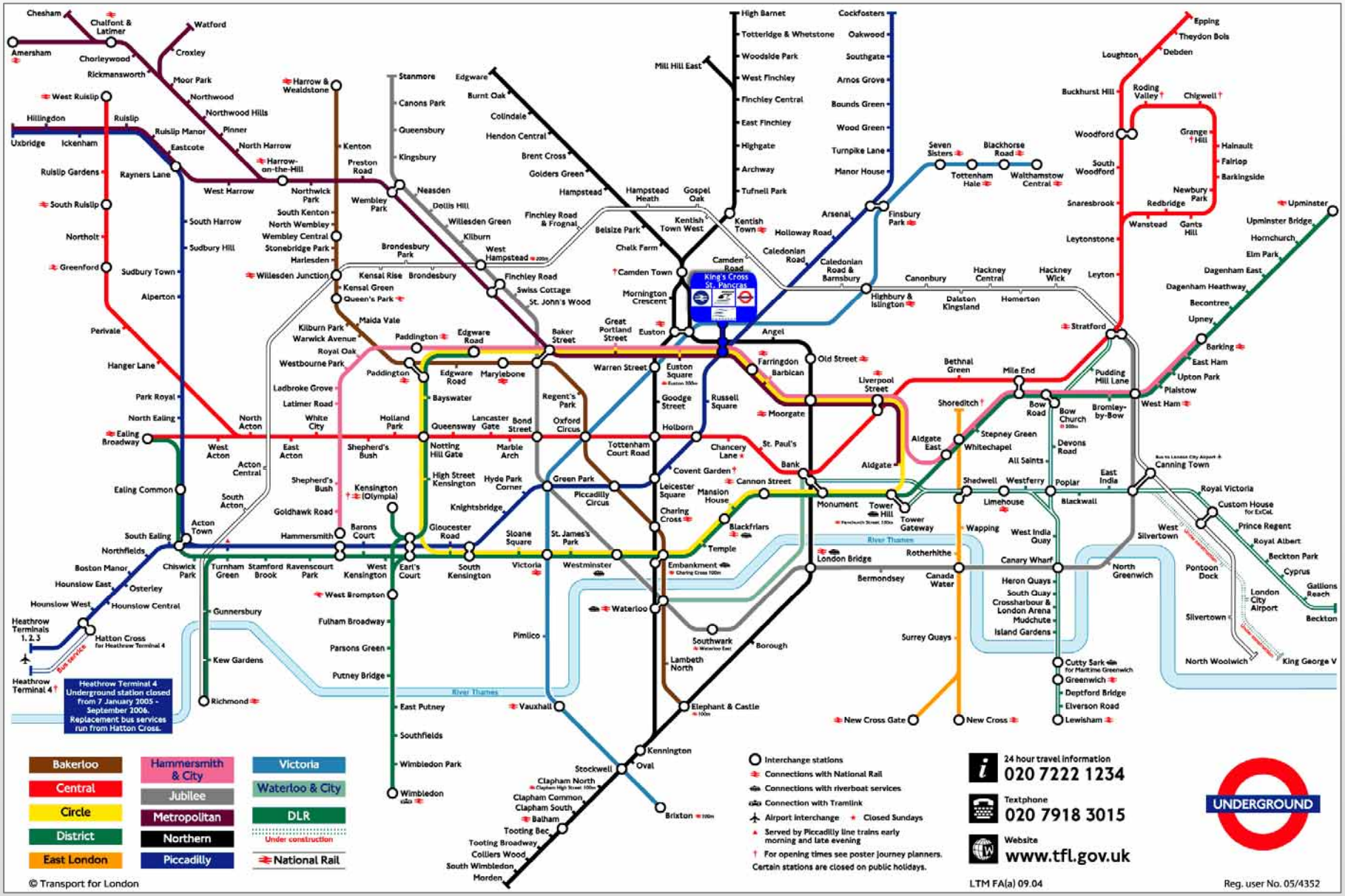
Here are some parameters of the system

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.

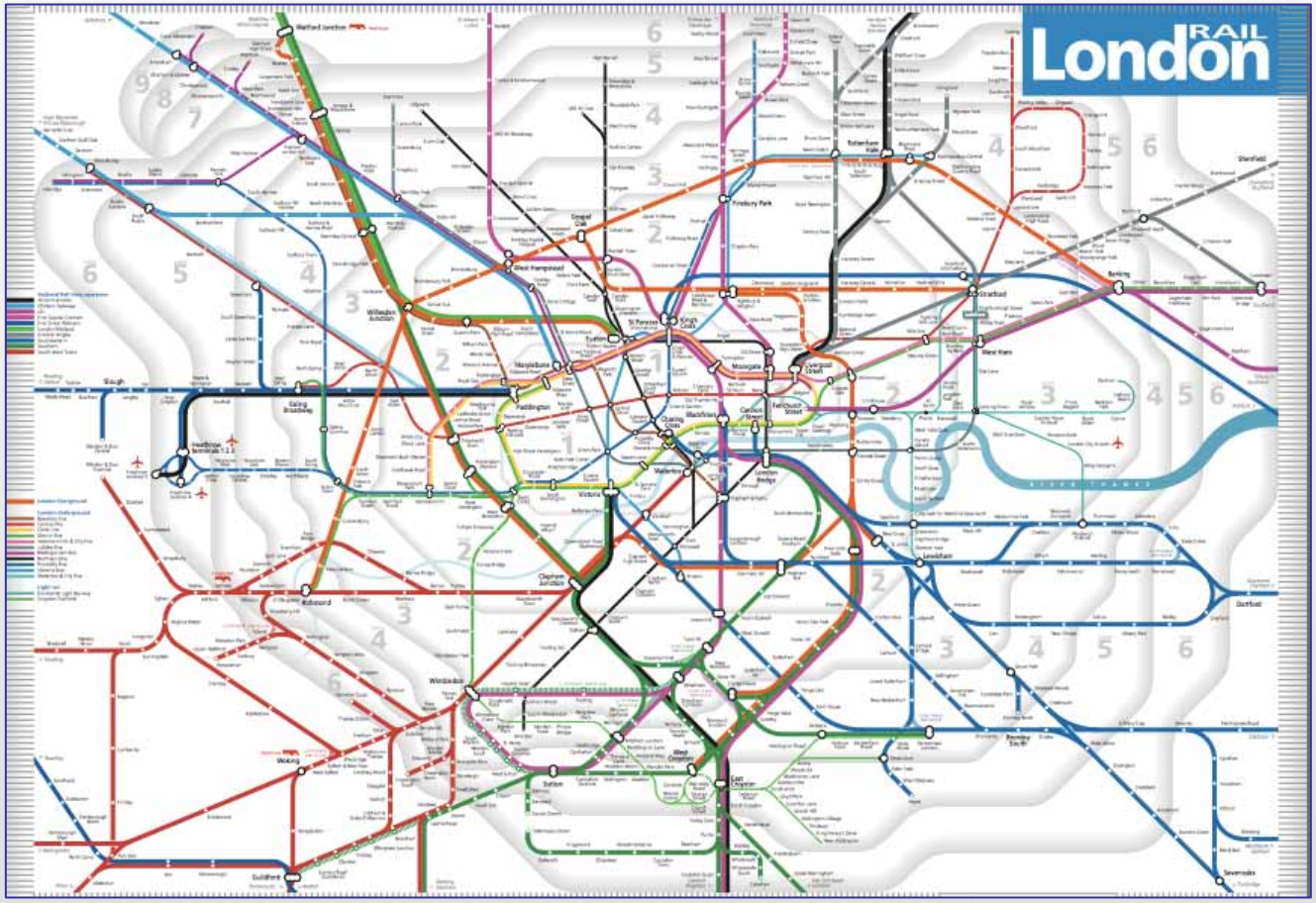
- The system is particularly large and complex – our focus on tube, DLR and overground has about 400 stations, tube about 270
- Approximately 640 stations across all modes
- 340 stations with Oyster Card readers served by National Rail trains
- 80 stations served by Overground trains
- 270 stations served 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)

Partial Views of London's Network




Bakerloo
Central
Circle
District
East London
Hammersmith & City
Jubilee
Metropolitan
Northern
Piccadilly
Victoria
Waterloo & City
DLR
Under construction
National Rail
Interchange stations
Connections with National Rail
Connections with riverboat services
Connection with Tramlink
Airport interchange
Closed Sundays
Served by Piccadilly line trains early morning and late evening
For opening times see poster journey planners.
Certain stations are closed on public holidays.
24 hour travel information
020 7222 1234
Textphone
020 7918 3015
Website
www.tfl.gov.uk
UNDERGROUND
Heathrow Terminal 4 Underground station closed from 7 January 2005 - September 2006. Replacement bus services run from Hatton Cross.
© Transport for London
LTM FA(a) 09.04
Reg. user No. 05/4352

The Total Rail Transport Environments in London



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

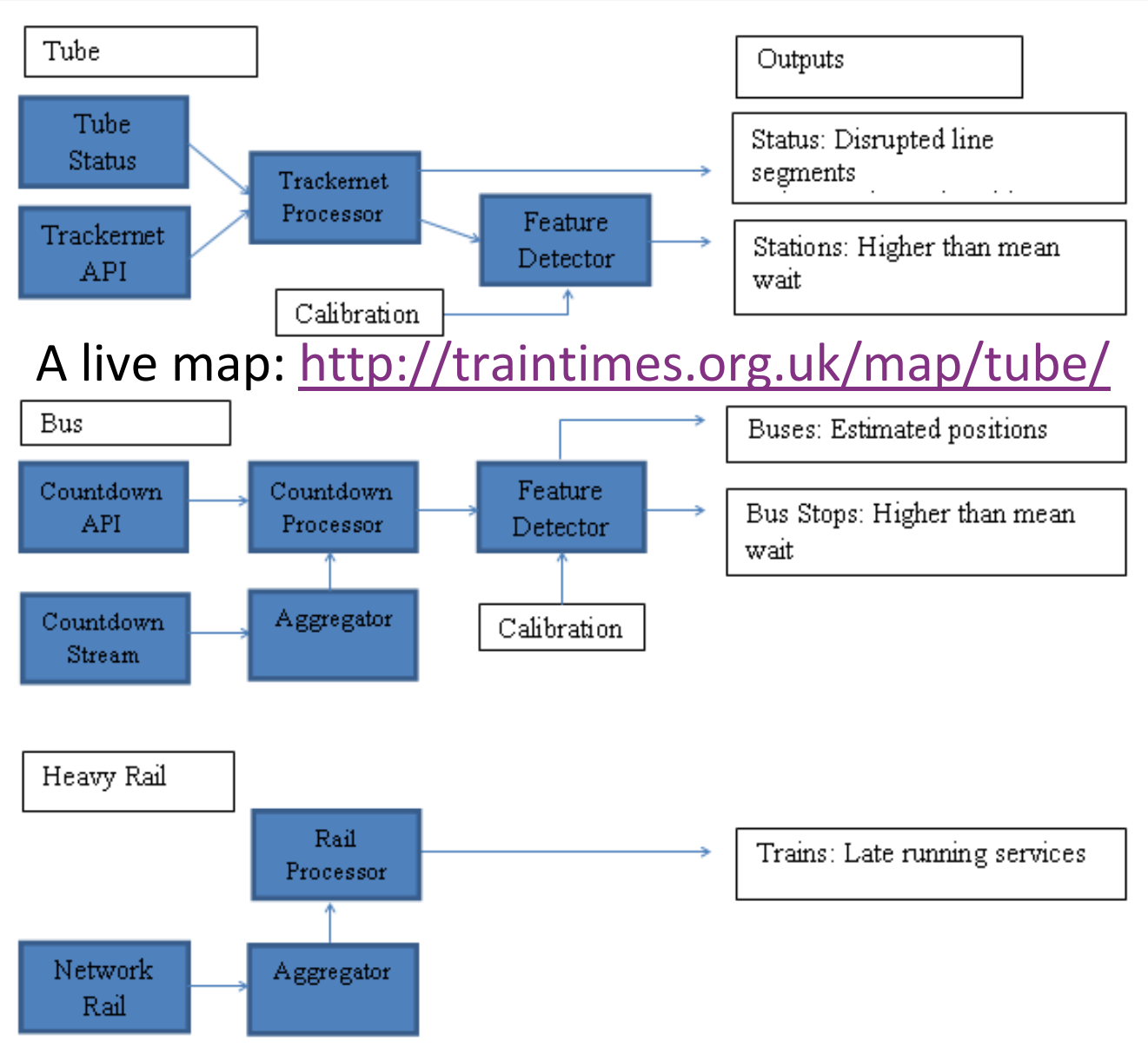
The Supply of Vehicles: Trains and Buses

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. 270 underground stations.

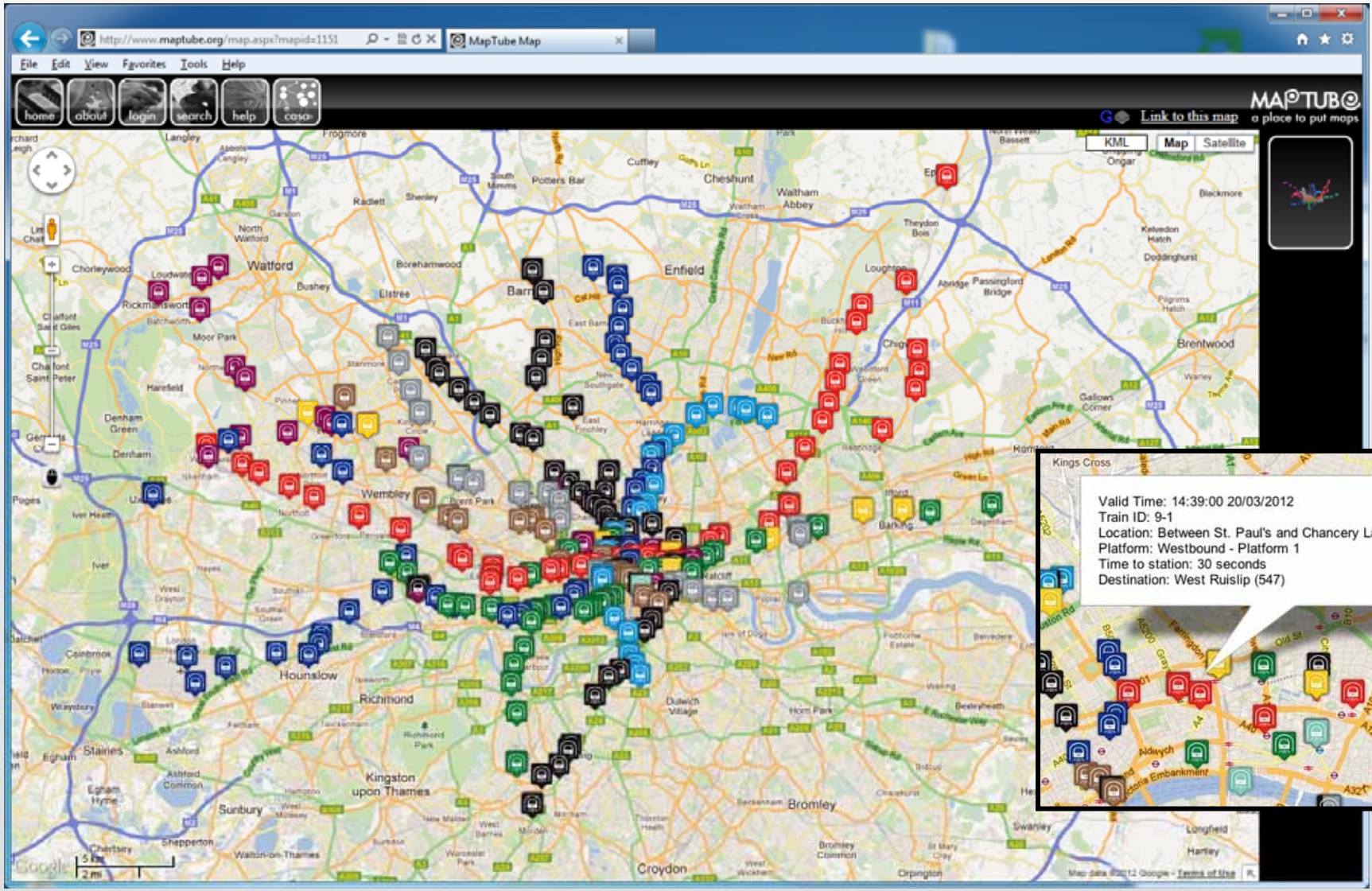
Delays for these transport systems were calculated by using an archive of historic data to find the mean wait time for every hour and every station or bus stop.

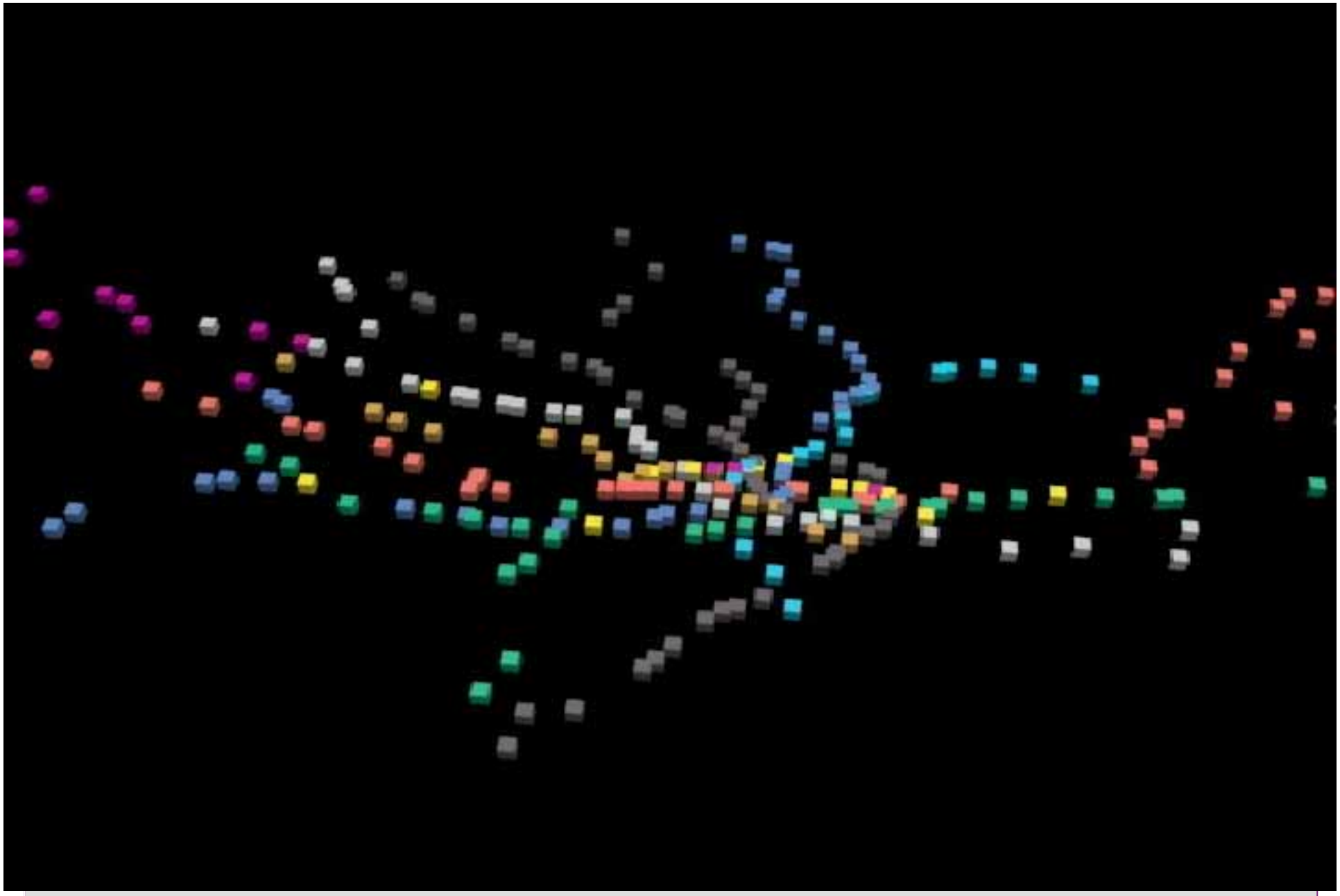
This can then be visualised in real time or after the event for further analysis. We show a mix of these visualisations in the figures that follow – here is first is the block diagram showing how we are assembling the data.

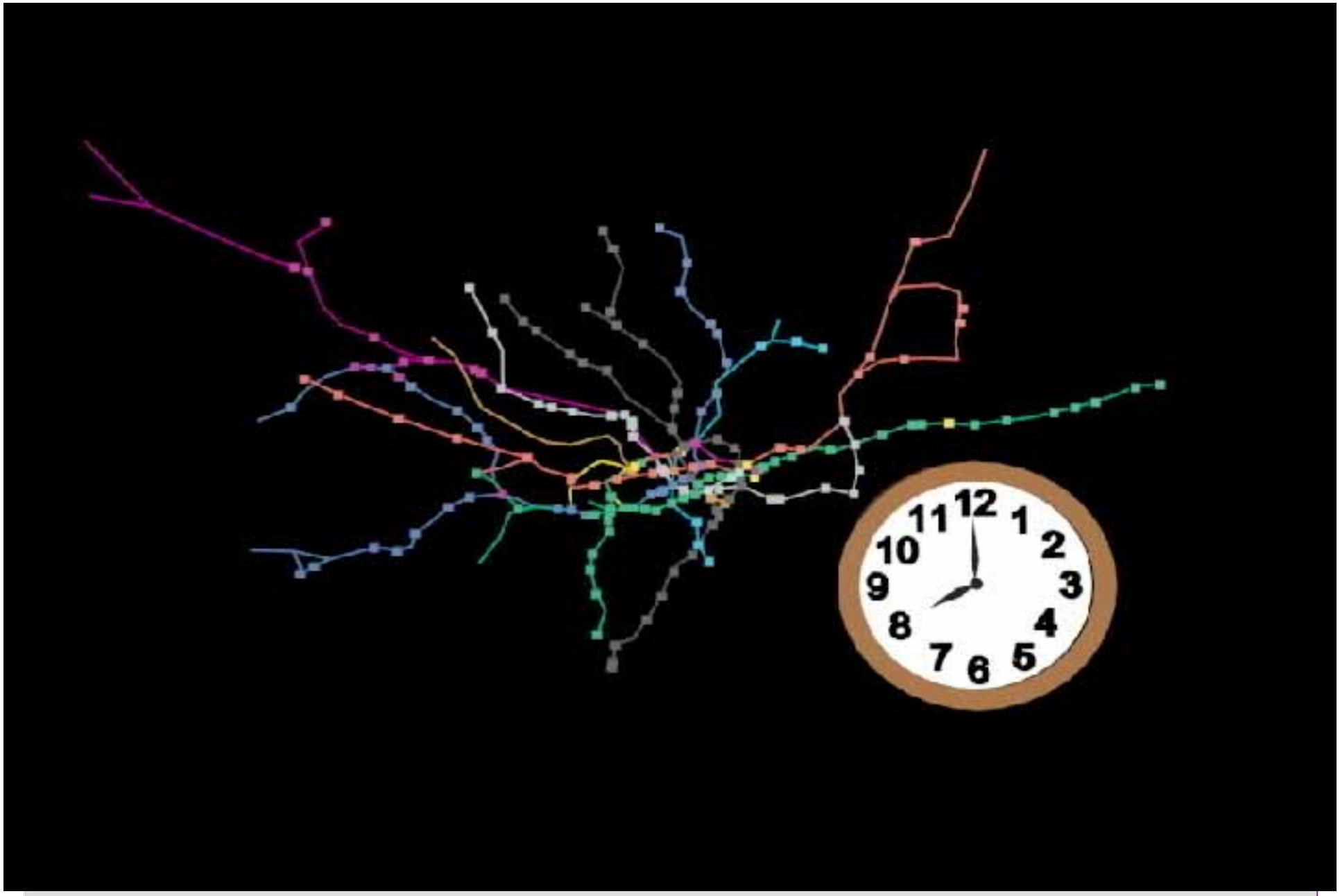


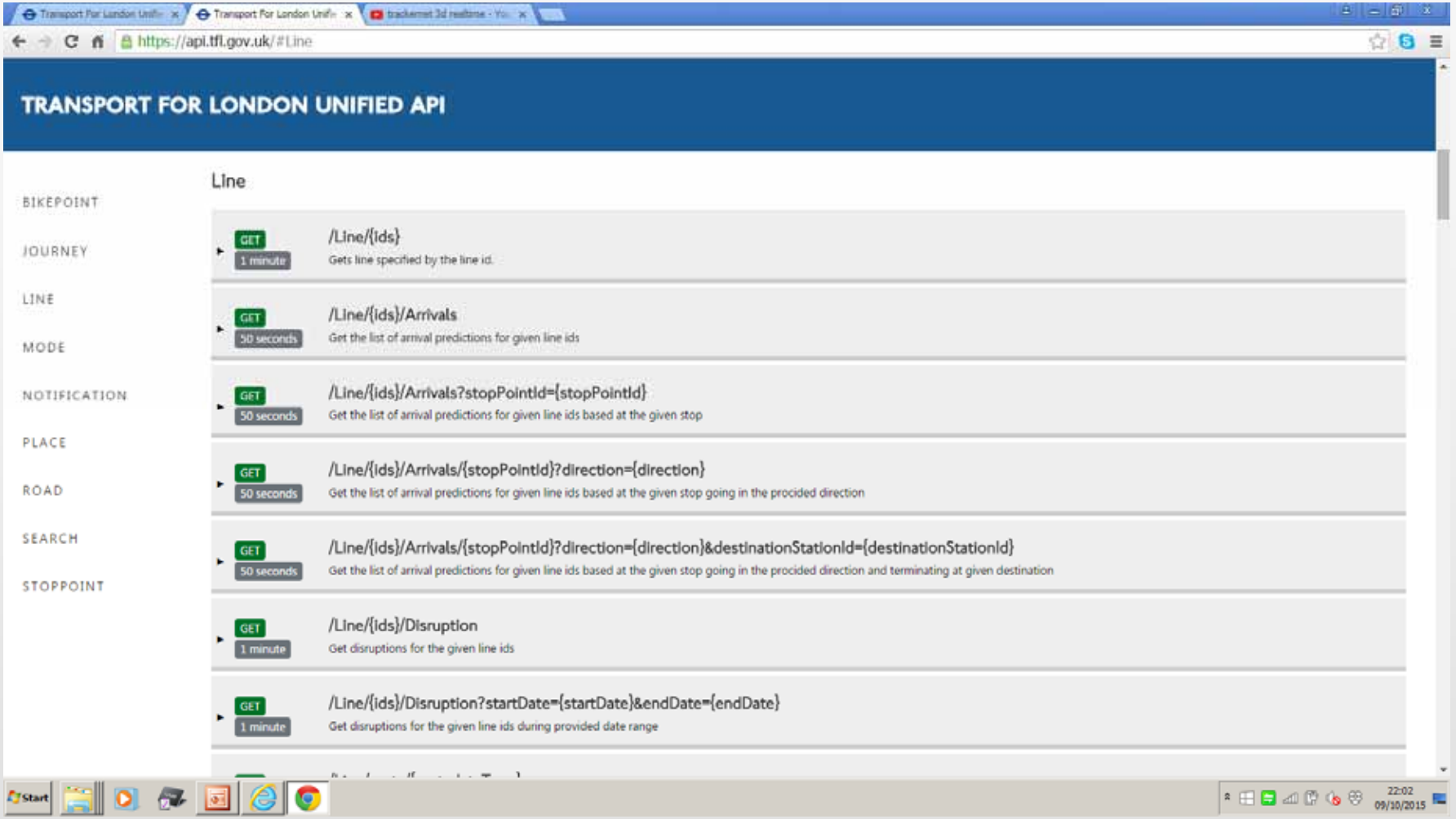
A live map: http://traintimes.org.uk/map/tube/

https://www.youtube.com/watch?v=3CnQ_ilx9RU&feature=youtu.be






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


https://api.tfl.gov.uk/#Line
TRANSPORT FOR LONDON UNIFIED API
BIKEPOINT
JOURNEY
LINE
MODE
NOTIFICATION
PLACE
ROAD
SEARCH
STOPPOINT
Line
GET
1 minute
/Line/{ids}
Gets line specified by the line id.
GET
50 seconds
/Line/{ids}/Arrivals
Get the list of arrival predictions for given line ids
GET
50 seconds
/Line/{ids}/Arrivals?stopPointId={stopPointId}
Get the list of arrival predictions for given line ids based at the given stop
GET
50 seconds
/Line/{ids}/Arrivals/{stopPointId}?direction={direction}
Get the list of arrival predictions for given line ids based at the given stop going in the procided direction
GET
50 seconds
/Line/{ids}/Arrivals/{stopPointId}?direction={direction}&destinationStationId={destinationStationId}
Get the list of arrival predictions for given line ids based at the given stop going in the procided direction and terminating at given destination
GET
1 minute
/Line/{ids}/Disruption
Get disruptions for the given line ids
GET
1 minute
/Line/{ids}/Disruption?startDate={startDate}&endDate={endDate}
Get disruptions for the given line ids during provided date range
Start
22:02
09/10/2015

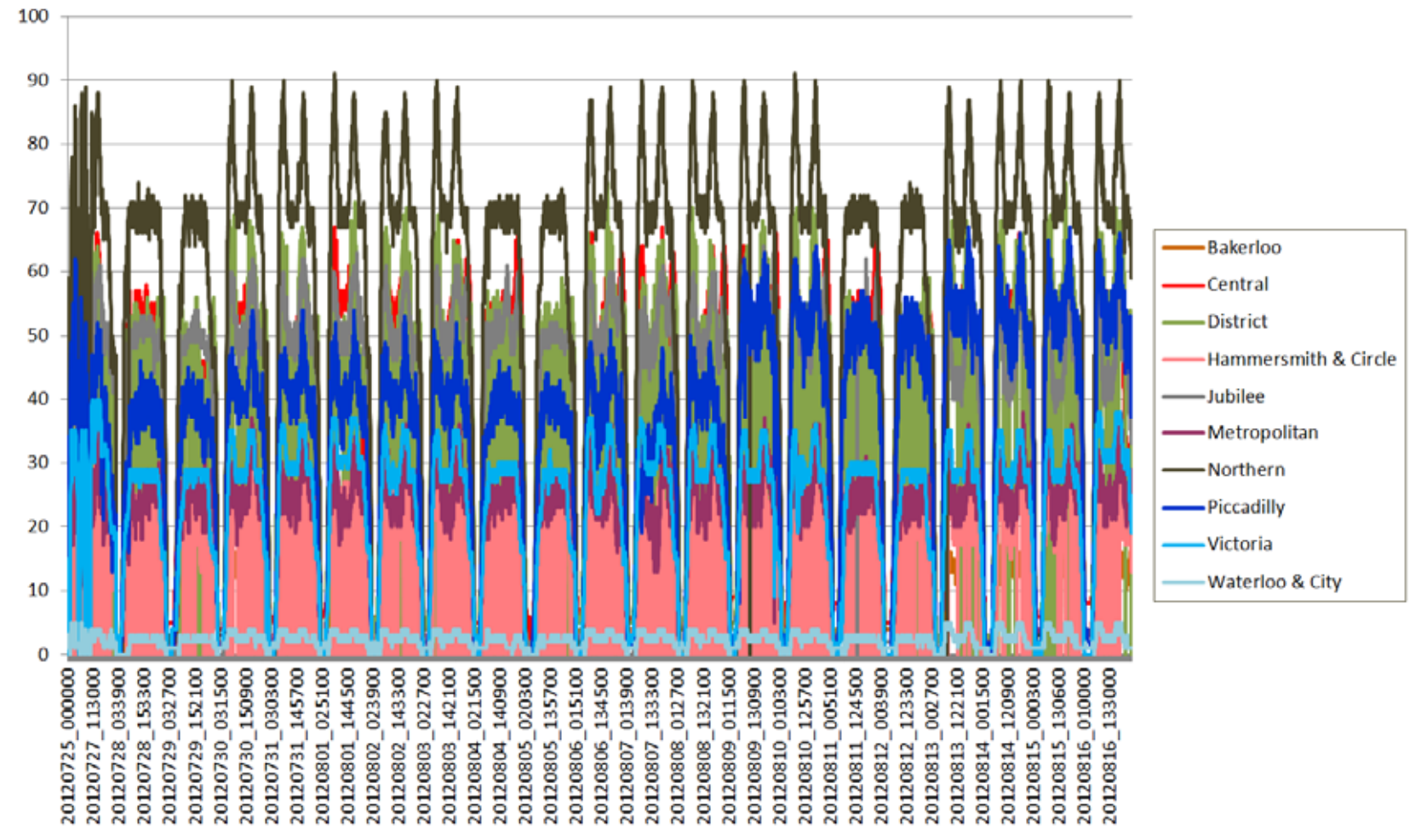
Our Unified API

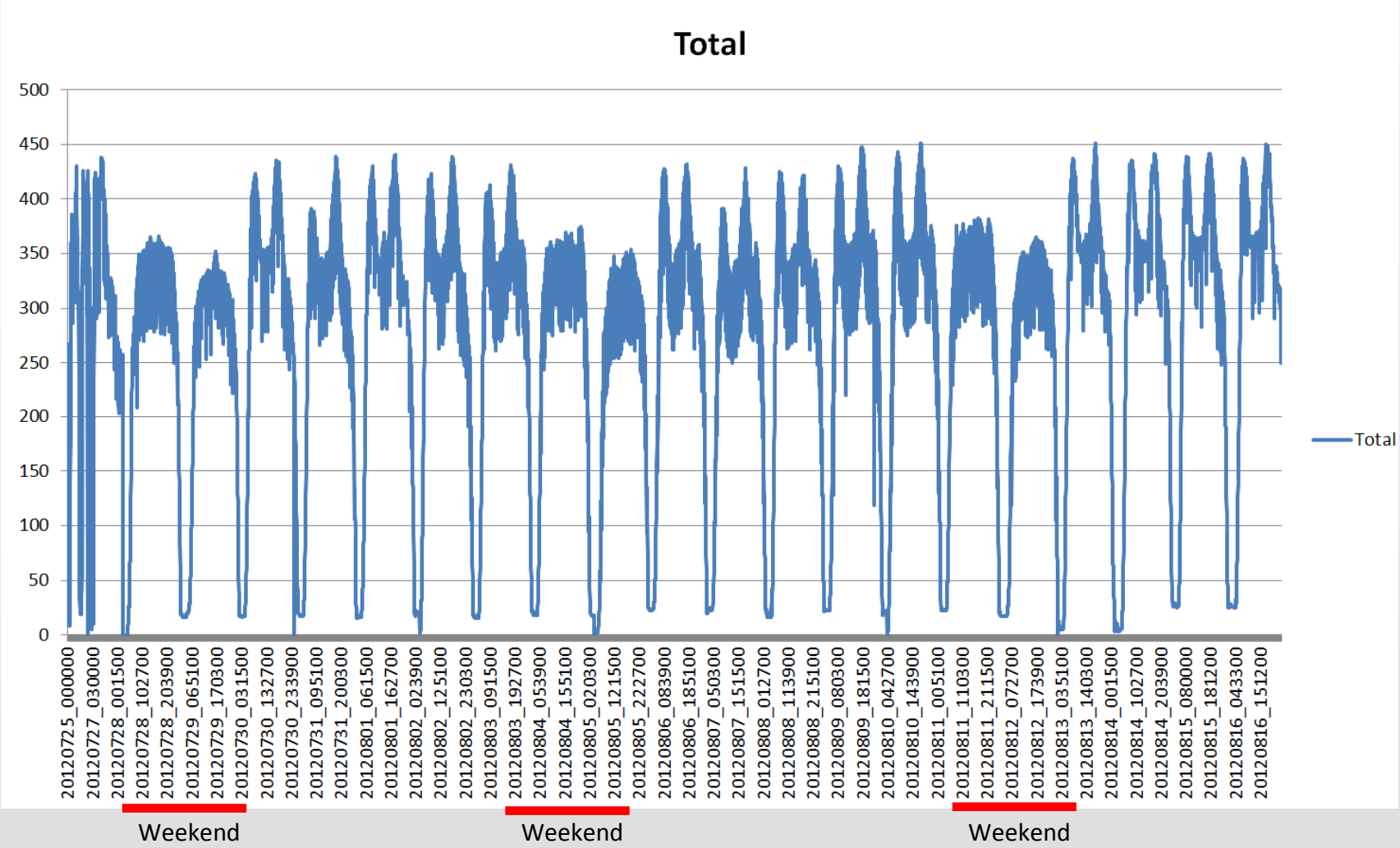
Our unified API brings together data across all modes of transport into a single RESTful API. This API provides access to the most highly requested realtime and status infomation across all the modes of transport, in a single and consistent way. Access to the devloper documentation is available at https://api.tfl.gov.uk

Example calls

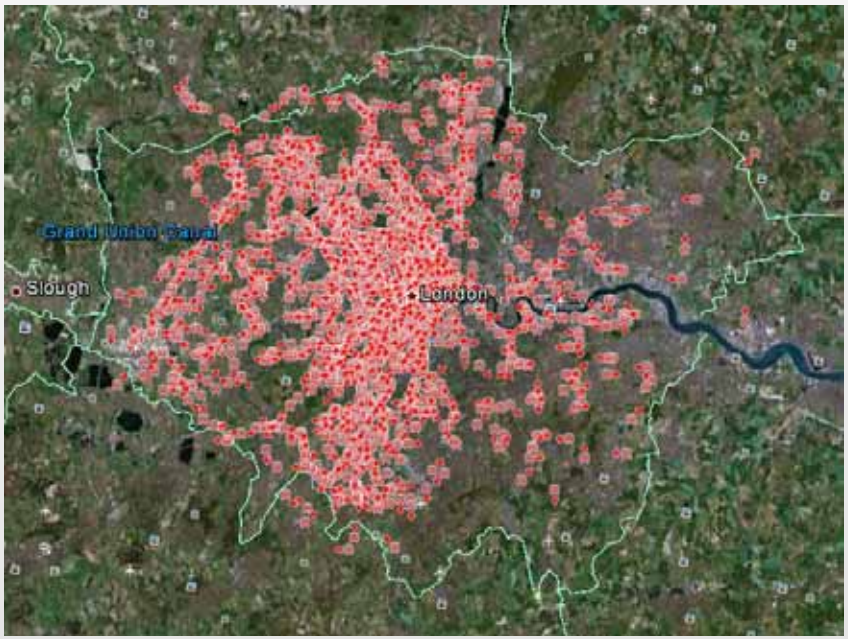
API area	Example calls	Description
Journey planning - Westminster to Bank (Disambiguation)	https://api.tfl.gov.uk/journey/journeyresults/westminster/to/bank	Journey planner supports many parameters to help find the right journey

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

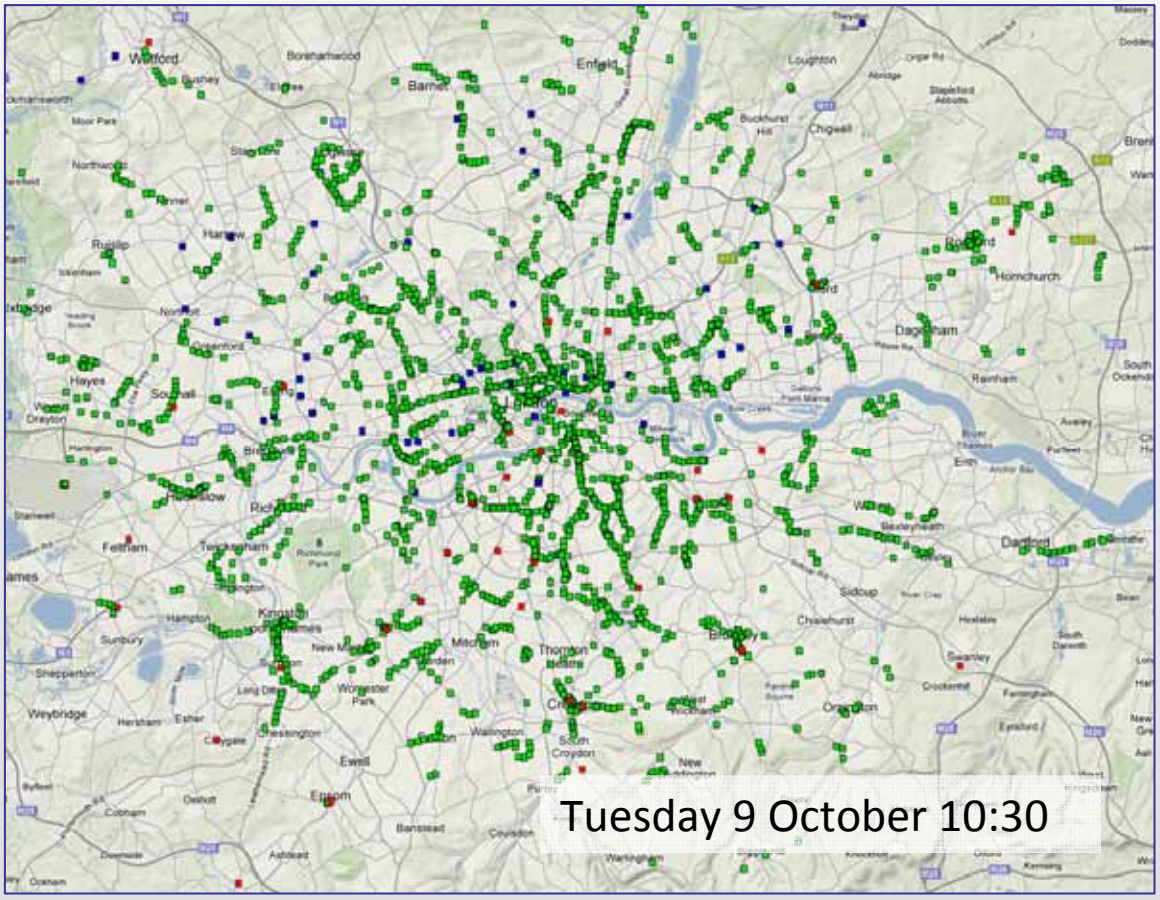


Tuesday 22nd May 2012, 09:00

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.

Delays from Tube, National Rail and Bus Fused

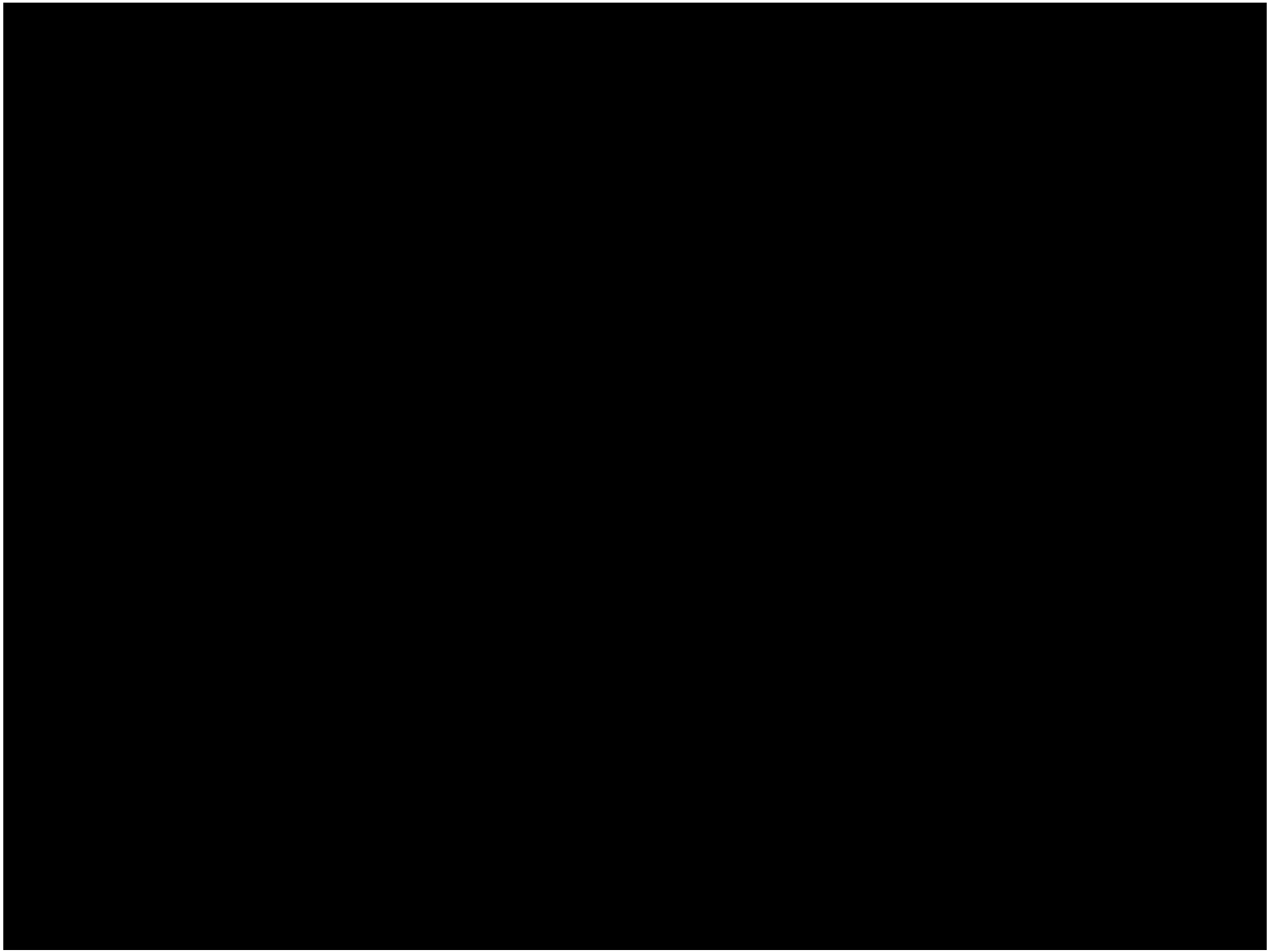


Key

- 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

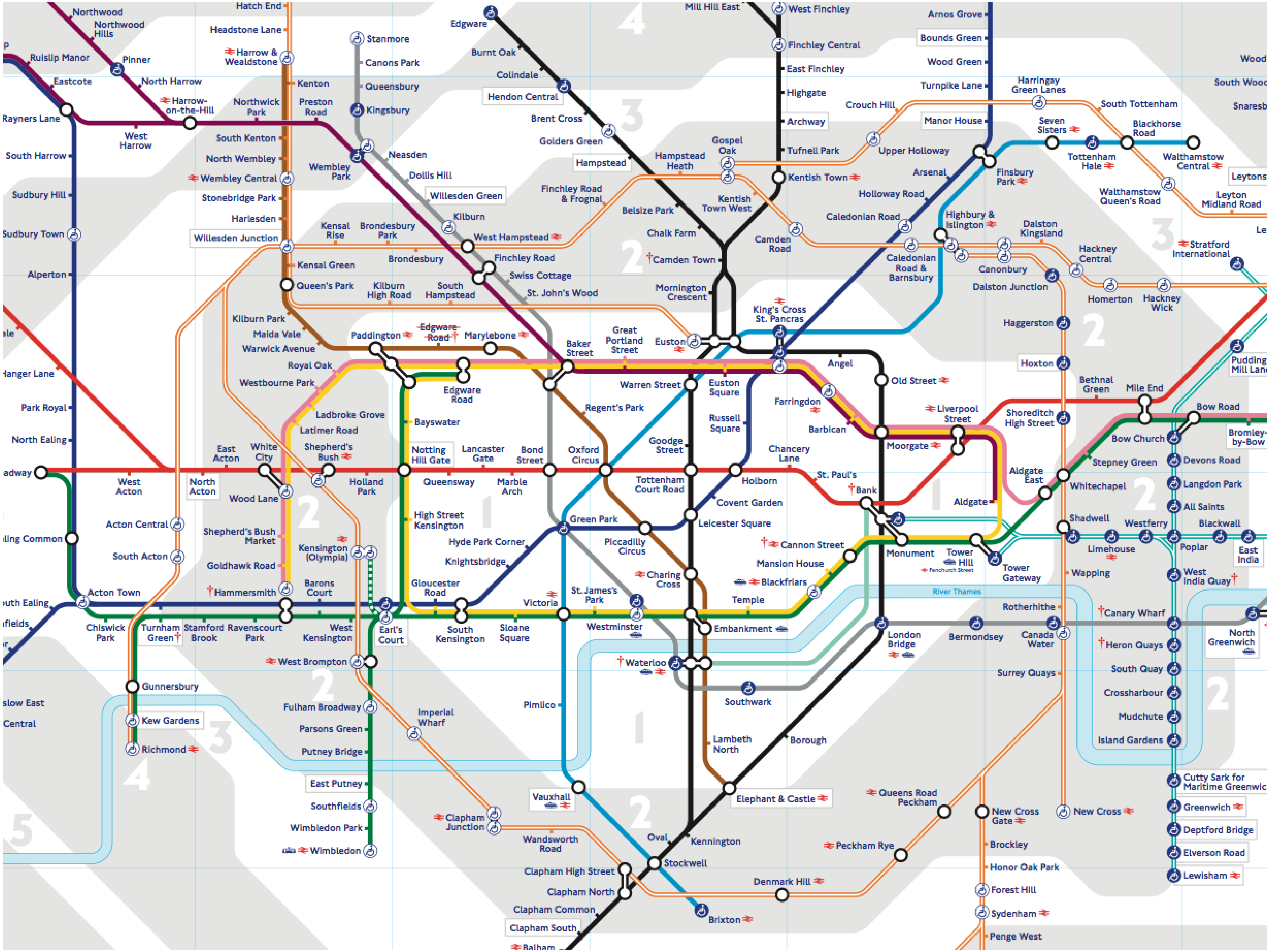
The Demand for Travel: Smart Cards, Oyster



Smart Card Data

Oyster Card Taps

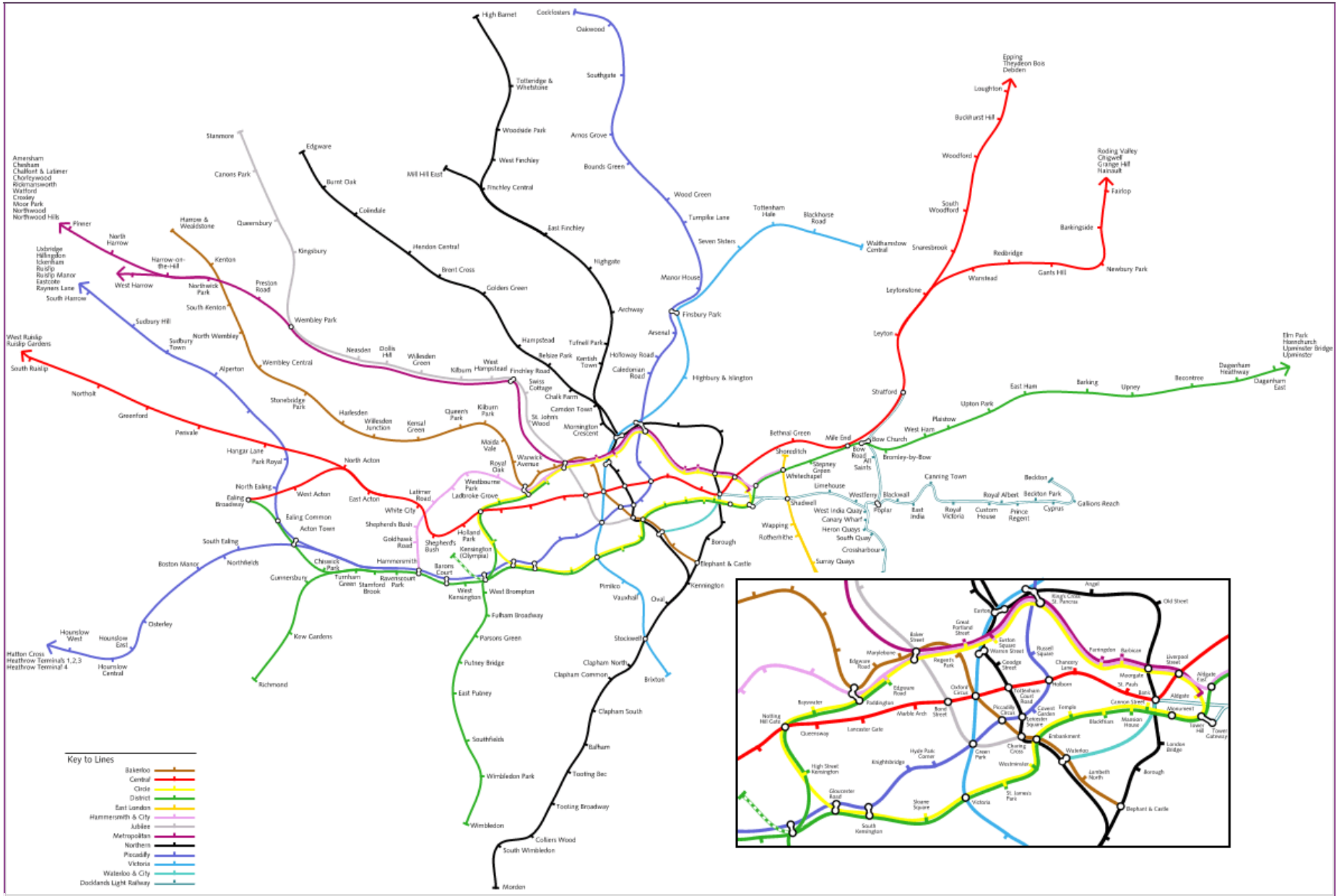
- Tap at **start** and **end** of train journeys
- Tap at **start only** on buses
- Accepted at 695 Underground and rail stations, and on thousands of buses. **1.053 billion** Oyster Card taps over July to September 201
- 762 million OD trips, 291 million UG and rail trips, 11.5m Oyster Cards. No routing data

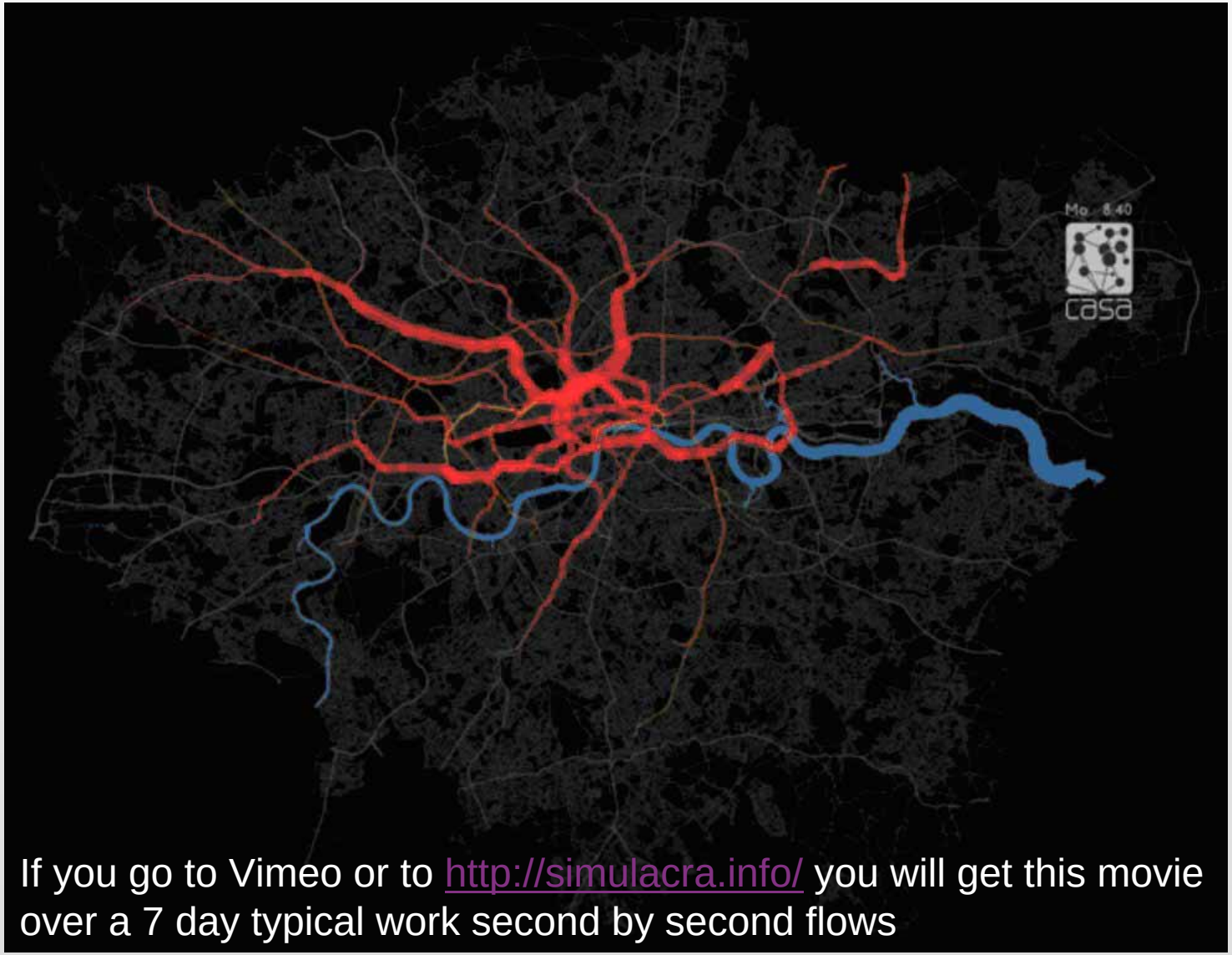


Northwood
Northwood Hills
Ruislip Manor
Pinner
Eastcote
North Harrow
Harrow-on-the-Hill
Rayners Lane
West Harrow
South Harrow
Sudbury Hill
Sudbury Town
Alperton
Hanger Lane
Park Royal
North Ealing
West Acton
North Acton
East Acton
White City
Wood Lane
Acton Central
Ealing Common
South Acton
Acton Town
Chiswick Park
Turnham Green
Stamford Brook
Ravenscourt Park
Gunnersbury
Kew Gardens
Richmond
Hatch End
Headstone Lane
Harrow & Wealdstone
Kenton
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
Shepherd's Bush
Shepherd's Bush Market
Goldhawk Road
Hammersmith
Barons Court
West Kensington
Kensington (Olympia)
Earl's Court
West Brompton
Fulham Broadway
Parsons Green
Putney Bridge
East Putney
Southfields
Wimbledon Park
Wimbledon
Stanmore
Canons Park
Queensbury
Kingsbury
Northwick Park
Preston Road
Wembley Park
Neasden
Dollis Hill
Willesden Green
Kilburn
West Hampstead
Kensal Rise
Brondesbury Park
Brondesbury
Kilburn High Road
South Hampstead
Finchley Road
Swiss Cottage
St. John's Wood
Edgware
Burnt Oak
Colindale
Hendon Central
Brent Cross
Golders Green
Hampstead
Hampstead Heath
Finchley Road & Frognal
Belsize Park
Chalk Farm
Camden Town
Gospel Oak
Kentish Town West
Mornington Crescent
Camden Road
Mill Hill East
West Finchley
Finchley Central
East Finchley
Highgate
Archway
Tufnell Park
Kentish Town
Crouch Hill
Upper Holloway
Paddington
Edgware Road
Marylebone
Baker Street
Great Portland Street
Euston
Edgware Road
Bayswater
Notting Hill Gate
Holland Park
Queensway
Lancaster Gate
Marble Arch
Bond Street
Oxford Circus
Regent's Park
Warren Street
Euston Square
Goodge Street
Tottenham Court Road
Russell Square
Holborn
Covent Garden
Leicester Square
Chancery Lane
High Street Kensington
Hyde Park Corner
Knightsbridge
Green Park
Piccadilly Circus
Charing Cross
Gloucester Road
South Kensington
Sloane Square
Victoria
St. James's Park
Westminster
Embankment
Temple
Blackfriars
Mansion House
Cannon Street
Imperial Wharf
Clapham Junction
Wandsworth Road
Pimlico
Vauxhall
Waterloo
Southwark
Lambeth North
Borough
Elephant & Castle
Kennington
Oval
Stockwell
Clapham High Street
Clapham North
Clapham Common
Clapham South
Balham
Brixton
Denmark Hill
Peckham Rye
Queens Road Peckham
King's Cross St. Pancras
Angel
Old Street
Farringdon
Barbican
Moorgate
Liverpool Street
St. Paul's
Bank
Monument
Tower Hill
Fenchurch Street
Tower Gateway
Aldgate
Aldgate East
London Bridge
Bermondsey
Canada Water
Rotherhithe
Wapping
Surrey Quays
New Cross Gate
New Cross
Brockley
Honor Oak Park
Forest Hill
Sydenham
Penge West
Arnos Grove
Bounds Green
Wood Green
Turnpike Lane
Harringay Green Lanes
Manor House
South Tottenham
Seven Sisters
Tottenham Hale
Blackhorse Road
Walthamstow Central
Walthamstow Queen's Road
Leyton Midland Road
Finsbury Park
Arsenal
Holloway Road
Caledonian Road
Highbury & Islington
Caledonian Road & Barnsbury
Canonbury
Dalston Kingsland
Dalston Junction
Haggerston
Hoxton
Hackney Central
Homerton
Hackney Wick
Stratford International
Pudding Mill Lane
Bethnal Green
Mile End
Bow Road
Bromley-by-Bow
Shoreditch High Street
Bow Church
Stepney Green
Whitechapel
Devons Road
Langdon Park
All Saints
Shadwell
Westferry
Blackwall
Limehouse
Poplar
East India
West India Quay
Canary Wharf
Heron Quays
South Quay
Crossharbour
Mudchute
Island Gardens
North Greenwich
Cutty Sark for Maritime Greenwich
Greenwich
Deptford Bridge
Elverson Road
Lewisham
River Thames


Service information
Date Saturday 19th January 2008
Time Planned Weekend Closure
No Service On The Entire Line
Between Hammersmith
And Kings Cross
Between Baker Street
And Aldgate Also Between Wembley Park
And Uxbridge / Northwood.
Between Earls Court
And Edgware Road
Between Green Park
And Stratford
Holloway Road Station Closed


UNDERGROUND

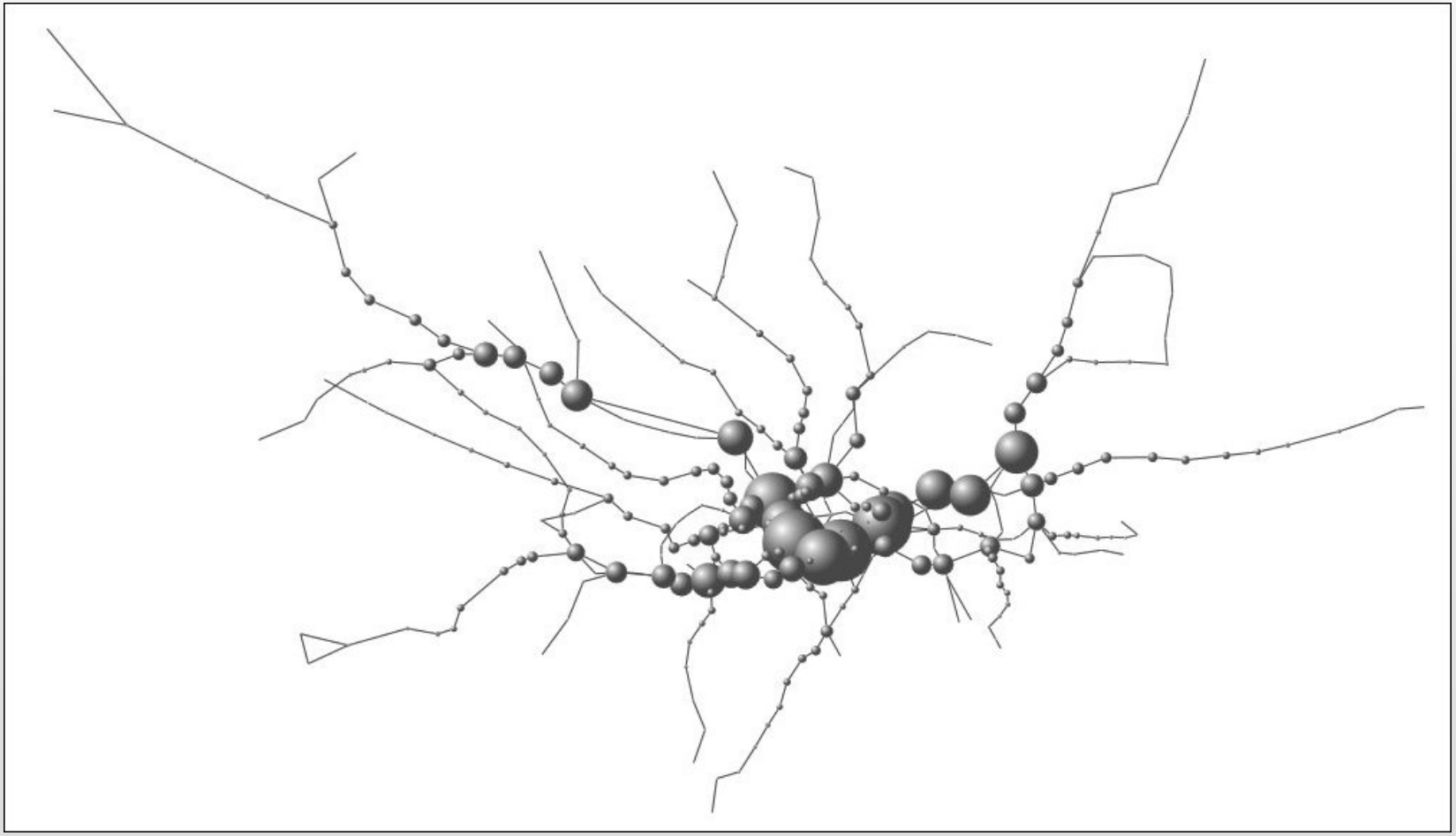

Key to Lines
Bakerloo
Central
Circle
District
East London
Hammersmith & City
Jubilee
Metropolitan
Northern
Piccadilly
Victoria
Waterloo & City
Docklands Light Railway



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

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

$$\left.\begin{aligned} T_i &= \sum_j T_{ij} \\ T_j &= \sum_i T_{ij} \end{aligned}\right\} \quad T = \sum_i T_i = \sum_j T_j = \sum_i \sum_j T_{ij}$$

Changes in
Trip Volumes

$$\left.\begin{aligned} \Delta_i &= T_i - T_i' \\ \Delta_j &= T_j - T_j' \end{aligned}\right\} \quad \sum_i \Delta_i = \sum_i \Delta_j = 0$$

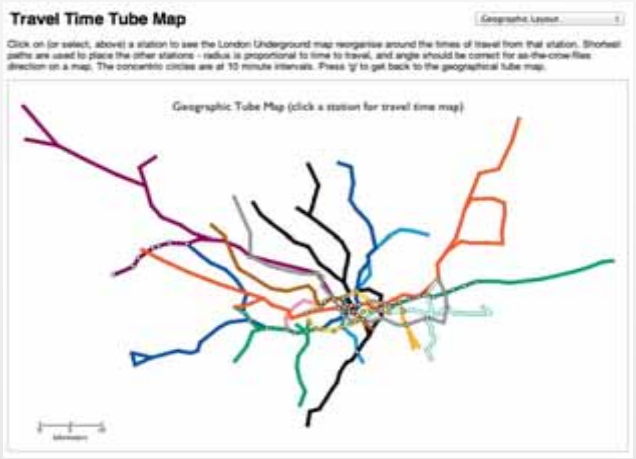
Weighted Betweenness
Centrality

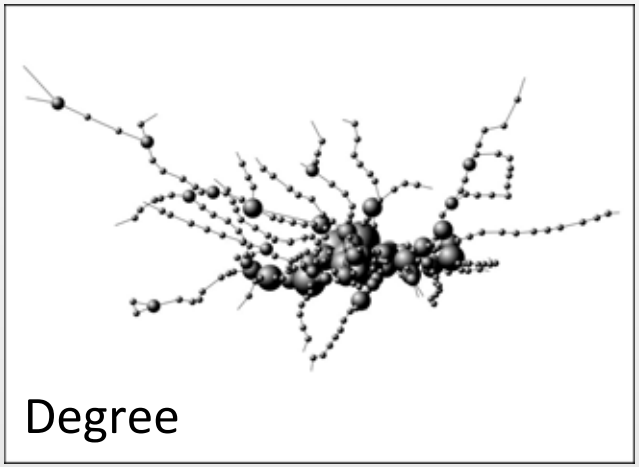
$$p_{ijk} = \frac{\sigma_{ikj}}{\sigma_{ij}} = \frac{\sigma_{ikj}}{\sum_\ell \sigma_{i\ell j}} \quad , \quad \sum_k p_{ikj} = 1$$

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

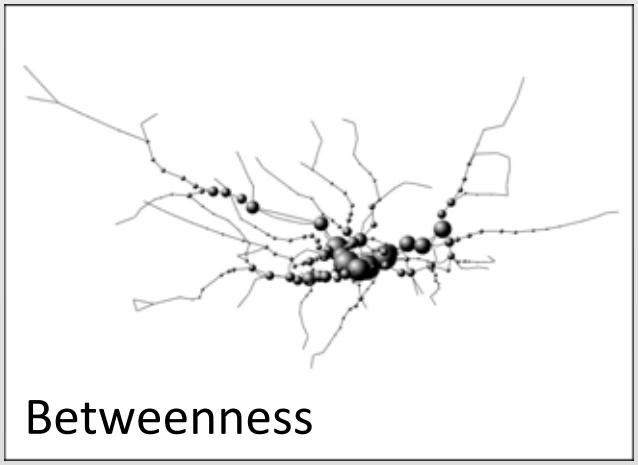
A Preliminary Analysis (1)

The Minimal Tube Network and the Three Centrality Indices

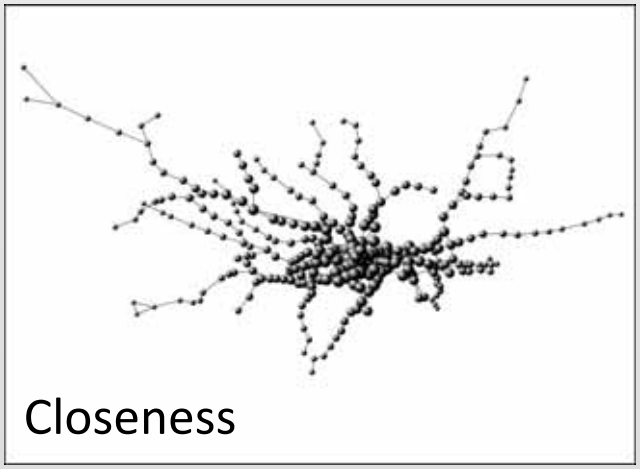




Degree



Betweenness



Closeness

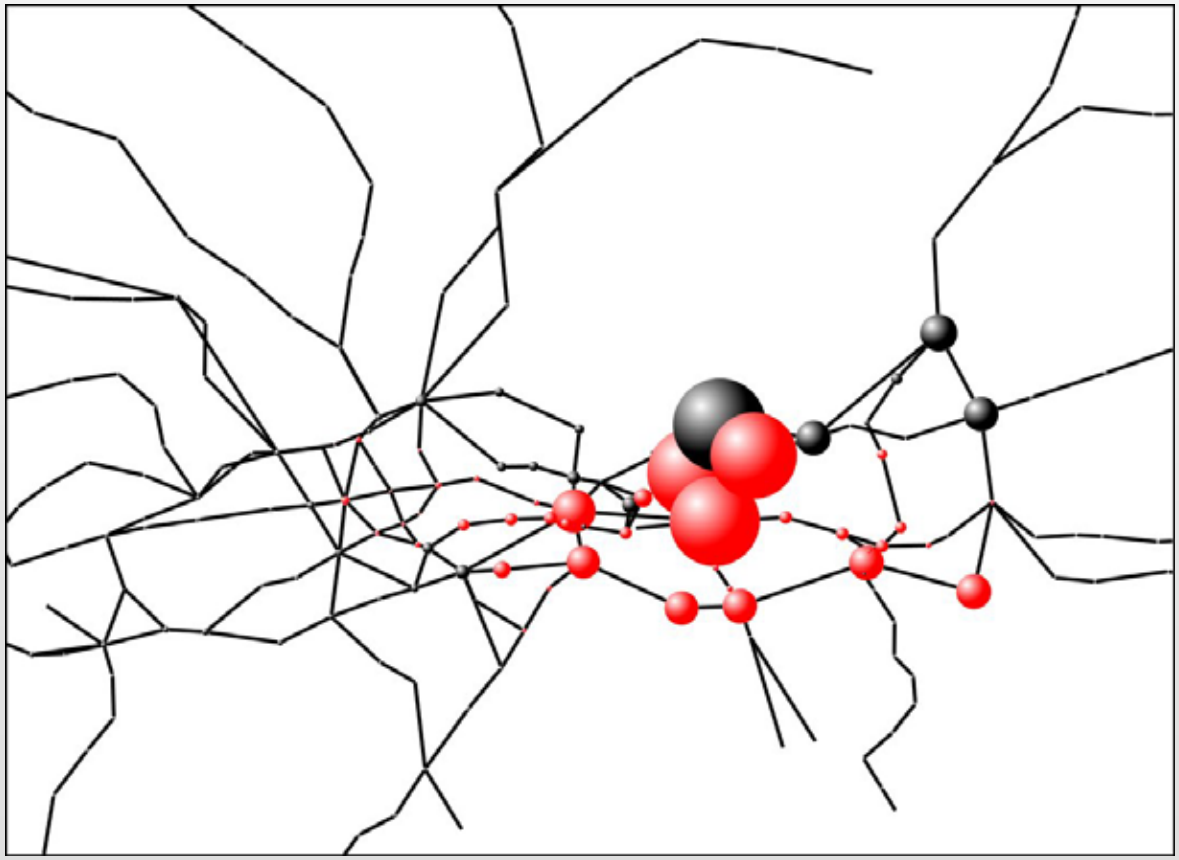
A Preliminary Analysis (2)

- 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

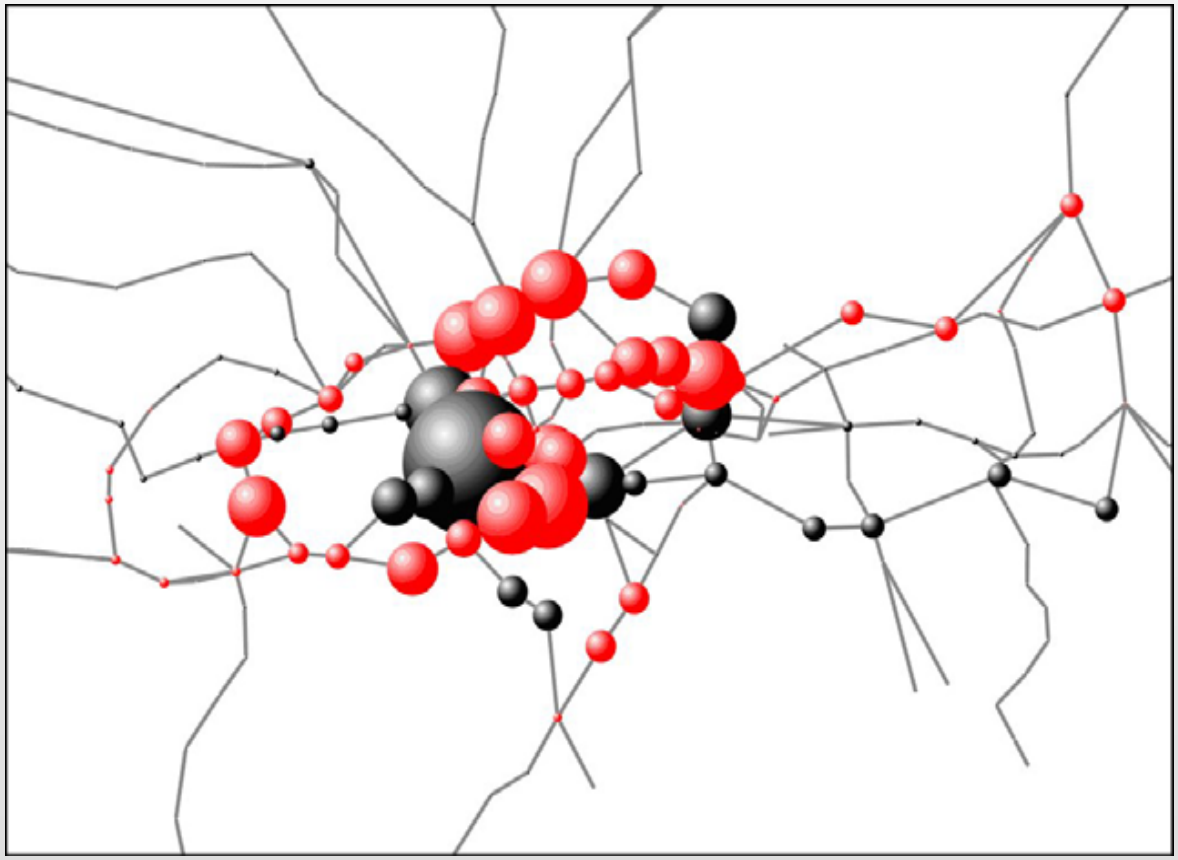
A Preliminary Analysis (3)

Closing Liverpool Street

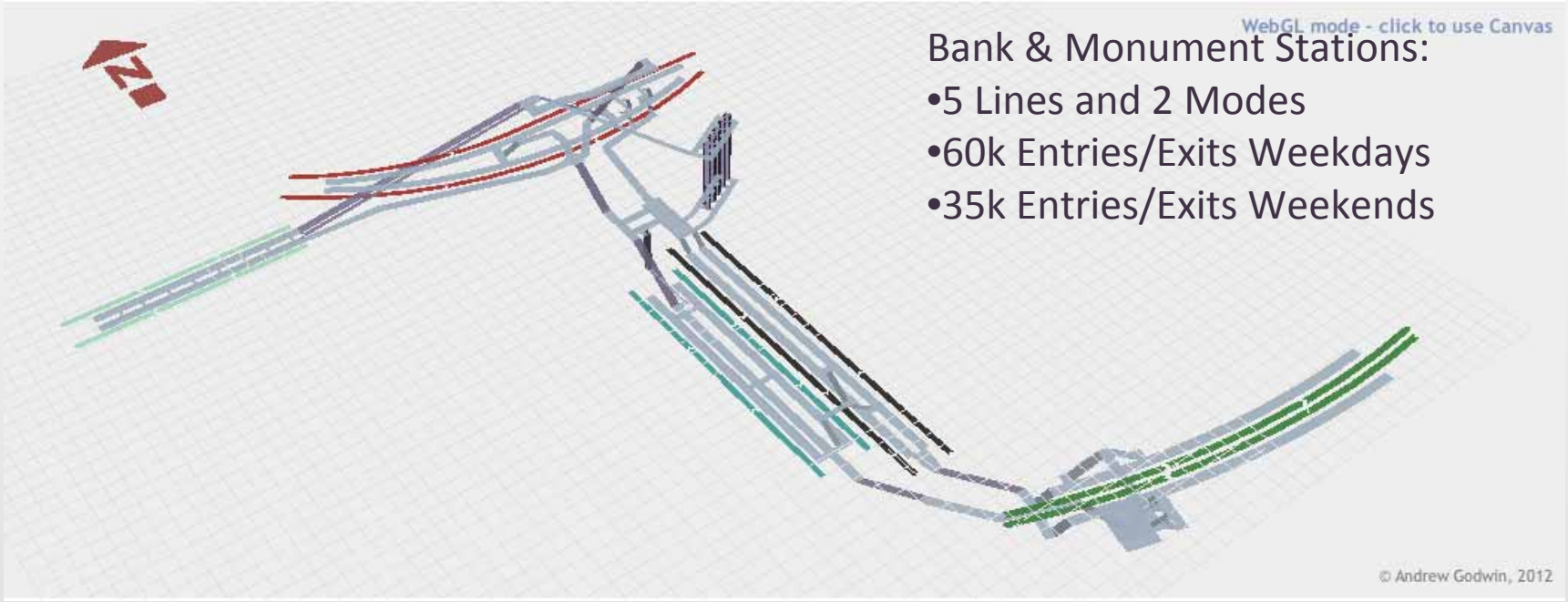


A Preliminary Analysis (3)

Closing Green Park



Looking at Station Closures and Shifts of Travellers Using The 'Shortest'-Path s

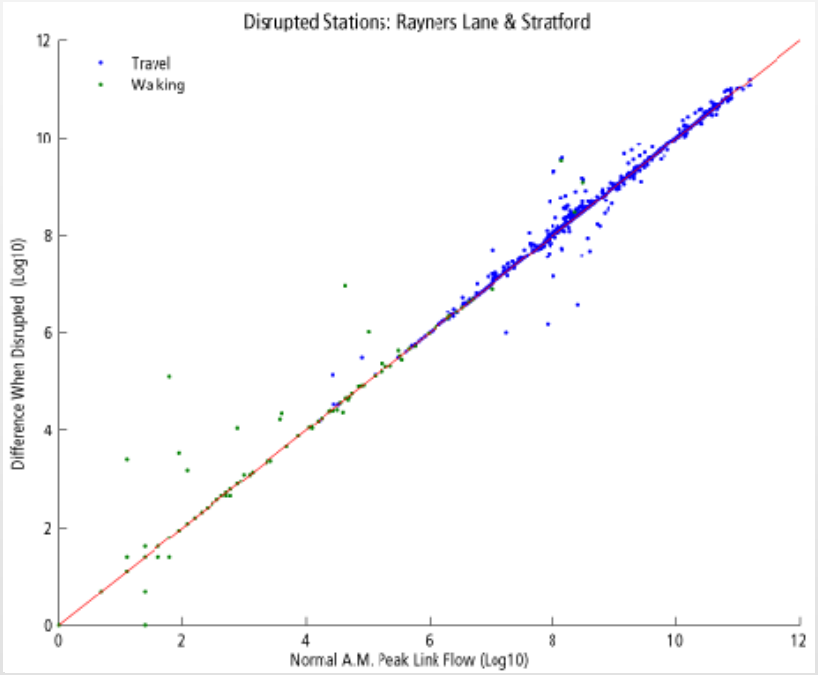


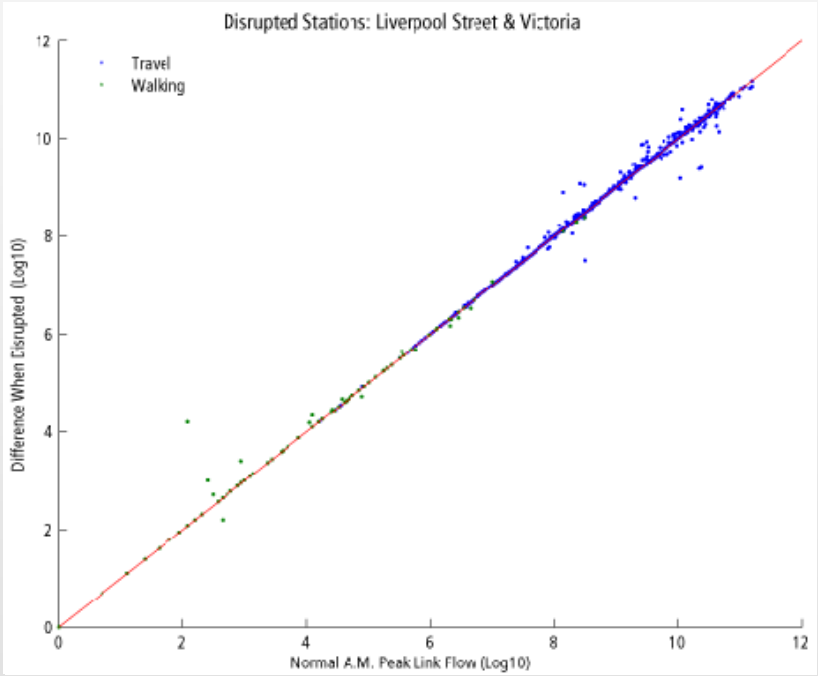
Although a simple station/line network may be sufficient for small cities, for 'Mega-Cities' such as London, New York, or Tokyo a much more detailed network is needed with interchanges measured down to the platform level. The 'penalties' for changing lines (and permitted Out-of-Station Interchanges) can be severe and should be included in a schematic network representation.

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

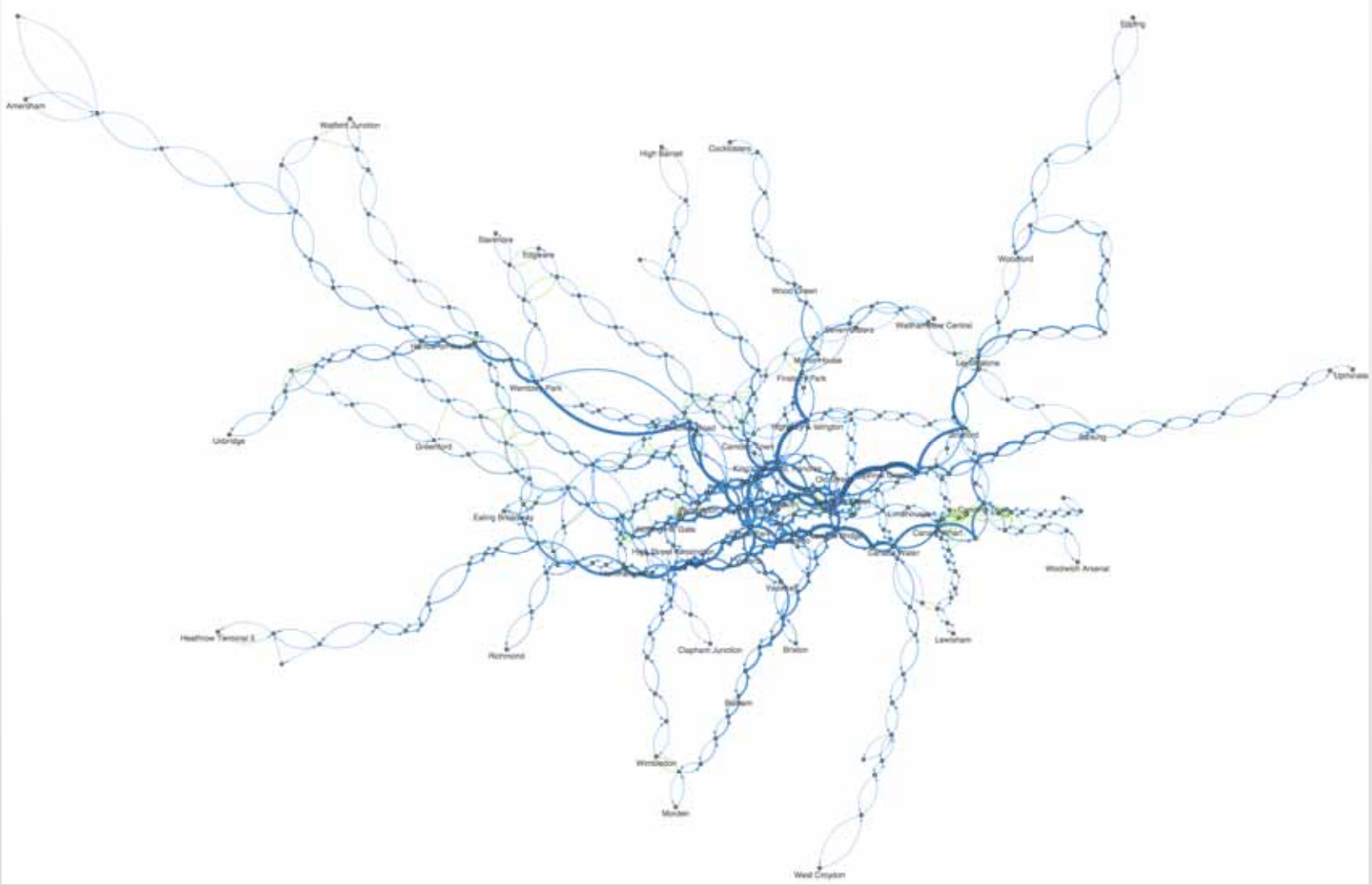
Link-Level Disruption



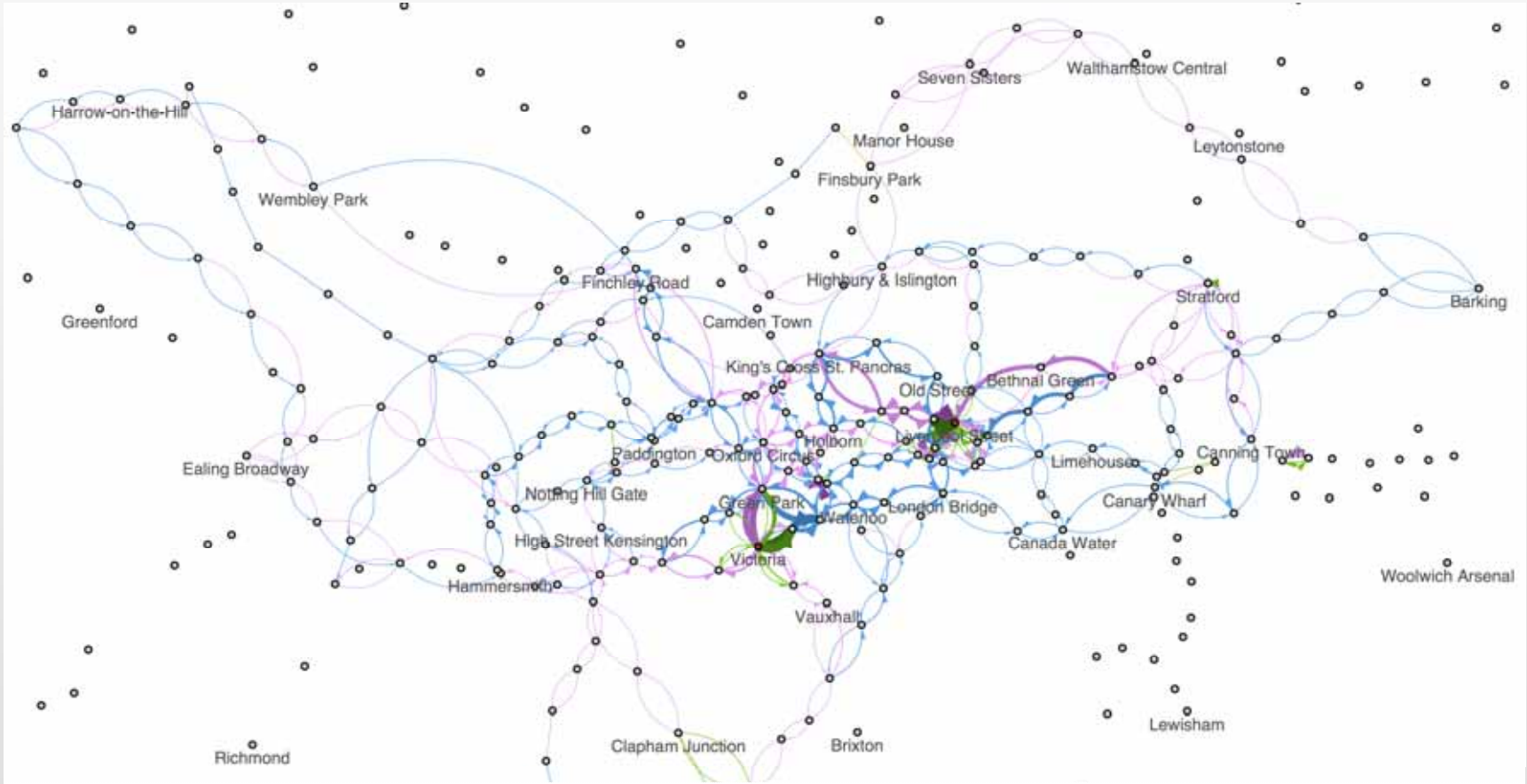


Single- and dual-station disruptions produce unexpected link-level interactions: changes in shortest-path typically cause some links to lose passengers, and gains are often less than expected. Moreover, it is not the biggest and most central stations that cause the largest shifts!

The Undisturbed Network

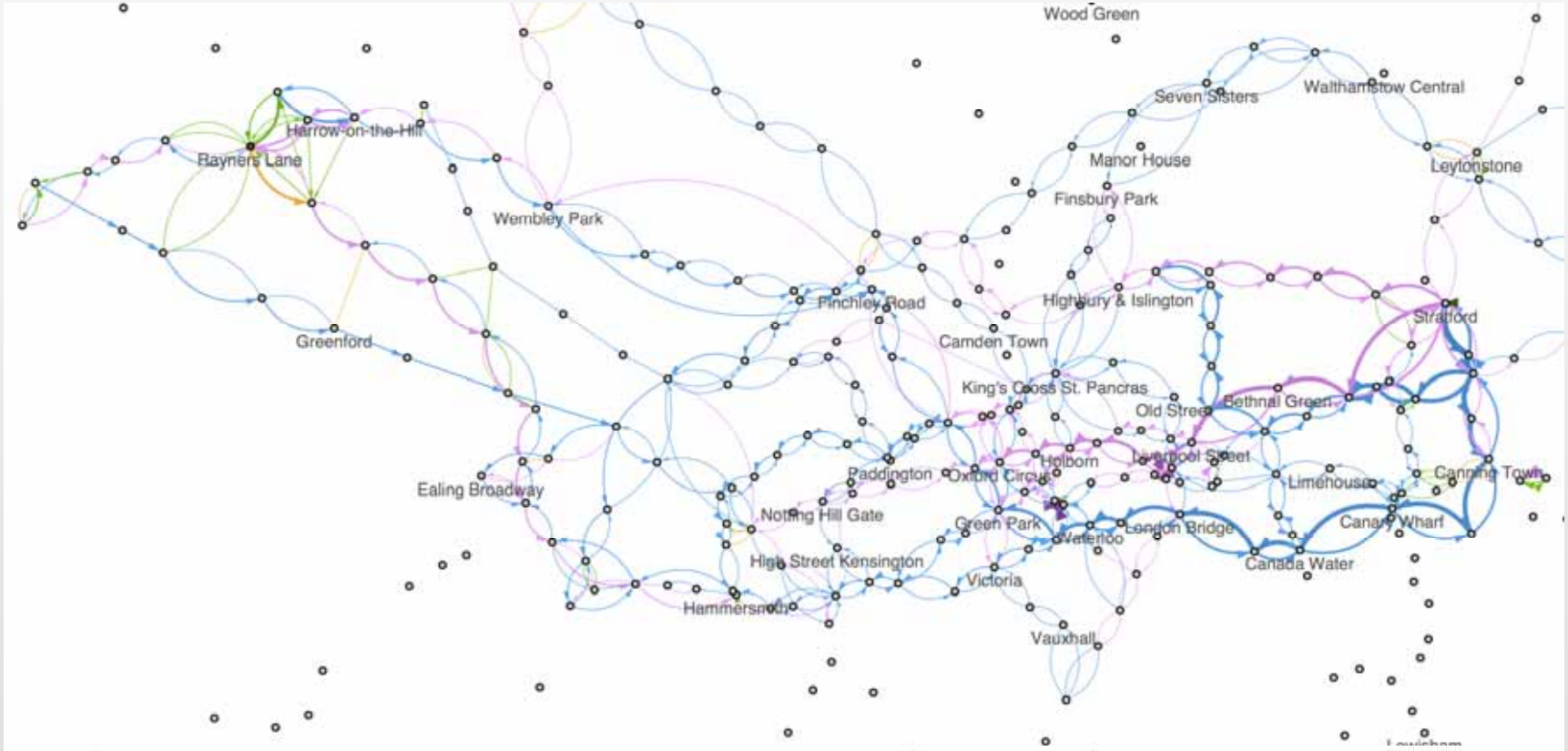


Liverpool Street & Victoria



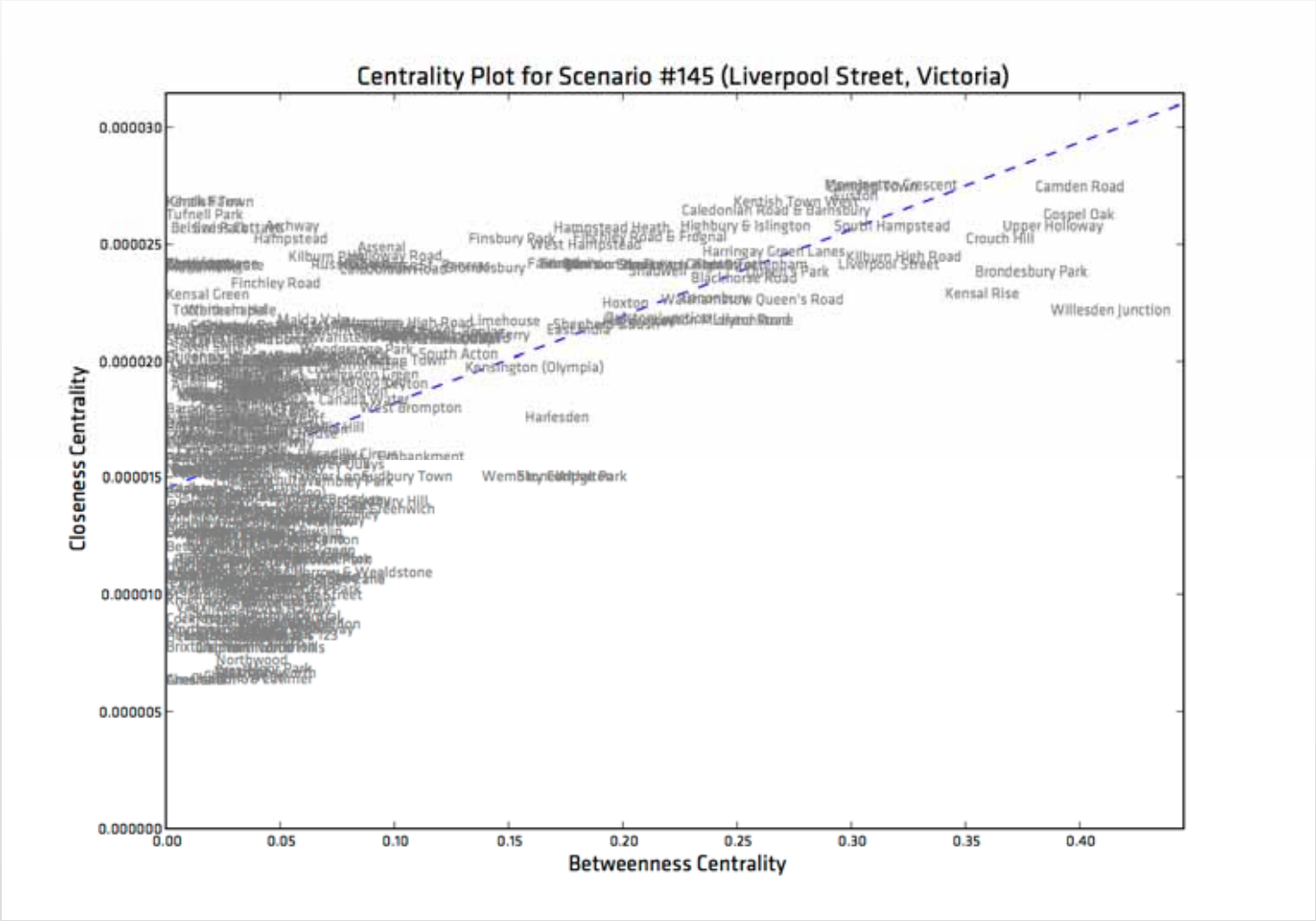
Two of London's busiest stations – because of connections to mainline rail – but if disruptions are localised to the Tube *alone* then there are many more local substitutes.

Rayners Lane & Stratford

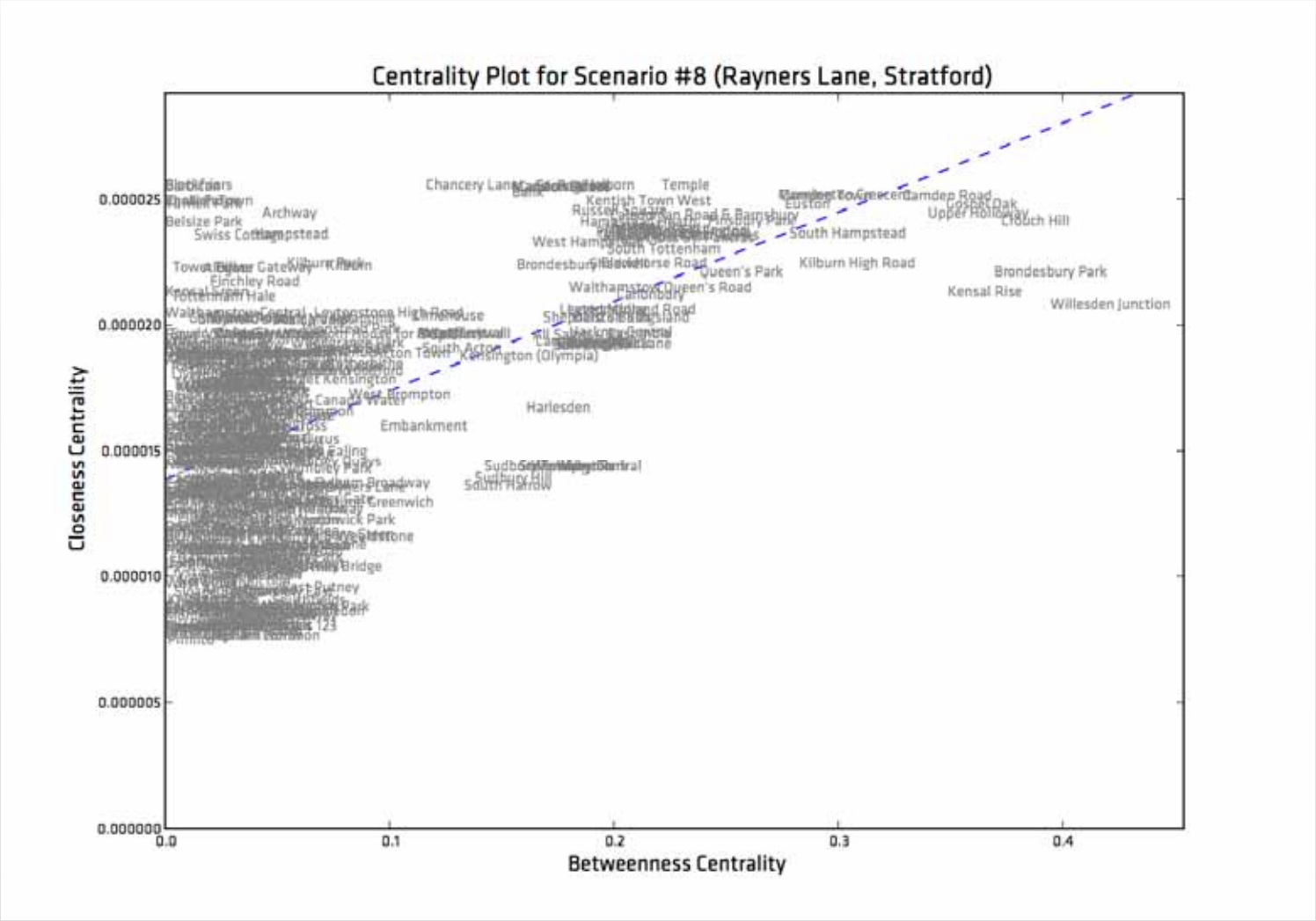


Secondary interchanges outside the core seem to cause greater disruption. Major re-routing required to complete journey, and time lost to walking long distances or travelling via more circuitous routes is much greater.

Betweenness & Closeness: Liverpool Street & Victoria

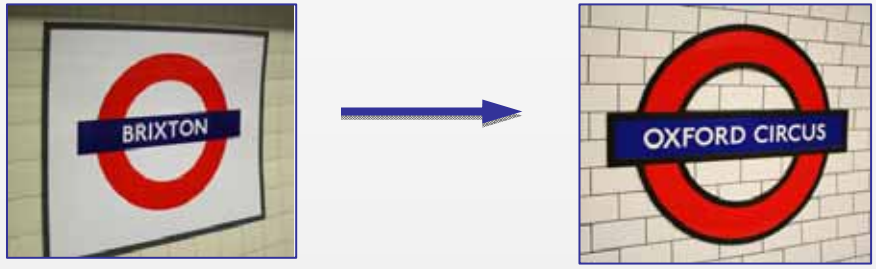


Betweenness & Closeness: Rayners Lane & Stratford

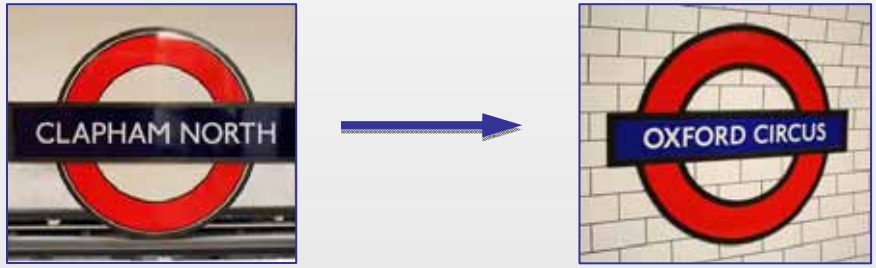


Disruption: *Examining Traveller Disruption: Stalled Trains*

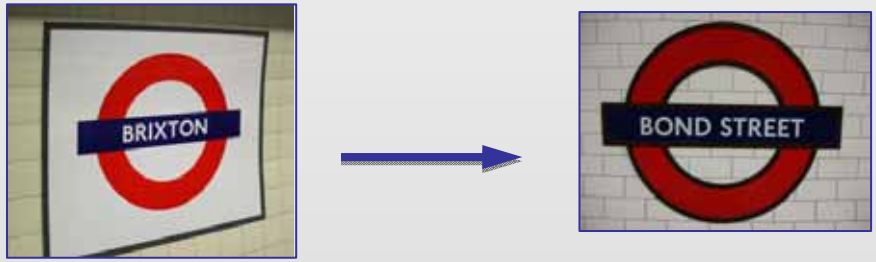
No Change



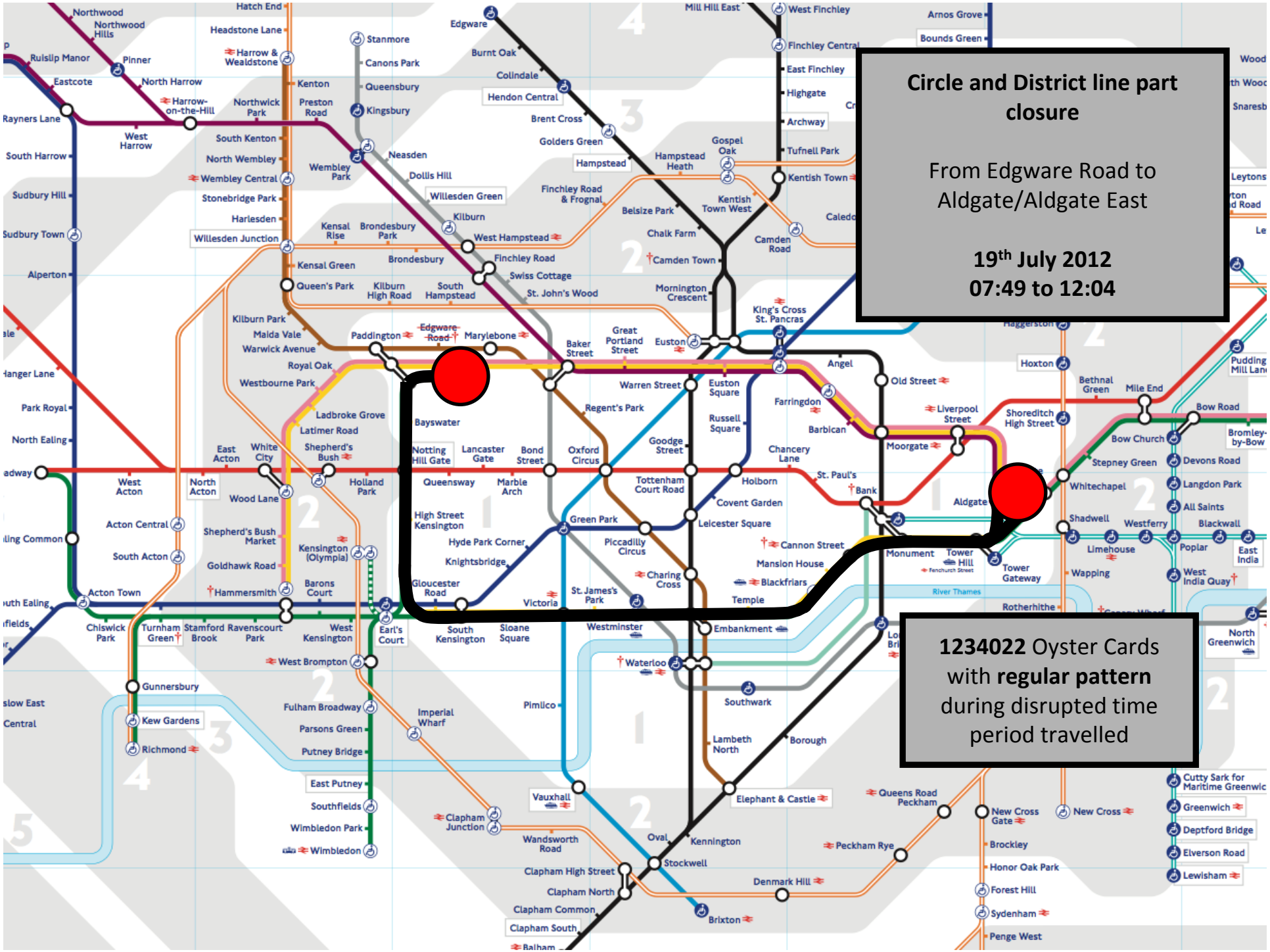
Change Origin

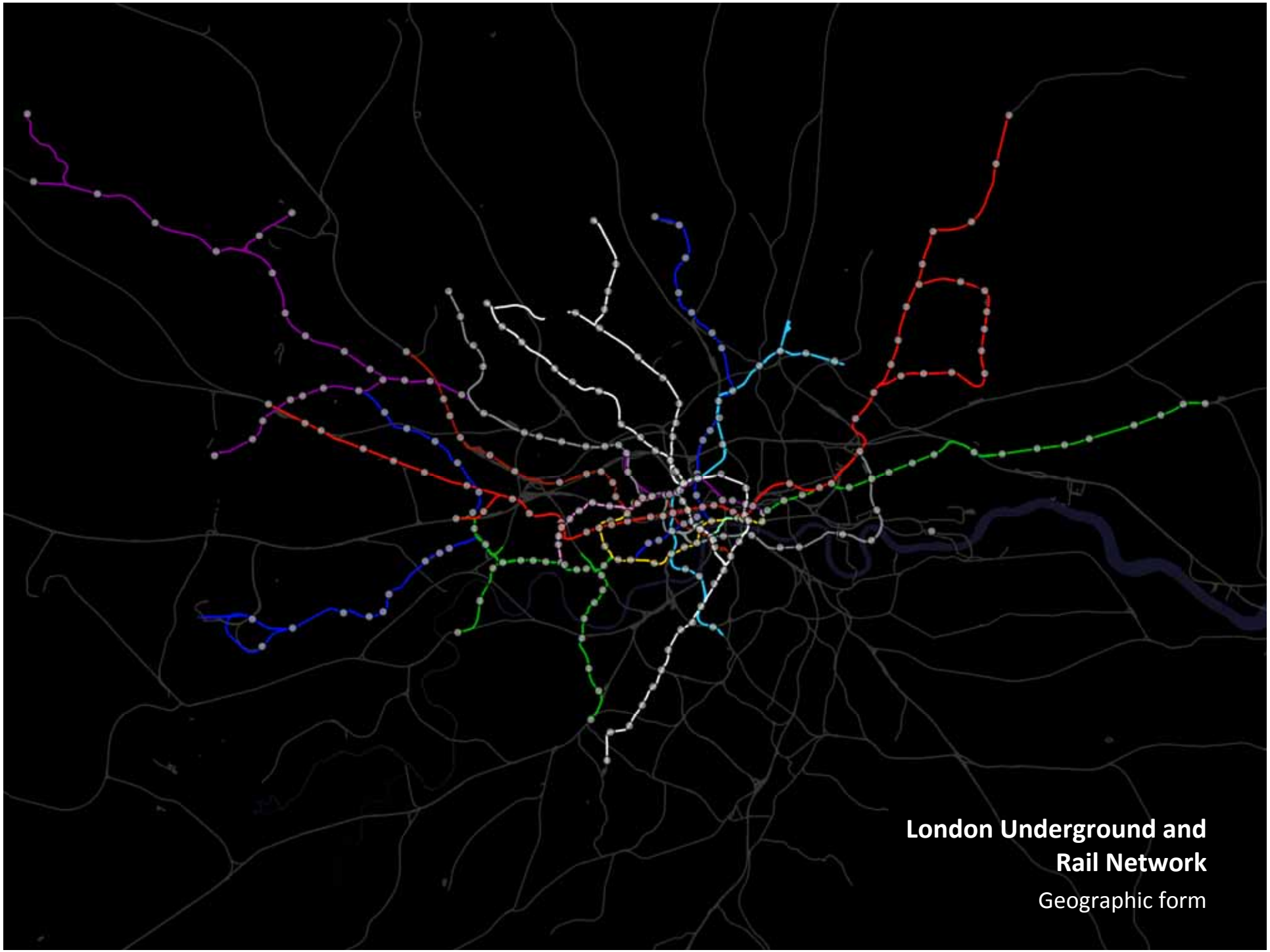


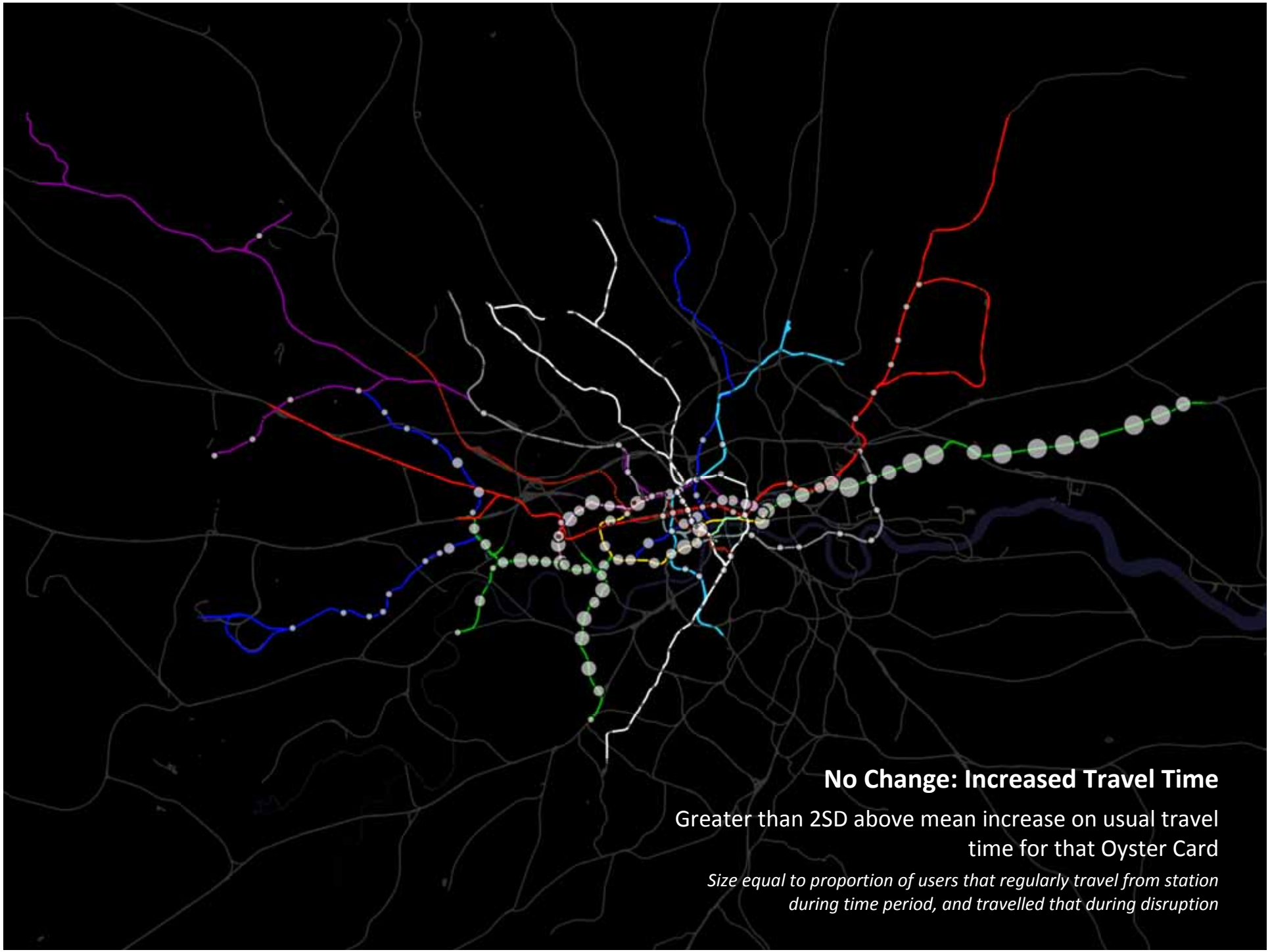
Change Destination

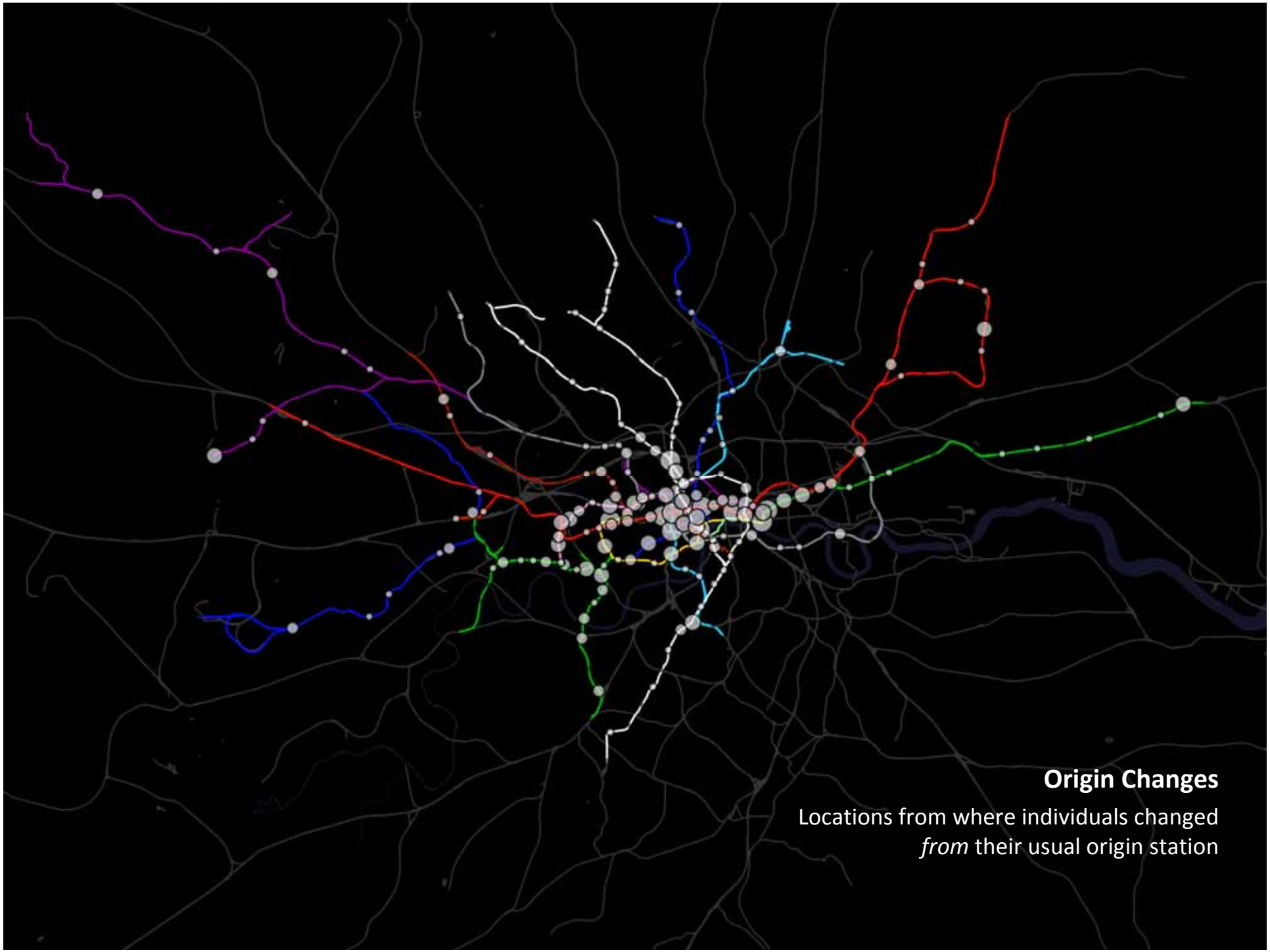


Change Mode


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

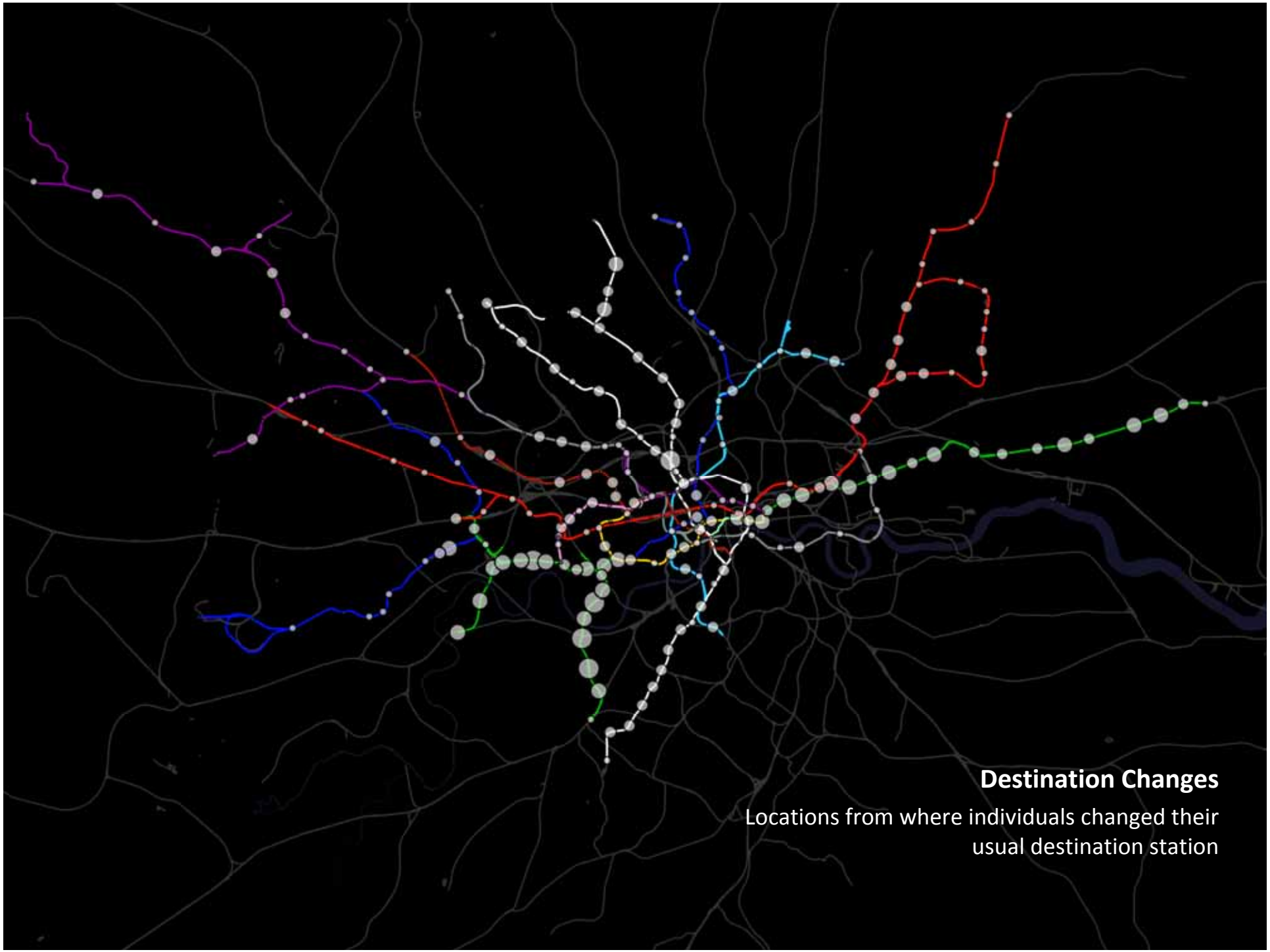

London Underground and
Rail Network
Geographic form


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



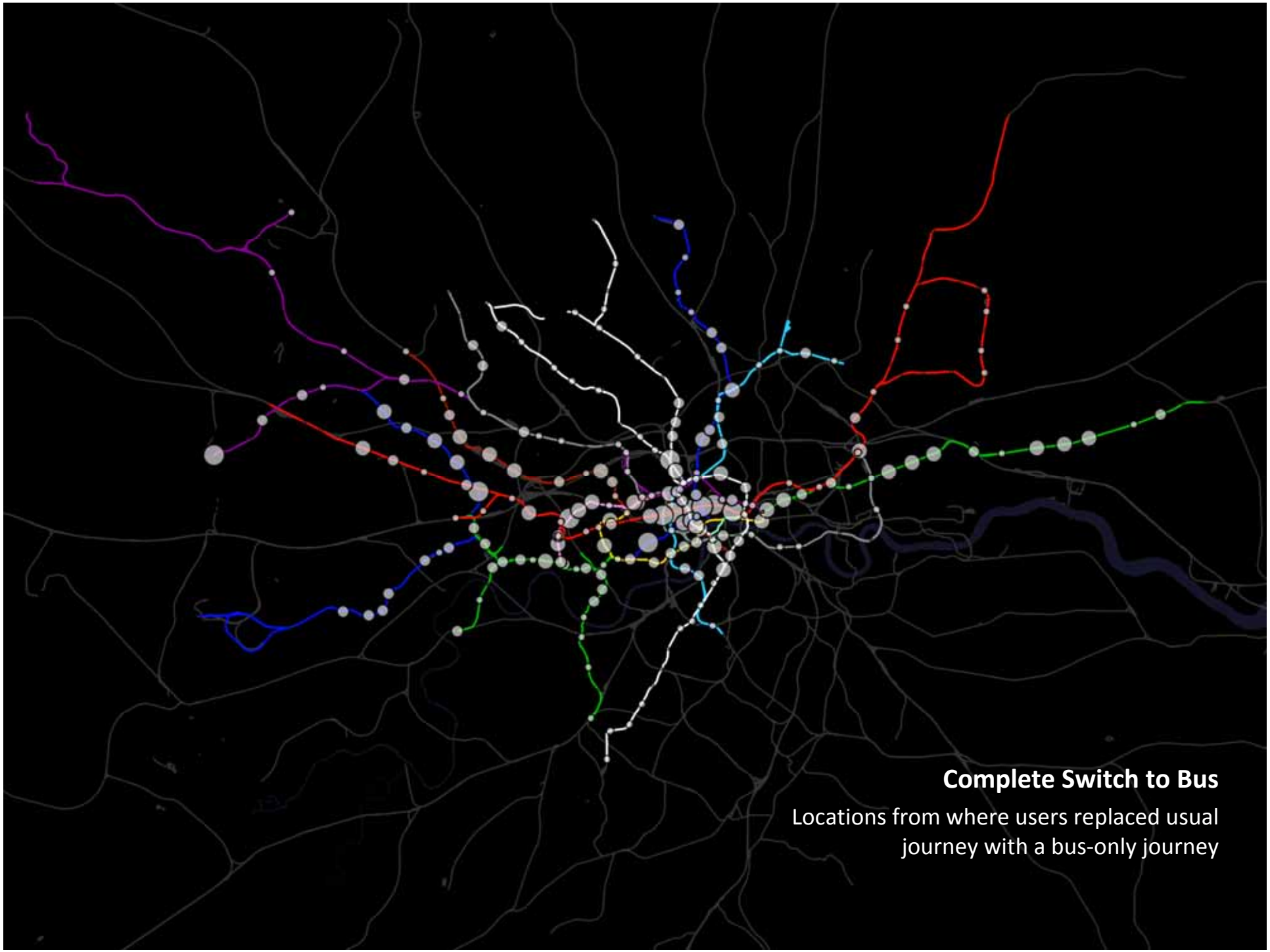
Origin Changes

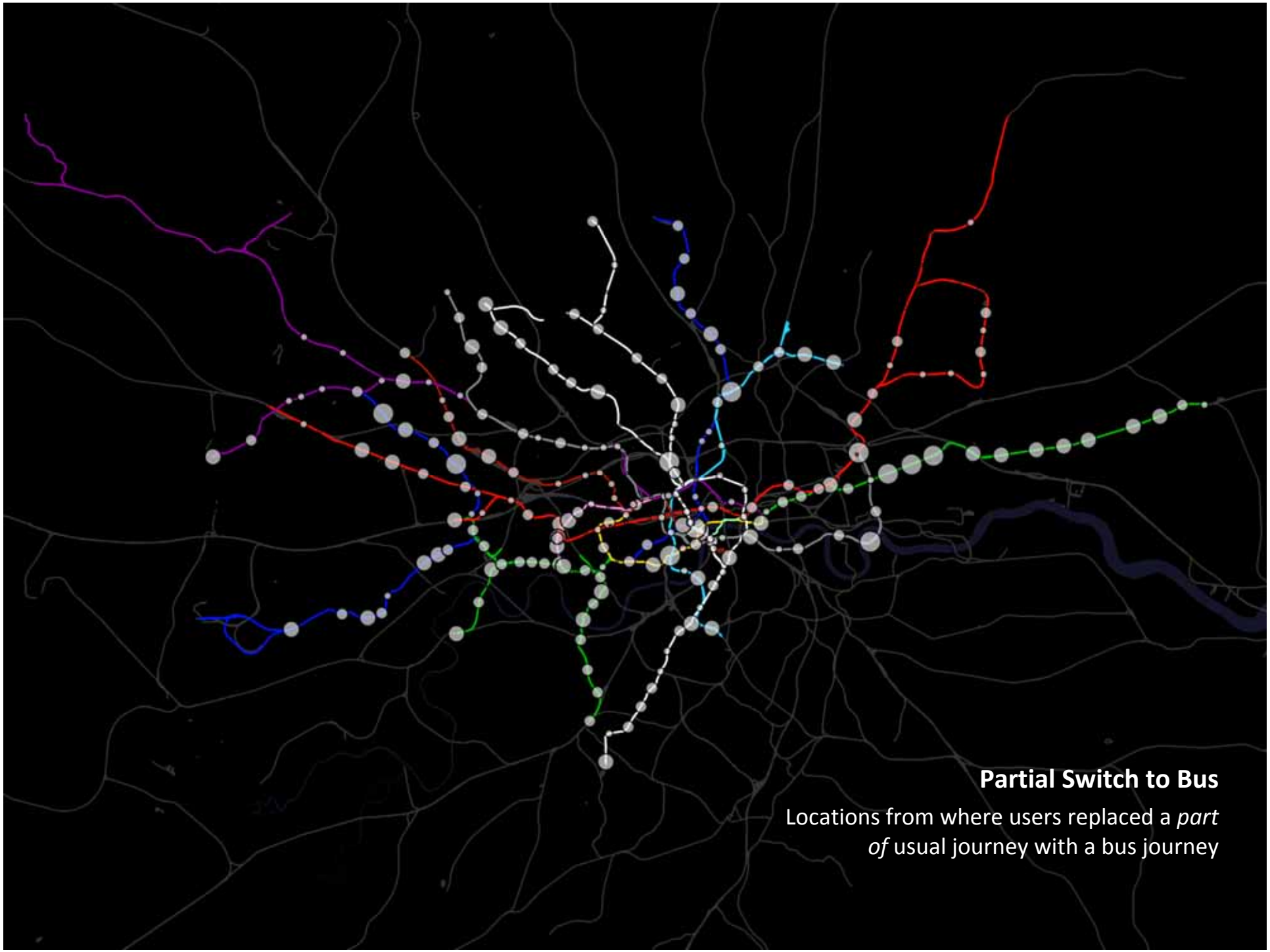
Locations from where individuals changed *from* their usual origin station



Destination Changes

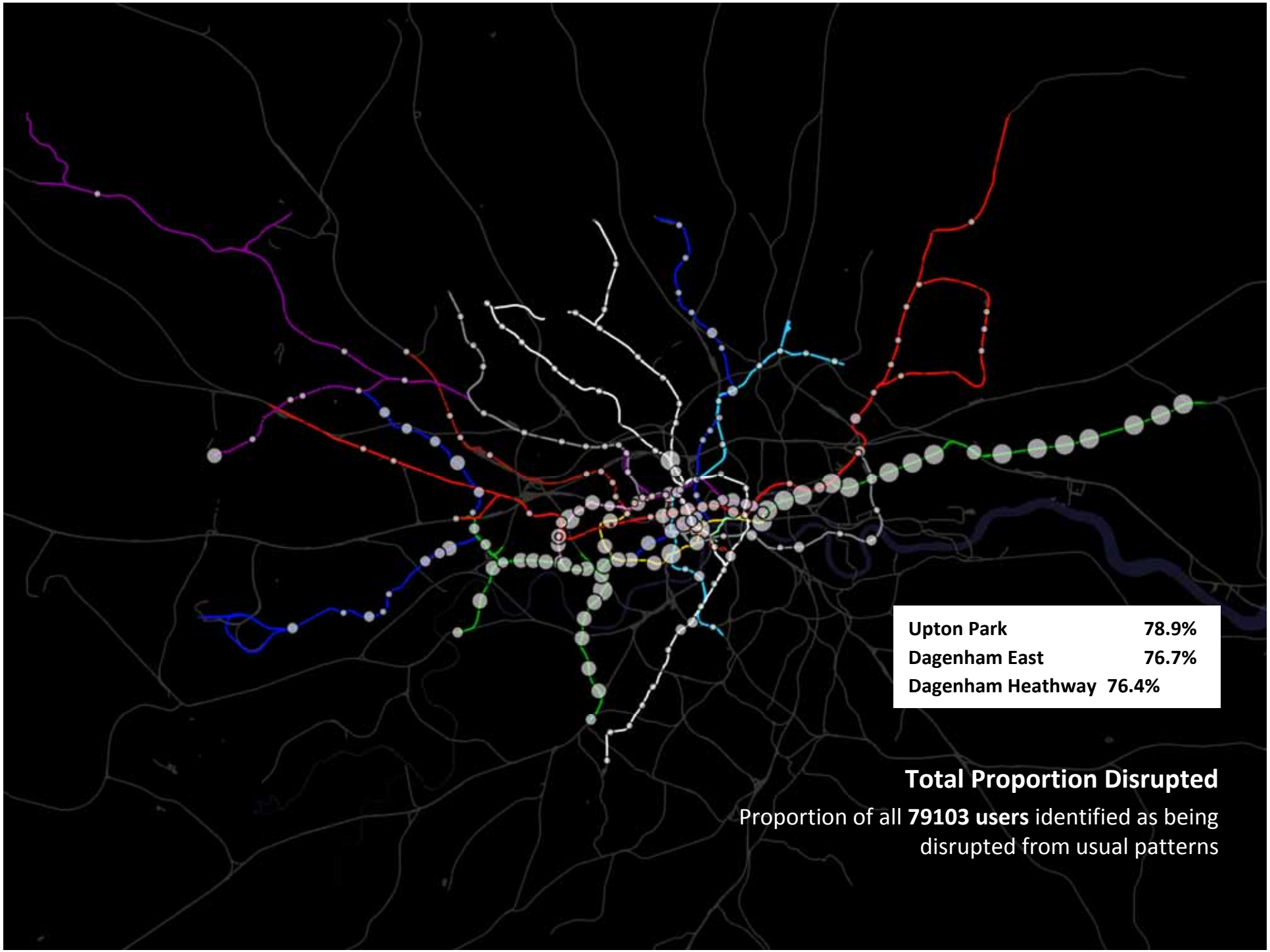
Locations from where individuals changed their usual destination station


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

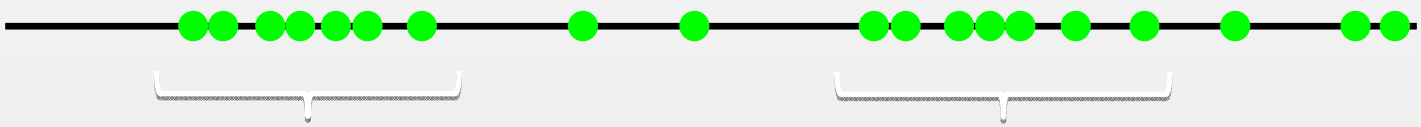
Disruption Analysis Some Key Points

- 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

Measuring Regularity

Version 2: DBSCAN Method

Oyster Card A – Origin 747



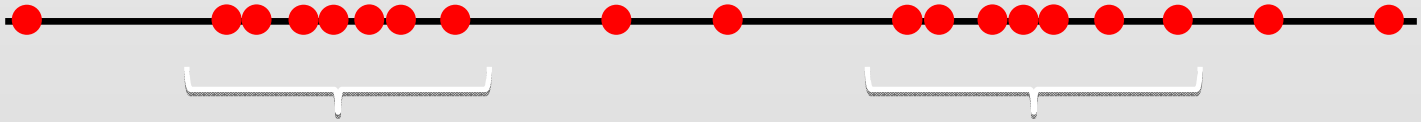
Cluster 1

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

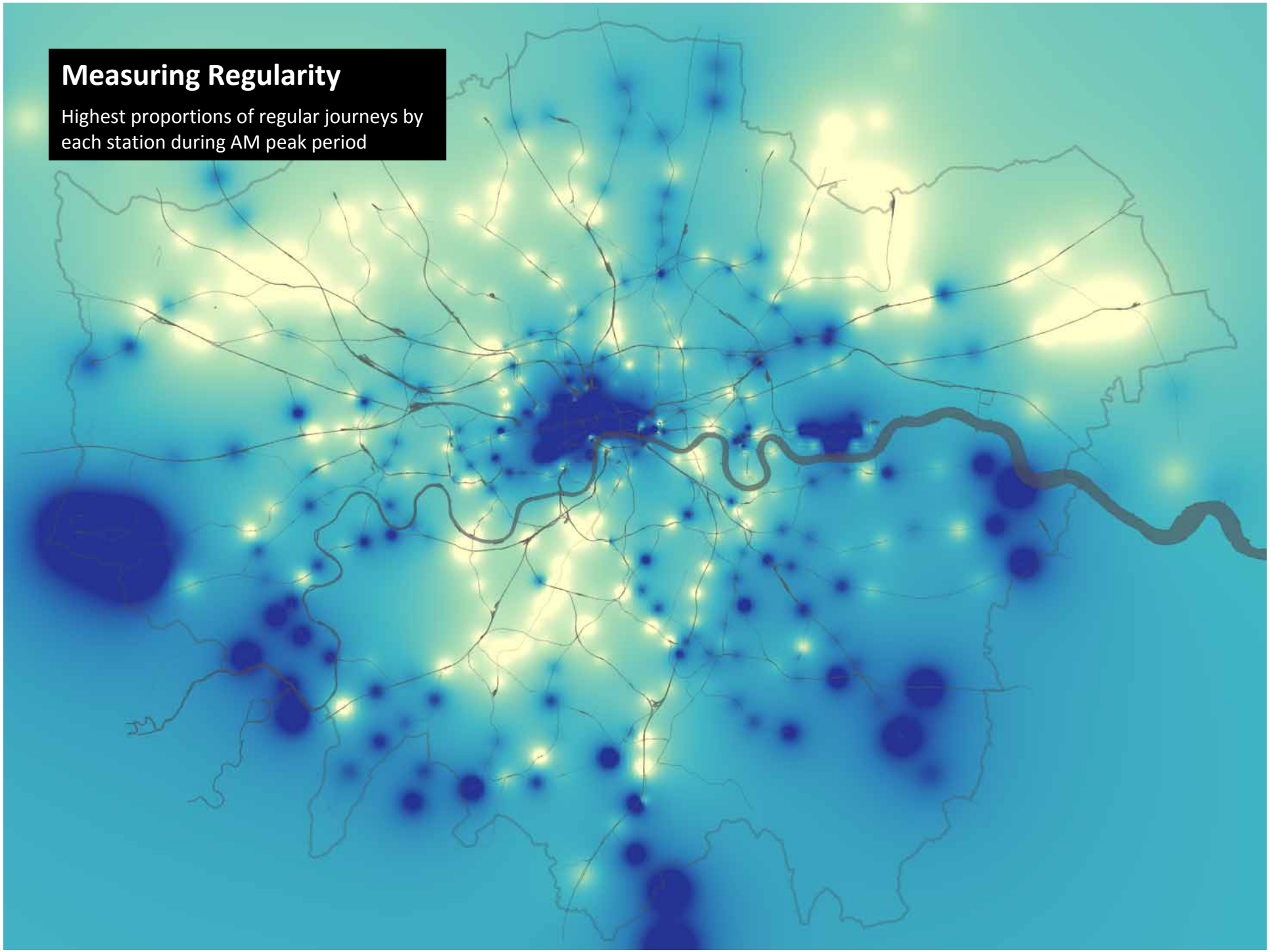


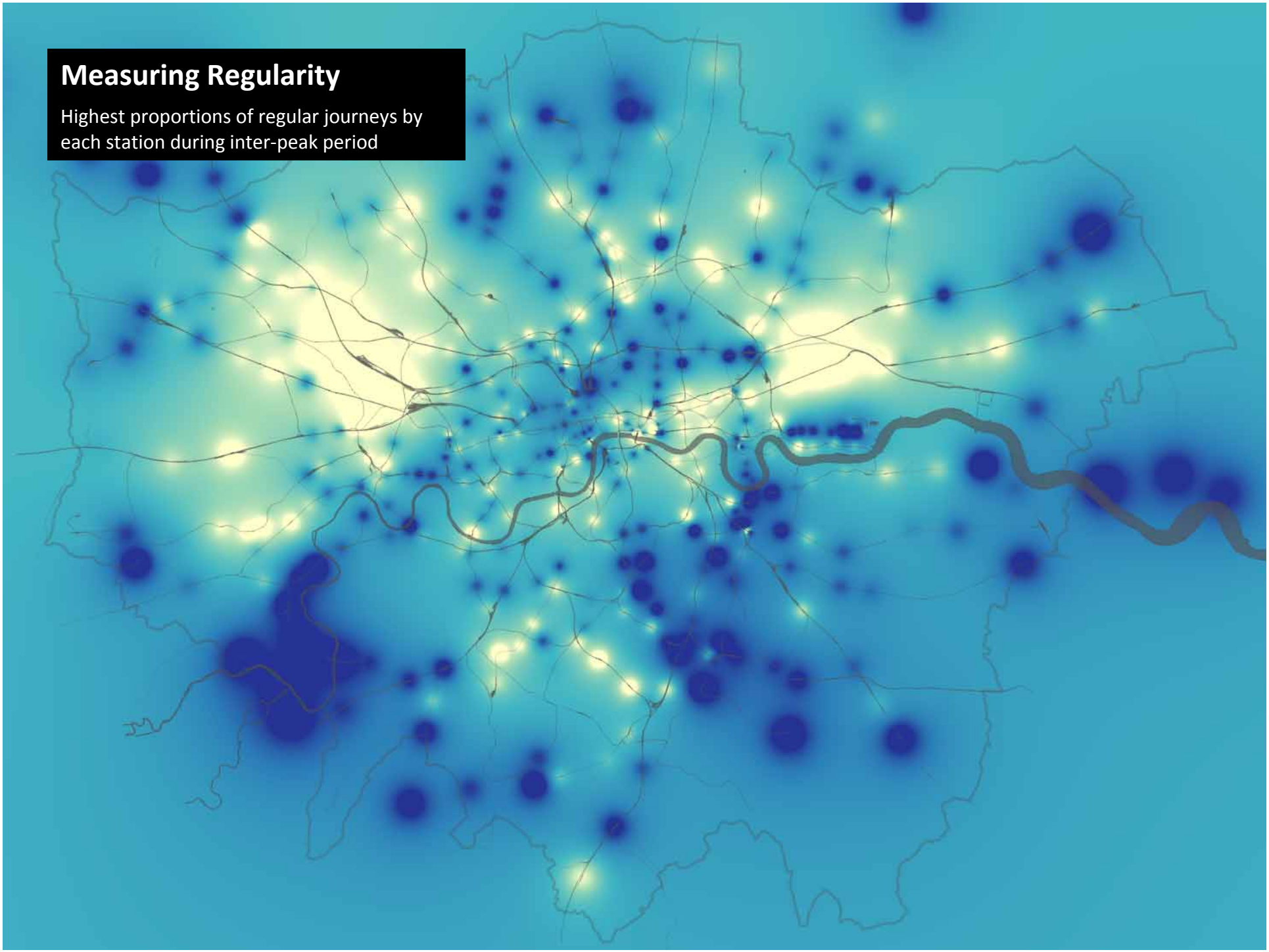
Cluster 3

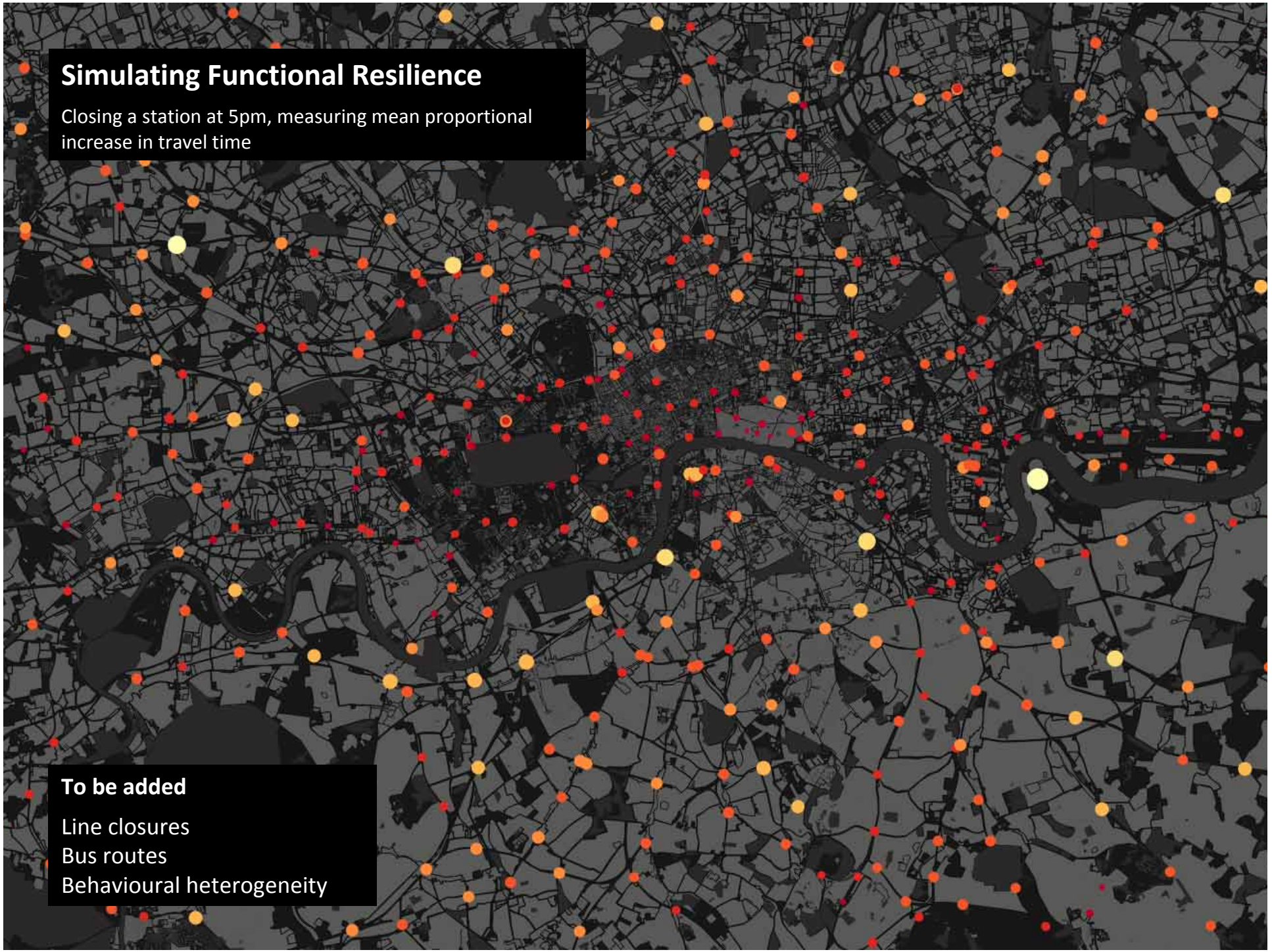
Mean: 08:37; SD = 7.4
Max: 08:13; Min: 08:48
Proportion of Days = 0.8

Cluster 4

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

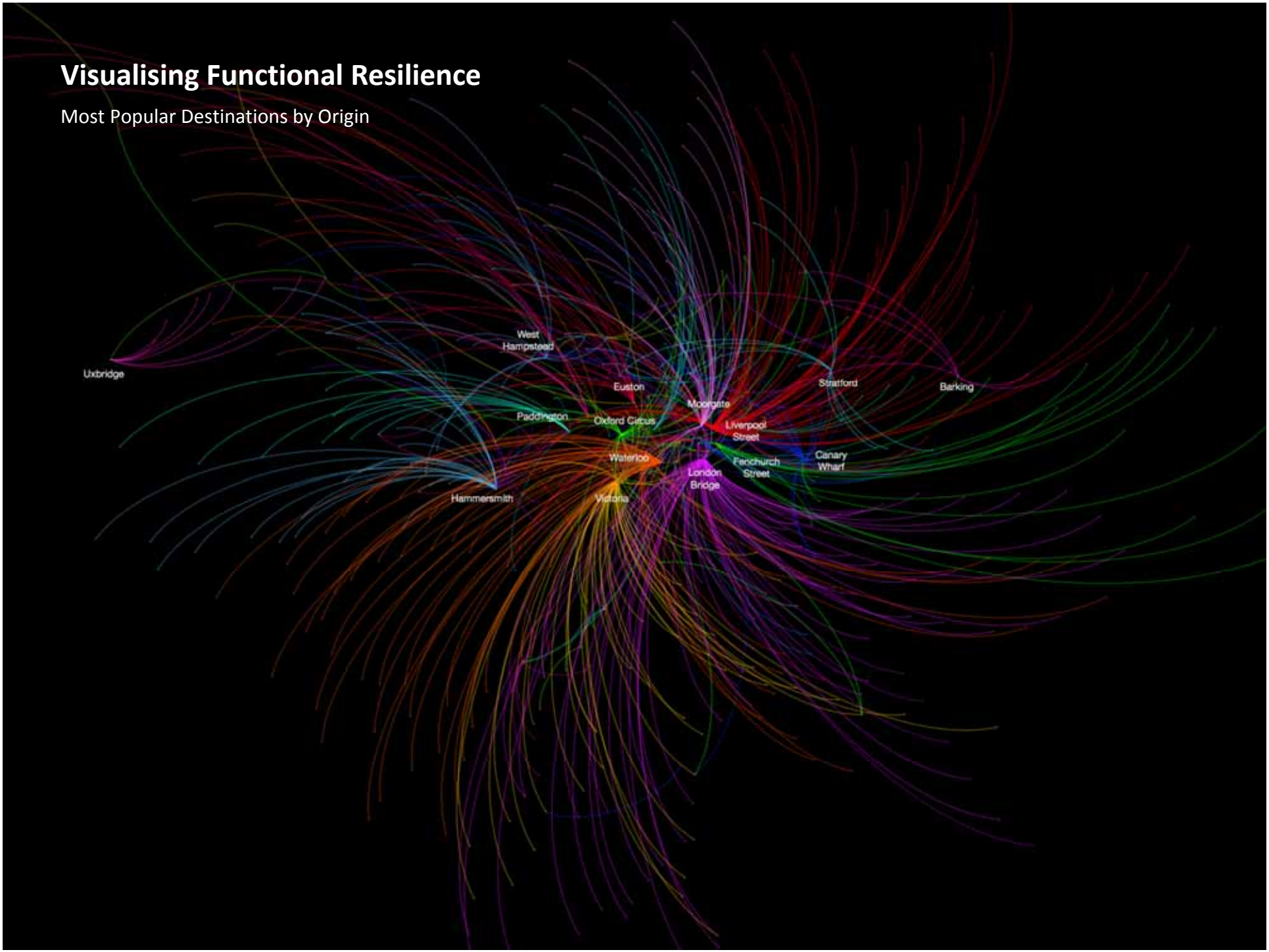

Measuring Regularity
Highest proportions of regular journeys by each station during AM peak period


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


Simulating Functional Resilience
Closing a station at 5pm, measuring mean proportional increase in travel time
To be added
Line closures
Bus routes
Behavioural heterogeneity

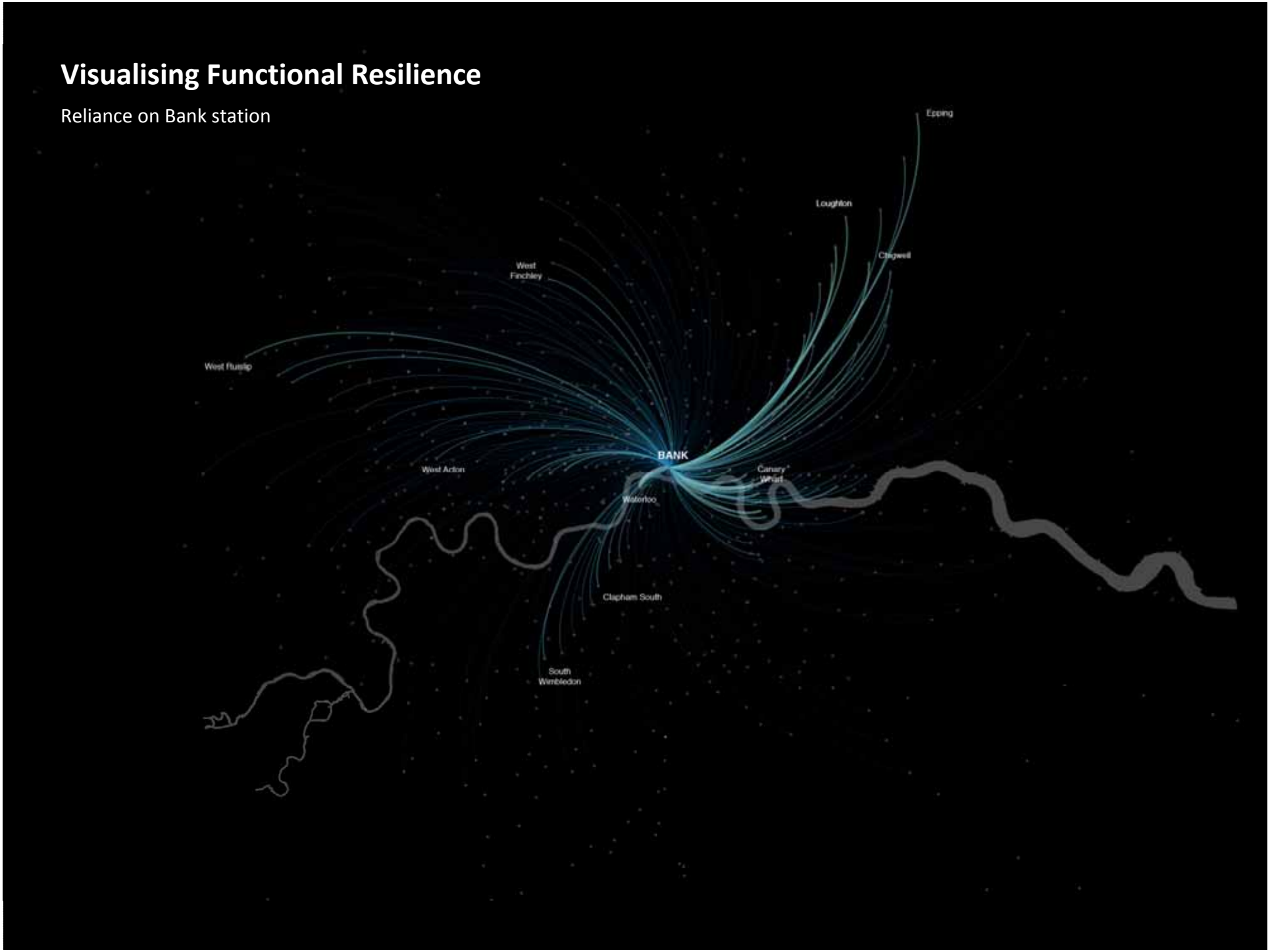
Visualising Functional Resilience

Most Popular Destinations by Origin



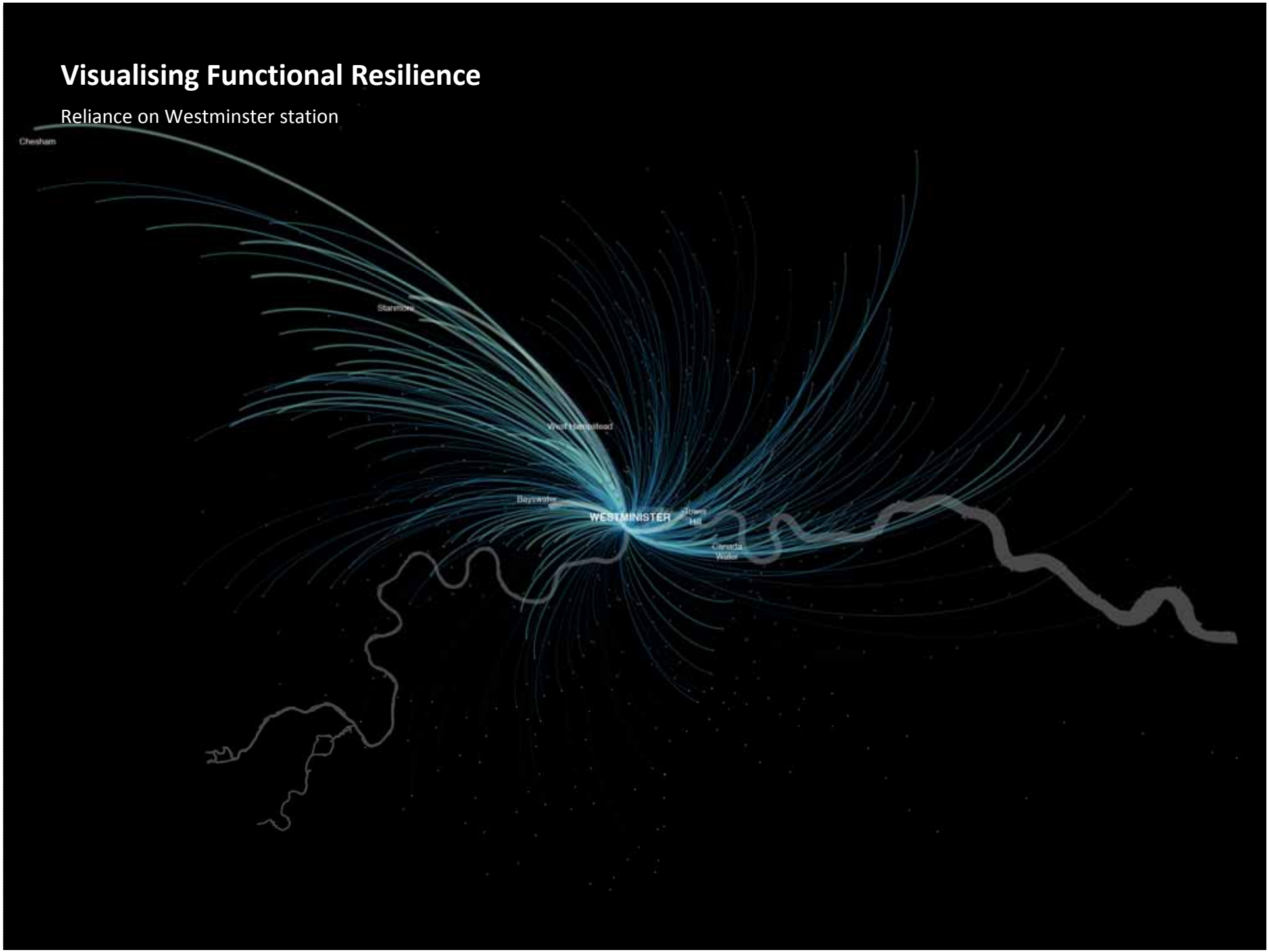
Visualising Functional Resilience

Reliance on Bank station



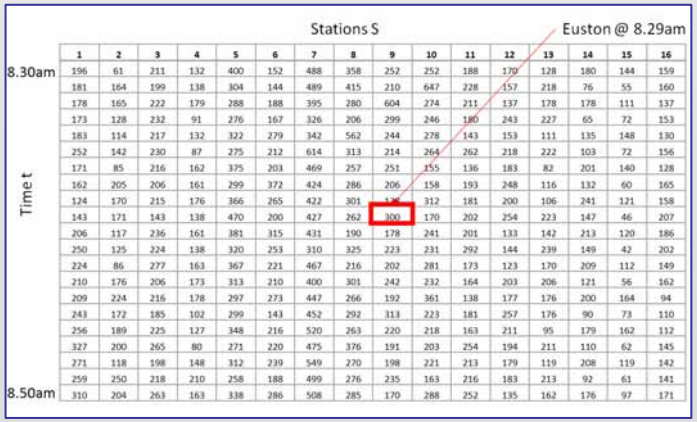
Visualising Functional Resilience

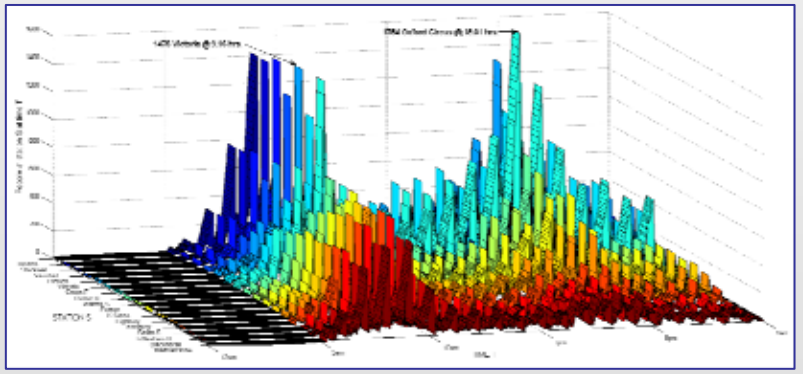
Reliance on Westminster station



Daily Polycentricity: Information Flow

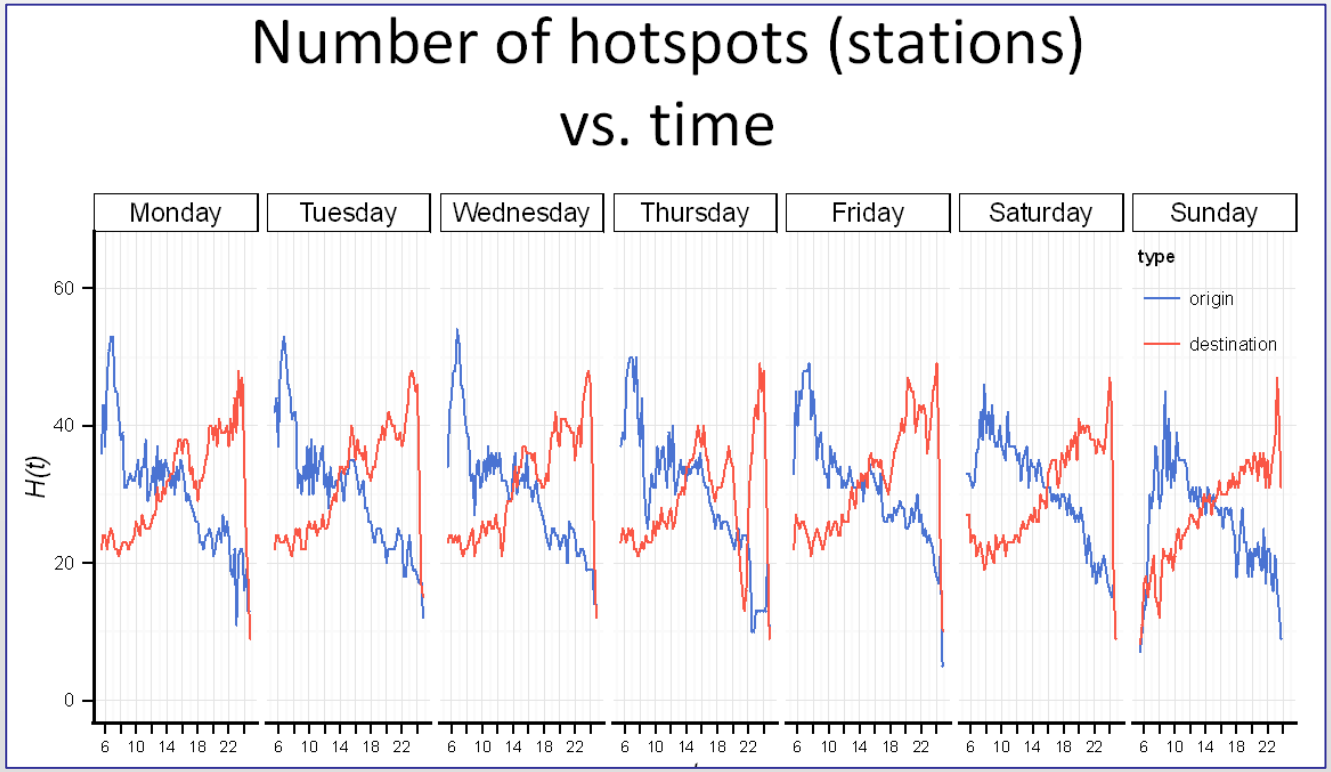
We are currently using information theory to figure out how much information from trips is transmitted from station to station through time by working out how many passengers are in stations or on trains in stations over time. We are using the concept of **transfer entropy** to do this. I don't have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

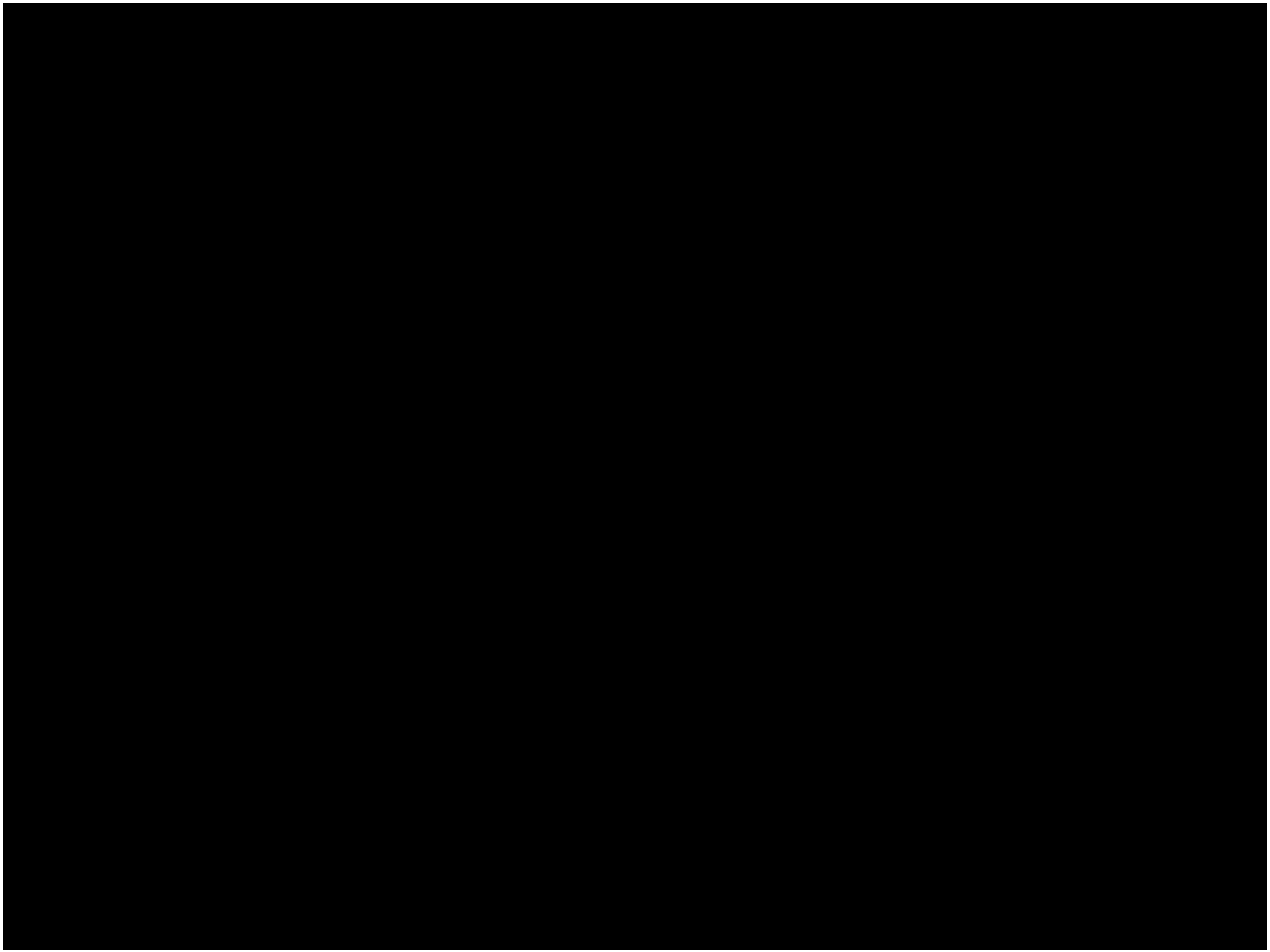




$$T_{YX} = \sum_{t=1} p(y_{t+1}, y_t, x_t) \log \frac{p(y_{t+1}|y_t, x_t)}{p(y_{t+1}|y_t)}$$

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.





And to finish this section – let me show you a movie where we combine this data with our 3D model and build a rather beautiful 3D moving environment that has all the problems of knowing how we should interpret different kinds of cycles in time and space

This is the kind of thing that is possible with this kind of big data and the current methods for visualisation – in fact in the live version you can query the movers and pull up some attribute data on them as the movie plays

Courtesy of Gareth Simons, MRes Student CASA, 2014