DEPARTMENT OF GEOGRAPHY

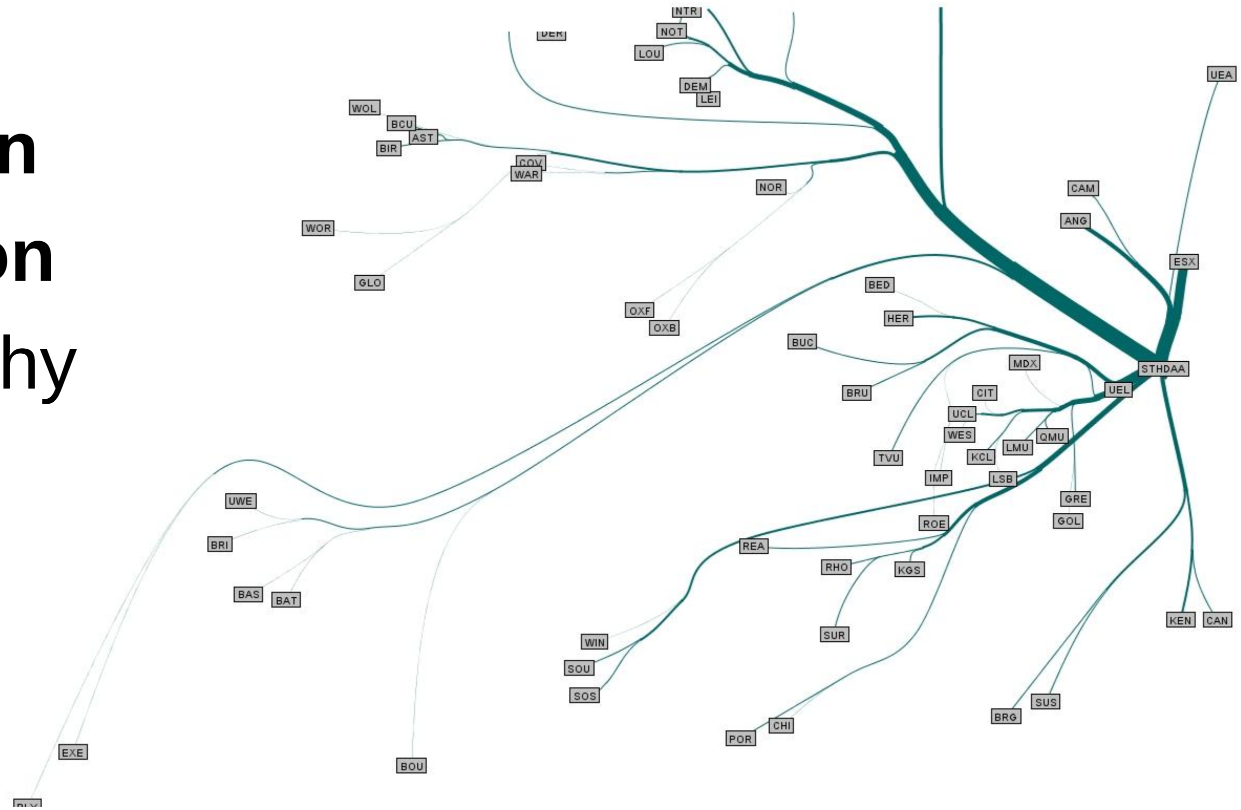
UCL

Spatial Interaction Models for Higher Education

Oliver O'Brien

Alex Singleton

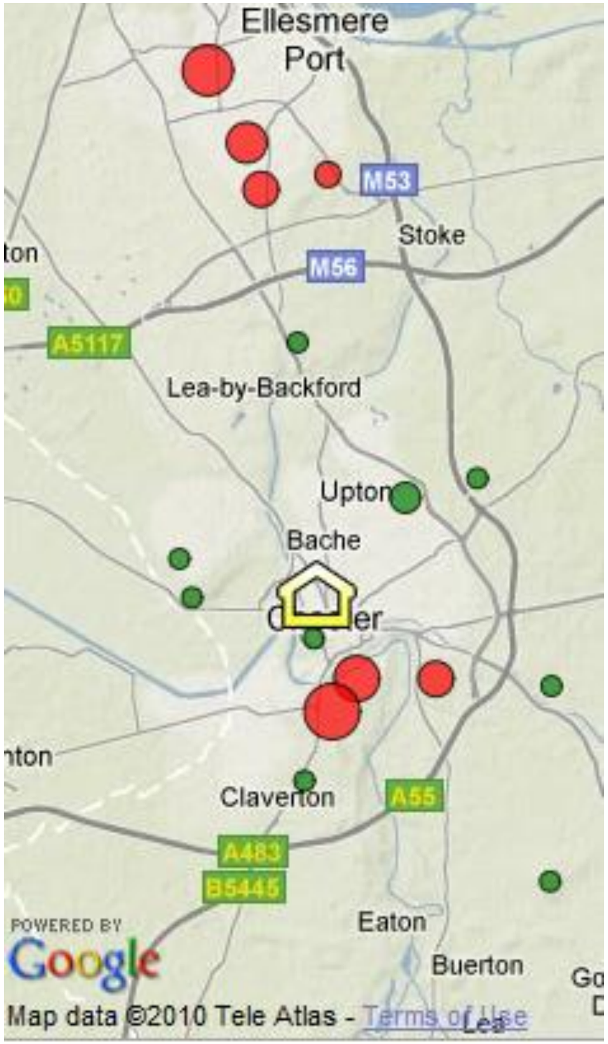
UCL Geography



E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Contents

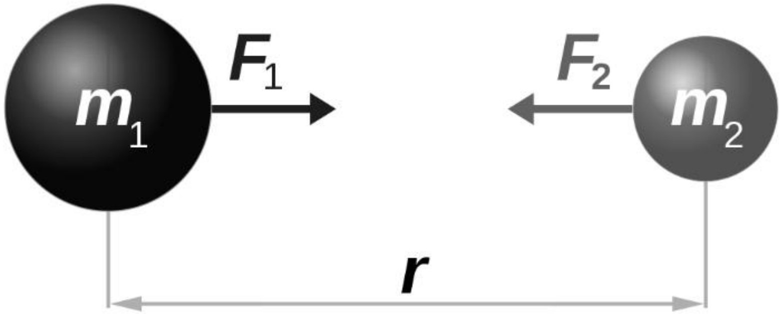
- Theory
  - Spatial Interaction Models
  - Geodemographics
- The Project
  - Putting them together
  - Simplifications
- Results
  - Interesting Anomalous Cases
  - Refinements



Spatial Interaction Models

- Modelling the flows from specific origin(s) to destination(s)
  - Commuting to work
  - Shopping at retail centres
    - *Exploring urban retail phase transitions* (Dearden & Wilson)
  - NHS G.P. Provision
  - Summer holidays

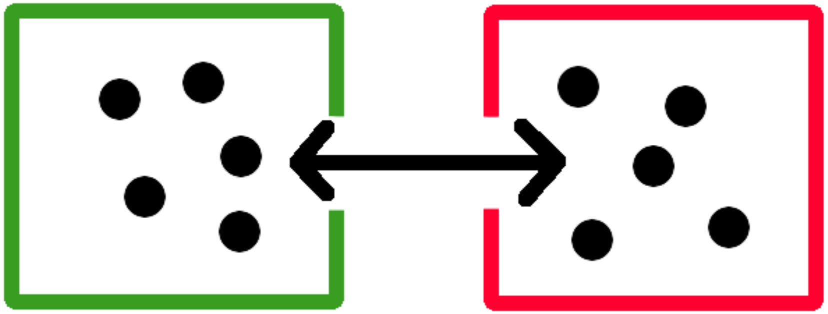
Spatial Interaction Modelling



$$F_1 = F_2 = G\frac{m_1 \times m_2}{r^2}$$

- A classic gravity model
  - Analogous to Newton's Law of Universal Gravitation
- Distance (or cost) decay is always a key component
  - Tobler's "first law of geography"

Spatial Interaction Modelling



- $F_{12} = G\, m_1\, m_2\, r_{12}^{-2}$
- $S_{ij} = k\, O_i\, D_j\, d_{ij}^{-\beta}$ "unconstrained"
- $S_{ij} = A_i\, B_j\, O_i\, D_j\, e^{-\beta c_{ij}}$ "doubly constrained"
  - Can also derive it from entropy-maximising theory
  - A_i depends on B_j which depends on A
    - Solve iteratively

Constraining the Model

- Doubly constrained model
  - A fully closed system
    - e.g World Travel
- Singly constrained model
  - A finite origin population or destination population
    - e.g. Retail - finite number of shoppers, but shopping centre will never want to be “full” and turning them away – particular if capacity is measured in $$$.
- Partially constrained model
  - A combination of the two
    - Some destinations full, others have spare capacity.
    - e.g. NHS doctor’s surgeries in a local authority.

Spatial Interaction Modelling for Higher Education

- The flows are from schools and F.E. colleges to universities
- Timescales are "different"
  - Flow is normally termly or one-way rather than daily or weekly
- Distances are "different"
  - Often intercity rather than intracity
- Distance is less important
  - Going to the "right" university is important for most people

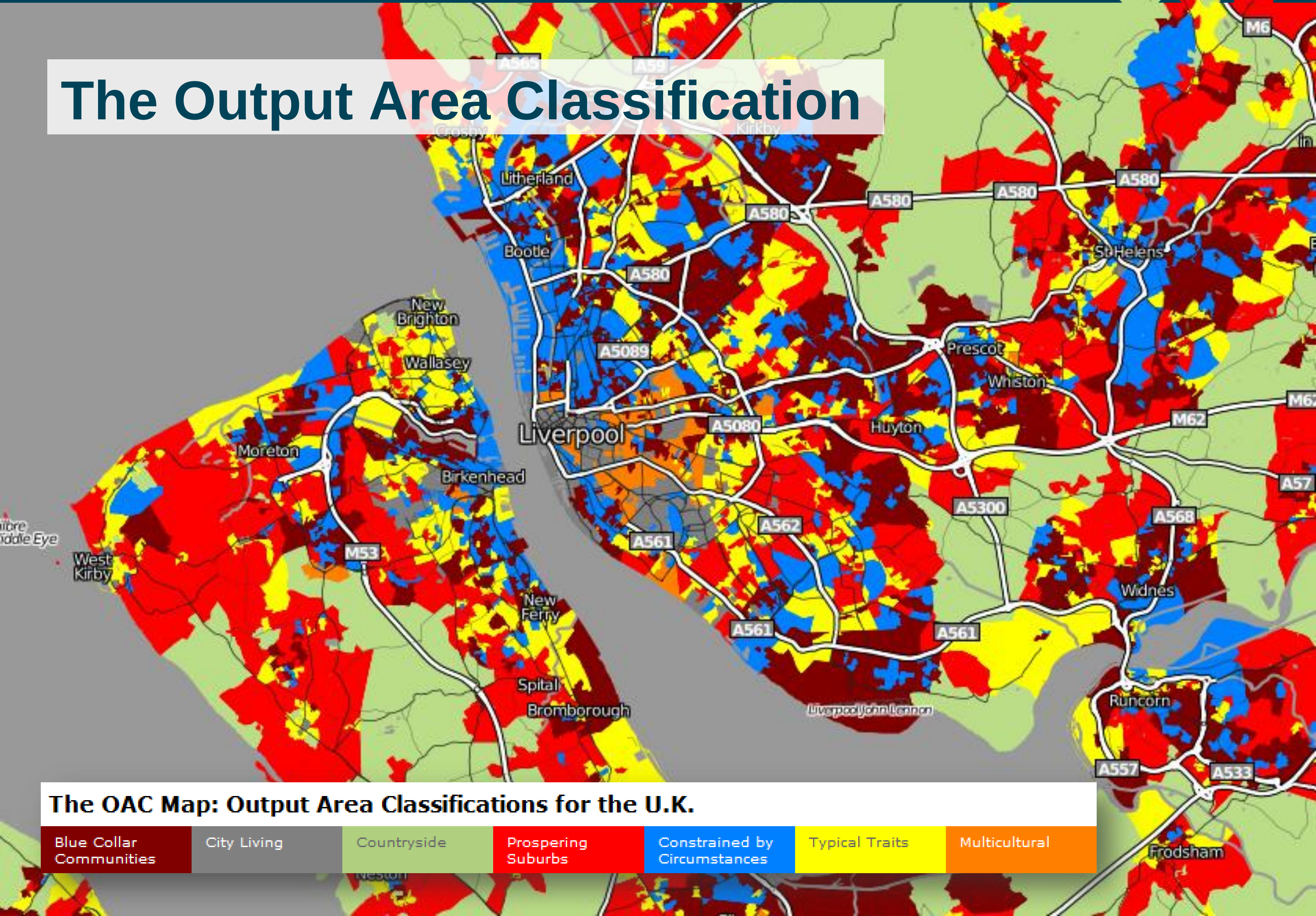
Partially Constrained Model

- Appropriate for modelling flows to higher education
  - More school pupils than university places but not every course at every university is fully subscribed
  - Have both “Selective” and “Recruiting” universities
  - Universities have quotas rather than operating in a fully unconstrained market
  - Many more universities have become selective recently
- Can treat singly-constrained and doubly-constrained flows separately
  - mark each flow appropriately in each iteration during the model run as the destinations “fill up”

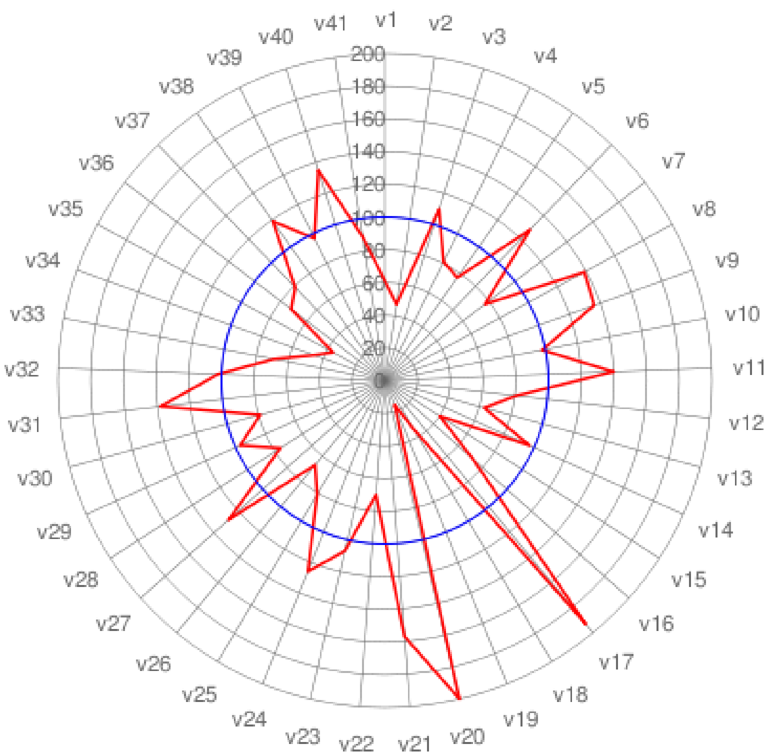
Geodemographics

- Demographic characteristics (age, ethnicity, housing type, occupation, marital status, facilities)
- Interested in how geodemographics affect the patterns of university choice
- Using the Output Area Classification (Vickers)
  - Generalised (not education specific)
  - Available for each output area (typically 10 postcodes)
- Other UK geodemographic classifications
  - Mosaic (by Experian), Acorn

The Output Area Classification



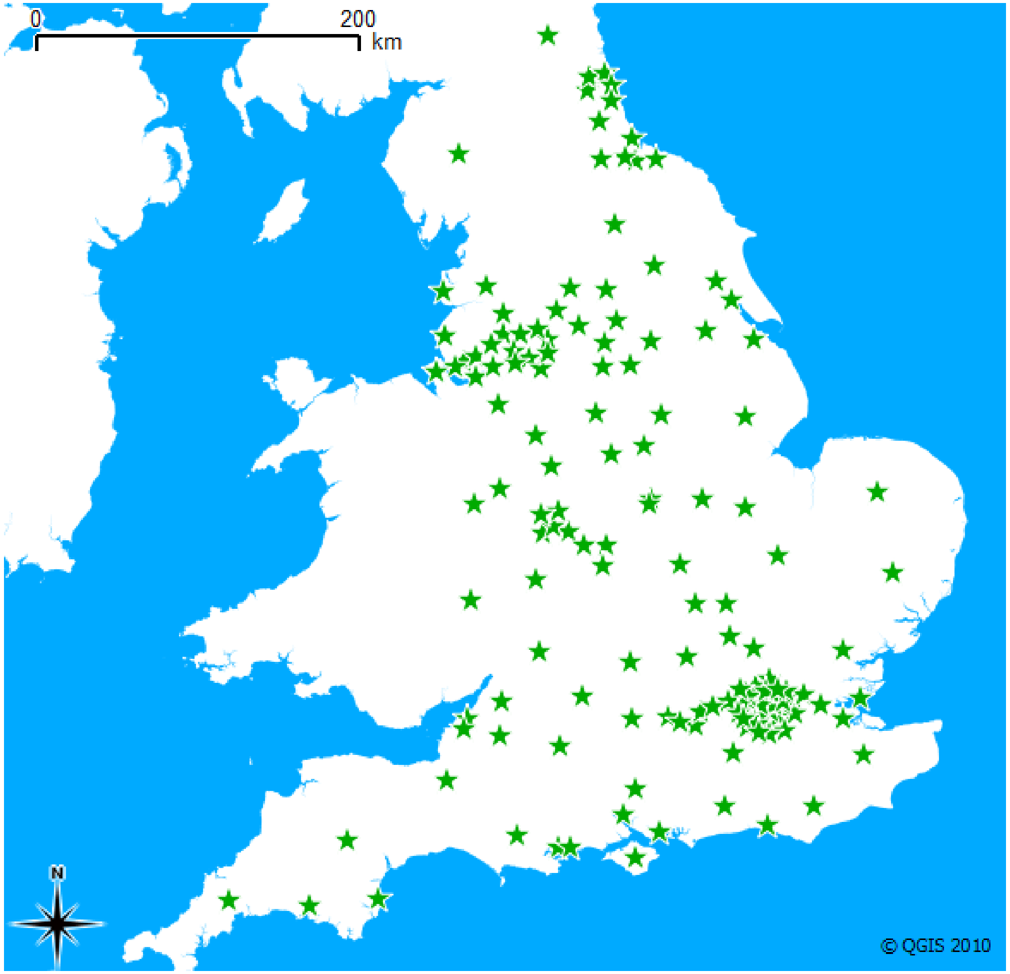
Output Area Classification 2A1 "City Living – Settled in the City 1"



- v1 Age 0-4
- v2 Age 5-14
- v3 Age 25-44
- v4 Age 45-64
- v5 Age 65+
- v6 Indian, Pakistani or Bangladeshi
- v7 Black african, Black Caribbean or Other Black
- v8 Born Outside the UK
- v9 Population Density
- v10 Divorced
- v11 Single person household (not pensioner)
- v12 Single pensioner household
- v13 Lone Parent household
- v14 Two adults no children
- v15 Households with non-dependant children
- v16 Rent (Public)
- v17 Rent (Private)
- v18 Terraced Housing
- v19 Detached Housing
- v20 All Flats
- v21 No central heating

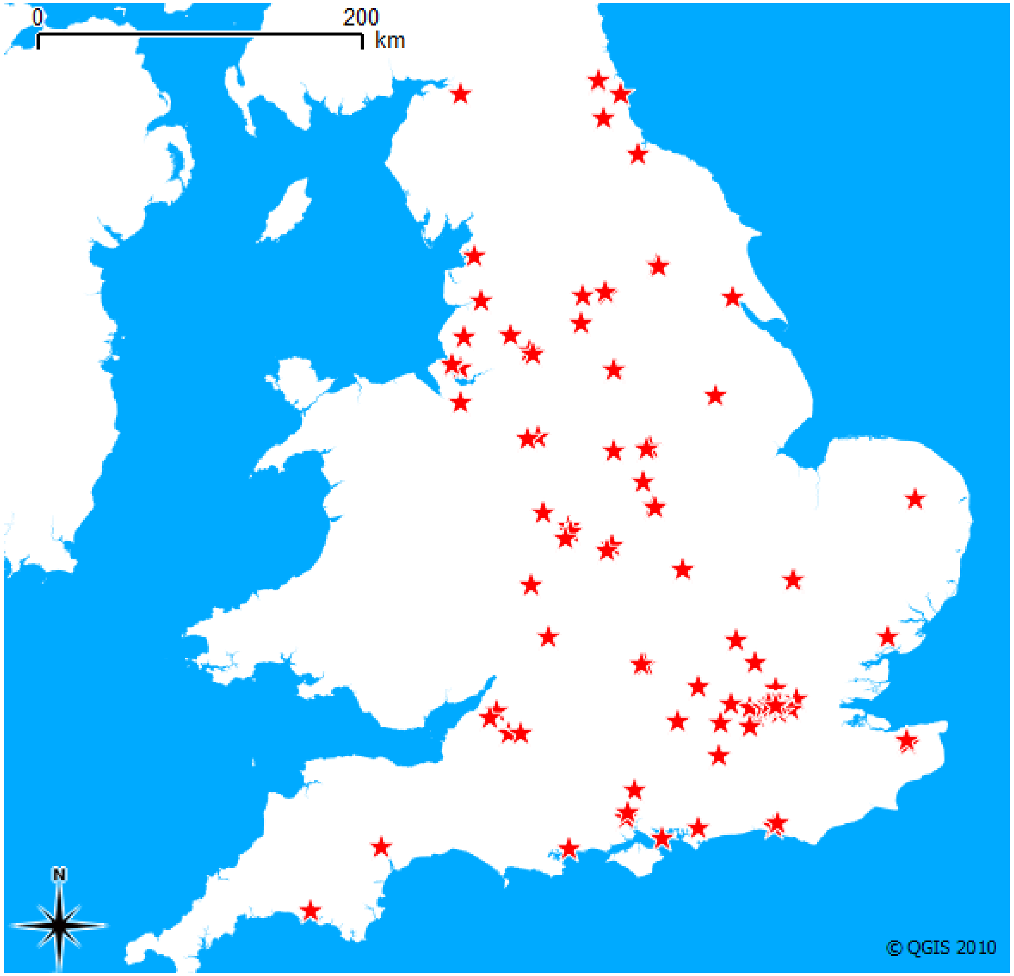
The Data – Origin Side

- National Pupil Database (NPD)
  - Home OAs (state only)
  - Used school OA for private schools
  - Includes attainment
- Individual Learning Records (ILR)
  - For sixth-form colleges
  - Home postcodes
  - Includes attainment
- OAC



The Data – Destination Side

- HESA Individual Student Records
  - Subjects
  - Home postcodes
  - A-Level point score
  - Nearly everything needed for modelling the flows, but excludes those who didn't go to university
  - Crucially, no theoretical capacity information



A Great Model – Modelling Reality

- Paper by Wilson (2002)
- $\mathbf{S_{ij} = A_i^{km}\, e_i^{km}\, P_i^{k}\, (W_j^{mh})^{\alpha^{km}}\, exp(-\beta^{km}\, c_{ij}^{k})}$
- This is the singly-constrained form
  - Finite number of school students go to university
  - No restriction on places at university
  - Doubly-constrained version is quite similar to look at
- W is the “attractiveness” of the institution

A Great Model – Modelling Reality

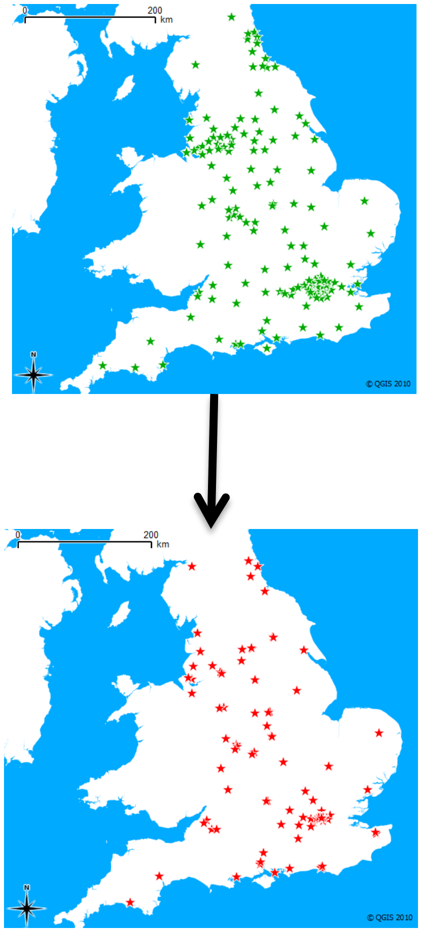
- 150 universities
- 3000 secondary schools + 500 F.E. Institutions
- 10 UCAS principal subject topics
  - e.g. Axxx – Medicine & Dentistry
- Multitude of possible attainments
  - A Level points scores, vocational qualifications, IB
  - Attainments are a useful additional factor for attractiveness

A Good Model – Simplifications

- In order to produce meaningful data on (relatively) small numbers (~300,000 annually) of students
  - use coarse categories
  - streamline the variables used
- Otherwise, the results would be a massive matrix with almost every value a fraction of a single person

A Good Model – Spatial Simplifications

- Assume universities are single-site
  - Generally using the "administrative HQ"
  - Some universities are fairly equally split
    - e.g. Angla Ruskin in Cambridge,Chelmsford
  - Ignore the Open University
- Assume English closed system
  - English schools and English universities only
- Make distance proportional to travel cost
- Assume schools and F.E. Institutions are a single institution at their LA's centroid
- **149 "super schools"**



A Good Model – Origin Simplifications

- Ignore school types
- For pupils without postcode information assume the pupil’s geodemographic is the same as the school’s
- Assume pupils don’t go to schools in a different local authority to that they live in
- Binary classification of attainment – “good”/”bad”
  - Based on A-level or equivalent points
- **2 attainment types**

A Good Model – Origin Simplifications

- Use the seven geodemographic “supergroups” from the Output Area Classification
  - Be aware of possible correlations between geodemographic and other factors included seperately in the model, such as attainment
  - Very different overall numbers (and proportions) of each demographic go to universities
- **7 demographics**

A Good Model – Destination Simplifications

- Ignore subjects
  - Assume all universities offer all subjects and admissions criteria does not differ
  - But some universities are selective for some subjects (e.g. Medicine) and recruiting for some subjects (e.g. Physics)
  - The nearest few universities to someone may not offer the subjects that the person wants to study
- Ignore universities with a specialist subject focus
  - University for the Creative Arts
  - London School of Economics
  - These are also generally “small” universities
- **89 universities, 1 “subject”**

A Good Model – Destination Simplifications

- Binary classification of attainment requirement
  - "good only"
  - "any"

- Account for students not going to university by a special catch-all "university of last resort"
  - No "distance" element
  - Adjust attractiveness of this university to see the relative popularity of the other universities in the model

A Good Model – Destination Simplifications

- Attractiveness
  - Very subjective – different people like different things
  - Was originally modelled as a university “type”
    - Ancient, 19^{th} century, Red brick, Plate glass, Post-1992
    - Funding type: Big research-focused institution & hospital, big research-focused, big teaching-based, small teaching
  - But difficult to categorise type and its relative effect on each of the origin geodemographics
  - Using Times Higher Education Score (range 200-1000)
    - Factor to modify its influence if necessary
- Attractiveness becomes less important and location more important, as more of the flows become doubly constrained (i.e. more universities fill to capacity)

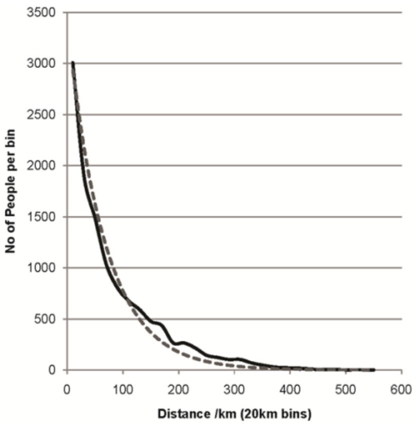
Simplified Form

- From: $\mathbf{S_{ij} = A_i^{km}\, e_i^{km}\, P_i^{k}\, (W_j^{mh})^{\alpha^{km}} \exp(-\beta^{km}\, c_{ij}^{k})}$
- To: $\mathbf{S_{ij} = A_i^{k}\, P_i^{k}\, (W_j^{h})^{\alpha} \exp(-\beta^{k}\, d_{ij})}$
  - No subject consideration
  - No “demand” factor
  - Cost is replaced by distance
  - Numbers of i, j locations greatly reduced
  - Attractiveness is not dependent on geodemographic
- Similar for the doubly-constrained version

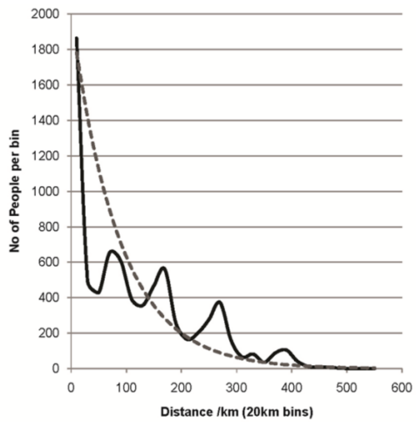
Calibrating the Model

- Find values for the constants in the equations
- $\mathbf{A_i}$ & $\mathbf{B_j}$ values are "balancing constants"
  - they converge on the correct values during iteration
- Calculate the $\boldsymbol{\beta}^k$ distance-decay with known flows
  - Overall distance decay for all pupils
  - Break down by geodemographic
    - Very unequal numbers within each geodemographic
  - Compare distance decay functions

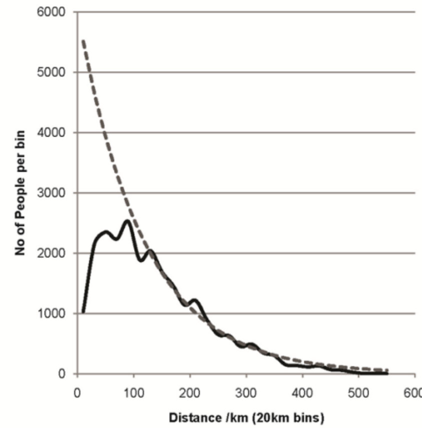
Calibrating the Model – Beta Decay



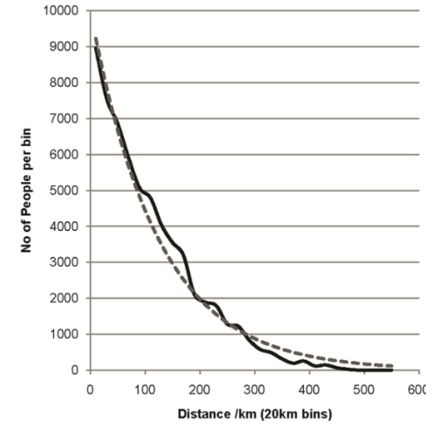
Blue Collar Communities



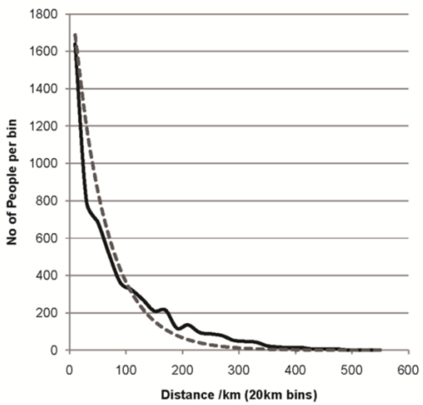
City Living



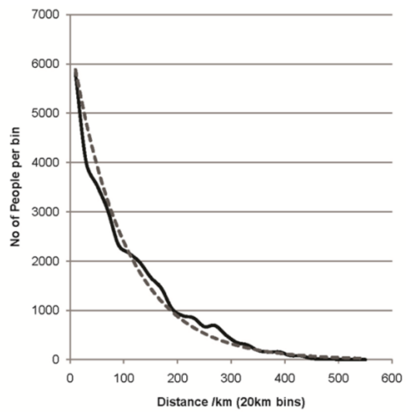
Countryside



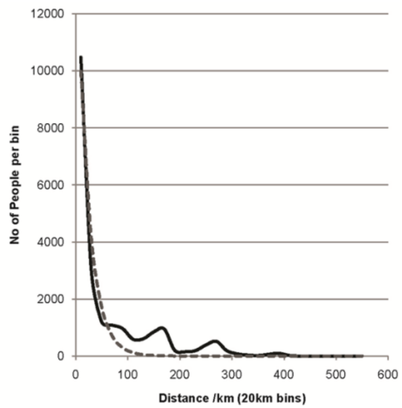
Prospering Suburbs



Constrained by Circumstances

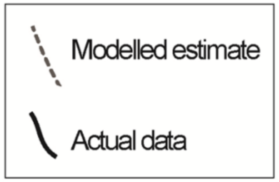


Typical Traits



Multicultural

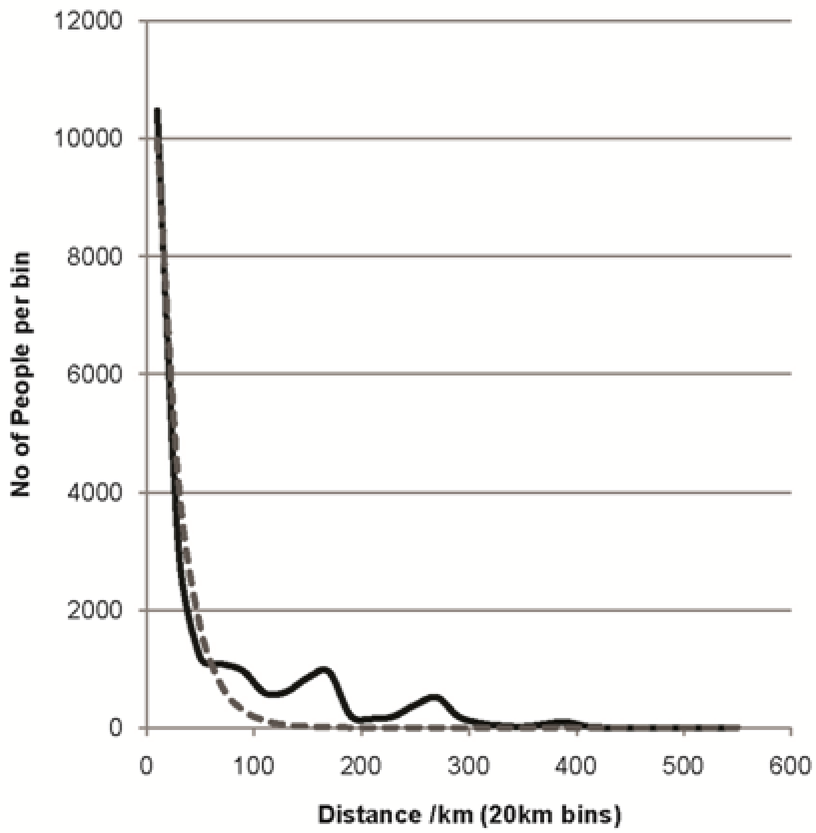
Legend



Calibrating the Model – Beta Decay

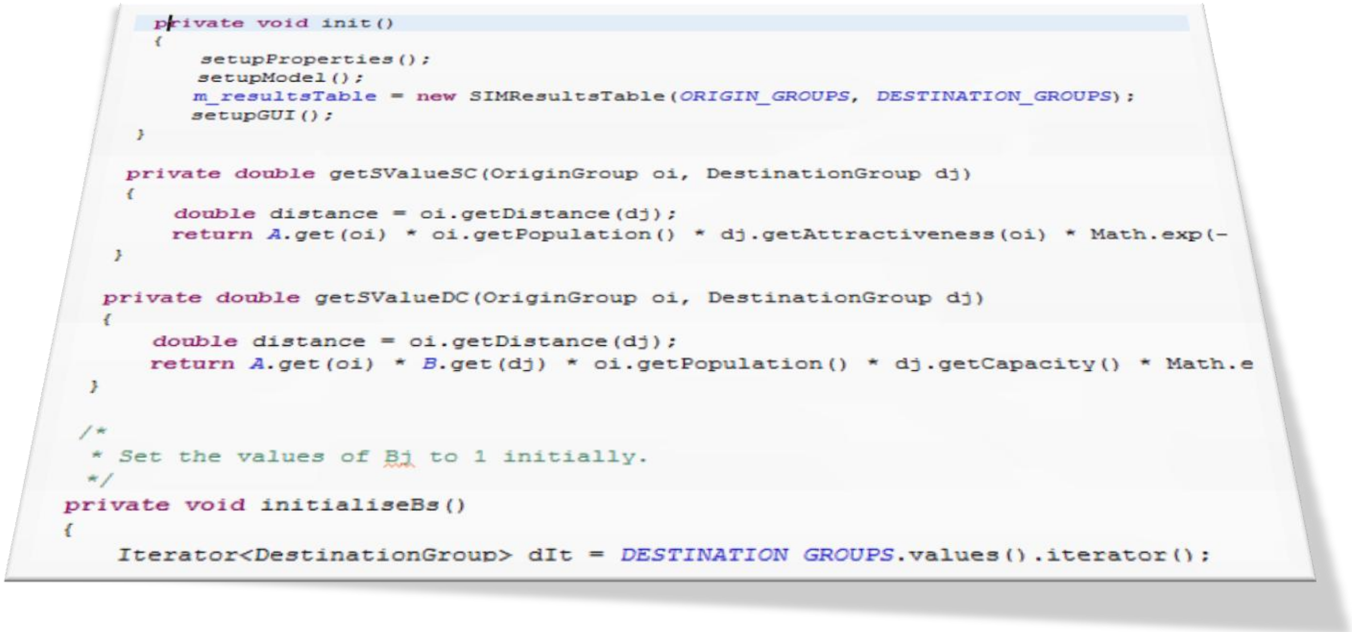
- London to Birmingham: 160 km
- London to Manchester: 260 km

- Distinctive pattern seen for the City Living & Multicultural demographics



Modelling

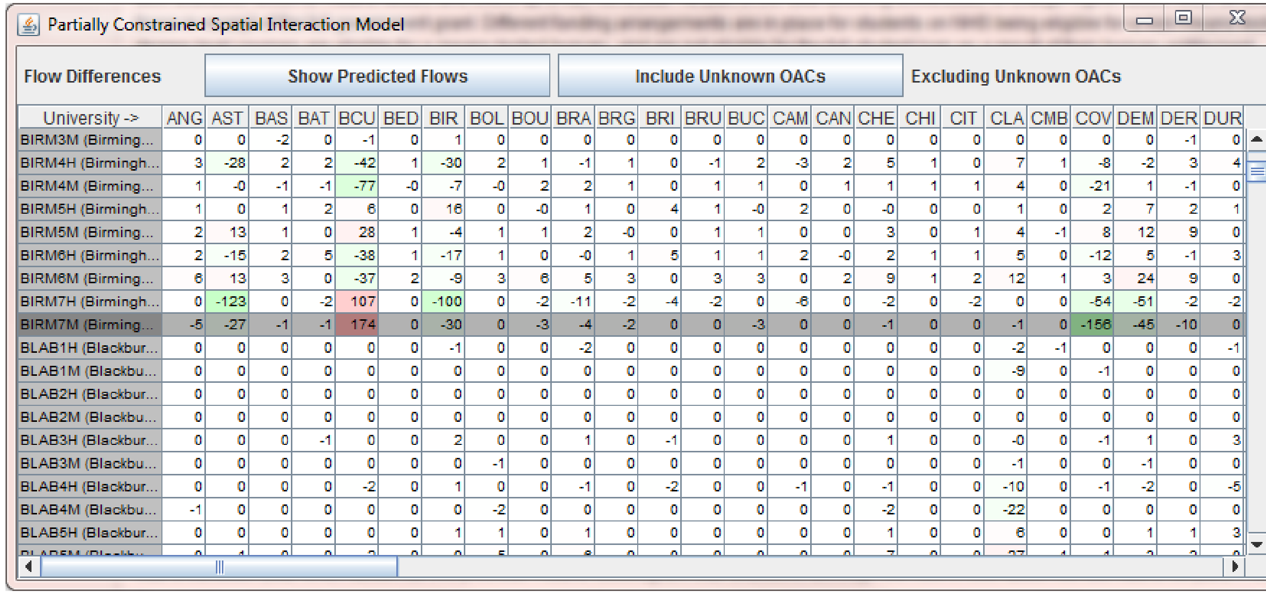
- Java



```
private void init()
{
    setupProperties();
    setupModel();
    m_resultsTable = new SIMResultsTable(ORIGIN_GROUPS, DESTINATION_GROUPS);
    setupGUI();
}

private double getSValueSC(OriginGroup oi, DestinationGroup dj)
{
    double distance = oi.getDistance(dj);
    return A.get(oi) * oi.getPopulation() * dj.getAttractiveness(oi) * Math.exp(-
}

private double getSValueDC(OriginGroup oi, DestinationGroup dj)
{
    double distance = oi.getDistance(dj);
    return A.get(oi) * B.get(dj) * oi.getPopulation() * dj.getCapacity() * Math.e
}

/*
 * Set the values of Bj to 1 initially.
 */
private void initialiseBs()
{
    Iterator<DestinationGroup> dIt = DESTINATION_GROUPS.values().iterator();
```

- Iterative process to calculate the normalising constants which depend on each other
  - Typically takes a minute to calculate the results

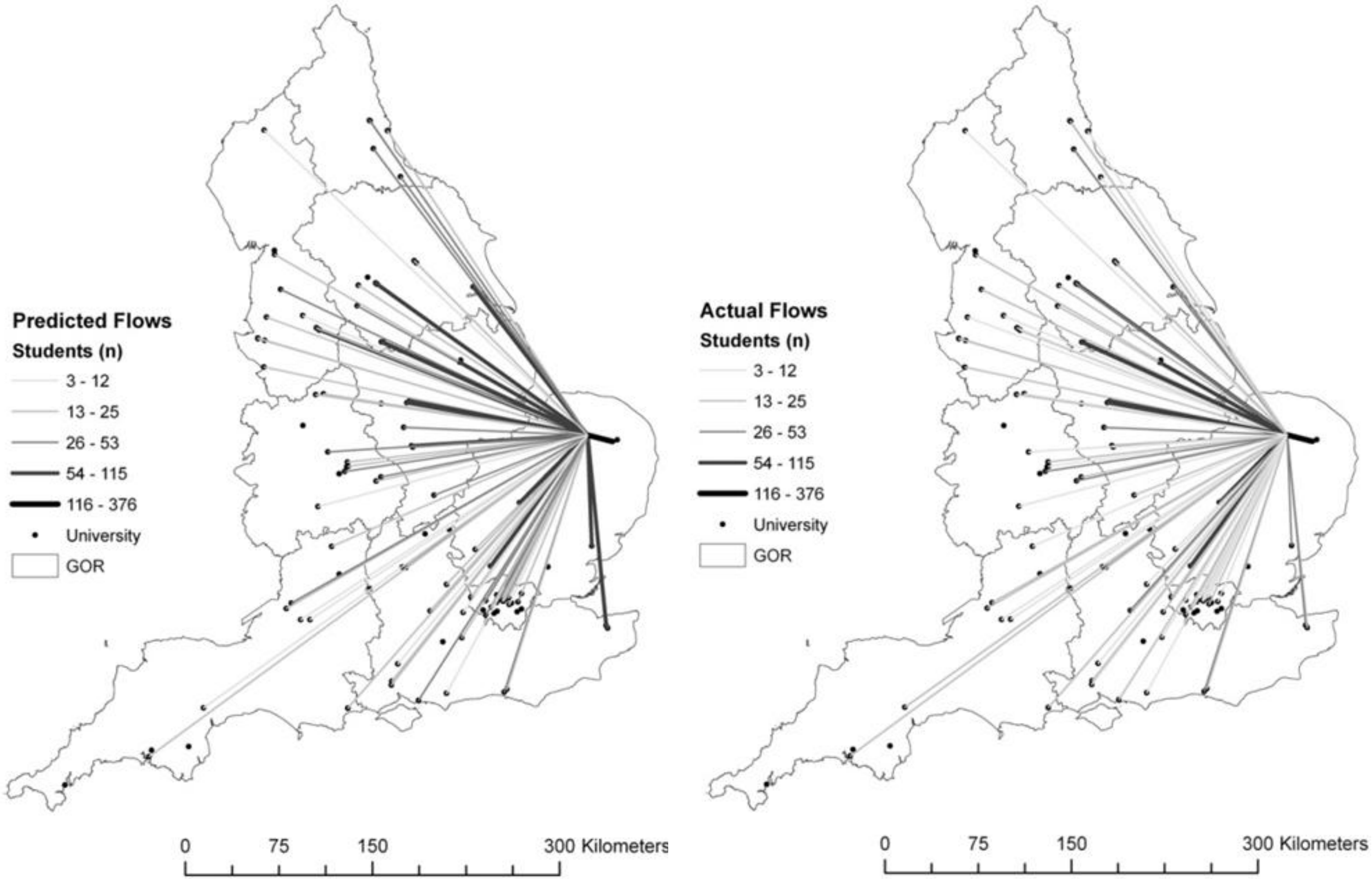
Results



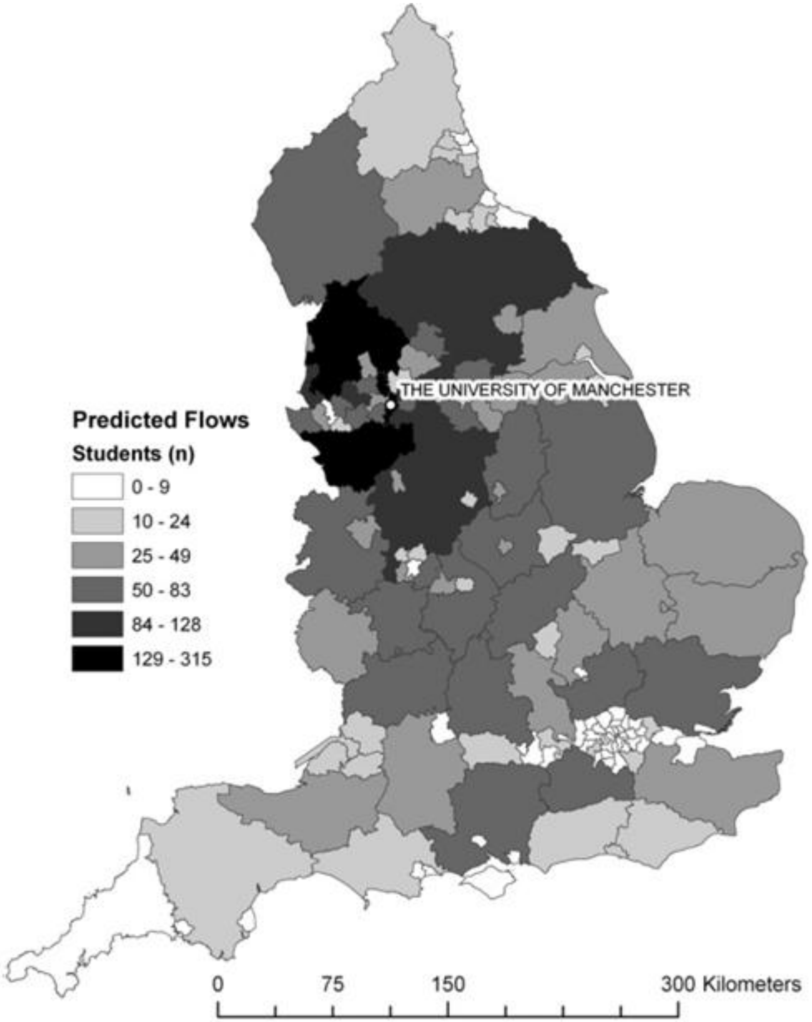
University ->	ANG	AST	BAS	BAT	BCU	BED	BIR	BOL	BOU	BRA	BRG	BRI	BRU	BUC	CAM	CAN	CHE	CHI	CIT	CLA	CMB	COV	DEM	DER	DUR
BIRM3M (Birming...	0	0	-2	0	-1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0
BIRM4H (Birmingh...	3	-28	2	2	-42	1	-30	2	1	-1	1	0	-1	2	-3	2	5	1	0	7	1	-8	-2	3	4
BIRM4M (Birming...	1	-0	-1	-1	-77	-0	-7	-0	2	2	1	0	1	1	0	1	1	1	1	4	0	-21	1	-1	0
BIRM5H (Birmingh...	1	0	1	2	6	0	16	0	-0	1	0	4	1	-0	2	0	-0	0	0	1	0	2	7	2	1
BIRM5M (Birming...	2	13	1	0	28	1	-4	1	1	2	-0	0	1	1	0	0	3	0	1	4	-1	8	12	9	0
BIRM6H (Birmingh...	2	-15	2	5	-38	1	-17	1	0	-0	1	5	1	1	2	-0	2	1	1	5	0	-12	5	-1	3
BIRM6M (Birming...	6	13	3	0	-37	2	-9	3	6	5	3	0	3	3	0	2	9	1	2	12	1	3	24	9	0
BIRM7H (Birmingh...	0	-123	0	-2	107	0	-100	0	-2	-11	-2	-4	-2	0	-6	0	-2	0	-2	0	0	-54	-51	-2	-2
BIRM7M (Birming...	-5	-27	-1	-1	174	0	-30	0	-3	-4	-2	0	0	-3	0	0	-1	0	0	-1	0	-156	-45	-10	0
BLAB1H (Blackbur...	0	0	0	0	0	0	-1	0	0	-2	0	0	0	0	0	0	0	0	0	-2	-1	0	0	0	-1
BLAB1M (Blackbu...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-9	0	-1	0	0	0
BLAB2H (Blackbur...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BLAB2M (Blackbu...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BLAB3H (Blackbur...	0	0	0	-1	0	0	2	0	0	1	0	-1	0	0	0	0	1	0	0	-0	0	-1	1	0	3
BLAB3M (Blackbu...	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	-1	0	0
BLAB4H (Blackbur...	0	0	0	0	-2	0	1	0	0	-1	0	-2	0	0	-1	0	-1	0	0	-10	0	-1	-2	0	-5
BLAB4M (Blackbu...	-1	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	0	-2	0	0	-22	0	0	0	0	0
BLAB5H (Blackbur...	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	6	0	0	1	1	3

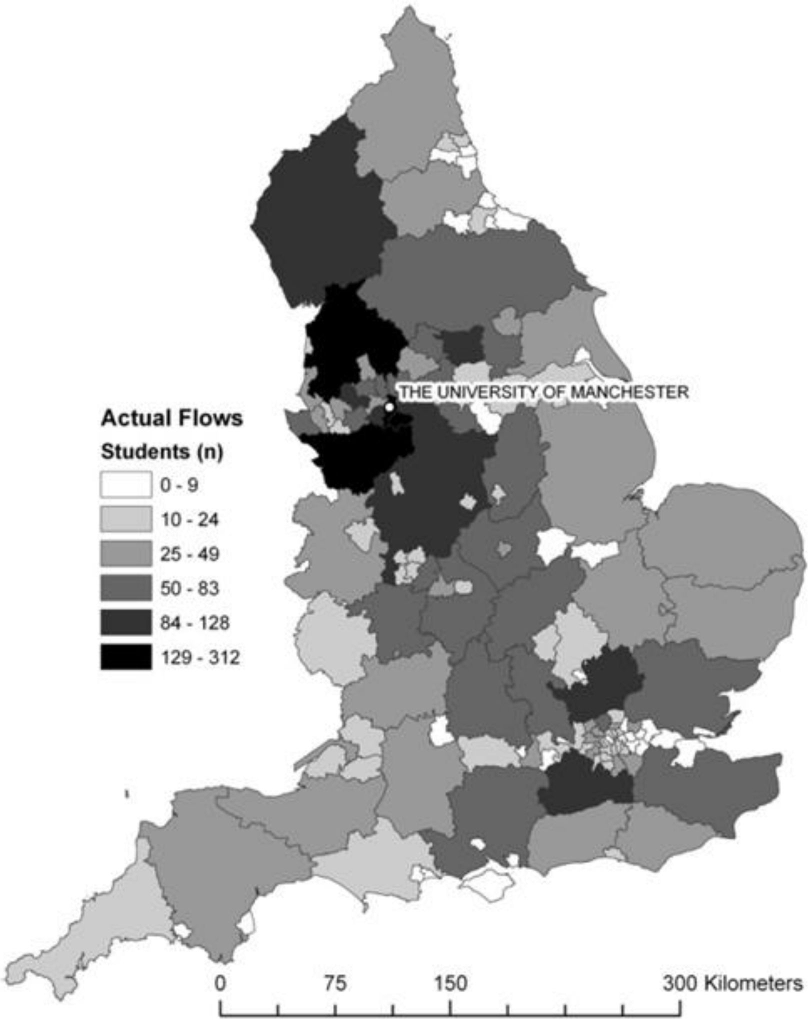
- Simple Java GUI to show the matrix of results
  - visually spot good/poor matches
  - refine model parameters
  - rerun

Results – Norfolk Schools to Universities



Results – Schools to University of Manchester



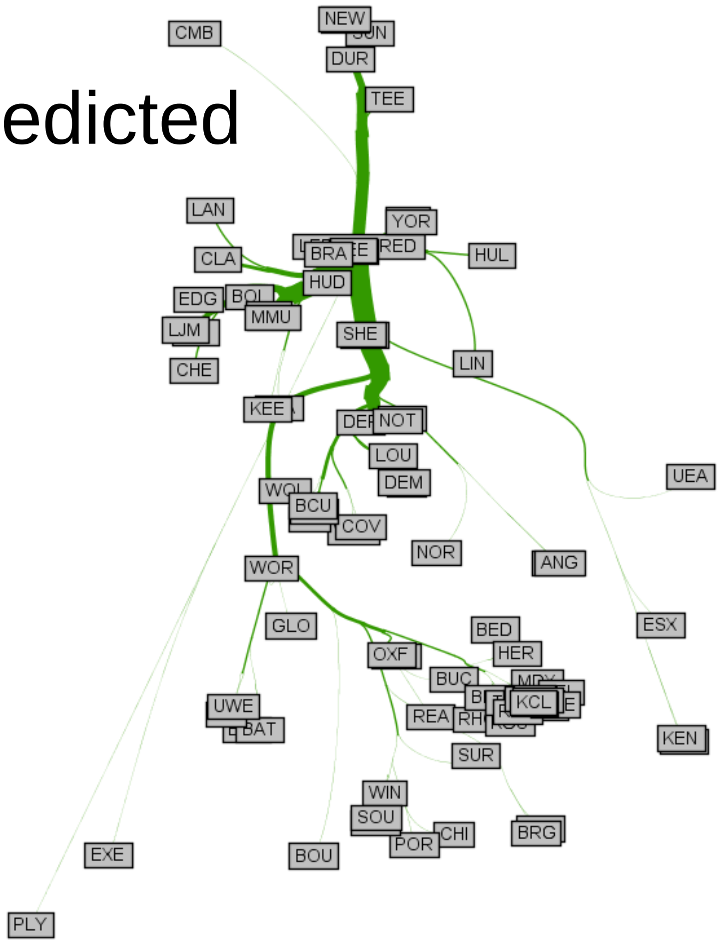


Results – Flow Maps

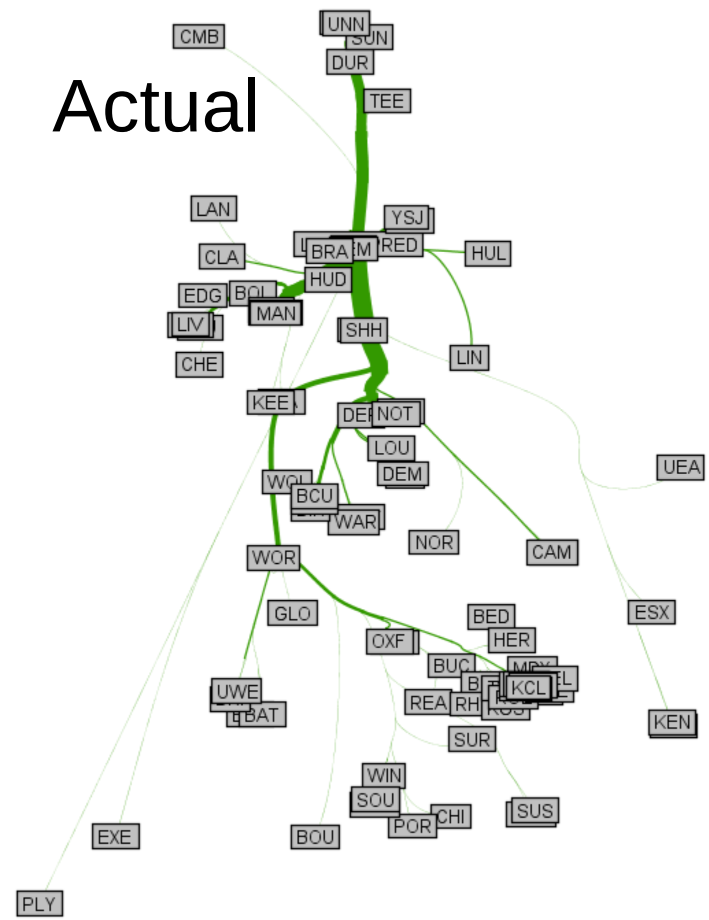
- *FlowMapLayout* Java application
  - Developed at Stanford for InfoVis 2005
- A more graphic & flexible presentation of the flows
  - Pseudo-spatial
  - Lengths and directions of the connecting lines are not meaningful

Results – Leeds Schools to Universities

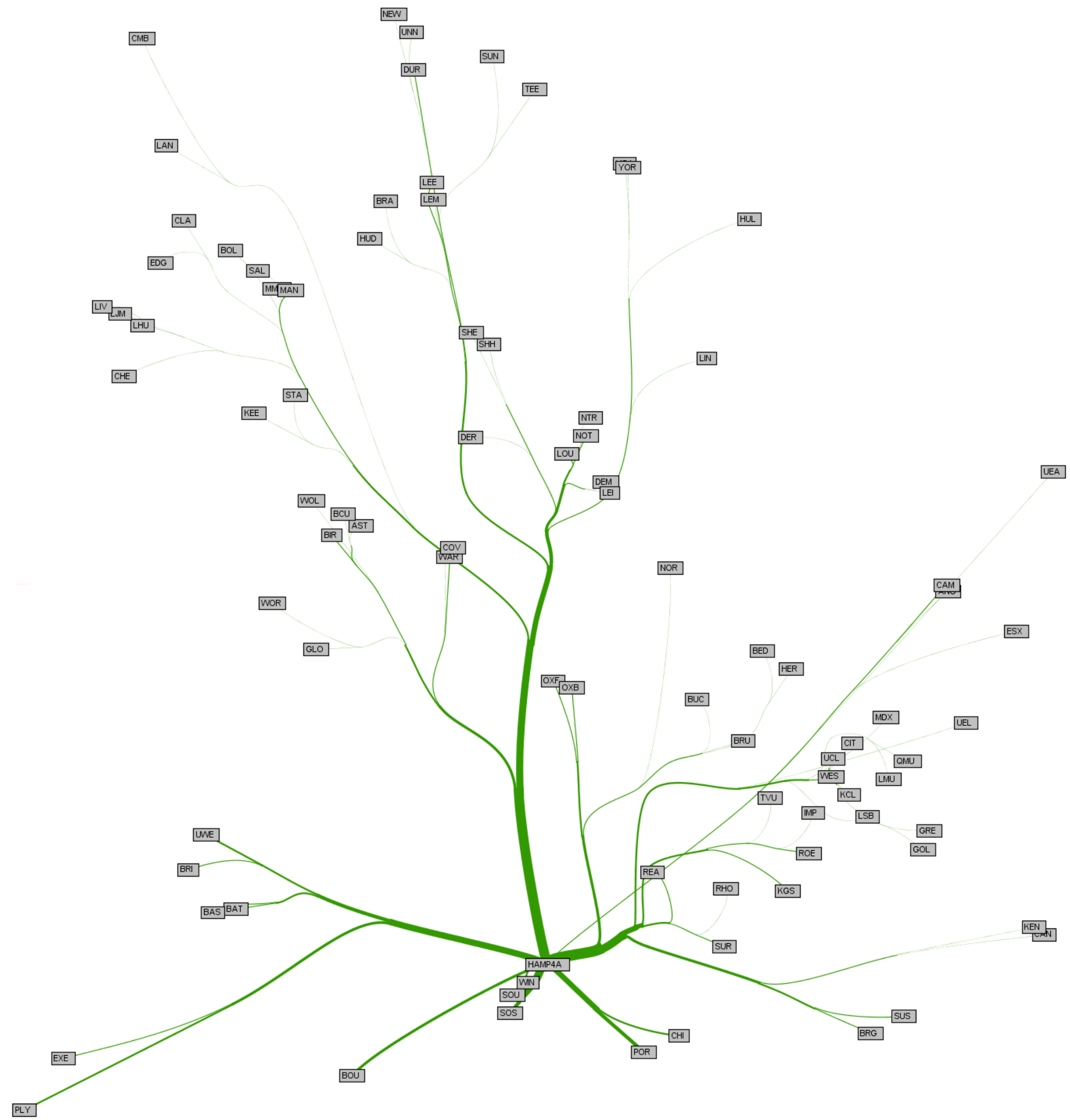
Predicted



Actual



Results – Hampshire Schools to Universities



Some Interesting Anomalous Results

- Flows significantly higher than expected
  - Flows from North-West London to Manchester
  - Essex to Exeter and Exeter to Essex
  - Small-distance flows to modern “metropolitan” universities, particularly paired with older institutions, such as Sheffield & Sheffield Hallam
- Flows significantly lower than expected
  - Yorkshire to/from Lancashire
  - Essex to/from Kent

Model Refinements

- Creating a genuinely partially-constrained model
  - University capacities are not known, instead we assume the actual enrolled numbers are all at capacity
  - This results in a completely doubly-constrained model, unless the “not at university” option is made attractive.
  - Possible solution would be to increase all capacities by a small % and adjust “not at university” attractiveness to rebalance the numbers
- The local “Metropolitan university” issue
  - Model adjusted to reduce the distances for these flows

Further Considerations

- Cost vs Distance – d_{ij} vs c_{ij}
  - Not necessarily linearly related for “life changing” spatial flows as such universities
- Straight-line distance is too simple
  - Natural barriers (e.g. mountain ranges, water)
  - Fast intracity & intercity transport networks
- Subject-specific analysis may be more revealing
  - e.g. flows for medicine courses only
- Including Scottish/Wales data
  - Potentially interesting with different fee requirements

References & Acknowledgements

- Wilson, A.G. (2000). The widening access debate: student flows to universities and associated performance indicators. *Environment and Planning A* ***32*** *pp 2019-31*
- Phan D. Et al (2005). Flow Map Layout. *http://graphics.stanford.edu/papers/flow_map_layout/*
- Dearden, J. and Wilson, A.G. (2008). An analysis system for exploring urban retail phase transitions – 1: an analysis system. *CASA Working Papers Series* ***140***
- Vickers, D.W. and Rees, P.H. (2007). Creating the National Statistics 2001 Output Area Classification. *Journal of the Royal Statistical Society, Series A*
- The paper for this project is currently in review.

Graphics Acknowledgements

- Gravity model graphic: Wikipedia
- Liverpool Street Station commuters: steve_way on Flickr
- Aberystwyth University examination hall: jackhynes on Flickr

Q&A

Oliver O'Brien
UCL Geography
Twitter: @oobr
http://www.oliverobrien.co.uk/

ESRC Funded Project