Mobile phone data and human mobility

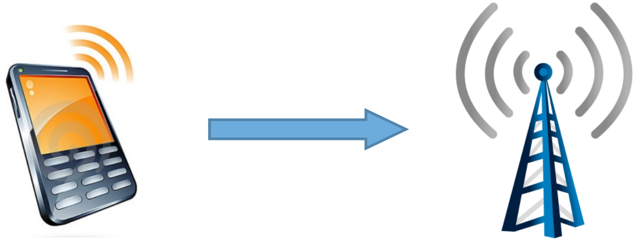
Maarten Vanhoof
Newcastle University/Orange Labs

M.vanhoof1@newcastle.ac.uk
@Metti_Hoof
MaartenVanhoof.com

Mobile Phone Data (Call Detail Records)

Metadata

- Caller (phone)
- Called phone
- Timestamp
- Type of event
- Duration of call/Length of text
- Location of celltower
- …



timestamp	caller	callee	event	duration	area id	tower id
2007/10/01 23:45:00	HJ123423	R482G9342	VO	3656s	1548	53571
2007/13/01 12:10:04	TR234S3	43FG3423	SI	125c	32768	53571
...	...	...	...	...	...	...

Call Detailed Records (CDR) data

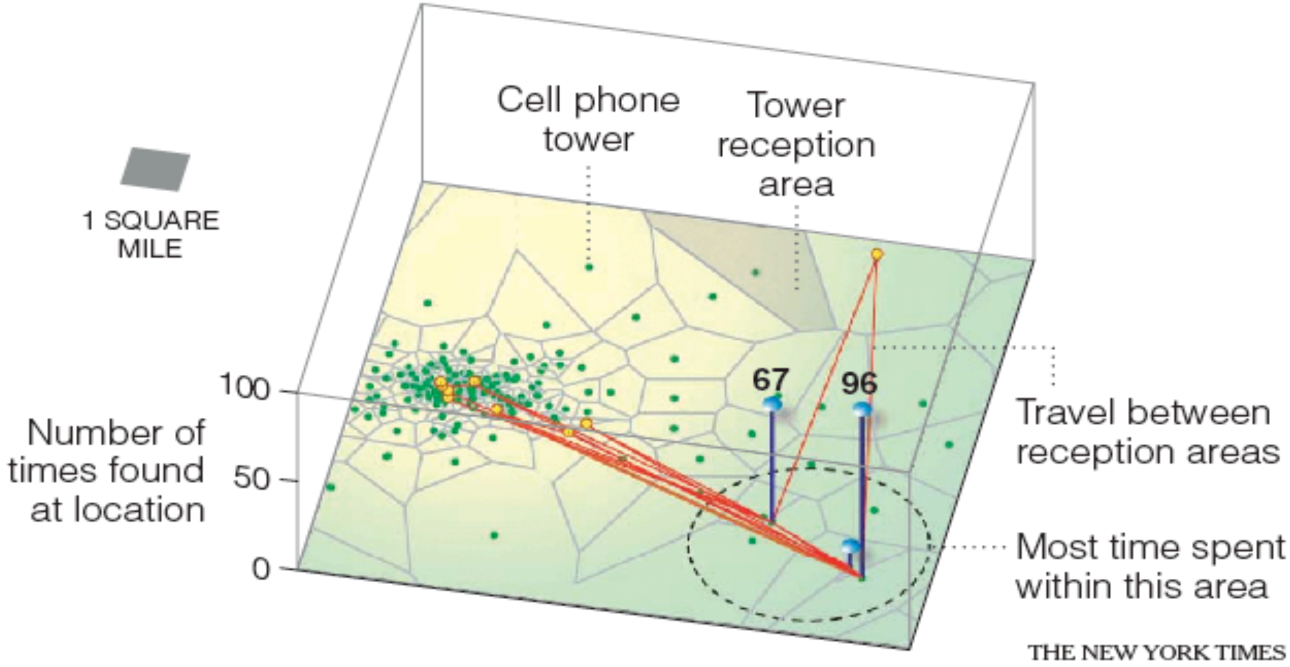
Advantages

- Very large populations (France 2007: ~30%)
- High penetration rates (France 2007: ~86%)
- Generally high spatio-temporal resolution (France 2007: 4 points in 3 distinct locations/day)
- Passive collection

Disadvantages

- Individually low spatio-temporal resolution
- User initiated collection, although passive
- Not designed for research purposes
- Poor in contextual information
- Absence of validation data

Individual mobility traces from CDR data



Gonzales et al. (2008) Understanding human mobility patterns.

State of the art

- Identification of stays and trips

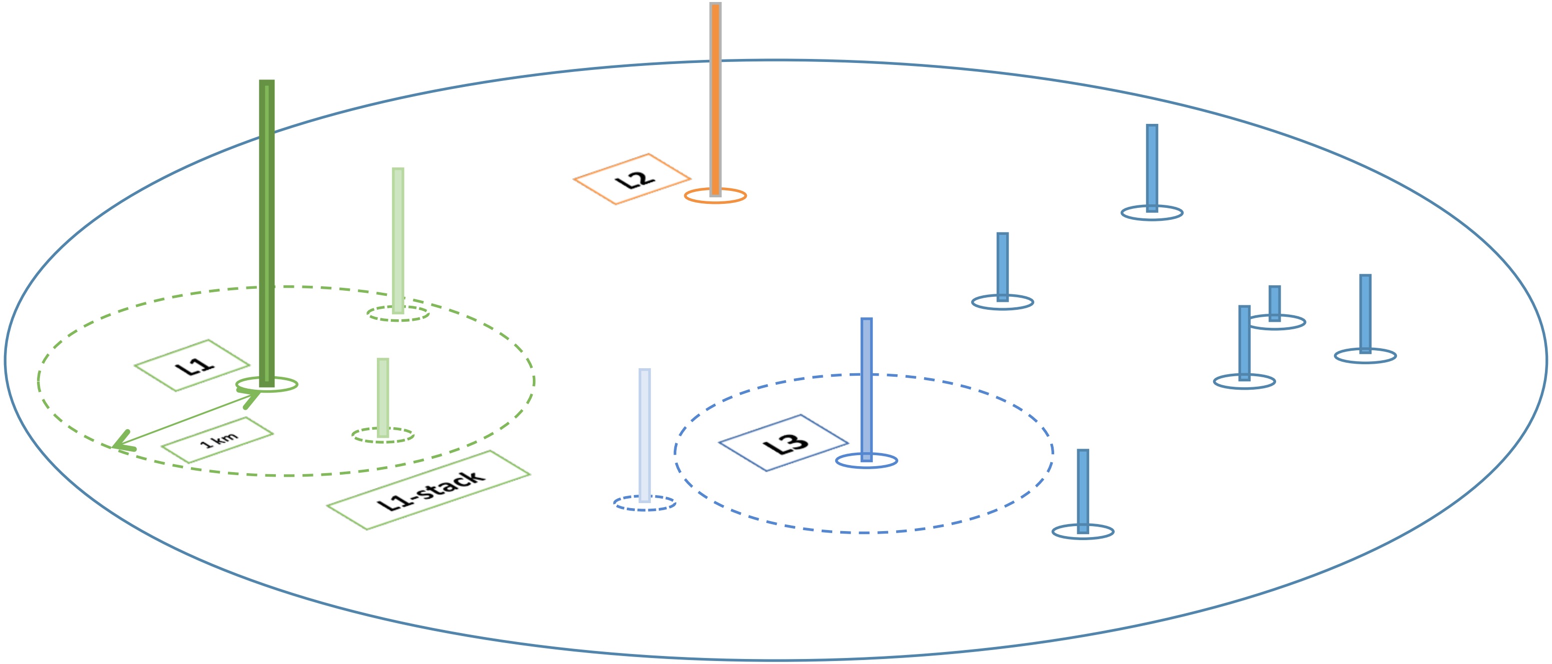
Wang et al. (In Press) Applying mobile phone data to travel behaviour research: a literature review

Stays and trips

Coarse temporal granularity forces us to use historical positioning data and consider aggregate frequency counts to derive typical stays and trips.

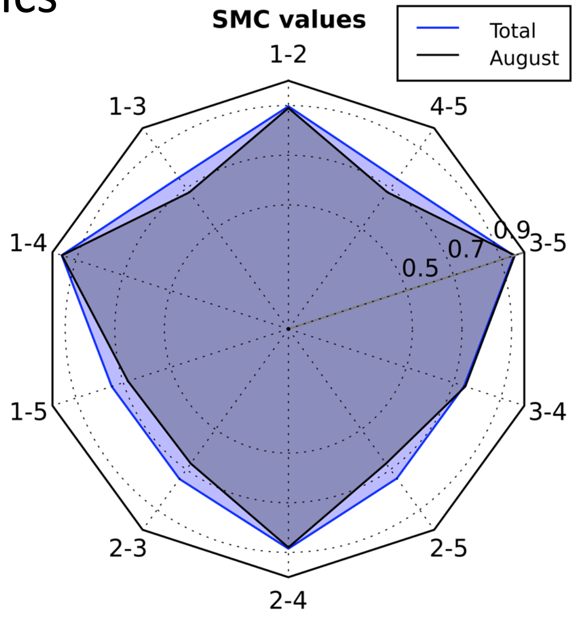
In other words, only stays and trips frequently taking place and preferably for/over a long period of time are well detectable

Stays and trips: home detection



Stays and trips: home detection

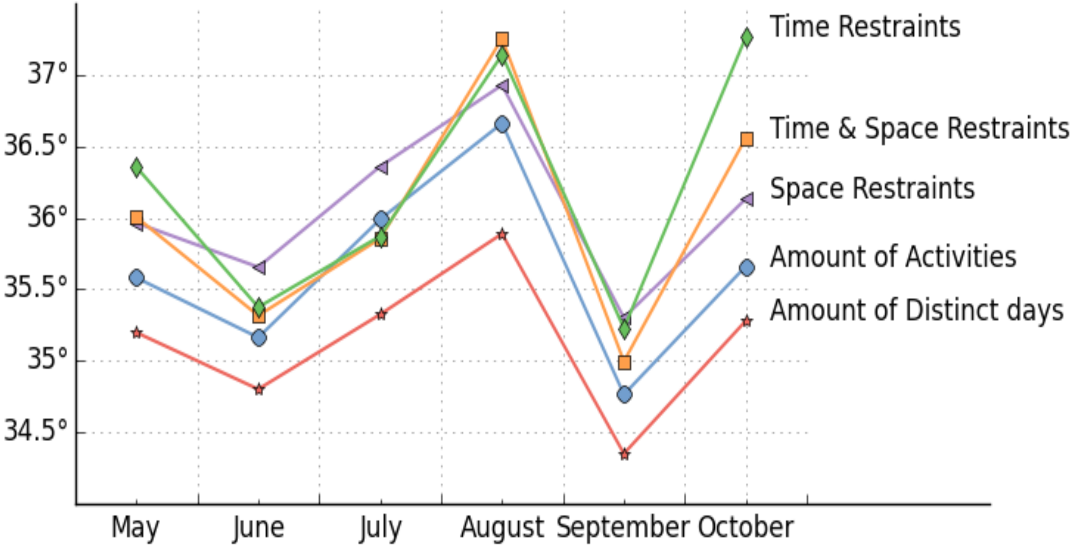
- 18 million users, 6 months, 5 simple heuristics
- Pairwise differences for up to 40% of cases
- Little influence of time on differences



Vanhoof et al. (in submission) Investigating Performance and Spatial Uncertainty of Home Detection Criteria for CDR data

Stays and trips: home detection

- All heuristics perform moderate
- Temporal variation is high
- Amount of distinct days is ‘best’
- Where does the 35° come from?
  - Mobile phone use?
  - Unevenly distributed market shares?
  - Census data itself?



Vanhoof et al. (in submission) Investigating Performance and Spatial Uncertainty of Home Detection Criteria for CDR data

Stays and trips: home detection

Uncertainty on home detection is still large on the general level and unknown on the individual level, given that no validation data exists there.

State of the art

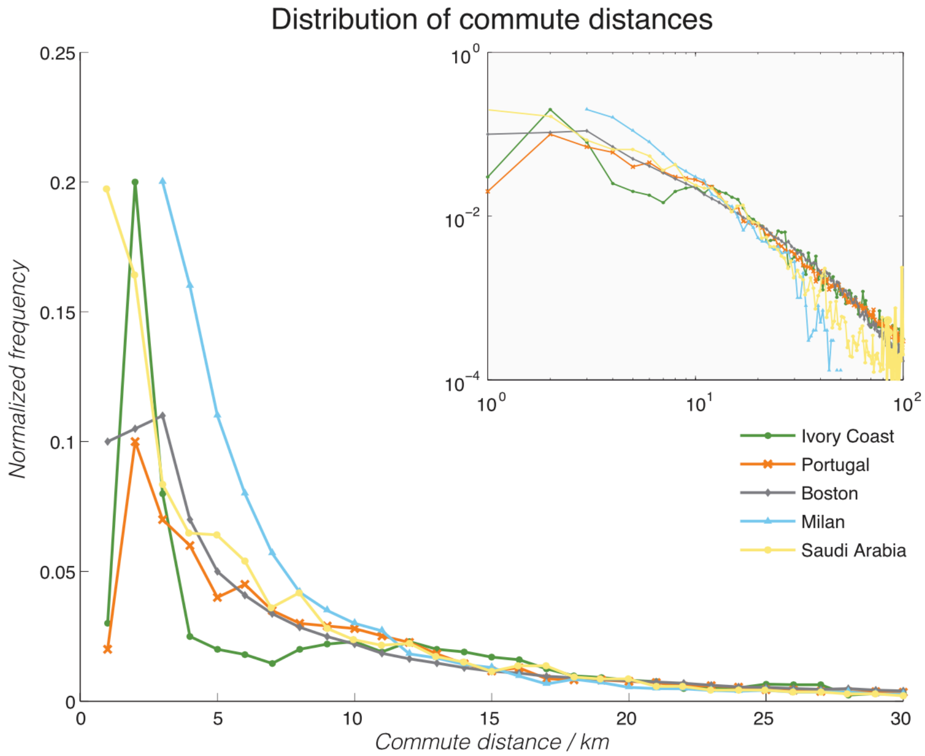
- Identification of stays and trips
- **Identification of activities and trips types**

Wang et al. (In Press) Applying mobile phone data to travel behaviour research: a literature review

Activity identification and trip types

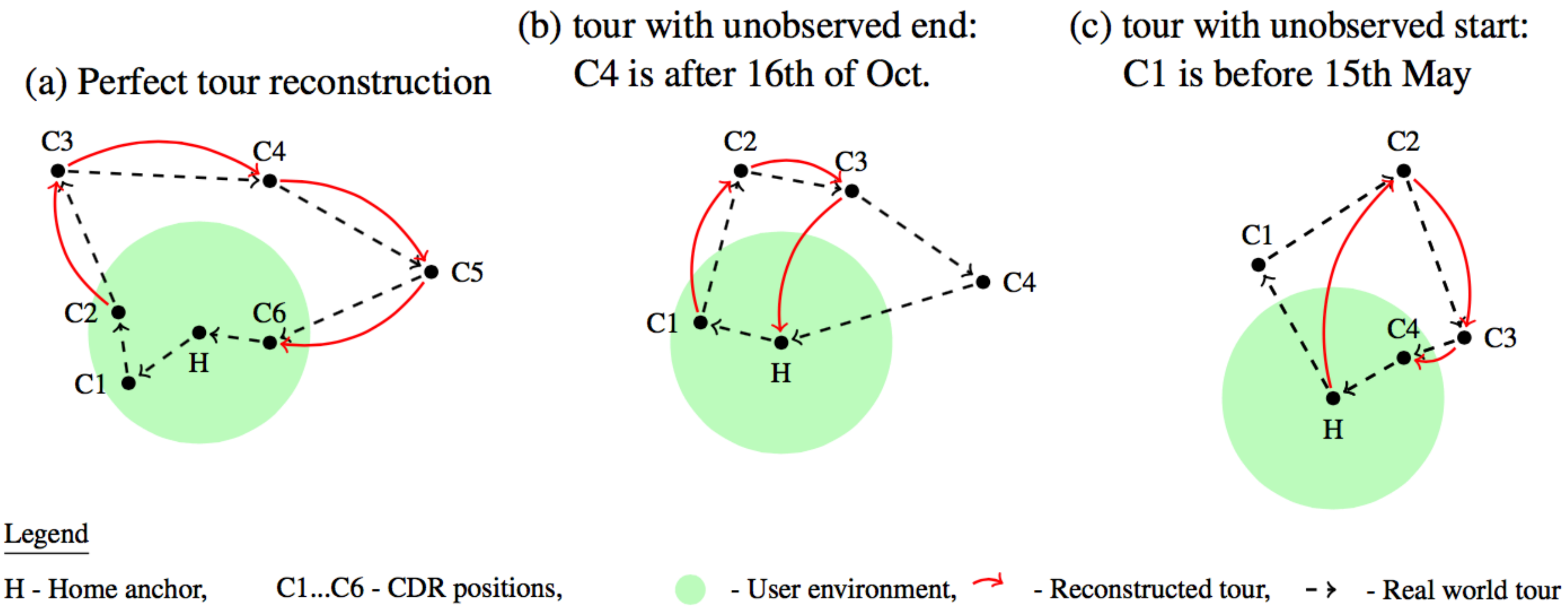
- Home-Work OD's has been widely investigated, although basing on uncertain detection of both locations.
- 'Third' location interpretations as 'leisure' are currently being pushed (but i remain skeptical)
- Other trip types might better suit the nature of CDR data:
  - Long distance trips
  - Seasonal movement
  - …

Activities and trip types: OD-matrices for home and work



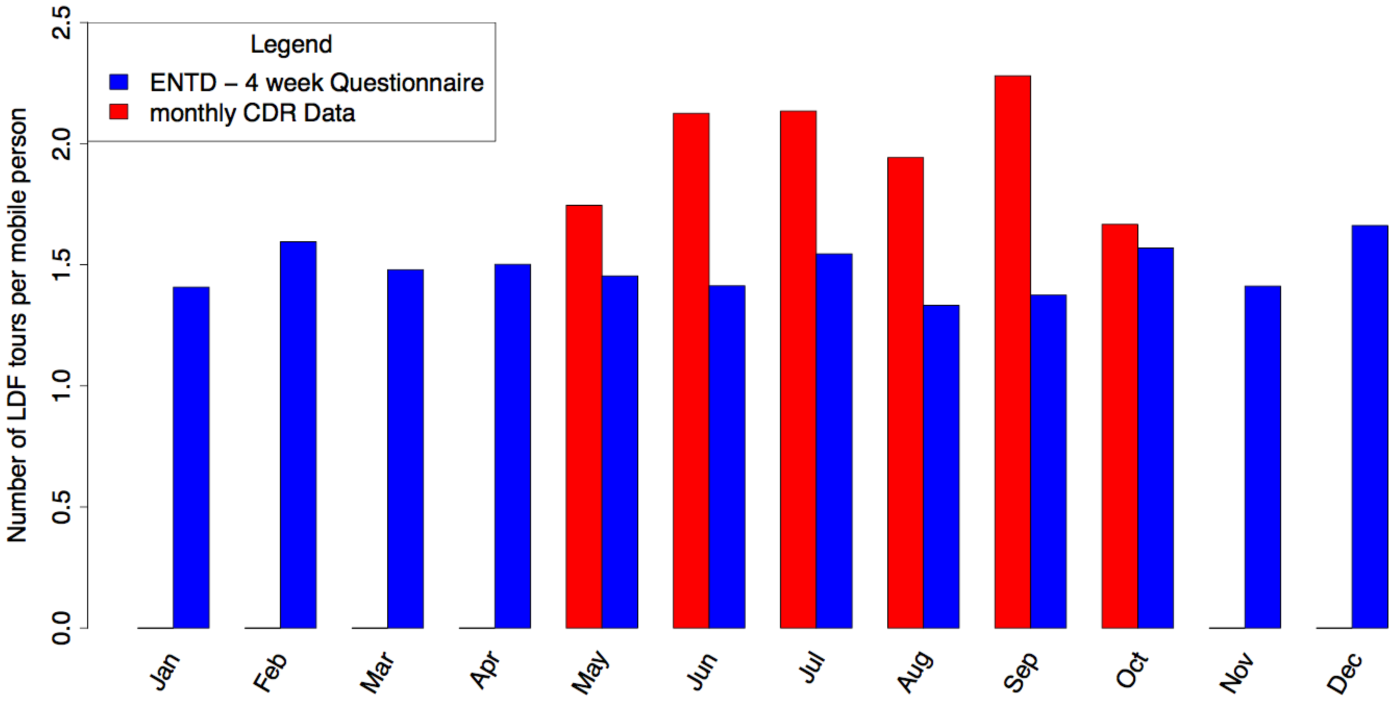
Kung et al. (2014) Exploring universal patterns in human home-work commuting from mobile phone data

Activities and trip types: Deriving long-distance trips



***Janzen** et al. (2016) Closer to the total? Long-distance travel of French Mobile Phone users*

Activities and trip types: Deriving long-distance trips



***Janzen** et al. (2016) Closer to the total? Long-distance travel of French Mobile Phone users*

Activities and trip types: Purpose imputation

	Commuting	Business	Vacation	Visits	Other	Uncertain
Total	18’733	87’418	64’738	231’800	136’968	38’957
Share	3.2%	15.1%	11.2%	40.1%	23.7%	6.8%
Share in ENTD 2008	3.5%	13.8%	22.0%	38.5%	22.2%	0.0%

***Janzen** et al. (Working real hard to get the job done)* Purpose Imputation for Long-Distance Tours without Personal Information

Activities and trip types: Purpose imputation

Attribute	Stage 1 Commuting	Stage 2 Business	Stage 3 Vacation	Stage 4 Visits/Other
Distance	Average	High	Average	Average
Duration	Average		Very High	High
Destination		Average	High	Average
Month		Average	Average	Low
Weekend-Share	Low	Very High		Average
Deviation	Average			Low
Frequency	Very High	Very High	Average	Low
Residence		Average	Average	Low

***Janzen** et al. (Working real hard to get the job done)* Purpose Imputation for Long-Distance Tours without Personal Information

State of the art

- Identification of stays and trips
- Identification of activities and trips types
- **Detection of travel mode**

Wang et al. (In Press) Applying mobile phone data to travel behaviour research: a literature review

Detection of travel mode

- Difficult.
- Based on estimation of speeds, time differences between locations
- Stationary, walking, vehicules distinction possible at small scale; application typically event based because of high mobile phone usage
- Exceptions for specific modes and trajectories, like airplanes, metro, and certain high-speed train trajectories

State of the art

- Identification of stays and trips
- Identification of activities and trips types
- Detection of travel mode
- **Identification of influencing factors for travel behaviour**

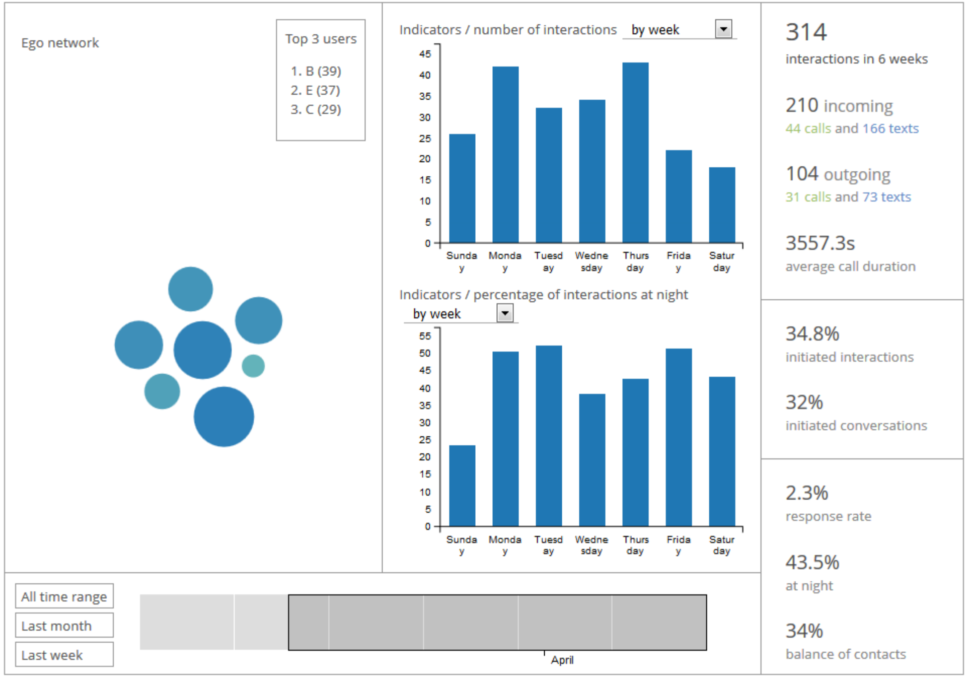
Wang et al. (In Press) Applying mobile phone data to travel behaviour research: a literature review

Identification of influencing factors

- Confrontation of mobility indicators with ‘internal’ indicators:
  - Temporal aspects (Wang et al. 2014)
  - Urban morphology (Kang et al. 2012)
  - Social networks (Cho et al. 2011)
- Confrontation of mobility indicators with ‘external’ indicators:
  - Socio-economic information (Pappalardo et al. 2016)
  - Gender and age (Yuan et al. 2012)
- Studies are descriptive, often covariates are not taken into account

Individual mobility indicators: Bandicoot.py

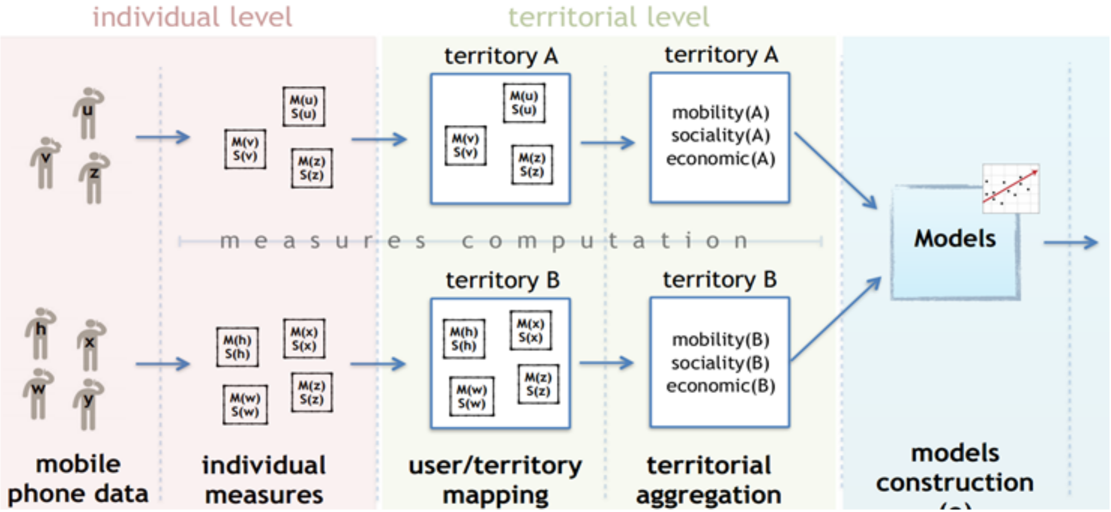
- active days
- number of contacts
- number of interactions
- call duration
- percent nocturnal
- percent initiated interactions
- response delay text
- entropy of contacts
- balance of contacts
- interactions per contact
- inter-event time
- percent pareto interactions
- percent pareto durations
- number of antennas
- entropy of antennas
- percent at home
- radius of gyration
- frequent antennas



http://bandicoot.mit.edu/demo/

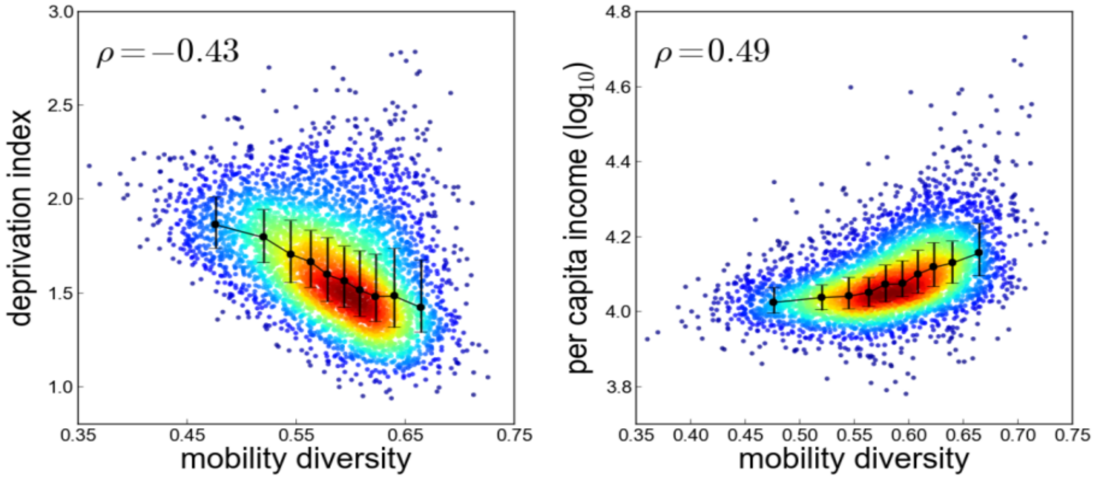
https://github.com/yvesalexandre/bandicoot

Confrontation with external indicators



Pappalardo,Vanhoof, et al. (2016) An Analytical Framework to Nowcast Well-Being using Mobile Phone Data.

Confrontation with external indicators



Pappalardo,Vanhoof, et al. (2016) An Analytical Framework to Nowcast Well-Being using Mobile Phone Data.

State of the art

- Identification of stays and trips
- Identification of activities and trips types
- Detection of travel mode
- Identification of influencing factors for travel behaviour
- **Statistical analysis of travel behaviour**

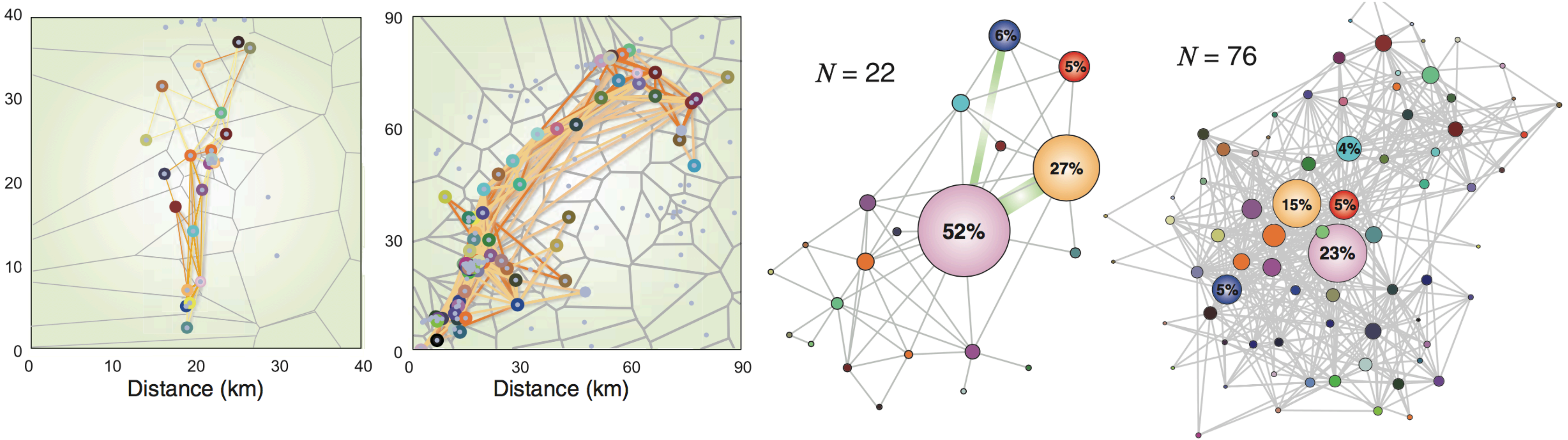
Wang et al. (In Press) Applying mobile phone data to travel behaviour research: a literature review

Statistical analysis of travel behaviour

- Statistical properties of large-scale human mobility, like power-law-like distributions of displacements, mobility motifs, visitation frequency, and staying time have been uncovered
- Travel behaviour is found to present a high regularity in space and time with people revisiting locations with high frequency while exploring others with a certain probability

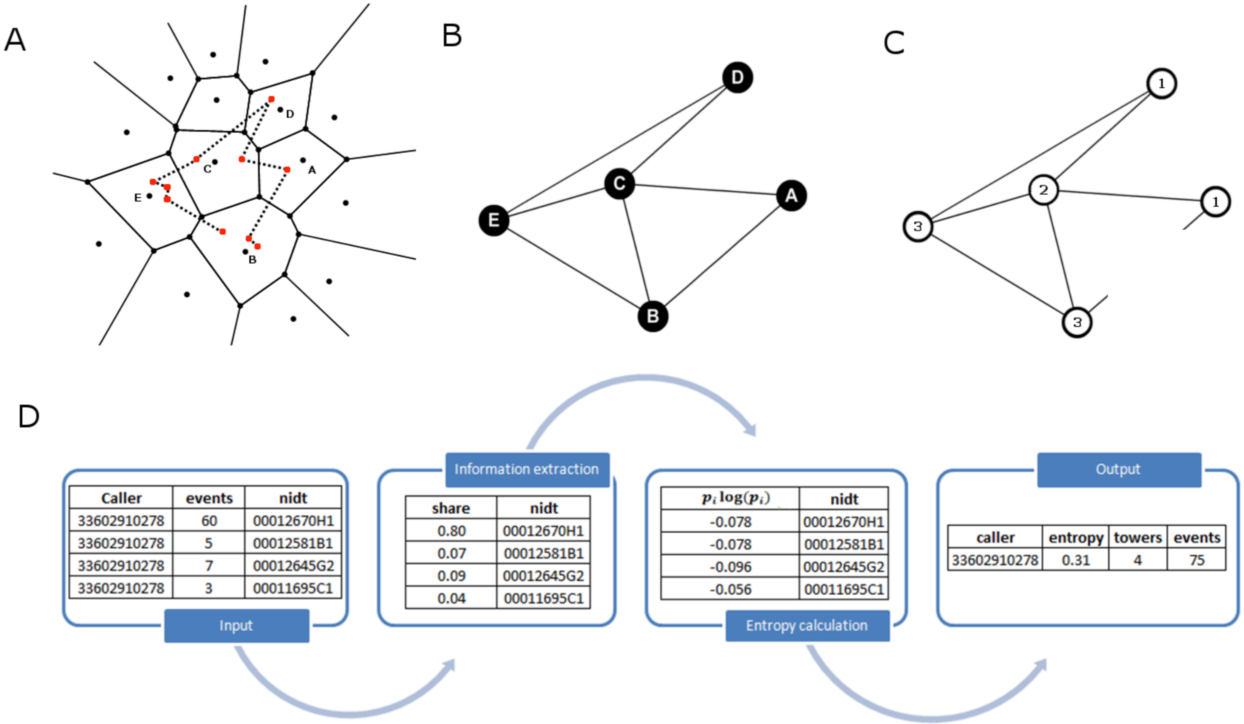
---> Potential predictability

Statistical analysis of travel behaviour



Song et al. (2010) Limits of predictability in human mobility.

Mobility Entropy



Caller	events	nidt
33602910278	60	00012670H1
33602910278	5	00012581B1
33602910278	7	00012645G2
33602910278	3	00011695C1

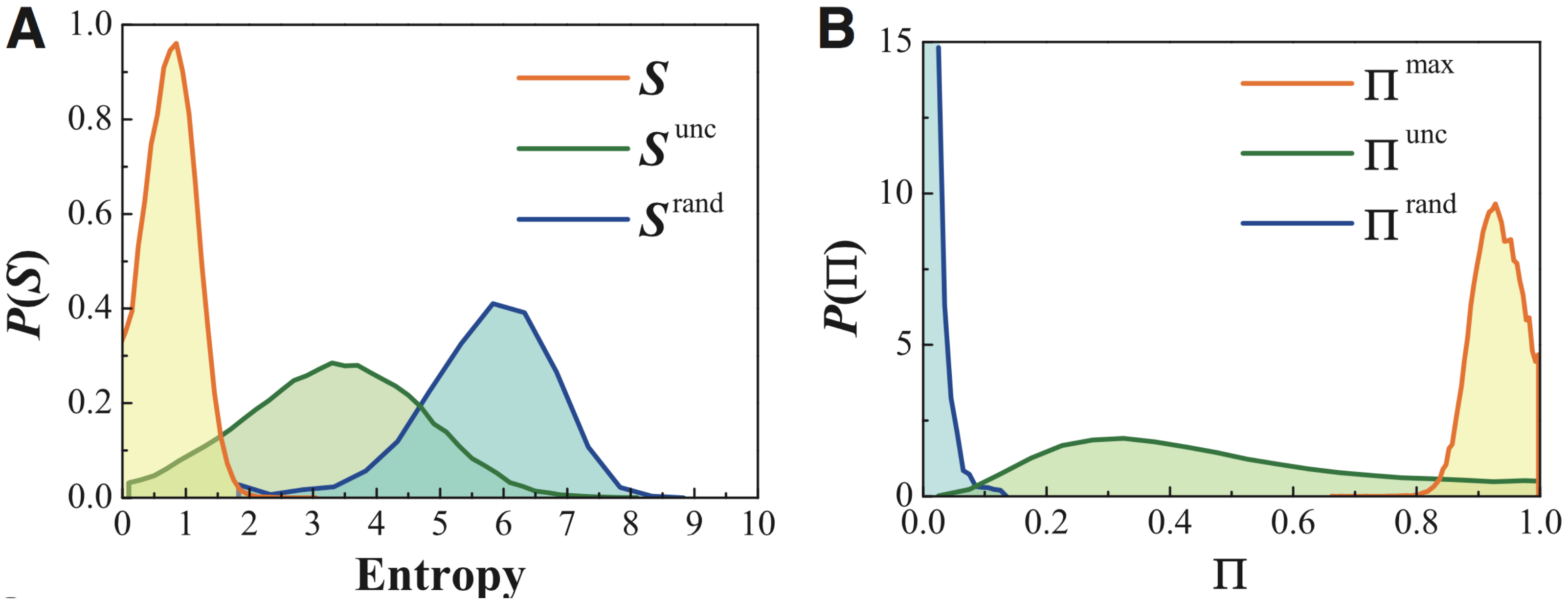
share	nidt
0.80	00012670H1
0.07	00012581B1
0.09	00012645G2
0.04	00011695C1

$p_i \log(p_i)$	nidt
-0.078	00012670H1
-0.078	00012581B1
-0.096	00012645G2
-0.056	00011695C1

caller	entropy	towers	events
33602910278	0.31	4	75

Vanhoof et al. (In revision) Correcting mobility entropy for regional comparison of individual movement patterns

Mobility entropy and predictability



Song et al. (2010) Limits of predictability in human mobility.

Statistical analysis of travel behaviour

"despite the apparent randomness of the individuals' trajectories, a historical record of the daily mobility pattern of the users hides an unexpectedly high degree of potential predictability"

Song et al. (2010) Limits of predictability in human mobility.

Statistical analysis of travel behaviour

Has lead to a thriving field in (social) physics, complexity science, computer science that is promoting a generalistic view of human mobility, fueled by publications in high-level journals

Statistical analysis of travel behaviour

Has lead to a thriving field in (social) physics and complexity science,
that is promoting a generalistic view of human mobility,
fueled by publications in high-level journals

Statistical analysis of travel behaviour

Has lead to a thriving field in (social) physics and complexity science,
that is promoting a generalistic view of human mobility,
fueled by publications in high-level journals

Statistical analysis of travel behaviour

- A universal model for mobility and migration patterns. Nature 2012
- Unraveling daily human mobility motifs. Journal of the R. S. Interface 2013
- Uncovering the spatial structure of mobility networks. Nature Communications 2015.
- A tale of many cities: Universal patterns in human urban mobility. PLoS ONE 2012
- Universal predictability of mobility patterns in cities. Journal of the R. S. Interface 2014
- Understanding individual mobility patterns. Nature 2009
- Natural Scales in Geographical Patterns. Scientific Reports 2017
- Limits of predictability in human mobility. Science 2010
- …

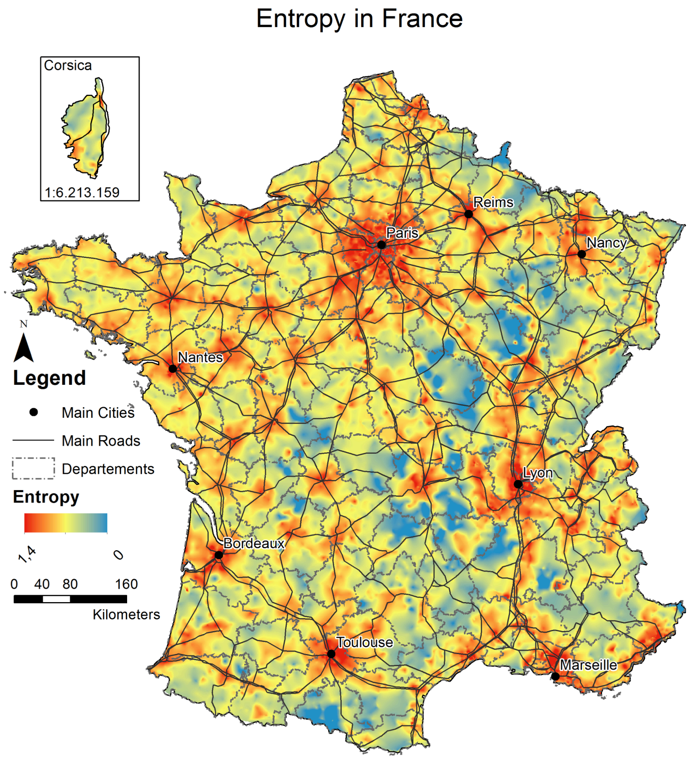
A critique

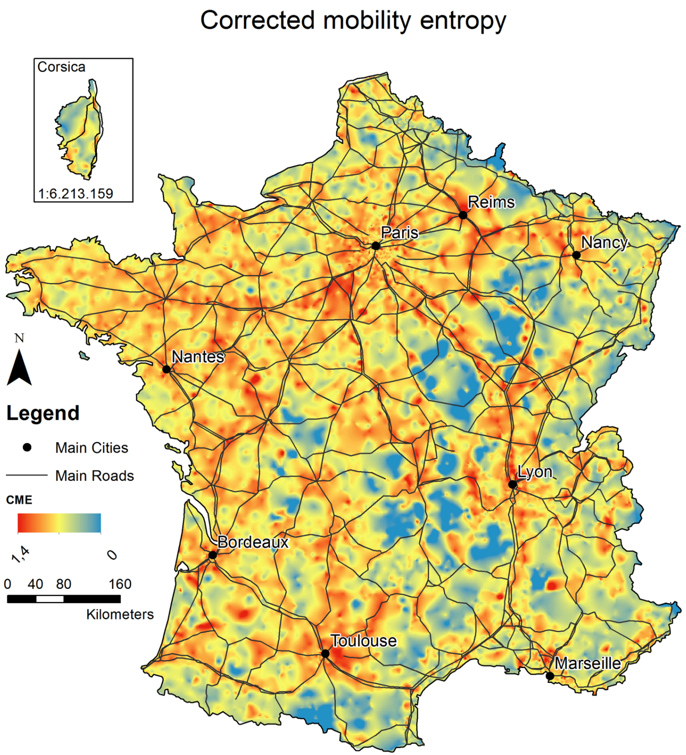
" … the greatest concern is that the prevalence of the big data meme might lead to a new scientistic, positivist, and quantitative turn that privileges generality over particularity "

Schwanen (2016, p.3)

Schwanen (2016) Geographies of transport II: Reconciling the general and the particular
Graham and Shelton (2013) Geography and the future of big data, big data and the future of geography
Wyly (2014) Automated (post)positivism

Generalistic view: Mobility entropy





Vanhoof et al. (in revision) Correcting mobility entropy for regional comparison of individual movement patterns

Statistical analysis of travel behaviour

Has lead to the ‘development of statistically self-consistent microscopic models to predict mobile phone users’ travel trajectory selection choices based on its selfquenching characteristics’.

(Wang et al. In Press)

--> ParadigmClash.

Paradigm clash: predictive models

Traditional

- Top down approach
- Based on theories and rules (e.g. utility)
- Incorporate many different factors
- Have high requirement concerning data accuracy and completeness because sensitivity to bias
- Difficult to capture inter-personal dependencies
- Don’t deal well with long time ranges

Big data

- Bottom-up approach
- Data driven
- Use only historical records
- Bias is reduced by large-scale samples
- Historical records incorporate great deal of inter-personal dependencies
- Capture long time ranges

Paradigm clash: predictive models

Traditional

- Top down approach
- Based on theories and rules (e.g. utility)
- Incorporate many different factors
- Have high requirement concerning data accuracy and completeness because sensitivity to bias
- Difficult to capture inter-personal dependencies
- Don't deal well with long time ranges

- **Reveal insights on underlying mechanism**

Big data

- Bottom-up approach
- Data driven
- Use only historical records
- Bias is reduced by large-scale samples
- Historical records incorporate great deal of inter-personal dependencies
- Capture long time ranges

- **Don't reveal much insight on mechanism**

Paradigm clash

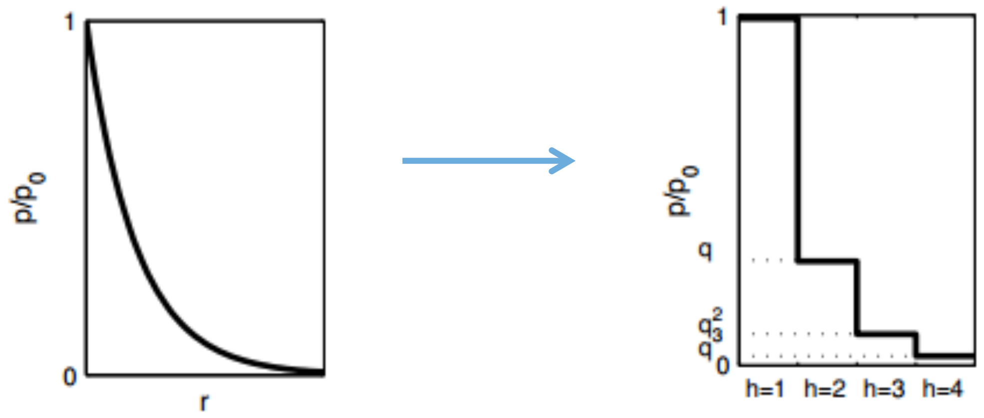
- Stimulates re-examination of existing theories and rules
- **Stimulates to find new ways to integrate available information**
  - **In existing frameworks**
  - With new theories / frameworks
- Opens up new research questions, fields of study and study areas.

Integrating mobile phone data in traditional models

- In four-step models
  - OD matrices for trip generation and trip distribution
  - Experiments with mode choice and route choice
- In agent/activity based models
  - Insights in long-distance behaviour
- **In standard spatial interaction models (gravity, radiation)**
  - **Re-definition of continuous distance into hierarchical distance based on mobile phone data (Grauwin et al.)**
- In entropy maximizing spatial interaction models (e.g. retail model)
  - Parameterization of attractivity factor (Wilson et al.)

Improving spatial interaction models

Replace continous distance effect by using a ‘hierarchical distance’ in which boundaries (q) between hierarchies (h) limit the probability of interaction between two places. Within boundaries, no differences in probababilty are imposed



Grauwin et al. (2017) Identifying and modeling the structural discontinuities of human interactions

Improving spatial interaction models

Calculate 'hierarchical distances' (h) from mobile phone data

Grauwin et al. (2017) Identifying and modeling the structural discontinuities of human interactions

Paradigm clash

- Stimulates re-examination of existing theories and rules
- Stimulates to find new ways to integrate available information
  - In existing frameworks
  - With new theories / frameworks
- **Opens up new research questions, fields of study and study areas.**

New research directions

- New data collection possibilities: data fusion, link with surveys,…
- Longitudinal aspect
  - Re-location tendencies, exploring patterns
  - Special days, events, emergencies, natural disasters
  - Long-term effects: climate change, impact of policies, behavioural change
- Social aspect is still under-developed but plausible
- Hard-to-reach groups
- Ethics and implementation of anonymising
- Biases (everywhere): measurement, contribution, sampling, …

The end.

Thank you!

M.vanhoof1@newcastle.ac.uk

@Metti Hoof

www.MaartenVanhoof.com