Dynamics of Urban Movement:

Changes in the Scaling of Hubs in the London Rail Network

Michael Batty & Jon Reades
University College London

m.batty@ucl.ac.uk
http://www.casa.ucl.ac.uk/

This is a work in progress.

It will be completed in time as we progress the research.

The pdf will be updated over the next few months

Outline

- What the project is all about – examining 24 hours of Digital Data concerning all rail flows on the London Tube (Underground) and Overground Systems
- Formalising the Network and Flow Problem
- Exploring the Temporal Scaling Profiles of Node Volumes
- Correlations and Comparisons of Profiles
- Dynamics of Rank-Size Profiles – Plots, Clocks, Graphs
- Further Work: Classification of Nodes/Hubs

What the project is all about

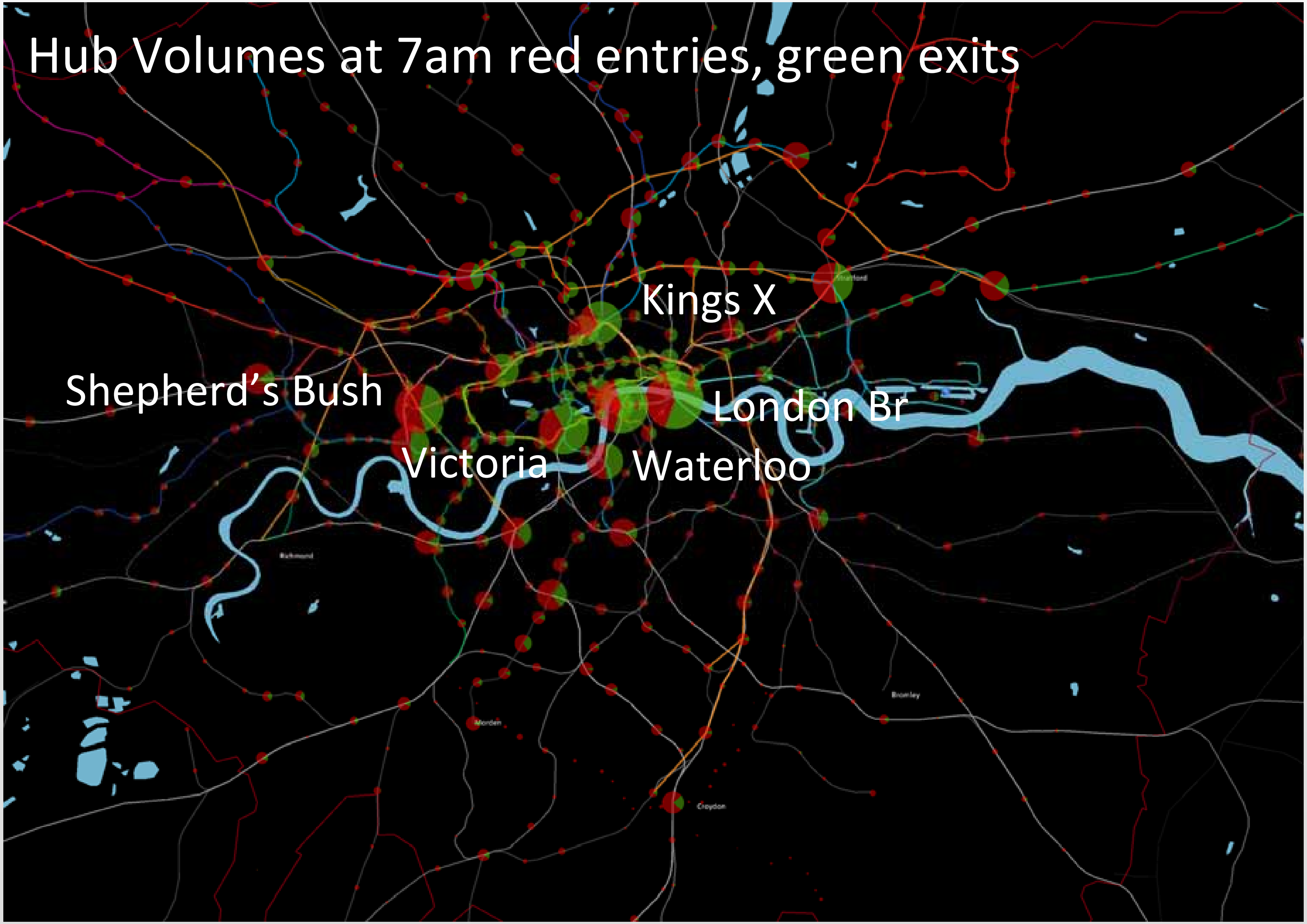
24 hours of Oyster Card data, for London (Transport for London area) on a Monday in November 2010 (as part of a much bigger data set for three weeks

Recorded Using Swipe in and Swipe Out which works only for Under- and Overground Rail

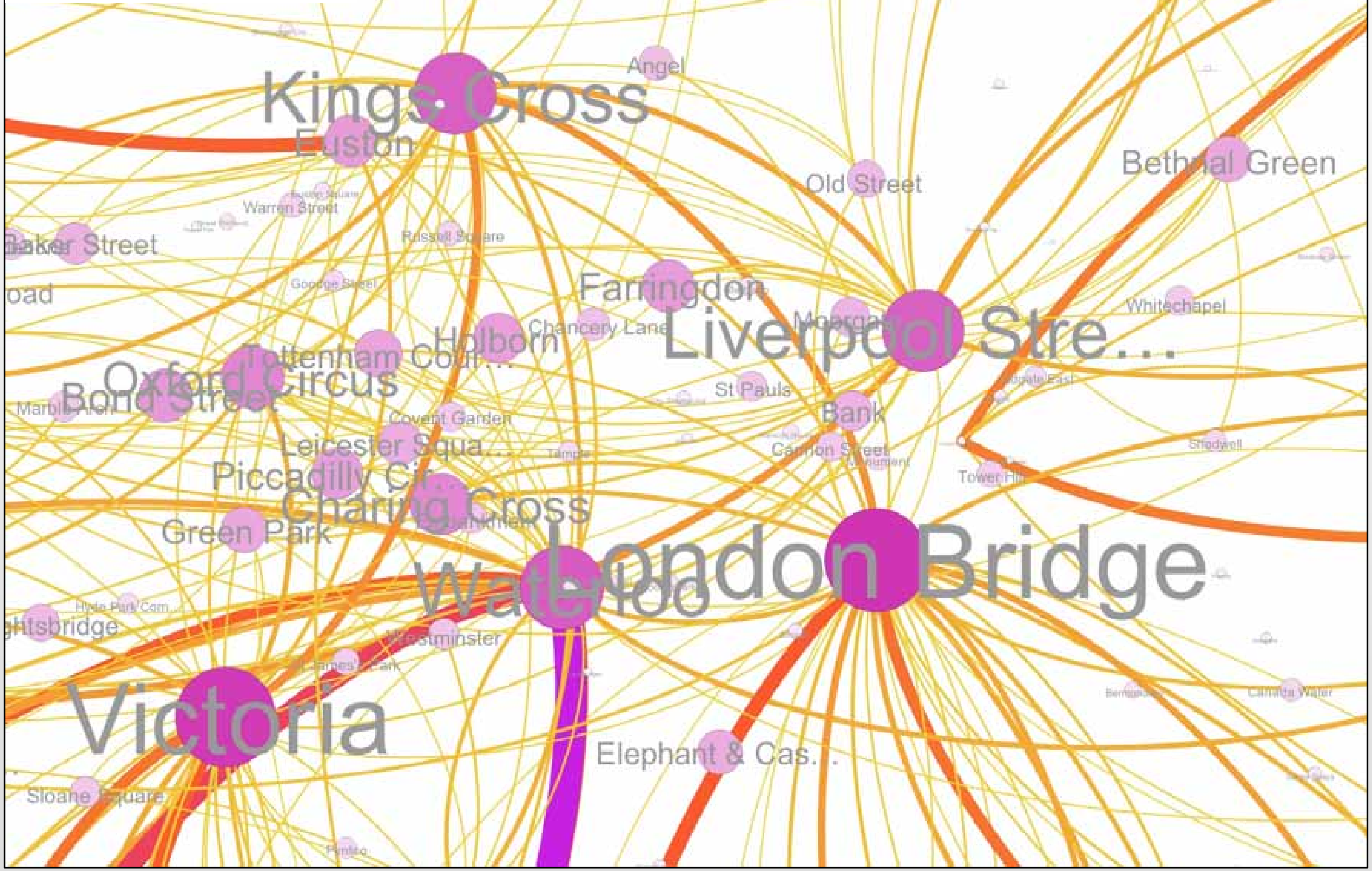
24 Hour day divided into 72 20 minute segments for the counts

666 nodes or hubs where passengers swipe their cards

6.2m individual entries/trips recorded over this day


Hub Volumes at 7am red entries, green exits
Kings X
Shepherd's Bush
London Br
Victoria
Waterloo

Flows at 7am


Kings Cross
Angel
Euston
Old Street
Bethnal Green
Warren Street
Baker Street
Russell Square
Goodge Street
Farringdon
Chancery Lane
Moorgate
Liverpool Stre...
Whitechapel
Holborn
Tottenham Cour...
Oxford Circus
Bond Street
St Pauls
Bank
Covent Garden
Leicester Squa...
Cannon Street
Tower Hill
Shadwell
Piccadilly Cir...
Charing Cross
Green Park
London Bridge
Waterloo
Hyde Park Corn...
Westminster
Victoria
Elephant & Cas...
Canada Water
Sloane Square
Pimlico

In fact, the volumes are tricky – there are 6.24m swipe-in entries and only 5.76m swipe-out exits. There is a leakage of some 480,000 trips due to barriers left open etc.

In fact we will call entries origins and exits destinations in the usual notation but they are not conserved as in a normal transport model

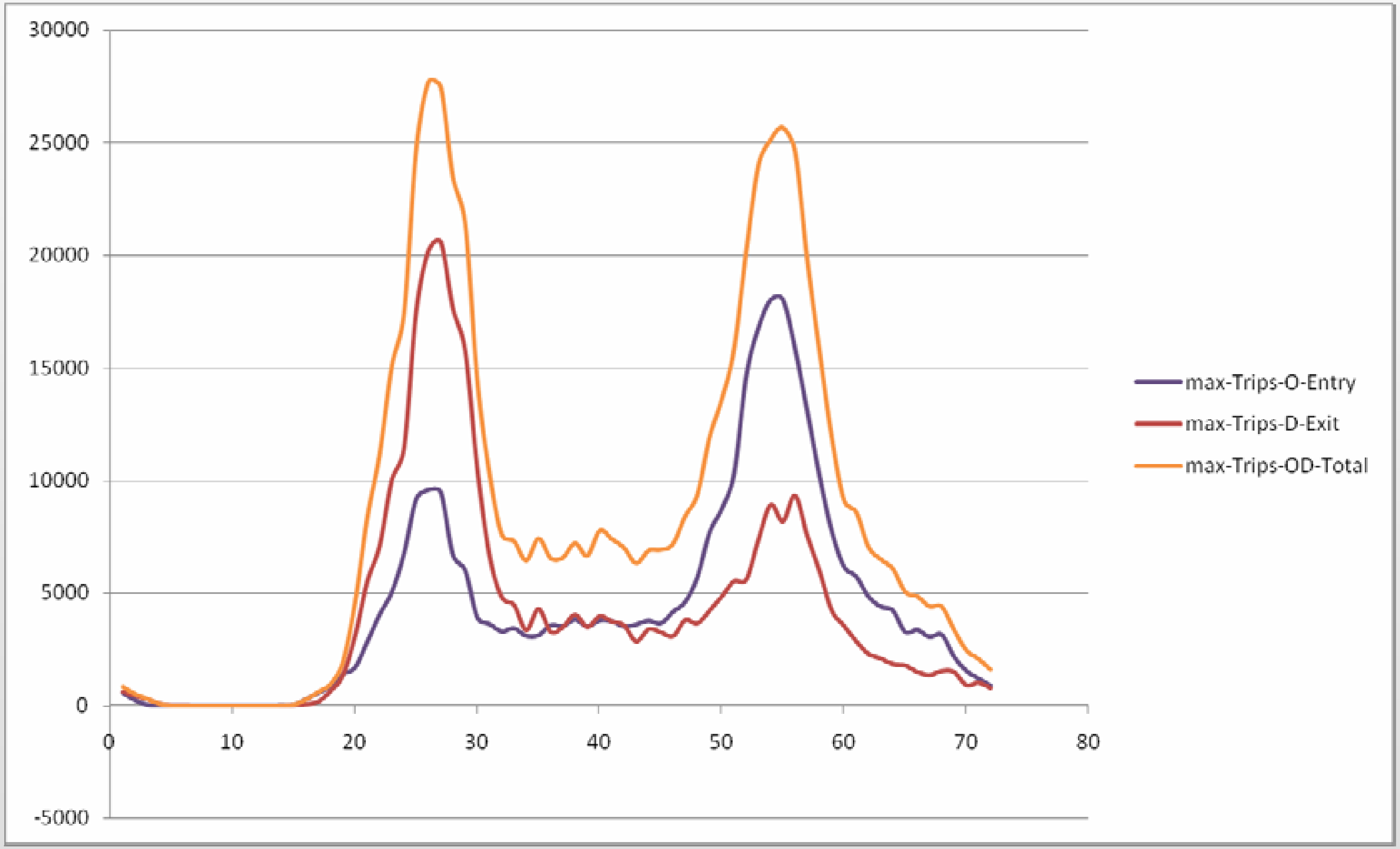
What we intend to do as a first analysis is to examine the profiles of the nodes across all 666 hubs for each of 72 time periods – the frequency distributions as rank size relation, but there are many other analyses yet to do – this is a first shot at the problem

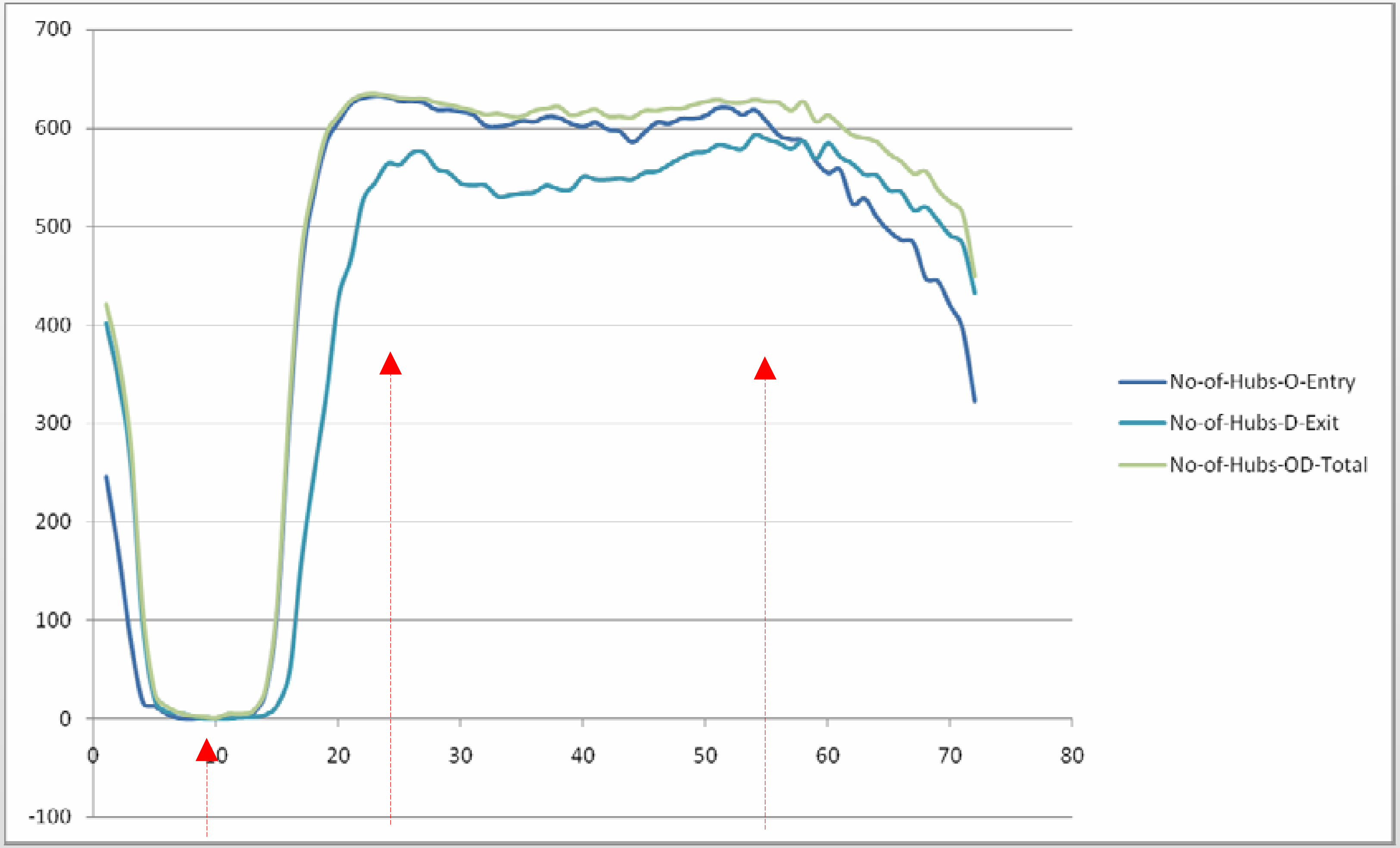
We can examine origins volumes, destination volumes separately and we are doing but in this paper we will simply add these together as total volumes – in this sense they will not have meaning any longer as trips

	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

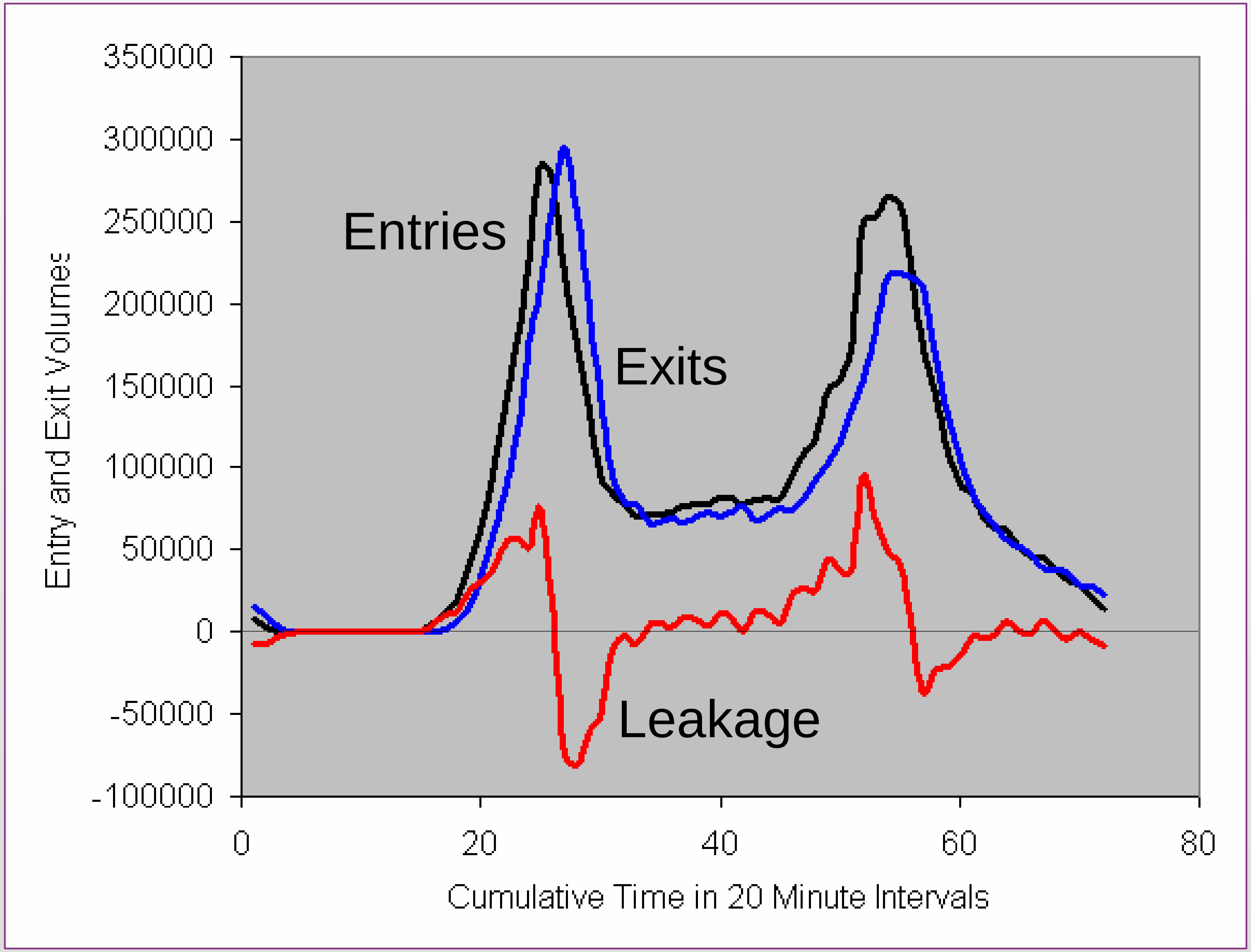
We will now examine the profiles of behaviour during the 24 hour day to provide some sense of the problem

Examining the Dynamics of the Hub Volumes





Night am peak pm peak



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

Formalising the Network and Flow Problem

We first define the trips in the conventional way as between origins k and destinations l which are nodes in the graph

$$O_k = \sum_{\ell} T_{k\ell} \qquad D_\ell = \sum_{k} T_{k\ell}$$

$$T = \sum_{k} O_k = \sum_{\ell} D_\ell = \sum_{k} \sum_{\ell} T_{k\ell}$$

Our problem is different as our nodes/hubs are interchange points not true origins and destinations

This makes a massive differences to the scaling

We call our flows F_{ij} and we define the interchange origins and destinations as (M_i is the leakage)

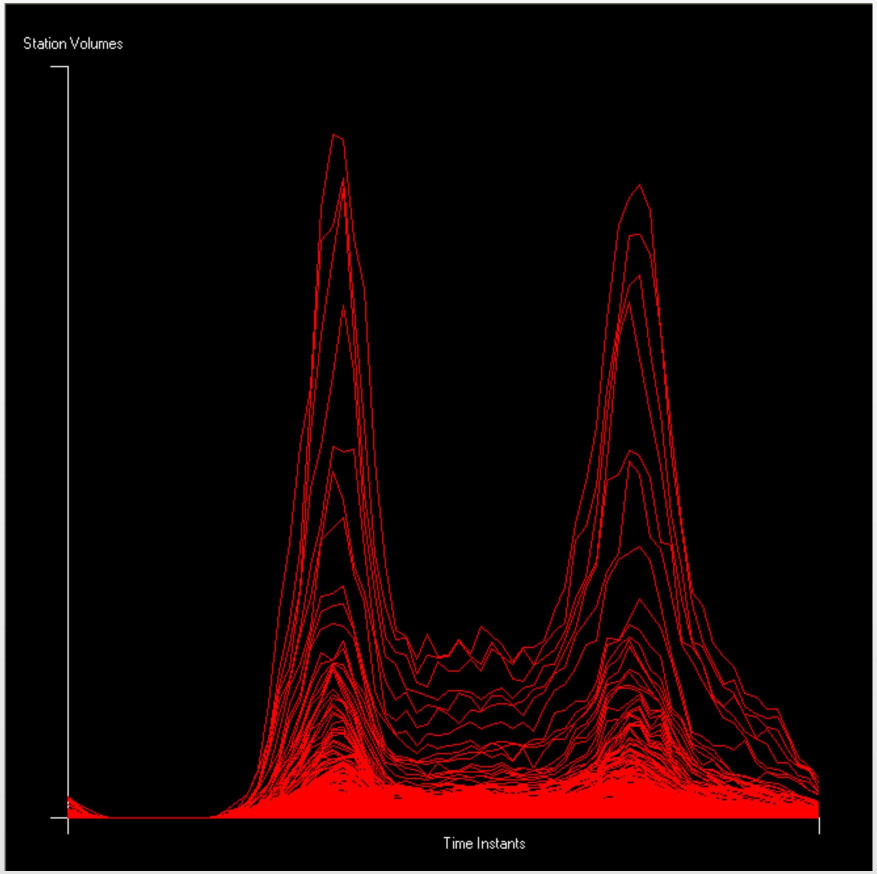
$$O_i = \sum_j F_{ij} \qquad D_j = \sum_i F_{ij}$$

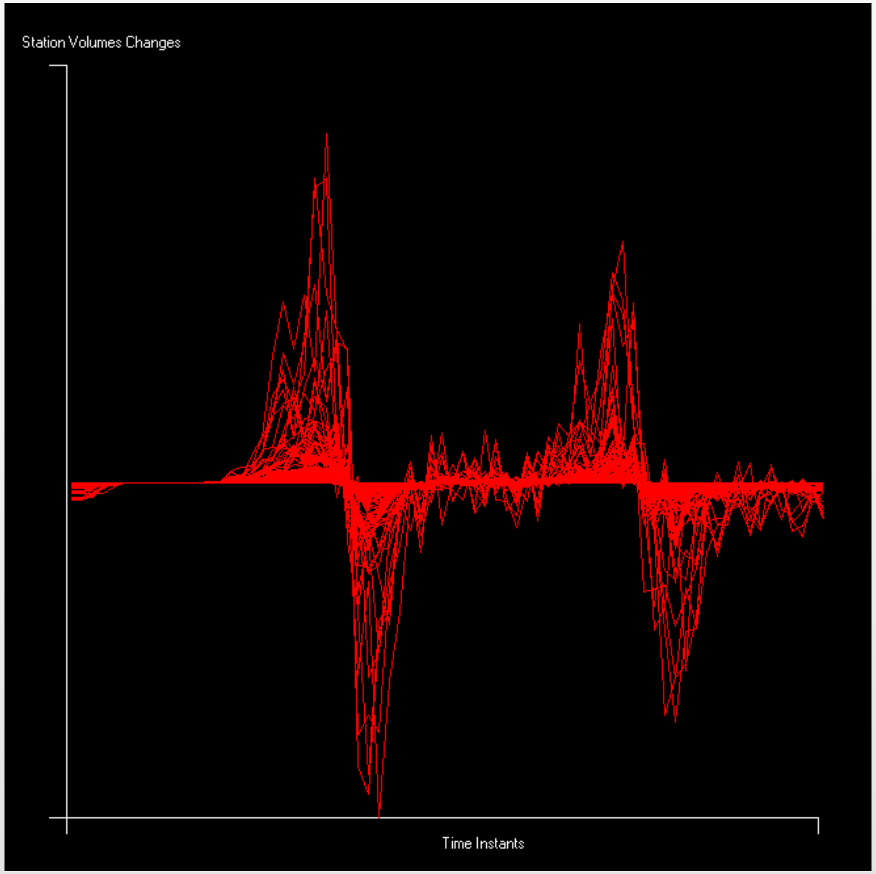
$$\sum_i O_i = \sum_i D_i + \sum_i M_i$$

Where we now define our node volumes as the sum of the origins and destinations

$$S_i = O_i + D_i$$

And we actually define each node in terms of time S_{it} which we show in terms of profiles as

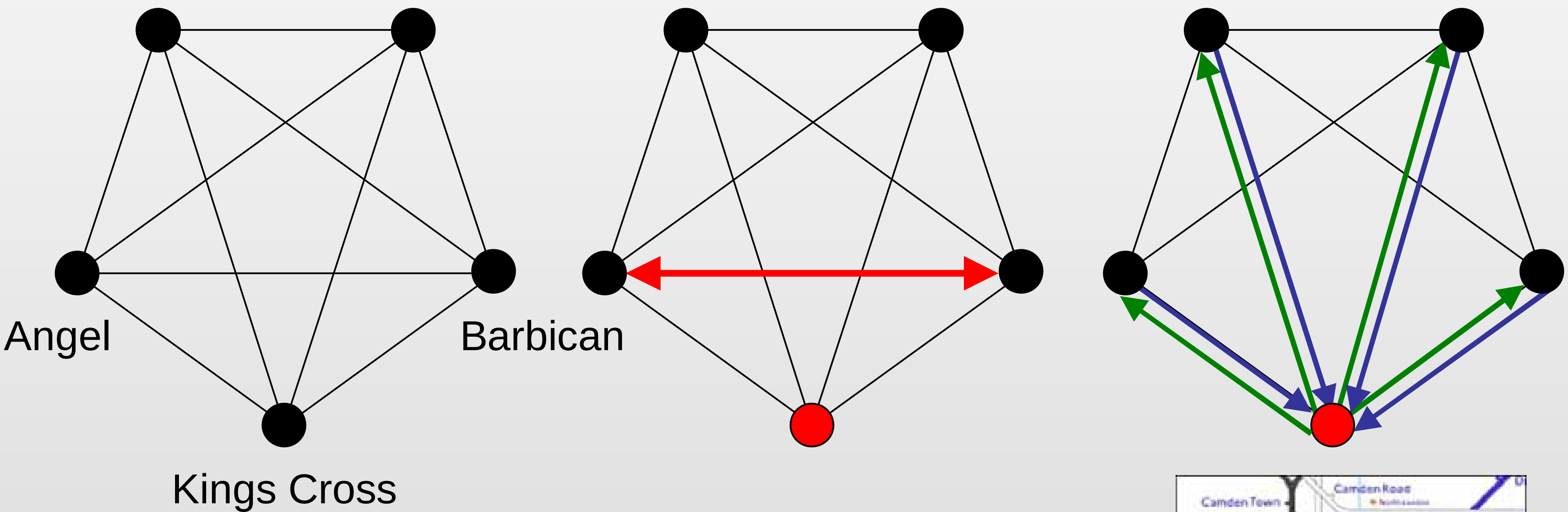




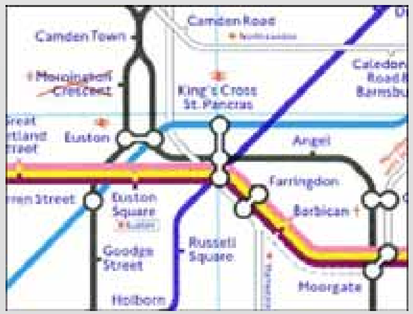
$$S_i(t) = O_i(t) + D_j(t) = \sum_{ij} F_{ij}(t)$$

$$\frac{dS_i(t)}{dt}$$

I need to digress slightly you tell you about what happens at a hub and I will use what I call the Strathclyde diagram to do this



Here is a segment of the real tube map



Exploring the Temporal Scaling Profiles of Node Volumes

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

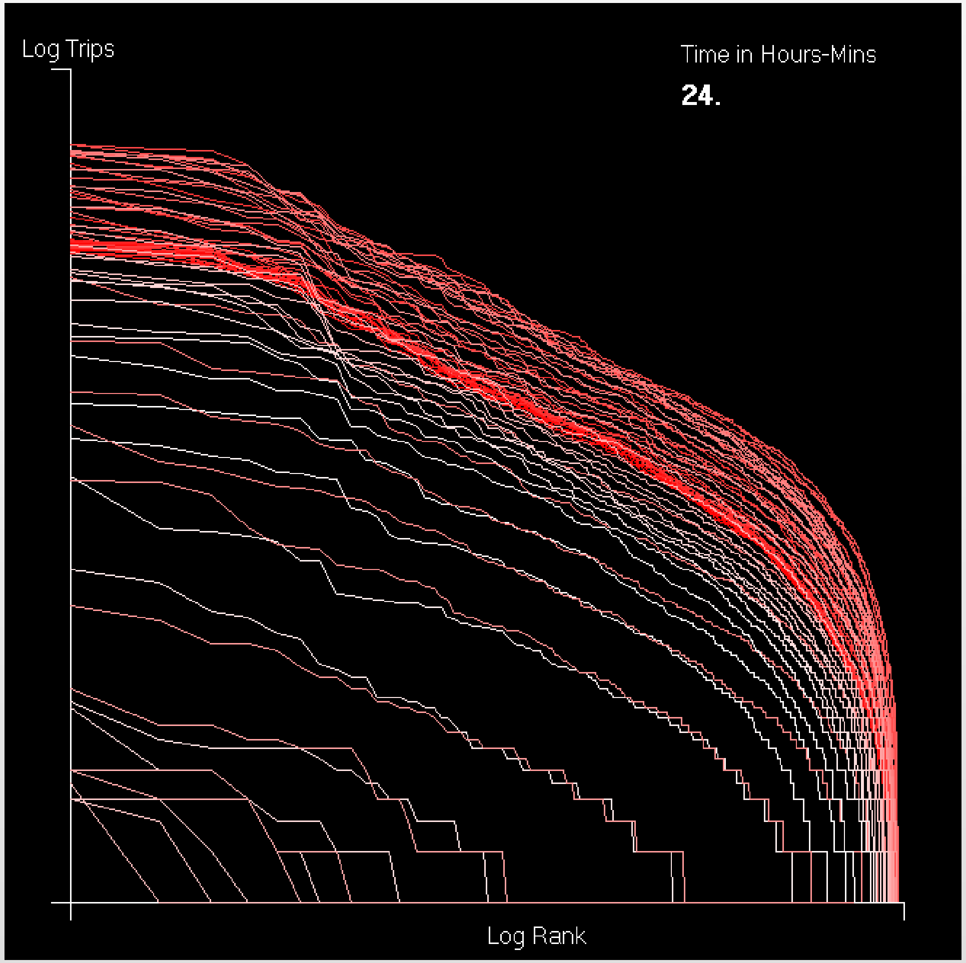
Because of their right skewness, we plot them on log log scales which if they follow power laws – which they don't for obvious constraints on their scaling – would appear as straight lines, as Marc showed

$$S_i(t) \sim \frac{1}{r_i^{\alpha}(t)}$$

All Hub Volumes Ordered as Rank Size Profiles

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

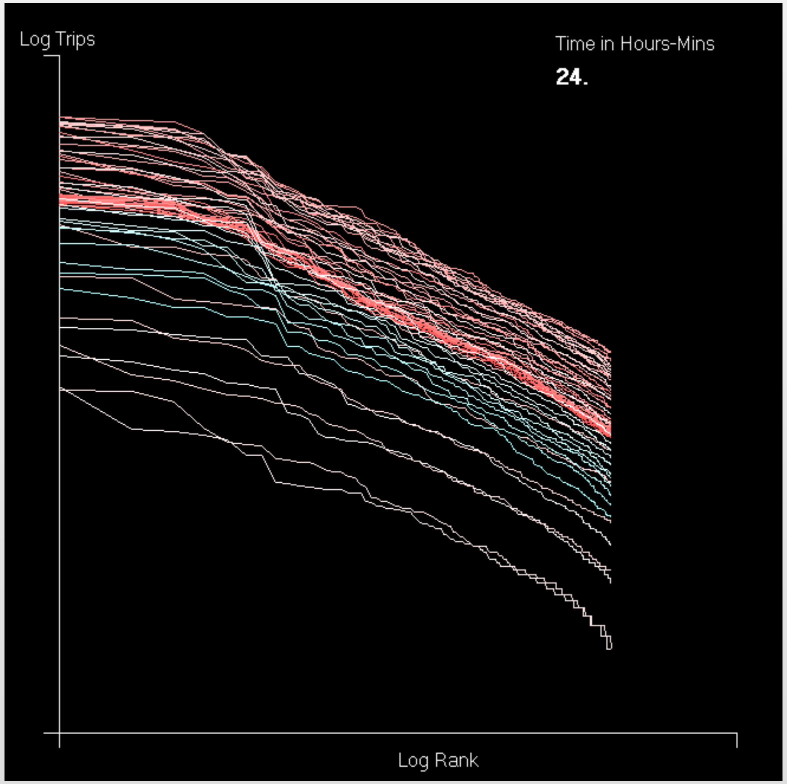
Let me load the program and run it as it is quite short and fast and gives you an idea of the dynamics



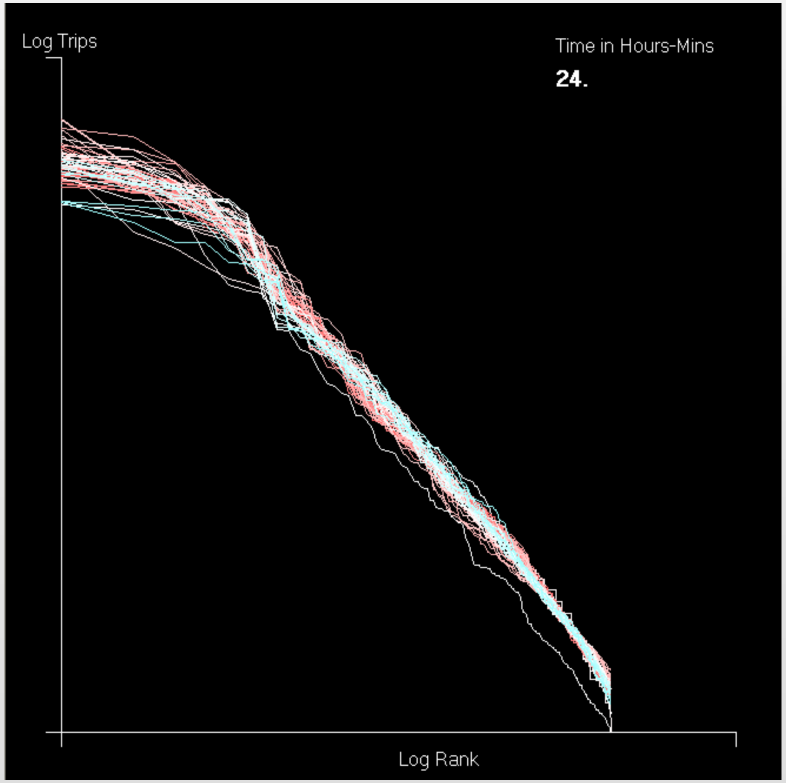
$$R_i(t)$$

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



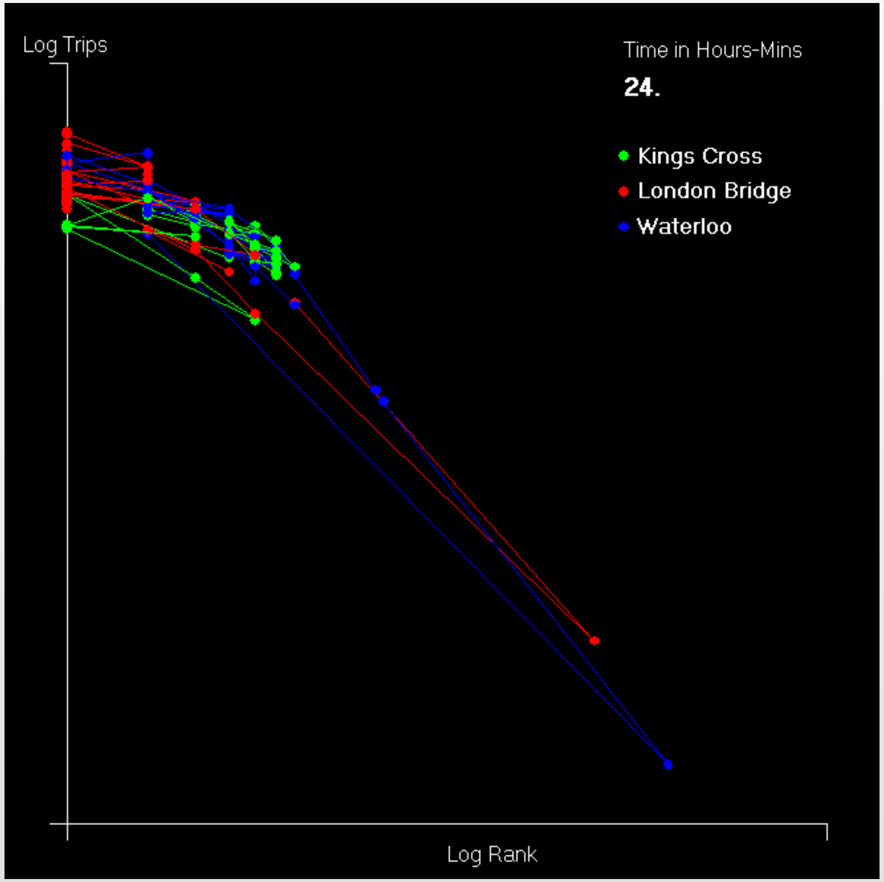
Collapsed/Standardised



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

Project1.exe
CASA-UCL



There are lot of dynamics I can show from the program

Correlations and Comparisons of Profiles

Note that we are not dealing explicitly with flows yet and will not for a while – all our focus is on hubs or nodes. The basic matrix we have is one of space-time – and if we think of this as follows

$$S_{it} = \begin{bmatrix} S_{11} \ldots\ldots\ldots S_{1\tau} \ldots\ldots\ldots \\ S_{21} \ldots\ldots\ldots S_{2r} \ldots\ldots\ldots \\ . \\ . \\ S_{n1} \ldots\ldots\ldots S_{n\tau} \ldots\ldots\ldots \end{bmatrix}$$

We can make comparisons **across space** or **across time**

We can form either of the possible cross correlations over space or time: these are general functionals

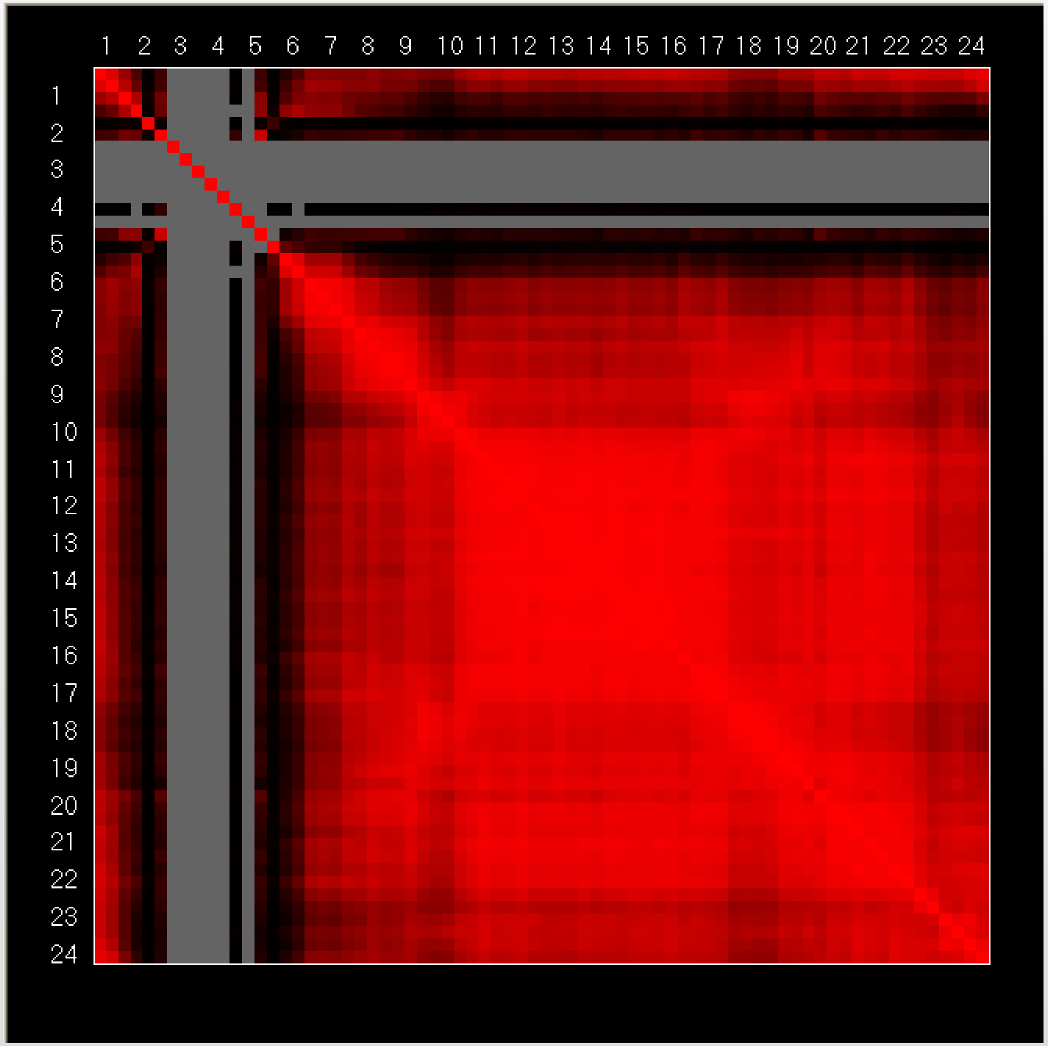
$$C_{ij} = \sum_{t} f(S_{it})g(S_{jt})$$

$$C_{t\tau} = \sum_{i} f(S_{it})g(S_{i\tau})$$

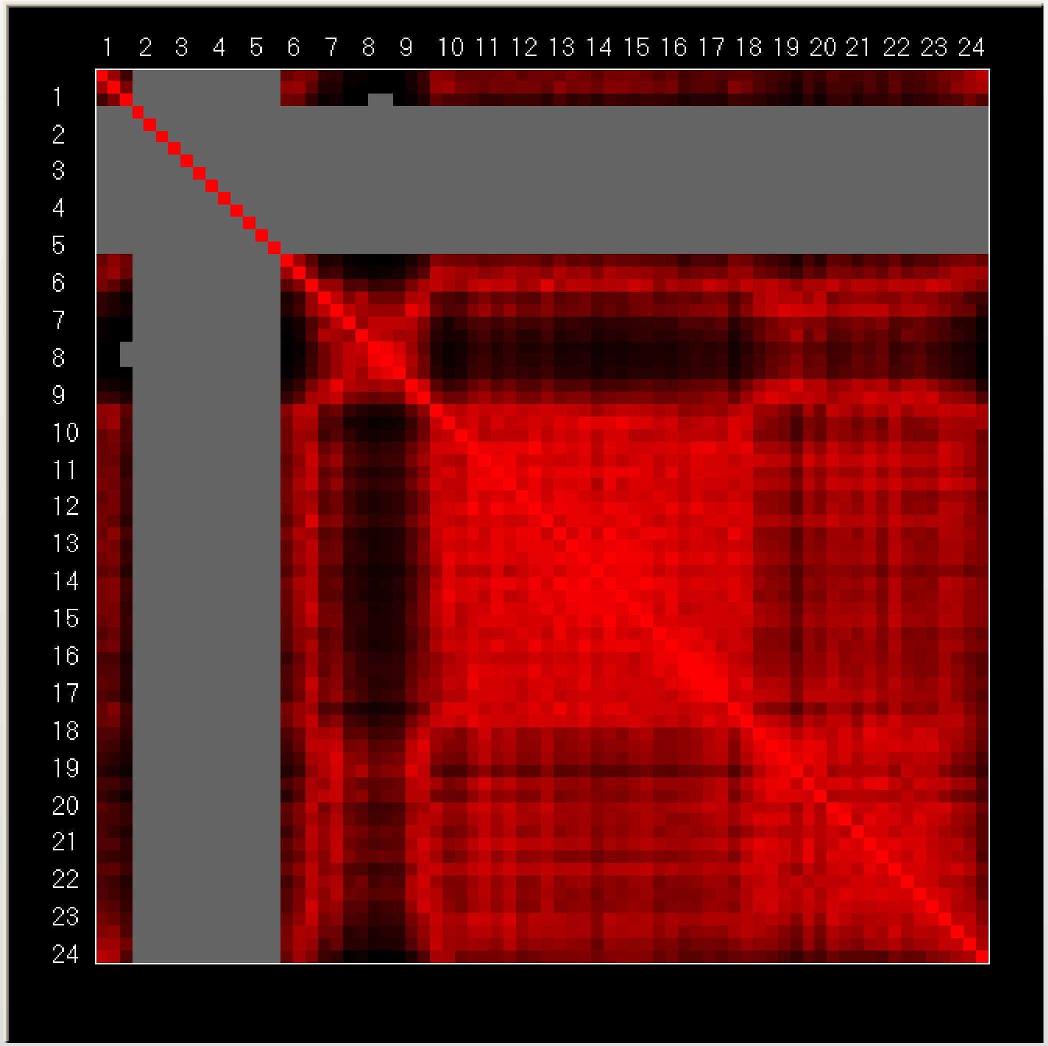
Here we will focus just on time. All we have done is as follows – we have looked at correlations and then we have looked at two types of comparison

We have not clustered these because what we want to see is how close the temporal periods are to one another across the hubs

First the raw correlations of volumes but in log form



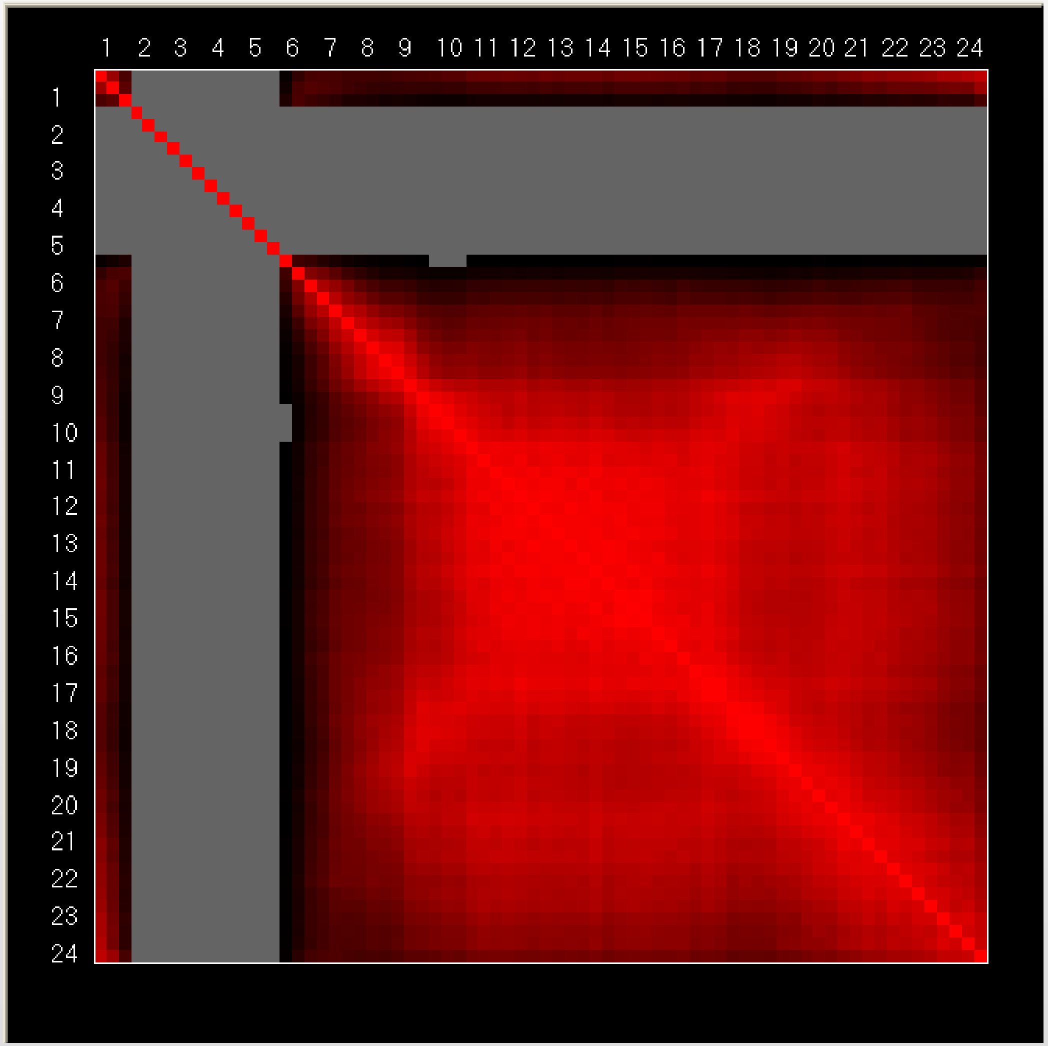
Second ……… the differences $\Omega_{t\tau} = \sum_i \left| S_{r_i(t)} - S_{r_i(\tau)} \right|$



This difference is based on comparing volumes which are at the same rank in different time periods

Third ……… the differentials

$$\Lambda_{t\tau} = \sum_i \left| S_i(t) - S_i(\tau) \right|$$



This differential is based on comparing the volumes at the same hubs but at different ranks in different time periods

Dynamics of Rank-Size Profiles – Plots, Clocks, Graphs

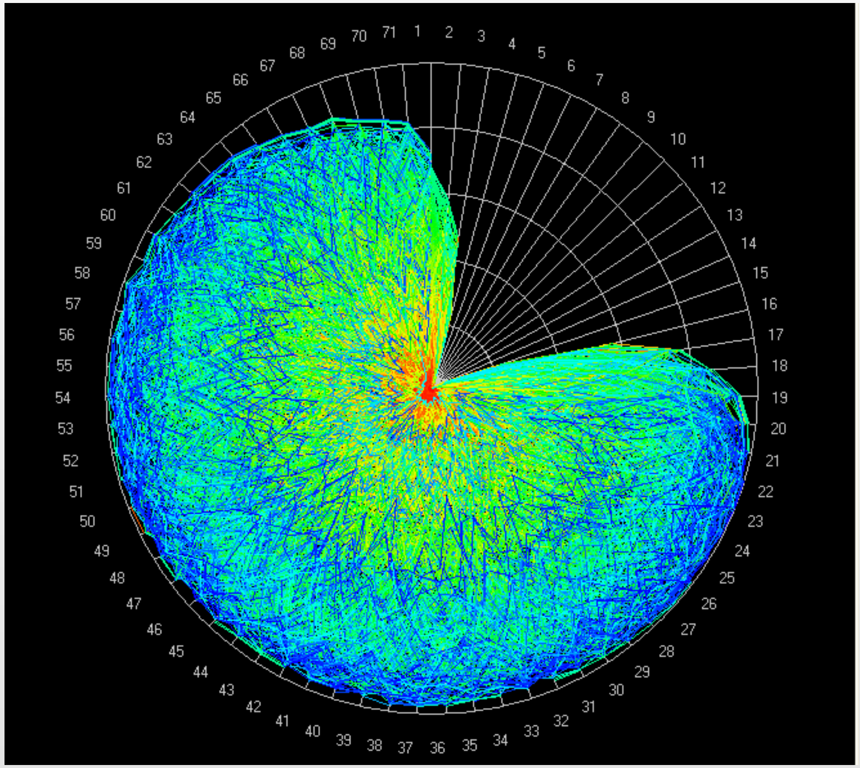
What we are searching for here is the kind of volatility that we see in city systems, firms sizes and so on.

But we can' really find it to the same extent – the profiles are highly consistent in terms of what we are saying so far

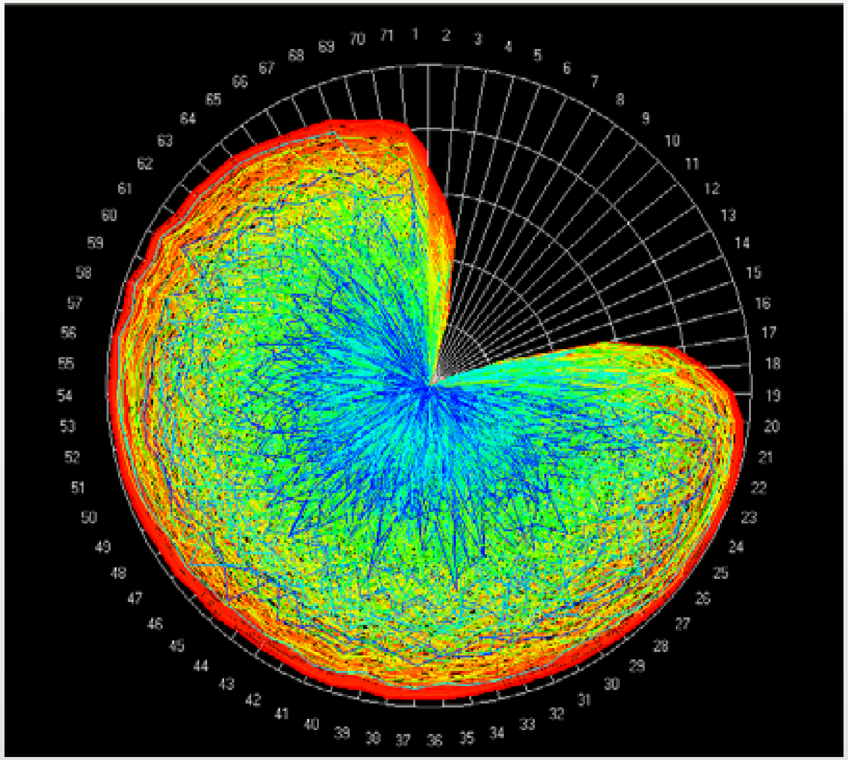
We might expect this from our first analysis although there are differences

However we have plugged all this into our rank clock program and the visualised the dynamics as follows

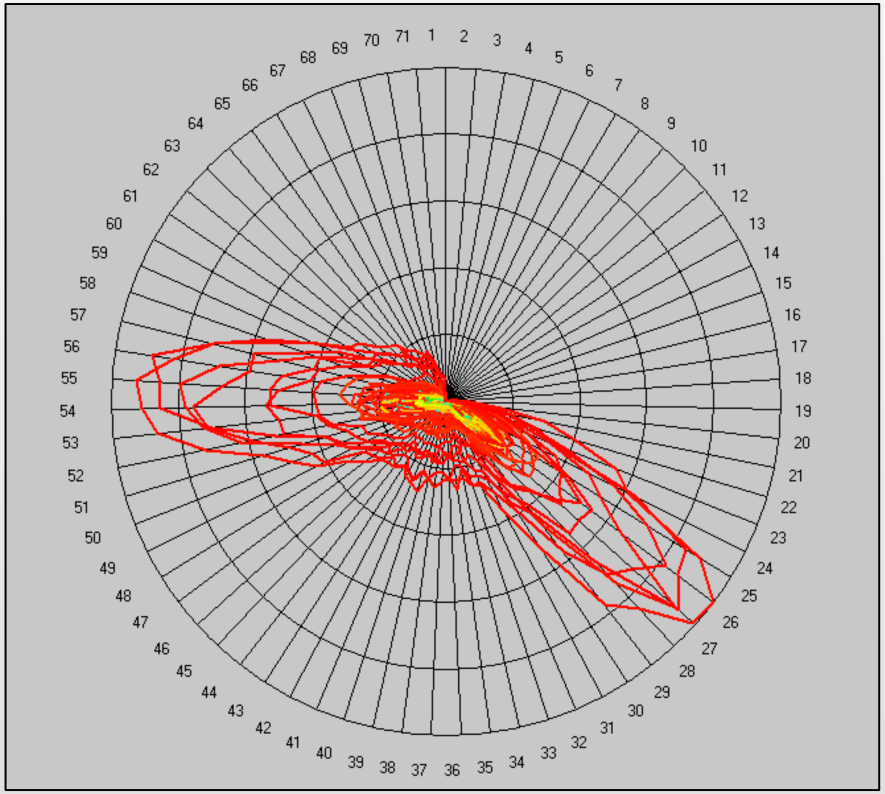
With rank 1 in centre



With rank 1 at edge



The clock not with rank but with volume changes – note the peaks show up rather well



Of course I need to figure out how to present a true clock which is the literal rank clock for 12 not 24 hours !

Future Work: Classification of Hubs

There are many things to do – I asked the question of my group – how should I cluster the cross correlations – they all said MATLAB and I saw a nice demo of this in the meeting – so that is the next step

But the real issue is how we relate this temporal analysis to space – there are some very fascinating pointers in the London data particularly to west London

And there is a lot going on the east too at the Olympics site and other parts of Docklands – watch this space

Acknowledgements and Resources

Thanks to Marc Barthelemy for getting us into the Oyster Card data which came from the lady who used to be at UCL but I have never met and is now at Cambridge via Soong Moon Kang at UCL

And Jon Reades whose contacts with Sensable City and Oyster Card got us the new Oyster Card data set.

Some of these things are on our Simulacra Blog

www.simulacra.blogs.casa.ucl.ac.uk

and my **A Science of Cities** site www.complexCity.info

Papers

OPEN ACCESS Freely available online PLoS one

Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows

Camille Roth[1,2], Soong Moon Kang[3], Michael Batty[4], Marc Barthélemy[1,5*]

1 Centre d'Analyse et de Mathématique Sociales (CAMS), CNRS/EHESS, Paris, France, 2 Institut des Systèmes Complexes de Paris-Ile de France (ISC-PIF), Paris, France, 3 Department of Management Science and Innovation, University College London (UCL), London, United Kingdom, 4 Centre for Advanced Spatial Analysis (CASA), University College London (UCL), London, United Kingdom, 5 Institut de Physique Théorique, CEA, IPhT CNRS, URA 2306, Gif-sur-Yvette, France

Abstract

RESEARCH ARTICLE

Visualizing Space–Time Dynamics in Scaling Systems

MICHAEL BATTY

Centre for Advanced Spatial Analysis (CASA), University College London (UCL), London WC1E 6BT, United Kingdom

Received May 6, 2010; accepted June 27, 2010

nature Vol 444 | 30 November 2006 | doi:10.1038/nature05302

LETTERS

Rank clocks

Michael Batty[1]

Many objects and events, such as cities, firms and internet hubs, scale with size[1–4] in the upper tails of their distributions. Despite intense interest in using power laws to characterize such distributions, most analyses have been concerned with observations at a single instant of time, with little analysis of objects or events that change in size through time (notwithstanding some significant exceptions[5–7]). It is now clear that the evident macro-stability in such distributions at different times can mask a volatile and often turbulent micro-dynamics, in which objects can change their position or rank-order rapidly while their aggregate distribution appears quite stable. Here I introduce a graphical representation

URBAN COMPUTING

Cellular Census: Explorations in Urban Data Collection

Analysis of cell phone use can provide an important new way of looking at the city as a holistic, dynamic system.

pervasive COMPUTING

www.computer.org/pervasive

Cellular Census: Explorations in Urban Data Collection

Jonathan Reades, Francesco Calabrese, Andres Sevtsuk, and Carlo Ratti

Vol. 6, No. 3
July–September 2007