

Big Data & Smart Cities

*how transport systems are being automated
through real-time sensing*

Michael Batty

m.batty@ucl.ac.uk

 @jmichaelbatty

<http://www.spatialcomplexity.info/>

<http://www.complexcity.info/>

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

Outline

I have far too much to say, so a whistle stop tour

- Smart Cities: A New Paradigm? Or Just Routine Business?
- The Merging of the Routine and the Longer Term
- An Old Exemplar 1: Land Use Transportation Modelling
- An Old Exemplar 2: Moving 2D into 3D – Symbolic into Iconic
- My Main Exemplars 3: Public Transport Networks & Flows
- & Exemplar 4: Public Bike Schemes: Local Routing and Local Models of Movement
- Many More exemplars – Crowdsourcing, Social media, but Conclusions – Short Term, Real Time or Long Term, More Abstract Time

http://simulacra.blogs.casa.ucl.ac.uk/ — Simulacra » Showcasing land use transport modelling, urban complexity and sustainability r... RSS Google

My Sites Simulacra 3 + New Howdy, Michael Batty

 Search

 **SIMULACRA**
Urban Modelling at UCL's
Centre for Advanced Spatial Analysis

People Projects Archive Resources



Pulse of the City (reboot)
As I get to better grips with the full richness of the Oyster data set and the complexity of the TfL network it's gradually getting easier to build better visualisations. One of the ones that I've wanted to revisit for ...
[View full post](#)

Recent Posts

- [Europe- a millennia in ten minutes](#)
- [Pulse of the City \(reboot\)](#)
- [A Week in the Life of London's Public Transit System](#)
- [Big Data, Complexity, Networks at the German Physical Society](#)
- [Understanding and Managing Complex Systems, 5 March 2012](#)

About Simulacra
This website showcases land use transport modelling, urban complexity and sustainability research from the Centre for Advanced Spatial Analysis, University College

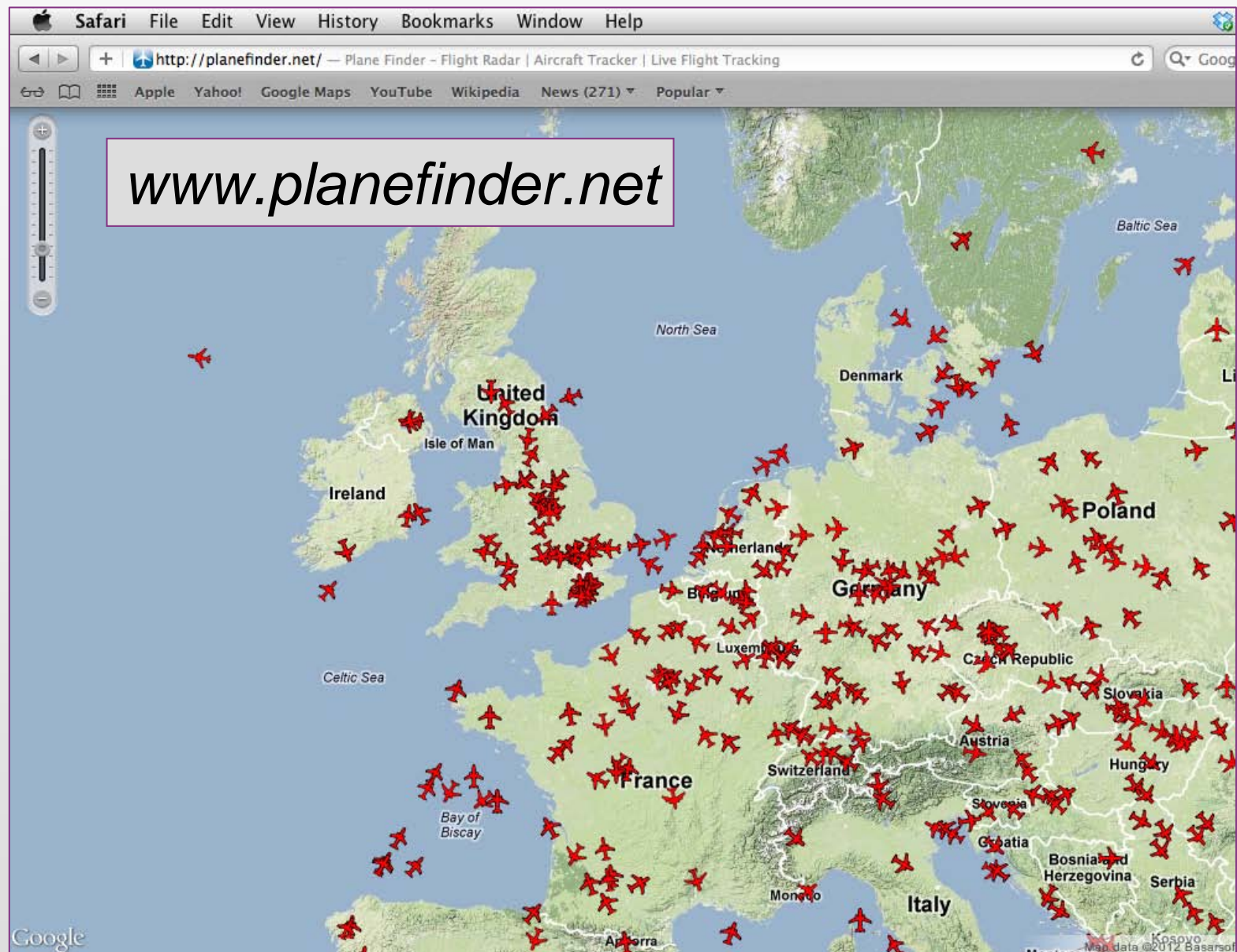
Pulse of the City (reboot)
by [Jon Reades](#) | May 8, 2012 ([Edit post](#))

<http://www.simulacra.info/>



Centre for Advanced Spatial Analysis





Smart Cities: A New Paradigm

I am going to look at several examples which various people in our centre are developing where we are able to use new online data sources – *big data* – with *new forms of model*, to make predictions about the smooth running of traffic and the economy.

At many scales: the regional-metropolitan, at area-wide network infrastructure scales, and at the local street scale

These examples show how we can fuse the strategic with the tactical and it also shows how visual media can inform simulation and vice versa.

One of our key themes is the focus now on time as well as space and on the short term rather than the long term.

First a little bit about the Smart City. It goes back a long way – well before the web – all about the wired city – installing fibre. This to some extent is our paradigm insofar as there is one



After the initial period of installing networks, the focus moved in smart cities (read intelligent cities, virtual cities, digital cities, ...) to the provision of services which still represents a main focus.

But the development of new data sets from real time or *near* real time sensing is now one of the key foci and linking our various technologies to understanding the city is providing a new momentum.

The key issues are interoperability, confidentiality and privacy, and participation using new media. There is a sub-theme involving better science, models, prediction. The evolution of the smart city idea is following the same line as computation:

First hardware, then software, then data and orgware

In fact, the smart city idea is joining with data capture, data mining and visualisation to generate a new momentum in our understanding & modelling cities

BIG DATA + SMART CITY = New Models

This is our focus really today. As a research group, we are not into applications that are robust enough to be implementable *en masse* but we have many proofs of concept, with potential for wider apps.

We see our research as informing and experimenting and exploring how we relate short term to long term

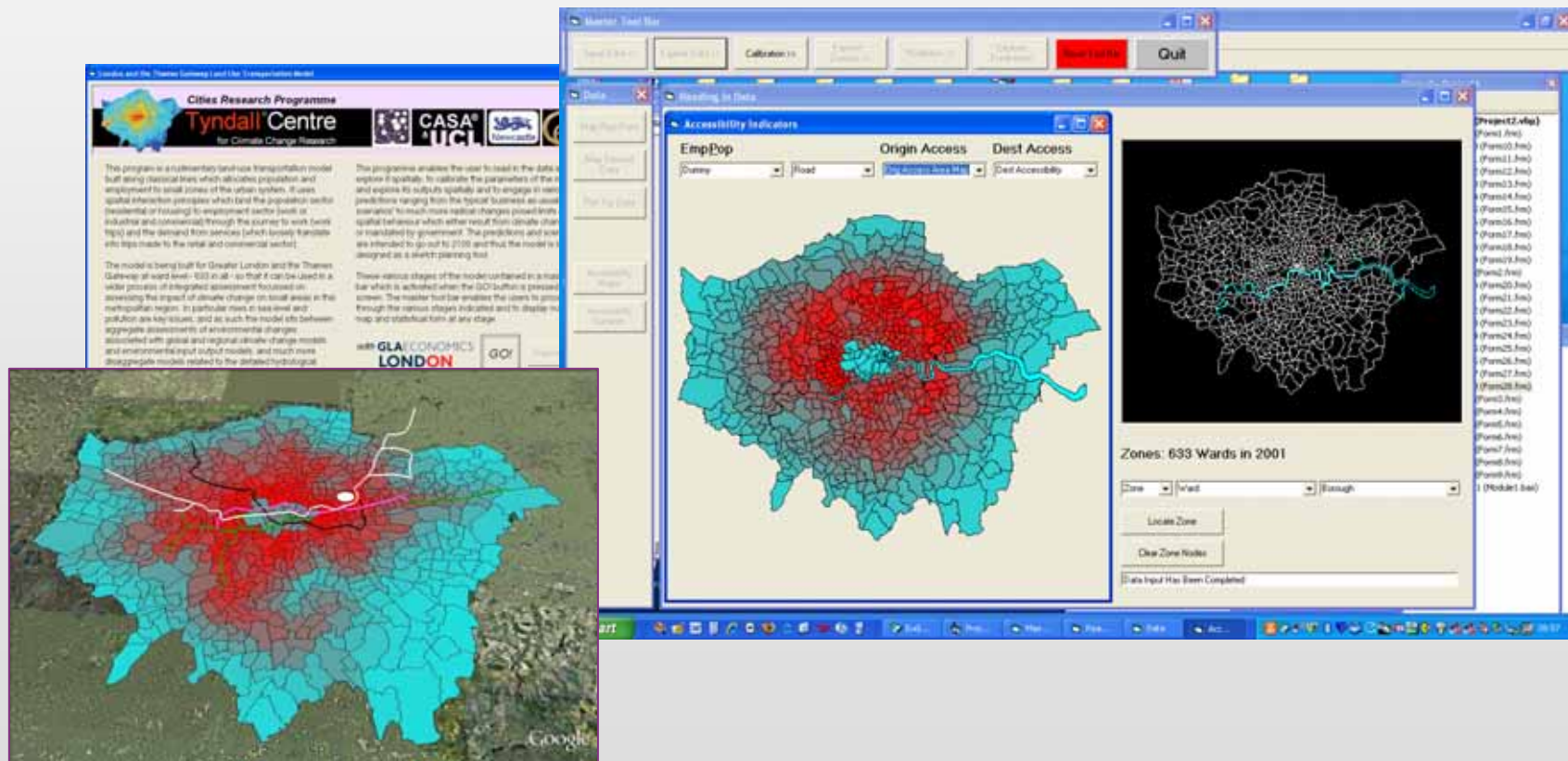
Our take on smart cities is about how smartness can enable longer term, less routine intelligence about our urban future. About strategy as well as tactics.

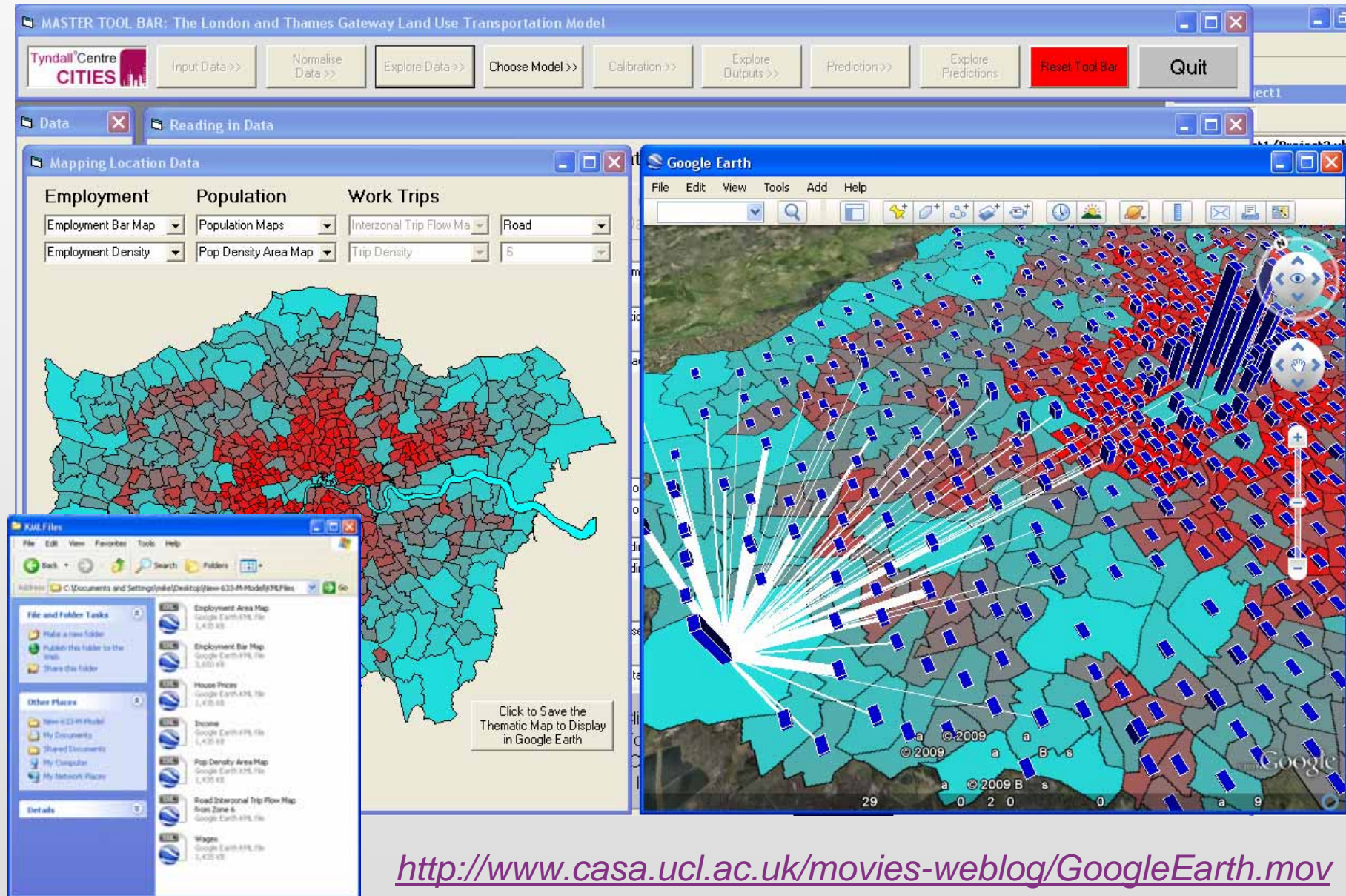
The longer term science of cities that we are concerned with involves computation to provide services

- initially in terms our understanding and long term planning
- and now complemented by providing more routine services across networks
- all of this involves online data, networks, simulations, optimisations and participation
- it involves treating the city as a online system, an integrated set of databases whose origins lie in the way we are able to sense what is happening
- This short term focus on real time data will eventually produce long term data series

1. Modelling Land Use Transportation, Energy, etc

Our core expertise is in land use transportation modelling and we have several such models for the London region:





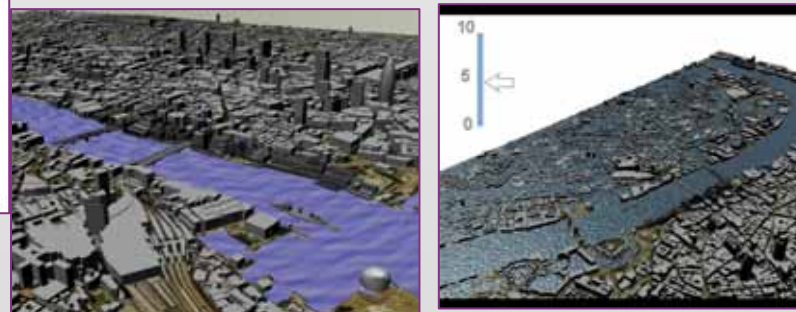
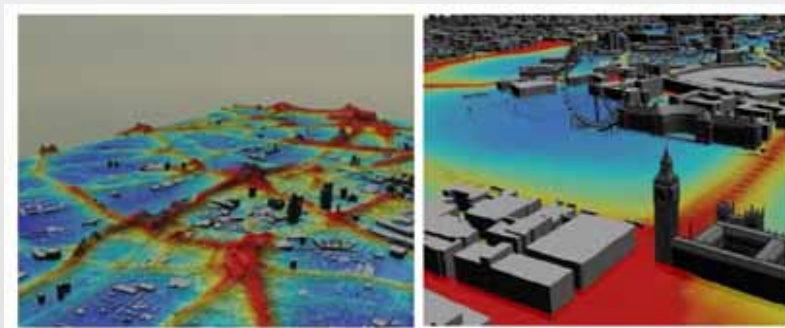
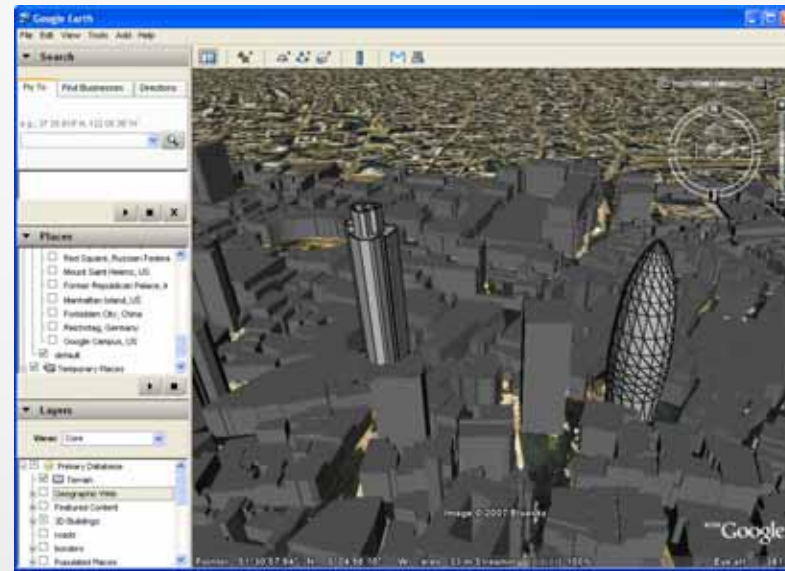
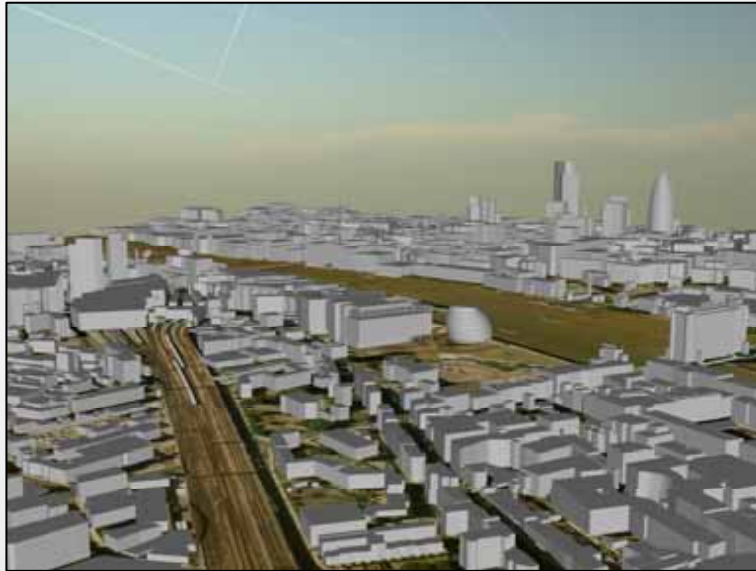
Modelling the Geometry of Cities: Virtual Cities

We have built a large scale 3-D model for London based on RS data at parcel levels. The model is different from our LUT models – requiring different skills

The models are being tagged with socio-economic data. We have used it for flooding, visualising air pollution, we have looked at the morphology of building form, and used it to visualise 2D to 3D design proposals.

When I last talked here in 2007, I think, in the technical university, I talked on these ideas – mainly on visualisation.

What is intriguing is the way *iconic and symbolic models are beginning to merge* – land use transport models with virtual city models.



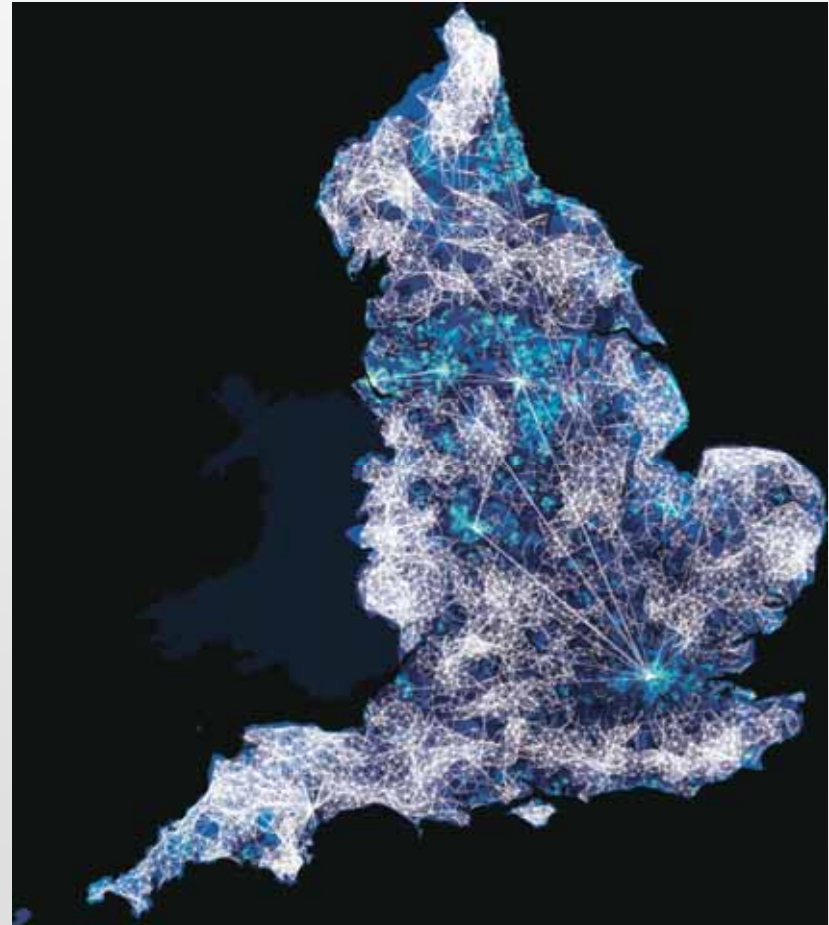
<http://www.londonair.org.uk/>

Representing Networks: Telecoms, Subways & Rail, Bikes

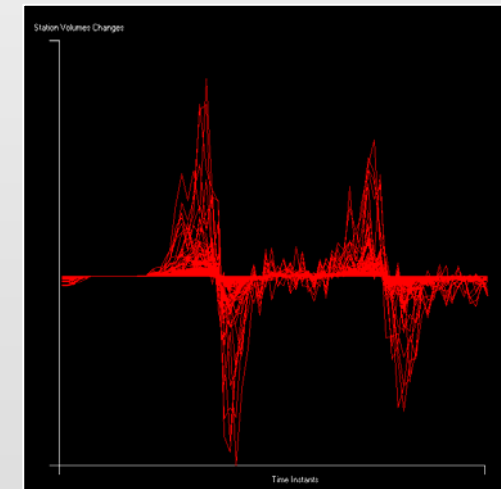
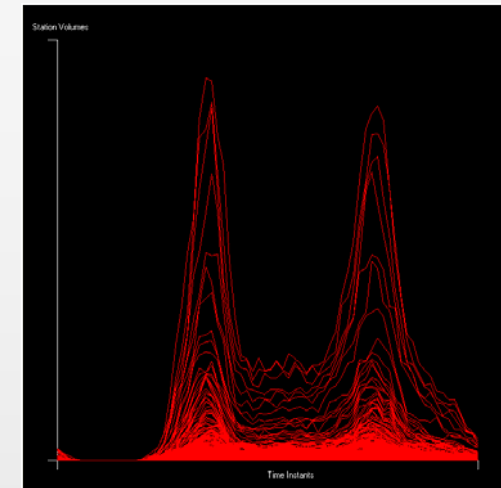
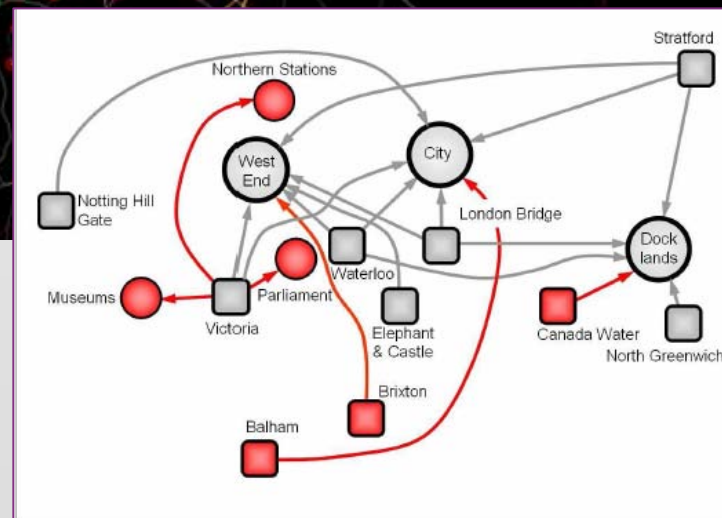
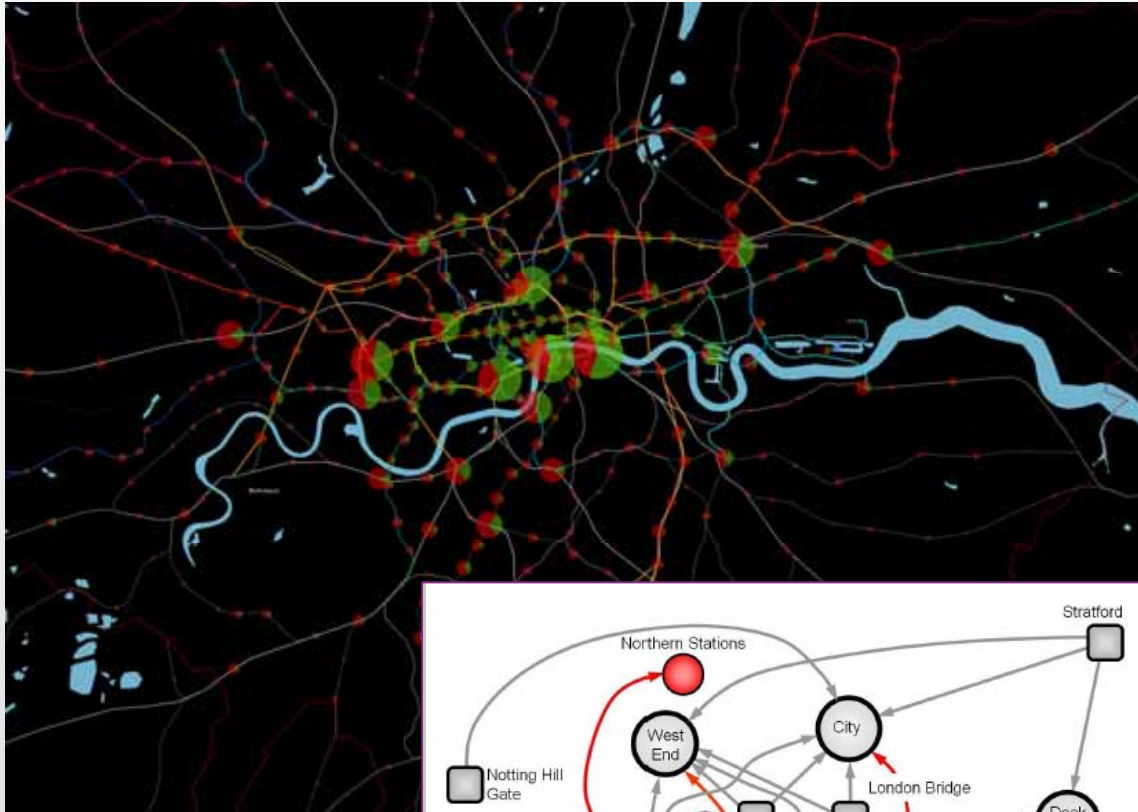
Many new sources of network data now exist, much of coming from digital sources and we are working with mining this data and extracting functionality from it

Our key data sets are telecoms data (landline) for the UK, the online travel card data (Oyster) for public transport schemes in London which is massive, really massive and the online bike movement data for the London bikes scheme. These are big data sets that record every phone call, trip etc over a period of days with each object time stamped. Let me show three static shots of this data.

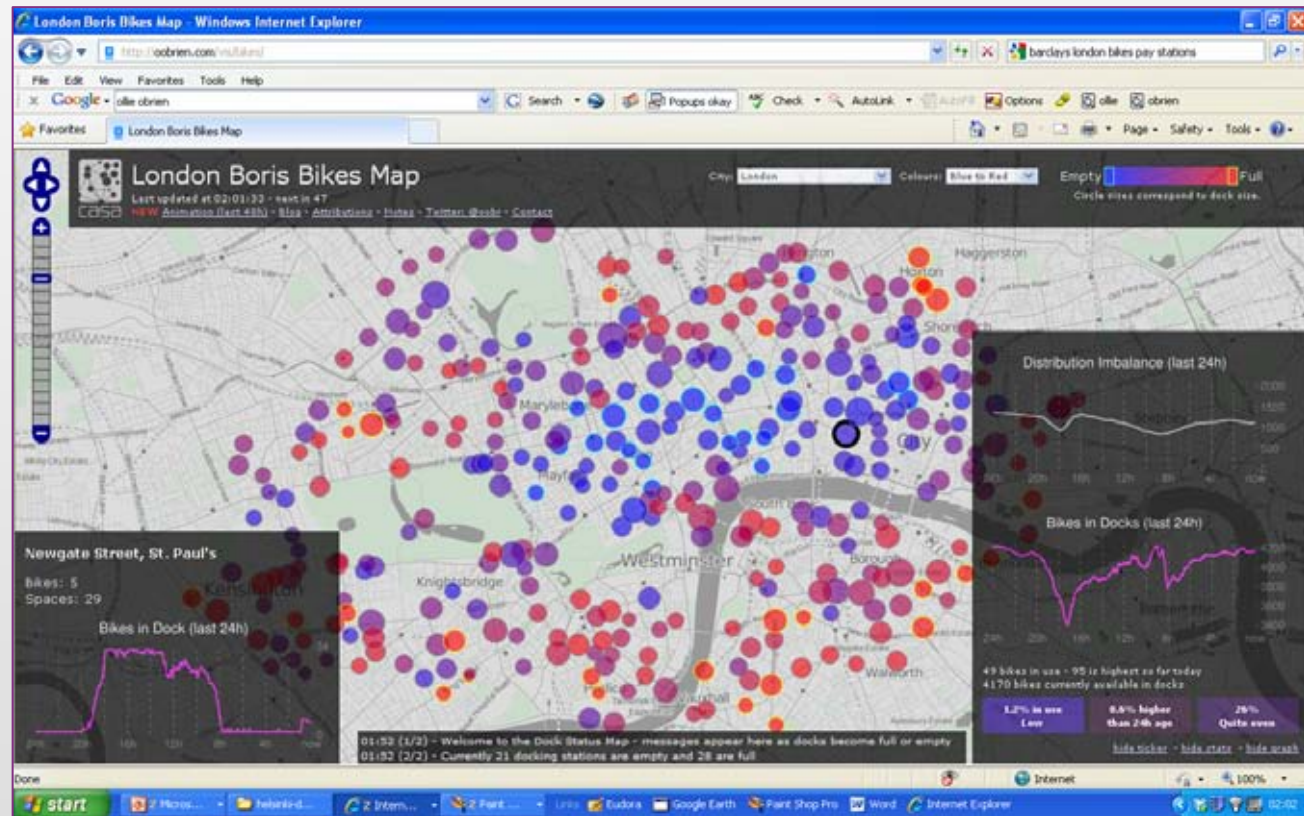
Telecoms – Jon Reades' work with a large UK telecoms provider and with Sensable City Lab at MIT



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



Bikes Data – 4200 bikes, started Nov 2010, all the data— everything – all trips, all times, all stations/docks



Our Oyster Card Project

We are hard at work with our Big Data set on London's public transport system. Let me emphasise yet again that we are talking here of 7-8 million swipes/trips a day, crudely – 40 million a week, 200 million a month, 2.4 billion a year, more than 10 billion in five years

As E. M. Forster in his short story of 1909 called “The Machine Stops” implies, this will only end when we or something switches off the machine ... the implications of data ***forever*** are interesting. We can mine this data a million ways but there are issues.

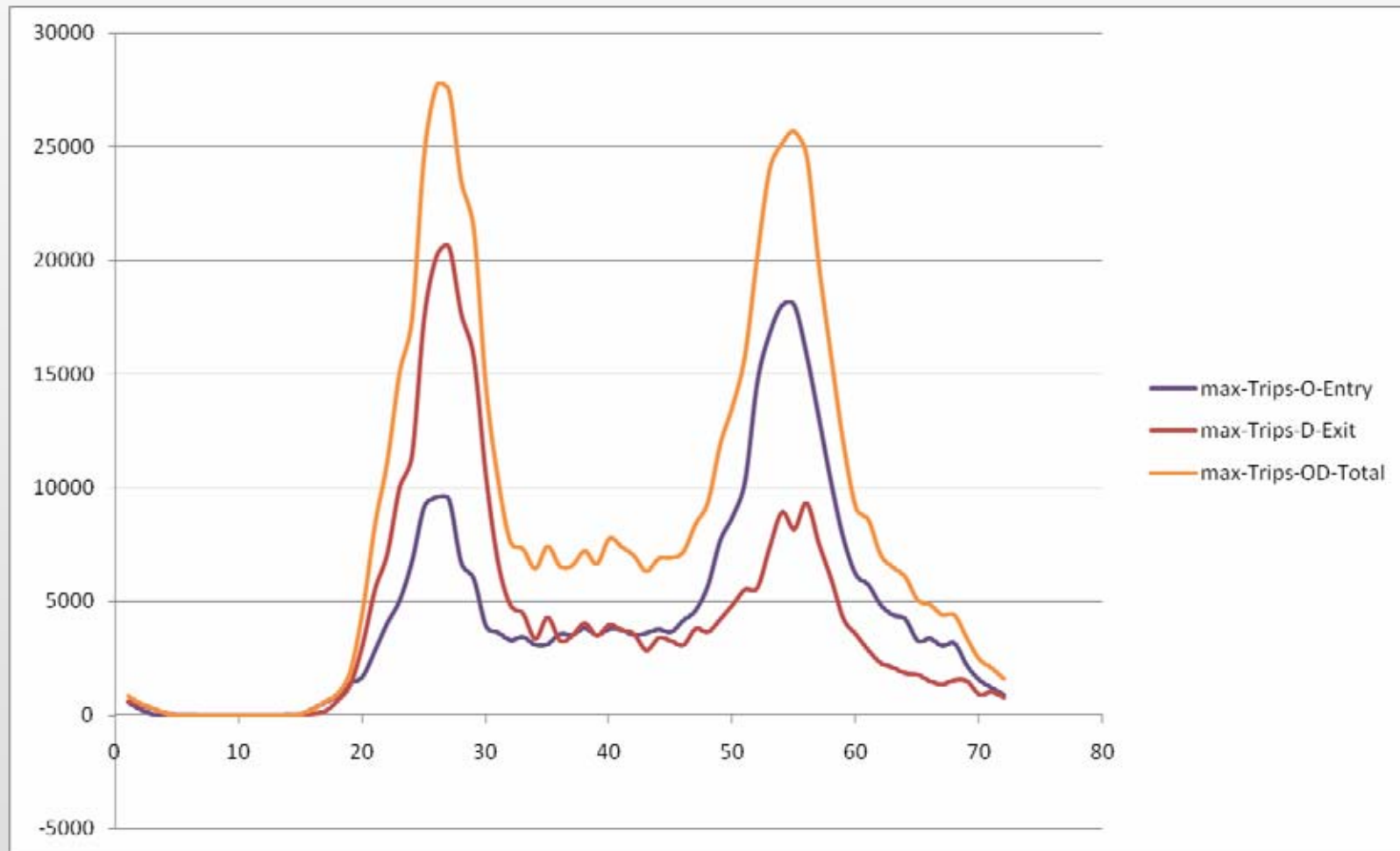
Let me give you a sense of what we are doing with our billion record data set

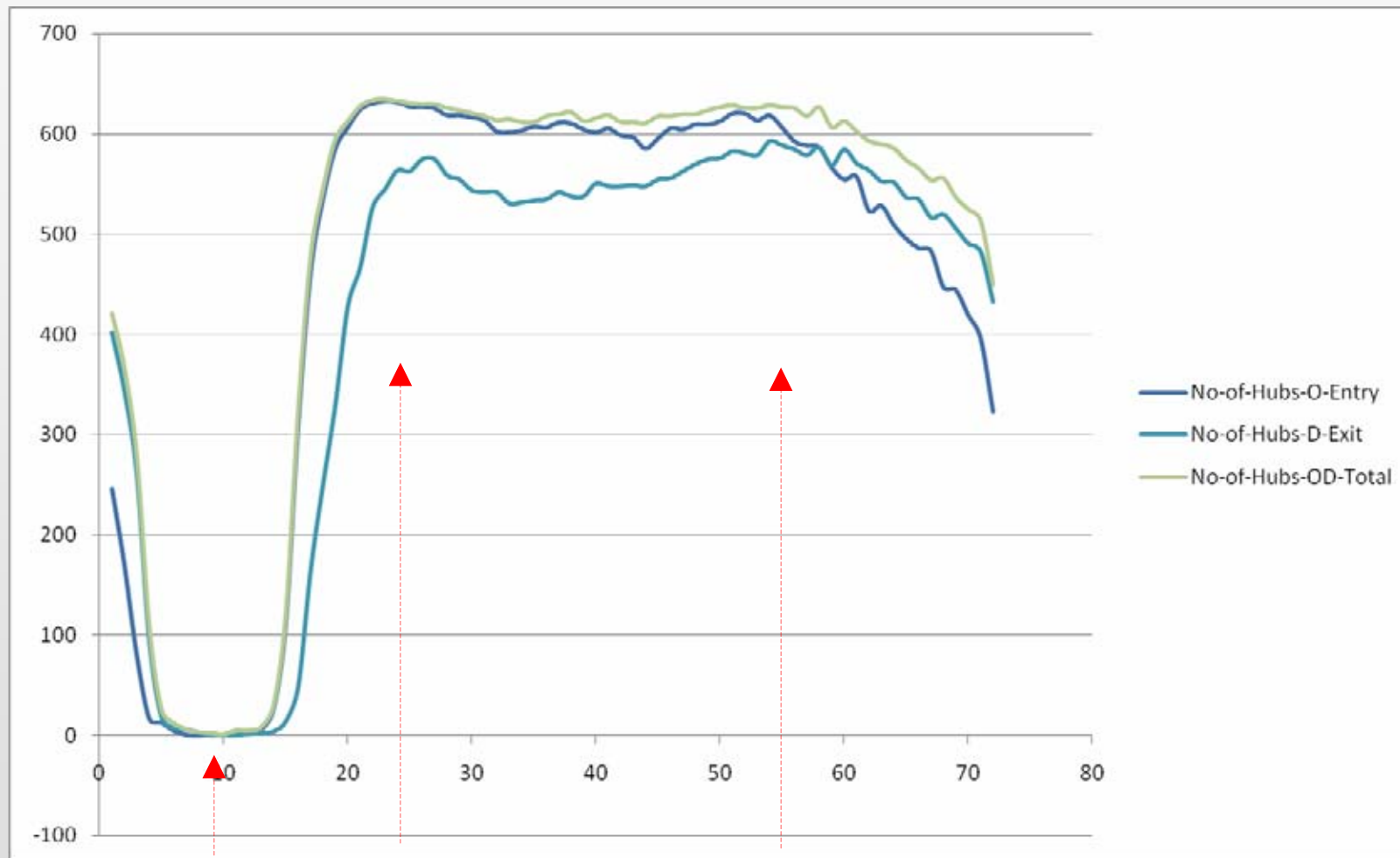
We can examine origins volumes, destination volumes separately and we are doing but here we will simply add these together as total volumes – in this sense they will not have meaning any longer as trips

	A	B	C
1	1	London-Bridge	599568
2	2	Victoria	502127
3	3	Waterloo	486861
4	4	Liverpool-Street	437658
5	5	Kings-Cross	395919
6	6	Shepherd's-Bush	346027
7	7	Hammersmith	274623
8	8	Wimbledon	198913
9	9	Paddington	196067
10	10	Vauxhall	180411
11	11	Stratford	177964
12	12	Oxford-Circus	150704
13	13	Charing-Cross	149290
14	14	Ealing-Broadway	139911
15	15	Euston	138394
16	16	Canary-Wharf	132206
17	17	Barking	112842
18	18	Balham	111090
19	19	Brixton	108814
20	20	London-Terminals	93026

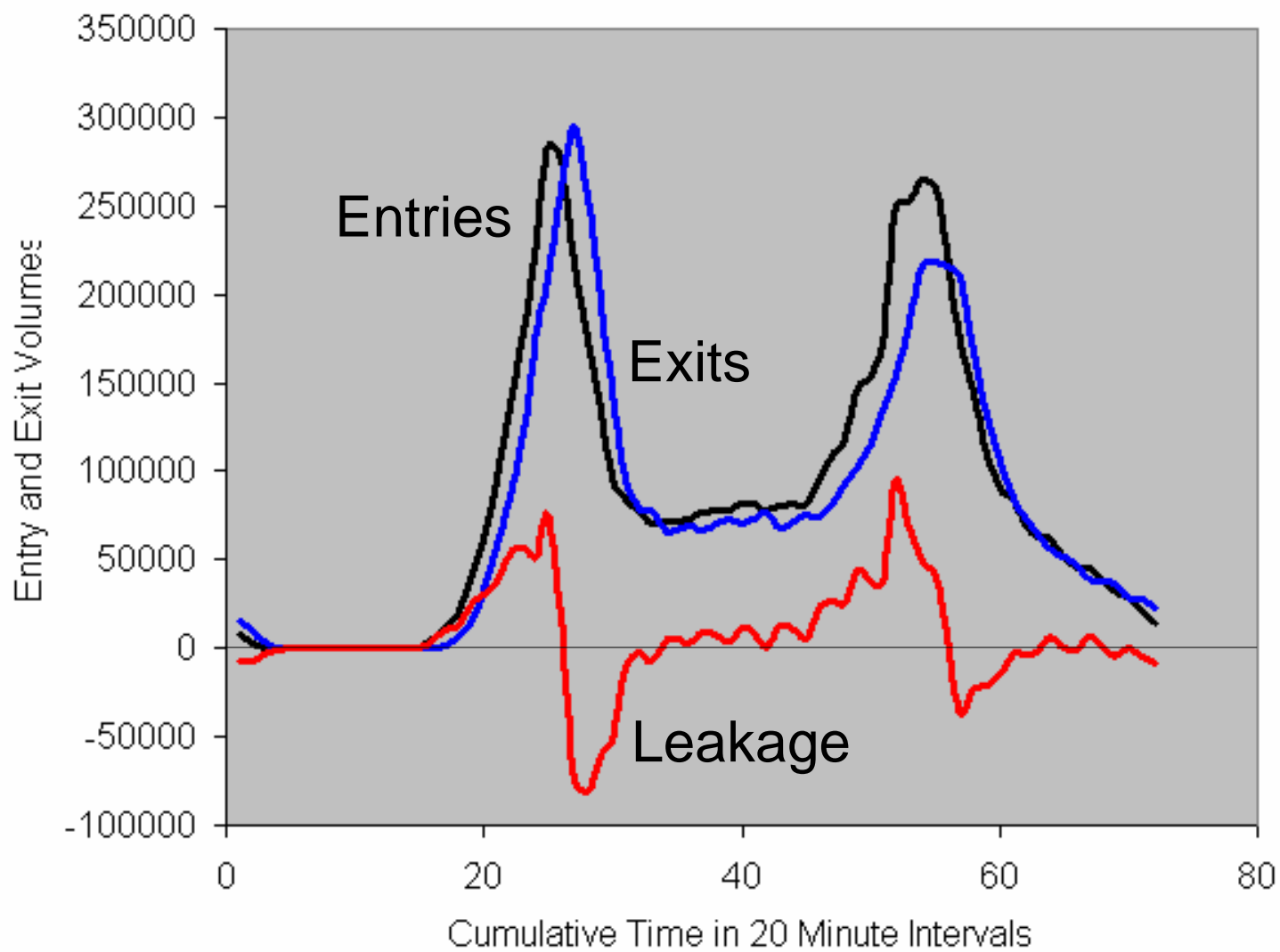
We will now examine the profiles of behaviour during the 24 hour day to provide some sense of the problem

Examining the Dynamics of the Hub Volumes





Night am peak pm peak



We will look at various comparisons between hub volumes as ordered from largest to smallest. We show these as counter-cumulative frequencies which are rank size plots.

We add swipe in to swipe outs. Because of their right skewness, we plot them on log log scales which if they follow power laws – which they don't for obvious constraints on their scaling – would appear as straight lines.

Here is the scaling

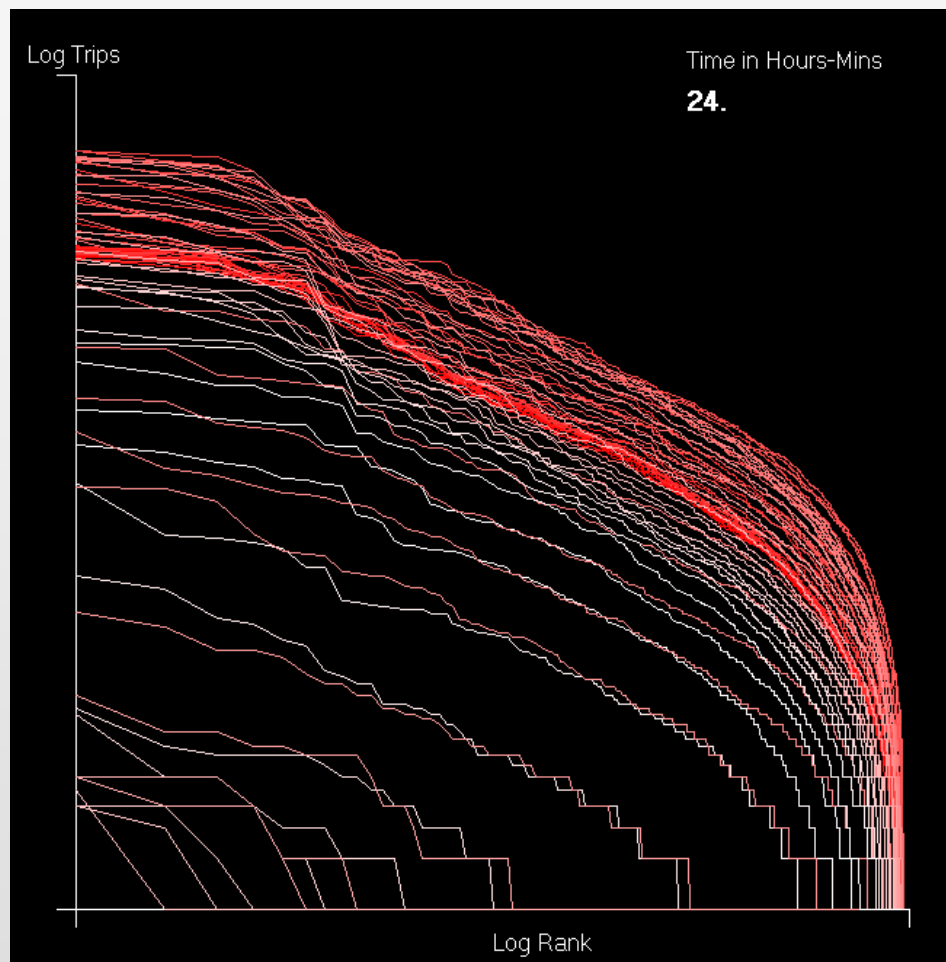
All Hub Volumes Ordered as Rank Size Profiles

$$O_i(t) + D_i(t) = \sum_j F_{ij}(t) + \sum_j F_{ji}(t)$$

Let me load the program and run it as it is quite short and fast and gives you an idea of the dynamics

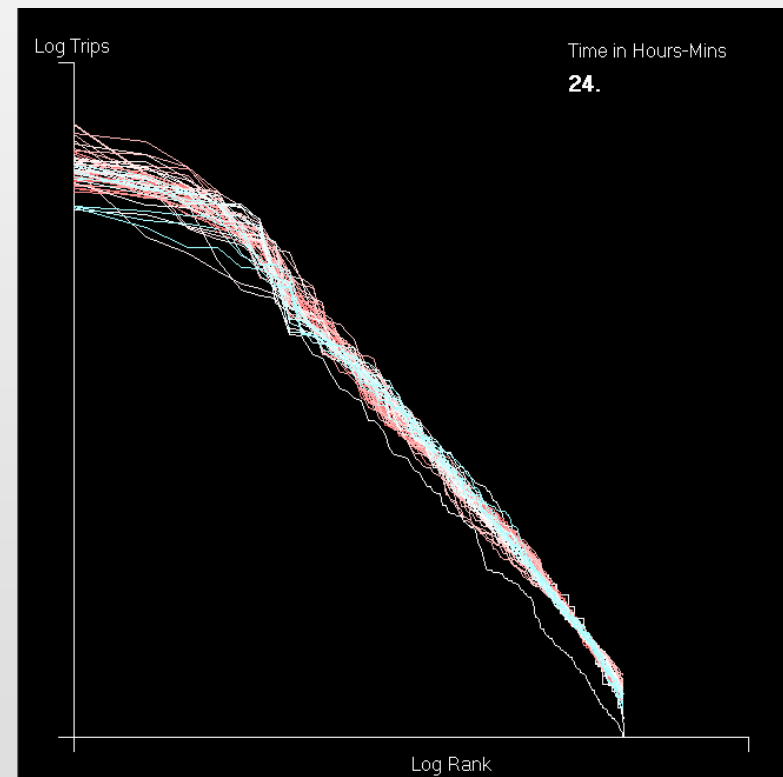
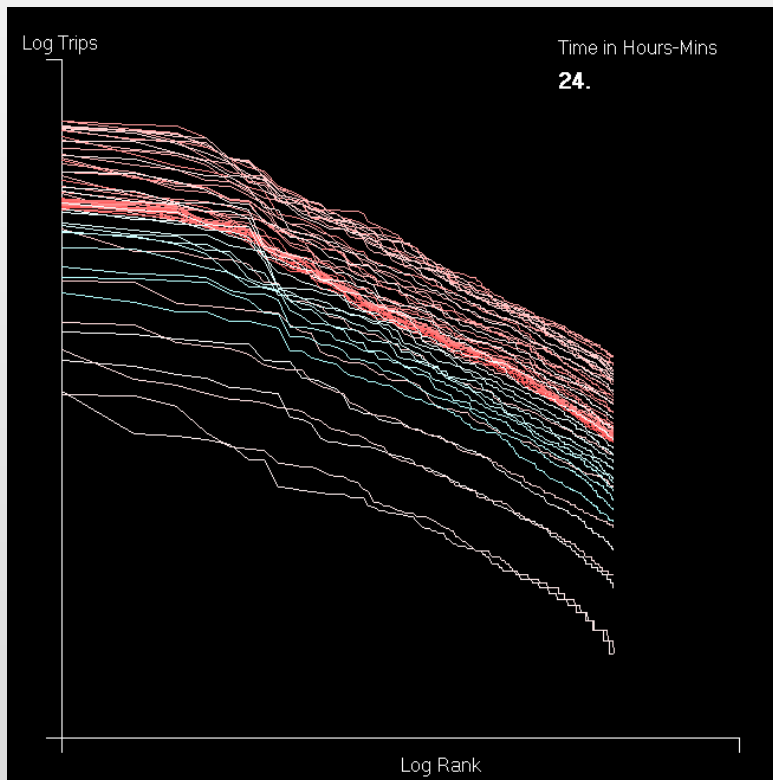


Project1.exe
CASA-UCL



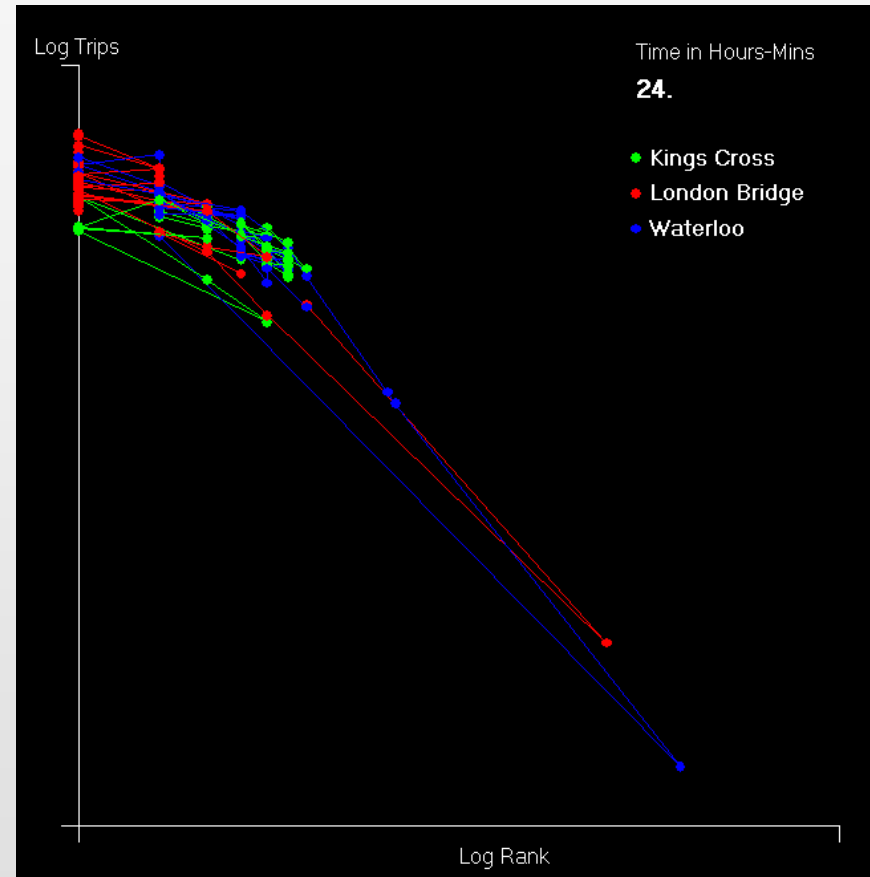
$R_i(t)$

We have a major problem as all hubs are not always active. To make good comparisons, we need to compare like with like – nos of hubs & volumes *Reduced to Top 200 Collapsed/Standardised*



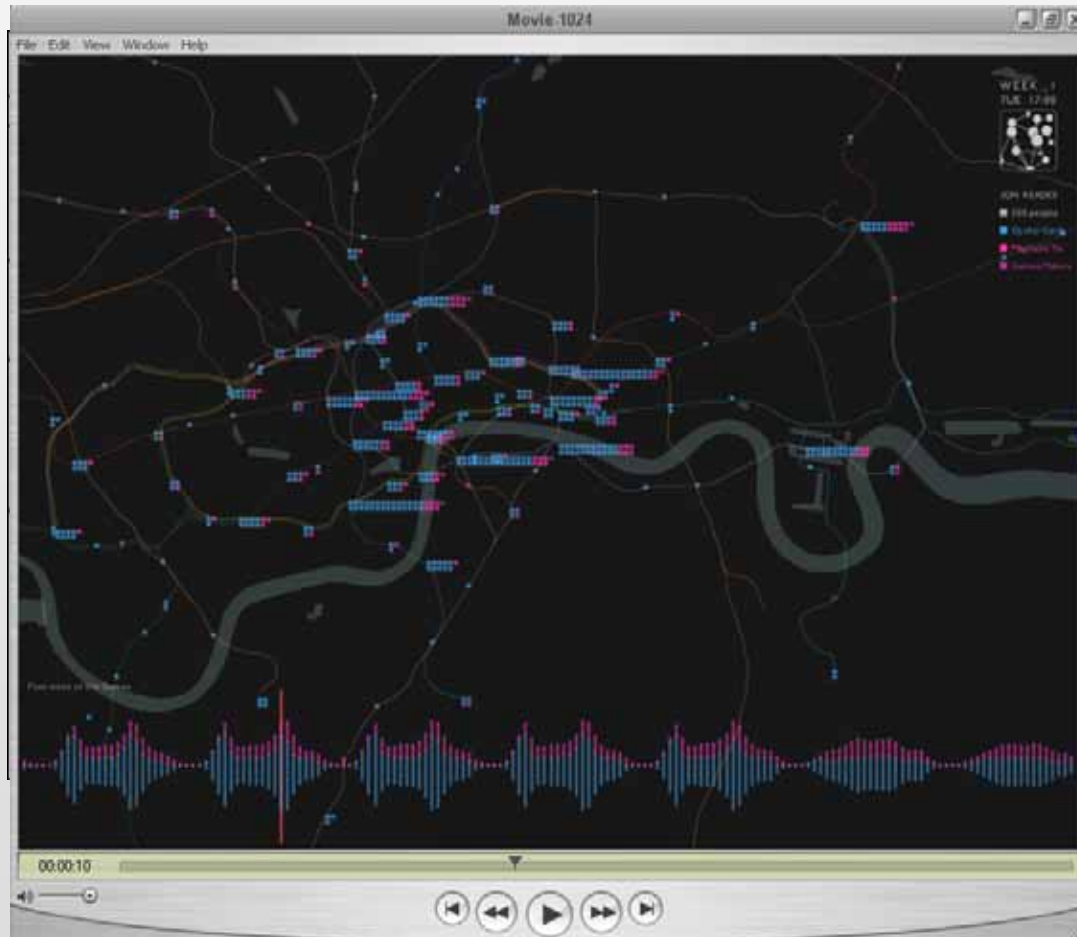
The Dynamics: Examining Individual Hubs: Trajectories

Here we show the movement in the hub volumes and ranks in the direction of the lines and dots



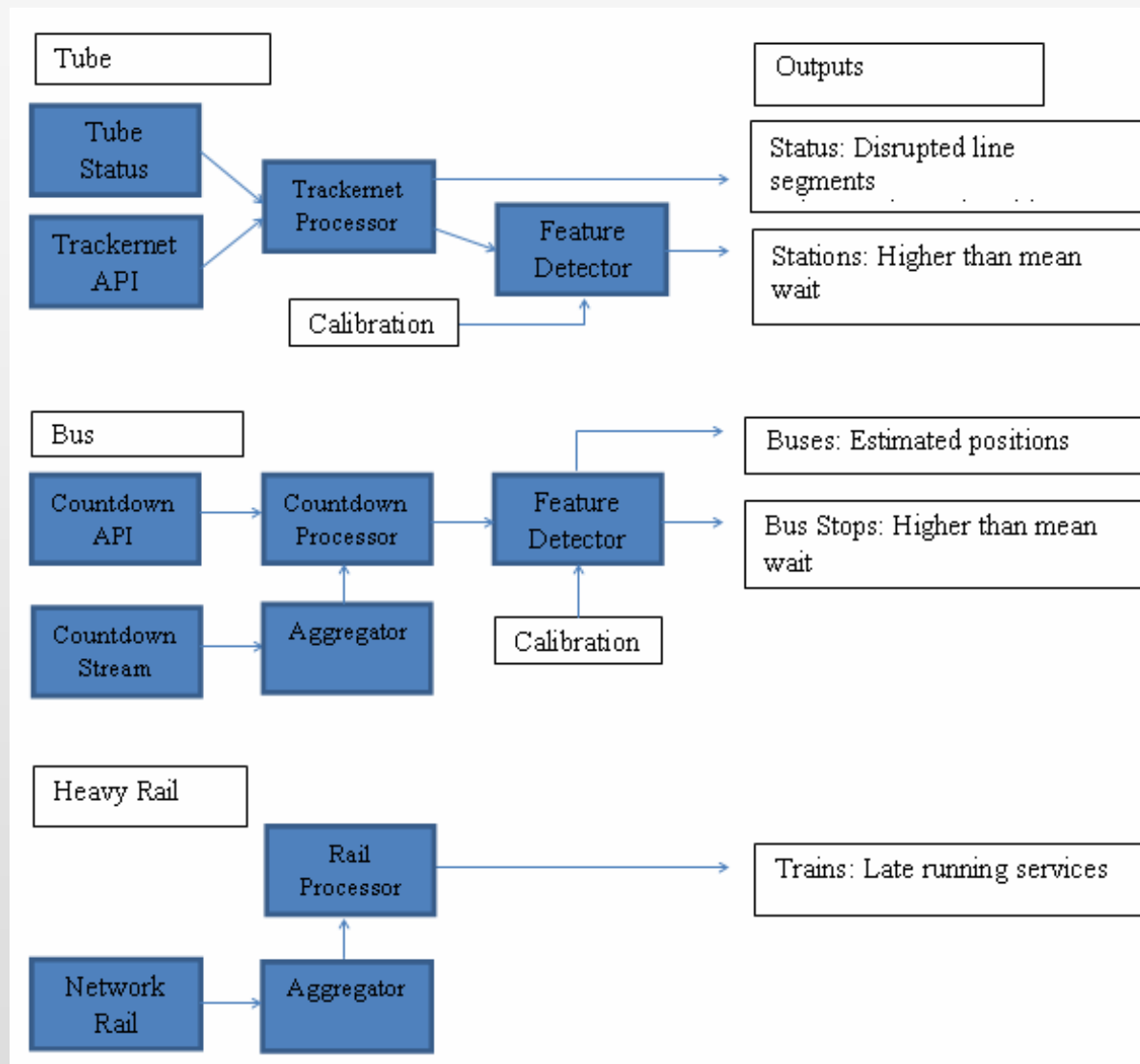
This gives you a sense of a tiny little bit of our data mining

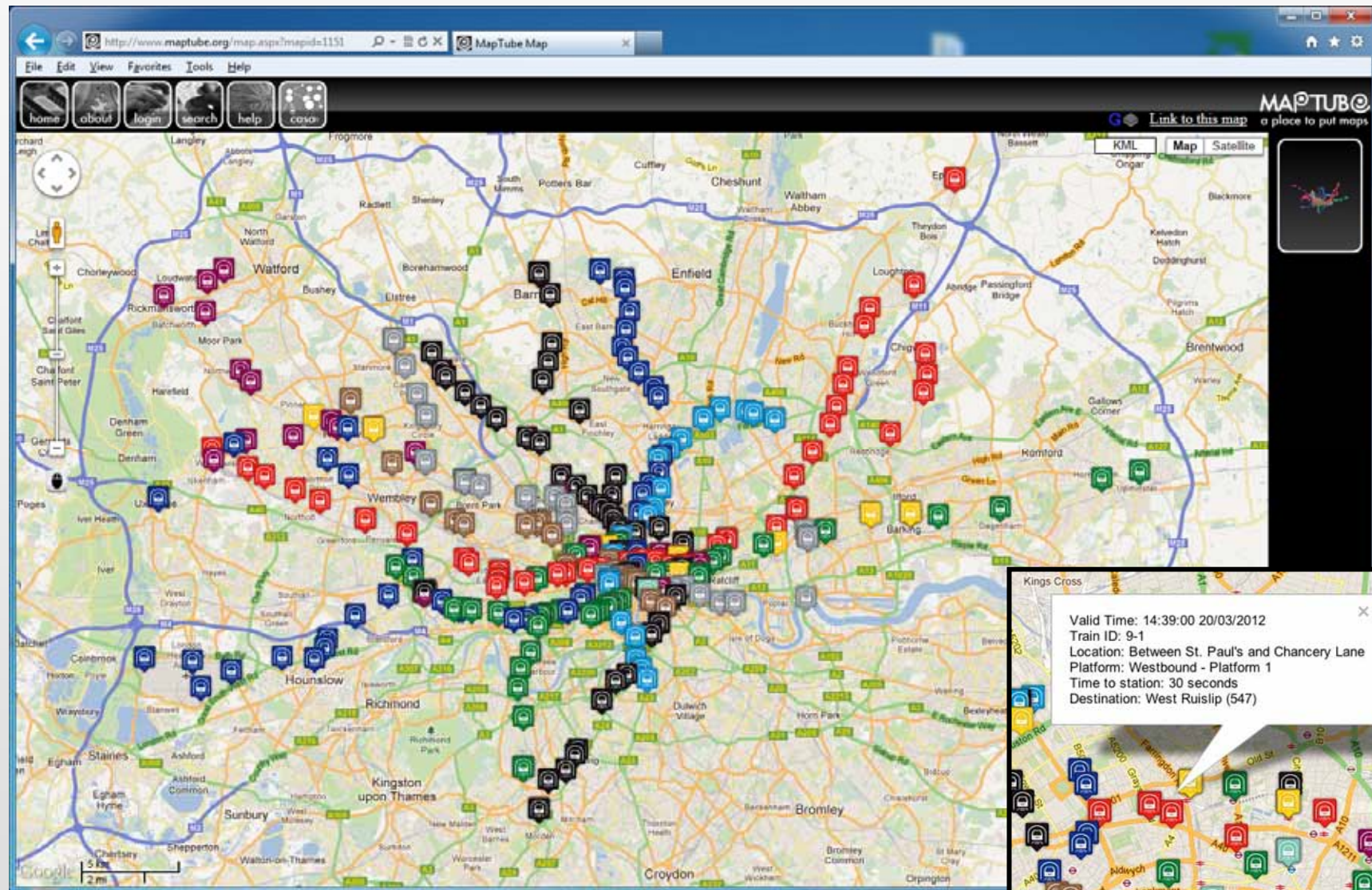
We are doing a lot of work on disruption – what happens if a subway station closes or line goes down – billions of possibilities that can provide information useful to travellers

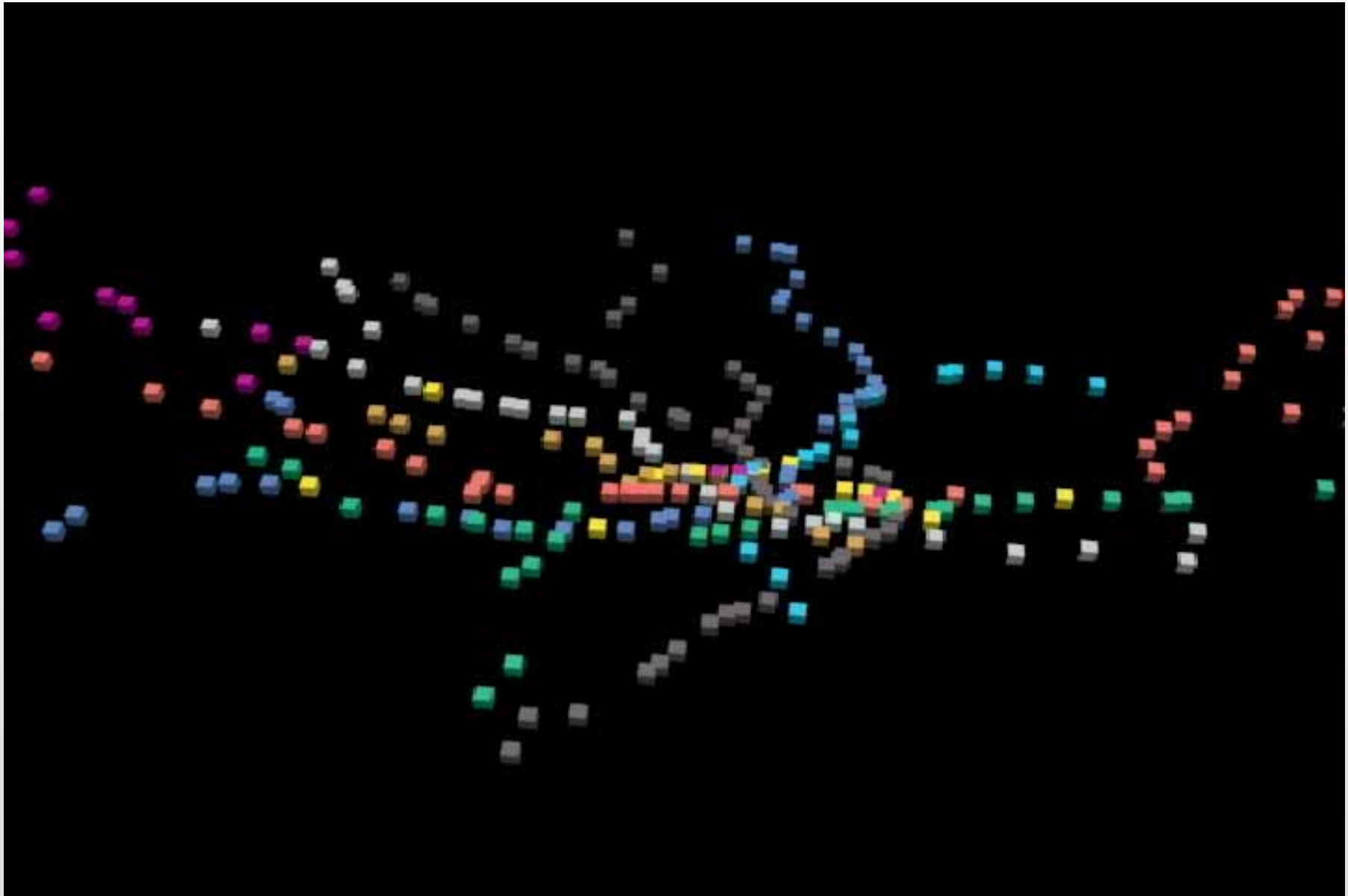


We are looking at the impact of the *Olympic Games*, and lots of stuff like this. And that would have been what my lecture would be on. But this is more generic and I will now turn to data sensing of vehicles

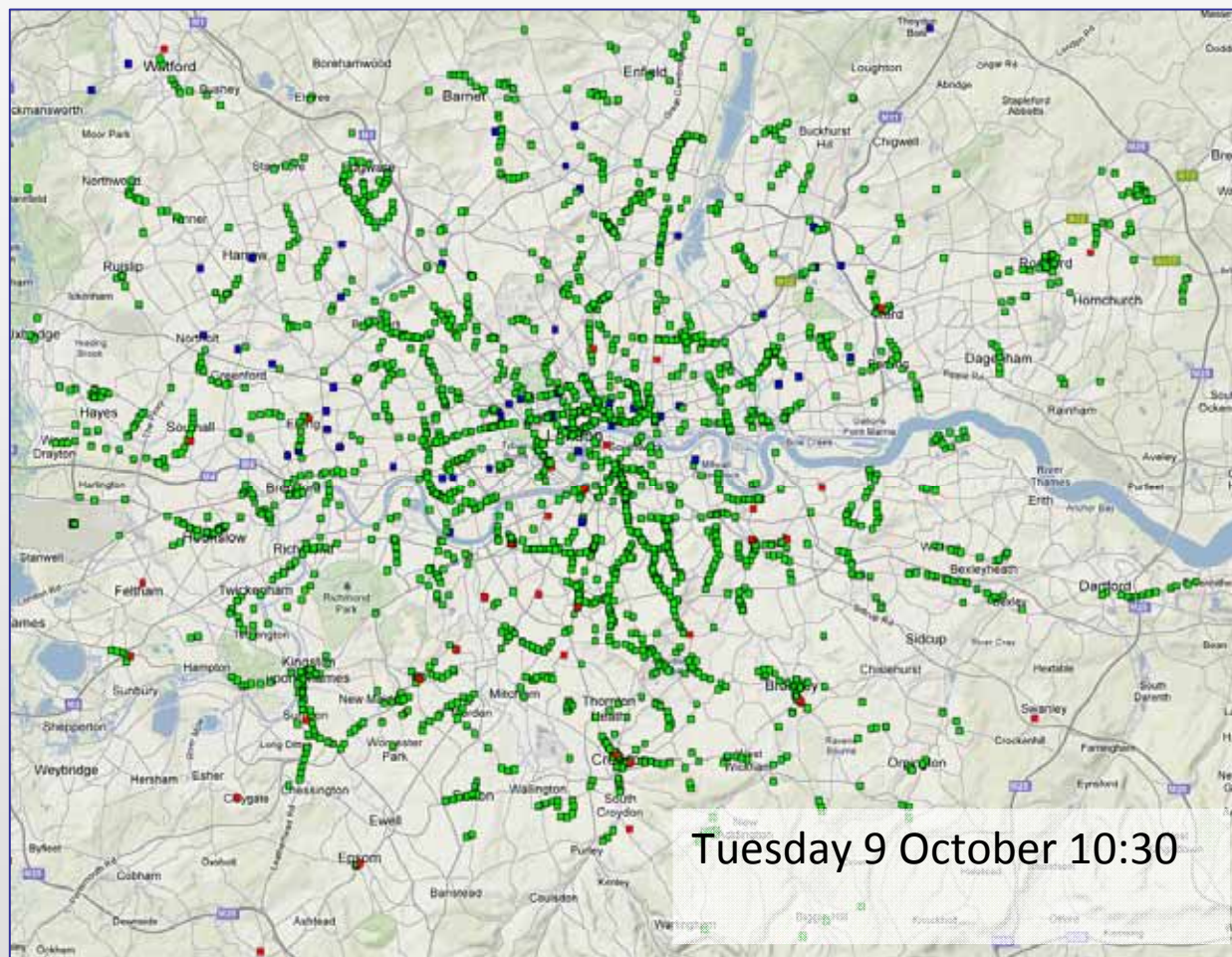
The Public Transport System in Terms of Vehicle Flows







Delays from Tube, National Rail and Bus Fused



Key



National Rail
more than 5
minutes late



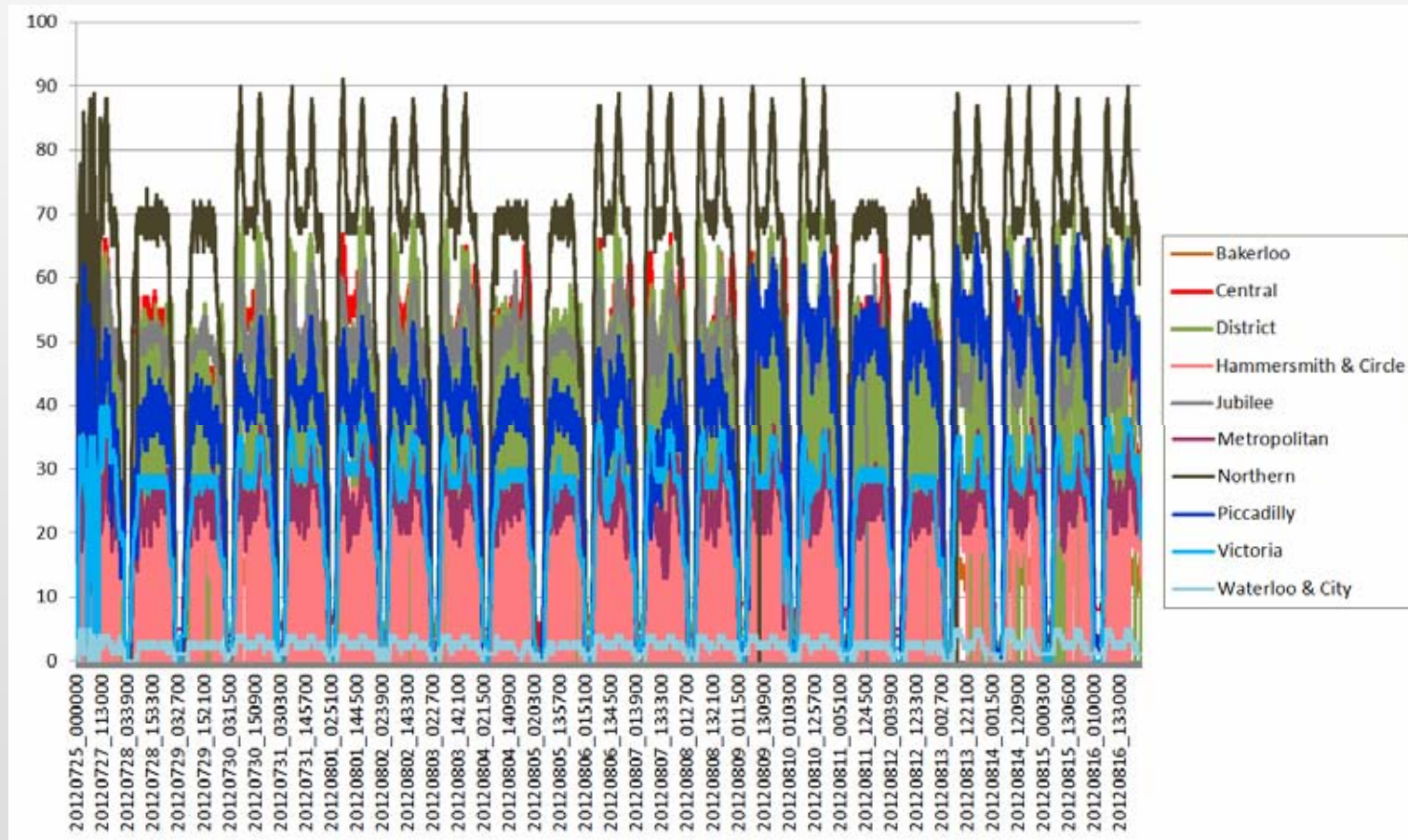
Tube stations
showing a wait
time 15% above
expected



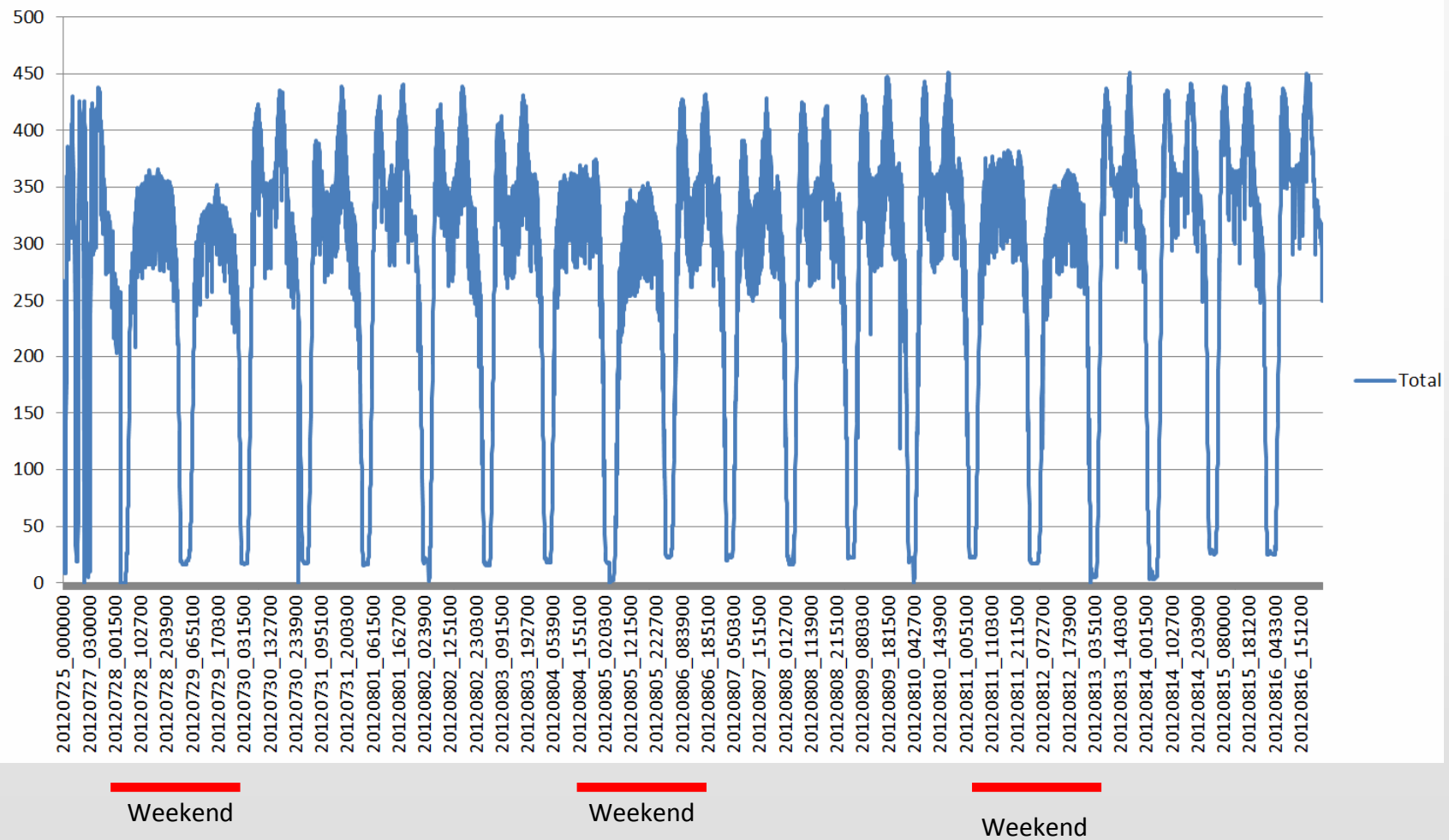
Bus stops
showing a wait
time 20% above
expected

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

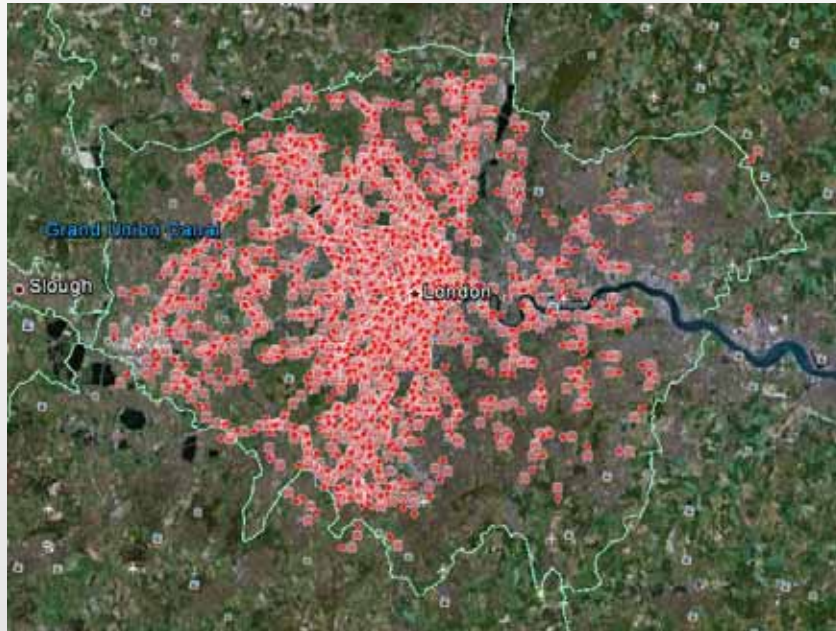
Flows During the Olympics – we are looking at this as a case study



Total



The Effect of Bus Strike



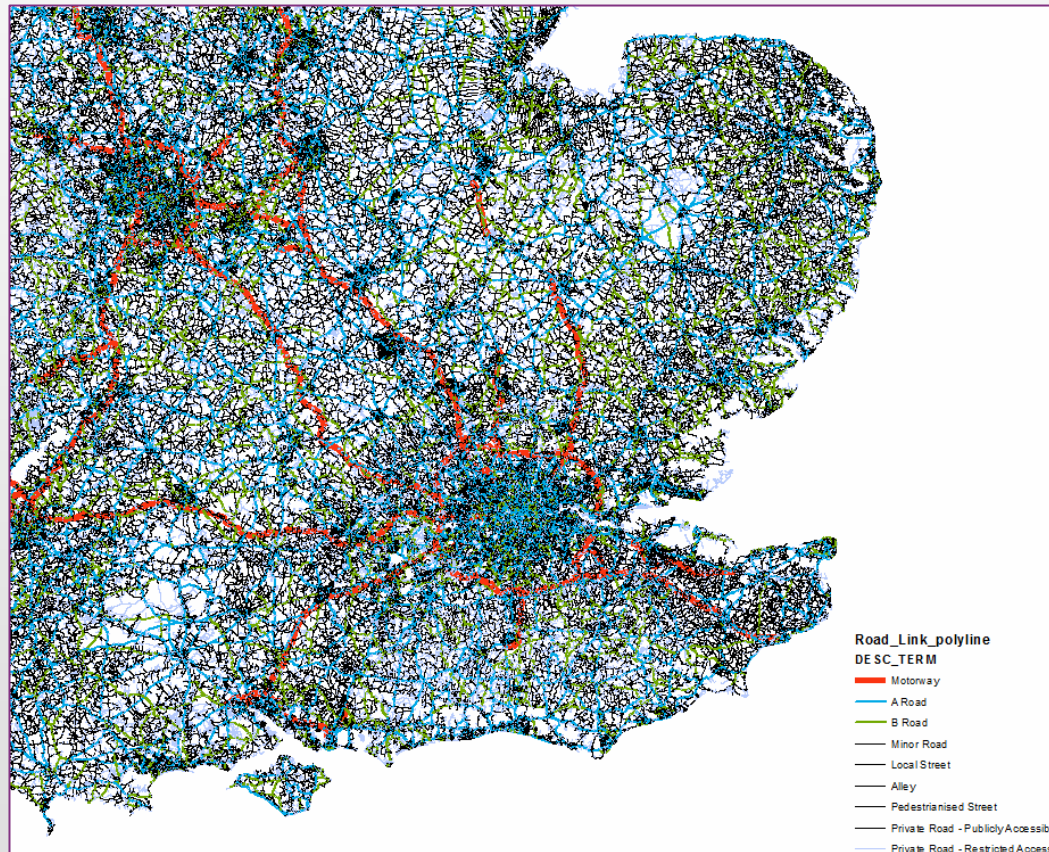
Tuesday 22nd May 2012, 09:00



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.

Public Transport Vehicle (not People) Flows from Timetable and OS Streetline Data by Joan Serras (CASA)



http://vimeo.com/21351143


View Favorites Tools Help

Search Popups okay Check AutoLink AutoFill Options

Google Earth Internet Explorer Paint Shop Pro Word Twitter

transport flows, UK on Vimeo

vimeo Me Videos Upload Tools Explore Help Search Videos

 **Public Transport flows, UK**
by Joan Serras PLUS
7 months ago

Couch Mode

07:49 AM

02:26

HD

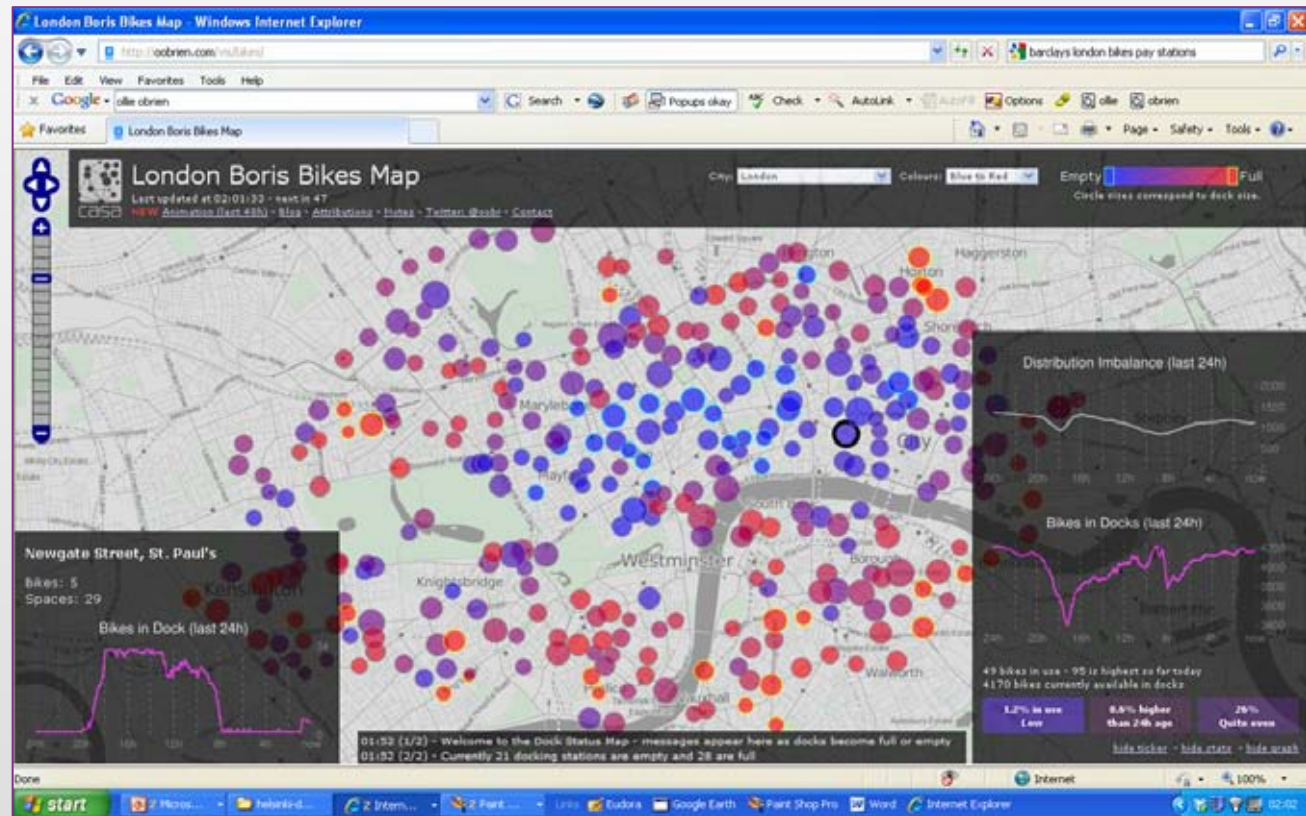
LIKE LATER SHARE EMBED

More Add

Show me Joan Serras' videos See all

- 3. **Public Transport flows, London**
by Joan Serras
7 months ago
- 2. **Bus flows, UK**
by Joan Serras
7 months ago
- 1. **Public Transport flows, UK**
by Joan Serras
7 months ago

Our Bikes Project: Bikes Data – 4200 bikes, started Nov 2010, all the data– everything – all trips, all times, all stations/docks





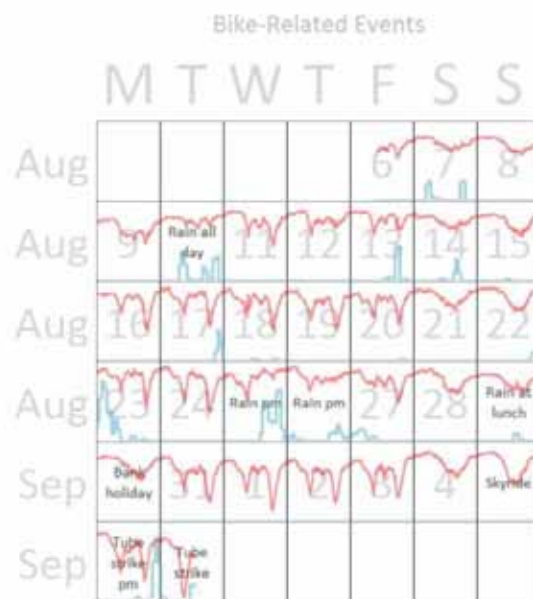
Animations of Public Bike Movements



Animations of Changes in the Bike Nodes: Docking

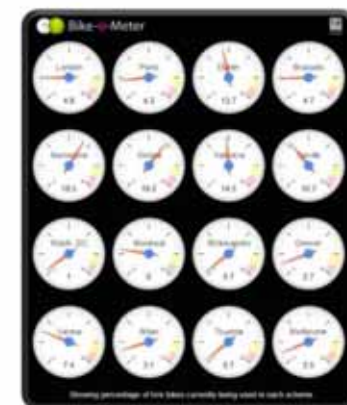
More Analysis

- **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 ☺



Flow Network Analysis - Martin Zaltz Austwick - Windows Internet Explorer


http://www.bartlett.ucl.ac.uk/casa/events/2012-02-29-Martin-Austwick

File Edit View Favorites Tools Help

Google Search

Google Earth Internet Explorer Paint Shop Pro Word

Flow Network Analysis - Martin Zaltz Austwick



Flow Network Analysis - Martin Zaltz Austwick

Centre for Advanced Spatial Analysis

School of Architecture

School of Construction & Project Mgt

Development Planning Unit

UCL Energy Institute

School of Graduate Studies

School of Planning

News

Events

Guest Lectures

Seminars

Blogs

Podcasts

Publications

Software

Bartlett TV

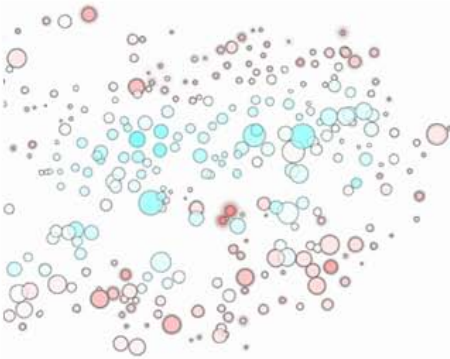
Subscribe

Twitter

17:00 29 February 2012

Location: Pearson Lecture Theatre (North East Entrance)

Flow Network Analysis by Martin Zaltz Austwick, Lecturer in Advanced Spatial Analysis



Related articles

This article has no keywords

Share



Centre for Advanced Spatial Analysis



A Framework for All of This: The Complexity Sciences

I will not labour this much longer but much of our work is informed by our general interest in understanding cities using the complexity sciences. Some of us in CASA have a bigger picture that involves how all this fits into urban theory and planning.

We are thus interested in understanding the social physics of the city, networks, flows morphology, dynamics, resilience, emergence and so on

I refer you to my own weblog – www.complexCity.info

Which I call **A Science of Cities** (because I believe there is more than one science – there are many)

Let me finish by listing some recent resources: first papers

Batty, M. et al. (2012) Smart Cities of the Future, **European Physical Journal Special Topics**, 214, 481–518.

Roth, C., Kang, S. M., **Batty, M.**, and Barthelemy, M. (2012) A Long-Time Limit for World Subway Networks, **Journal of the Royal Society Interface**, online 16 May 2012

Johansson, A, Batty, M., Hayashi, K., Al Bar, O., Marcozzi, D., and Memish, Z. A. (2012)) Crowd and Environmental Management During Mass Gatherings, **The Lancet Infectious Diseases**, doi:10.1016/S1473-3099(11)70287.

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

Batty, M., Hudson-Smith, A., Milton, R. and Crooks, A. (2010) Map Mashups, Web 2.0 and the GIS Revolution, **Annals of GIS, 16(1)**, 1 – 13.

Ratti C, Sobolevsky S, Calabrese F, Andris C, **Reades J.**, Martino, M., Claxton, R., Strogatz, S. (2010) Redrawing the Map of Great Britain from a Network of Human Interactions. **PLoS ONE 5(12)**: e14248. doi:10.1371/journal.pone.0014248

And some of our blogs

A Science of Cities <http://www.complexcity.info/>

Spatial Complexity <http://www.spatialcomplexity.info/>

Big Data ToolKit <http://bigdatatoolkit.org/>

Digital Urban <http://www.digitalurban.org/>

GIS and Agent-Based Modelling <http://gisagents.blogspot.com/>

Simulacra <http://simulacra.blogs.casa.ucl.ac.uk/>

Sociable Physics <http://sociablephysics.wordpress.com/>

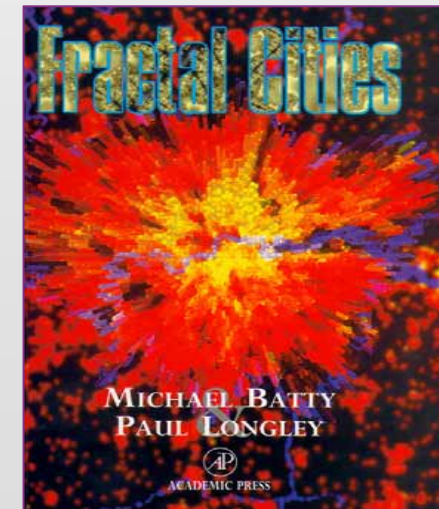
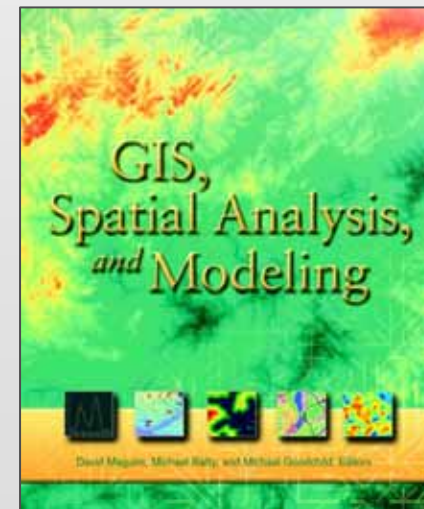
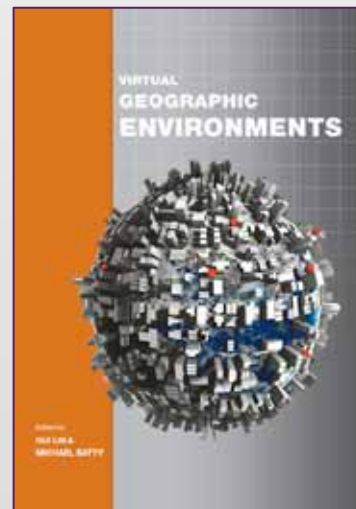
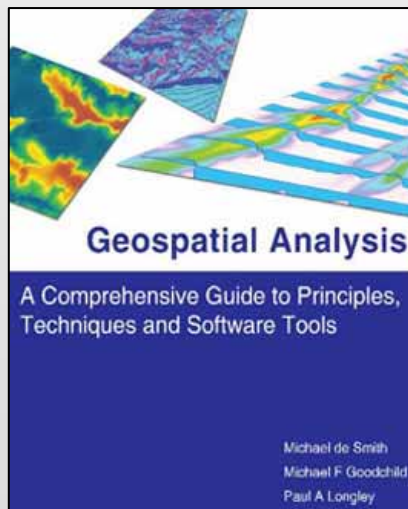
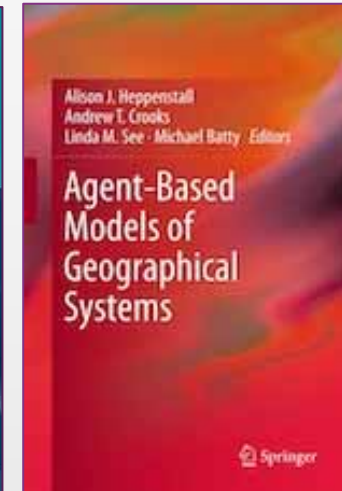
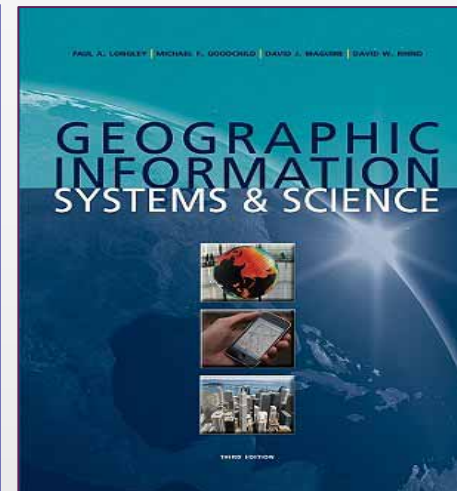
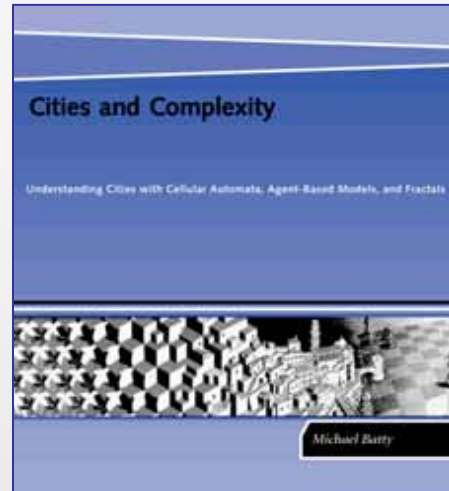
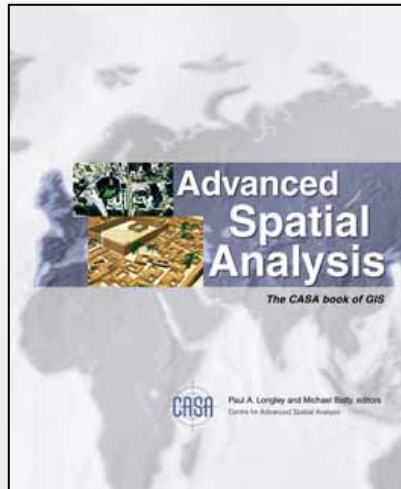
Spatial Analysis <http://spatialanalysis.co.uk/>

Suprageography <http://oliverobrien.co.uk/>

The Mapping London Blog <http://mappinglondon.co.uk/>

Urban Tick <http://urbantick.blogspot.com/>

And some of our books



<http://blogs.casa.ucl.ac.uk/>

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

Thanks, Questions?