Gravity Models & Social Media

Tanane and



daniele quercia













mobility data +

-

787

LL)



... + listings of social events







recommend social events





Recommending Social Events from Mobile Phone Location Data [ICDM]





offline & online

twitter

P

ERSTWITTERI

top-using city London

North Weald Tracking Gross Community Happiness from Tweets [CSCW] r=.350 word count r=.365 MaxEnt Rickmansworth Moot Park Oxhey Buckhurst h Brentwood Chalfont Chigwel St Reter inner Romford Brent Islington Ive Ockendon London Lambeth Greenwich Grays Hounslow Wandsworth Lewisham Tilbur Bexley Dartford Swanscomb Staines Ashford Kingston Darenth Egham Merton Upon Thames Bromley Sunbury-on-thames Swanley Beckenham Shepperton ter Chertsey! Addlestone Weybridge Croydon Sutton Esher West Byfleet Epsom Cobham Banstead Woking Ashtead Warlingham

Talk of the City [ICWSM]









What is a gravity model?







Proposal



Proposal



Housing Living Environment Income Employment Health Crime Education

build a gravity model ~ flow of passengers









gravity works: r= .72!



Social deprivation might be connected to:

HP1: "unexpected" mobilityHP2: use of bus (compared to tube use)HP3: low social/geographic diversity

Index of Multiple Deprivation (IMD)

- 1. Income
- 2. Employment
- 3. Health
- 4. Education
- 5. Housing
- 6. Crime
- 7. Living Environment



Hypothesis 1

Where the model **fails to fit** well unexplained bit: prevailing socioeconomic factors



Hypothesis 1

Where the model <u>fails to fit</u> well unexplained bit: prevailing socioeconomic factors We look at (gravity)**Residuals**



Hypothesis 2

#tube passengers proportional to population
 "Bus/Car bias": residual between #passengers and population

Hypothesis 1+2 (prediction of actual values)

| Gravity Residuals | * * * |
|----------------------------------|---------|
| Bus Bias | * * |
| Socio/Geo Diversity | n.s.s. |
| | |
| R^2= 9% IMD | |
| R ² = 34% Living Envi | ronment |

Hypothesis 1+2 (binary predictions. top+bottom only)

| Gravity Residuals | * * * |
|-----------------------------------|---------|
| Bus Bias | * * |
| Socio/Geo Diversity | n.s.s. |
| | |
| R^2= 27% IMD | |
| R ² = 54% Living Envir | ronment |

Predicting...



(a) Real composite IMD score

(b) True class

(c) Predicted Class

Predicting...



Predicting...



So what?

Strong Passenger flow—urban deprivation
(timely & effective & longitudinal)

Situation



(Its already 75% in the USA)

Situation



By 2025 another **1.2** billion living in urban areas

Situation



Cities in developing countries: **5M** new inhabitants each month

Problem

Inequality!



Timely allocation of scarce resources



Reading List

Finger on the Pulse: Identifying Deprivation Using Transit Flow Analysis. CSCW 2013.

Questions

And social media? Only works **within** a city?

Source: stuart_beard

one year later...



Twitter ain't Without Frontiers. CSCW 2014.

10

ruth garcia @upt.edu (Ph.O survivor)

The End



| | Active in 2011 & 2013 | | |
|-------------------|-----------------------|-------------|--|
| | 2011 | 2013 | |
| Users | 1,315,313 | 1,125,968 | |
| English Tweets | 406,719,999 | 256,330,241 | |

Inactive < 1 tweet per day Hyperactive > 22 per day



- **5K** country country pairs
- 111 countries
- **13M** Geolocated users
 - **3B** Geolocated Tweets





clustered according to their geography





10 weeks



-

Communication Volume

Culture

Distance

Economic

Social

| Predictor | eta(%) | P-value |
|-------------------------|--------|---------------------------------------|
| Trade | 6.34 | *** |
| Cultural Dimension | 3.91 | *** |
| Gravity Model x Exports | 3.78 | ** |
| Gravity Model | 2.79 | *** |
| Language | 2.70 | · · · · · · · · · · · · · · · · · · · |

Reading List

Finger on the Pulse: Identifying Deprivation Using Transit Flow Analysis. CSCW 2013.

Twitter ain't Without Frontiers. CSCW 2014.



Questions

- How about penetration biases?
- How is it possible that such a simple model works?