Research proposals for CASA master thesis

Dr. VANHOOF Maarten
February 2021

The supervisor

Dr. Maarten Vanhoof is a honorary research associate| who previously worked with prof. Mike Batty on the Quant project. Maarten holds a PhD in Computing Science, and masters in Geography and Cultural Studies.

Maarten's research focuses on bridging mobility research, geographical research, and (big) data analytics and he is an expert in the analysis of mobile phone data. He has collaborated with some of the most prestigious institutions and universities in the world like the MIT Senseable City Lab, the NYU Centre for Urban Science and Progress and the Big Data division of the European Commission.

In the past, Maarten has supervised two highly successful theses (both resulting in paper publications), and he has supervised around 10 internships working on big mobile phone data.

Maarten is Belgian, meaning he understands the struggles of not being a native English speaker (after French and Dutch, English is his 3rd language). He is looking for students who like to come up with their own cool ideas but at the same time are keen on finishing things on time.

How it works

In the following pages you find two specific topics I think would make interesting master theses, based on datasets that I have available. They form an indication of the work I would be interested in supervising, but please, feel free to come up with you own ideas, datasets, or any combinations of them. I will be happy to talk thing trough with you during surgery sessions or other moments.

Just to make things clear: The dataset in the first topic is a mobile phone dataset from France that I have available from a collaboration with the French telecom provider Orange. It is a nation-wide network in which nodes are cell phone towers and edges are the amount of people that have been calling to each other. The data is based on almost 18 million people's activities, so there is a lot of information available in here.

The dataset in the second topic is one we have produced ourselves in the Quant project for the UK. These are datasets at the MSOA level in the UK, stating how many people commute between MSOAs for different transport modes (car, bus, and rail) as well as the times it typically takes to do this commuting. You can think of this dataset as very high number of small networks from regions to nearby, other regions in the UK.

Summary

Mobile phone calling patterns and city systems

Investigate whether there is an influence of cities on the mobile phone calling patterns in France.

Geographies of sustainable commuting in the UK

Investigate large-scale commuting patterns and define which places have the highest potential to promote alternative, more sustainable commuting

Your own proposal

Come and have a chat with me to discuss your own proposal or a proposal proposed by our industry partners.

Mobile phone calling patterns and city systems

Investigate whether there is an influence of cities on the mobile phone calling patterns in France.

Mobile phone data, Networks, City systems, Spatial interaction models, Visualisation

Concept

Thanks to new technologies and extensive passive data gathering from mobile phones, it has now become possible to construct the geographical patterns of calling networks for an entire nation. The question now becomes can we properly understand and visualise these calling networks?

Using spatial interaction networks, for example, researchers have discovered that calling patterns are influenced by distance (if places are further away, there will be less communication between them), and that they can be used to create regional delineations of a countries (regions that call more with each other tend to be the same as some administrative regions).

One outstanding question, however, is whether the effects of city systems can also be seen in these calling networks. Do we observe combinations of large-medium-small cities to communicate more or less with each other? What would drive these patterns? Can we visualise or maybe even predict them?
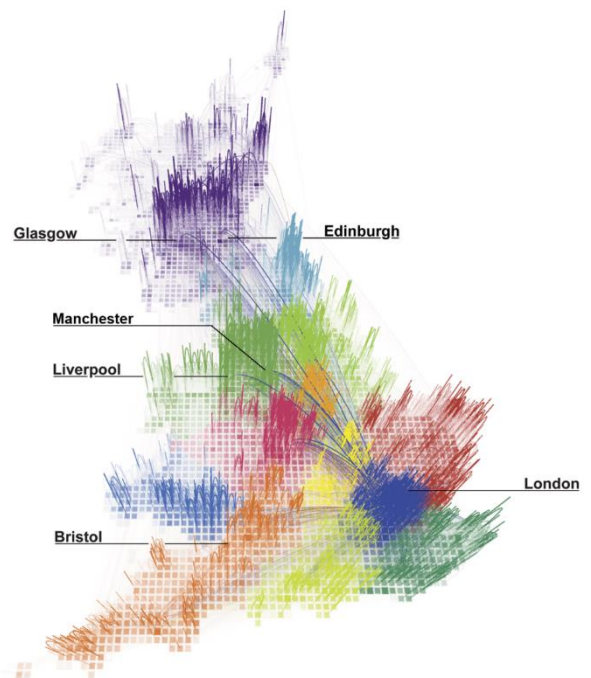
Example from Ratti et al. 2010 where calling networks where used to identify regions in the UK

Input data

- We provide a calling network for France
- The network consists of 18.000 nodes, which are cell towers located in the country and capture the amount of calls made between cell towers
- Calling networks are available for different periods in 2007, but more recent data could be applied for.

Reading material

- Maarten Vanhoofs blogpost on using calling networks: https://hellofuture.orange.com/en/drawing-boundaries-of-social-interaction/
- Sobolevsky et al. (2013) Delineating Geographical Regions with Networks of Human Interactions in an Extensive Set of Countries
- Ratti et al. (2010) Redrawing the Map of Great Britain from a Network of Human Interactions

Mise en œuvre

- Coupling of the call network with information of cities (for example, identification of small, medium and large, cities)
- Data analysis on the communication patterns between cities and within the city system (python or R)
- Testing of a simple interaction model to see whether it can re-produce empirical observations
- Visualisation of the data in maps.
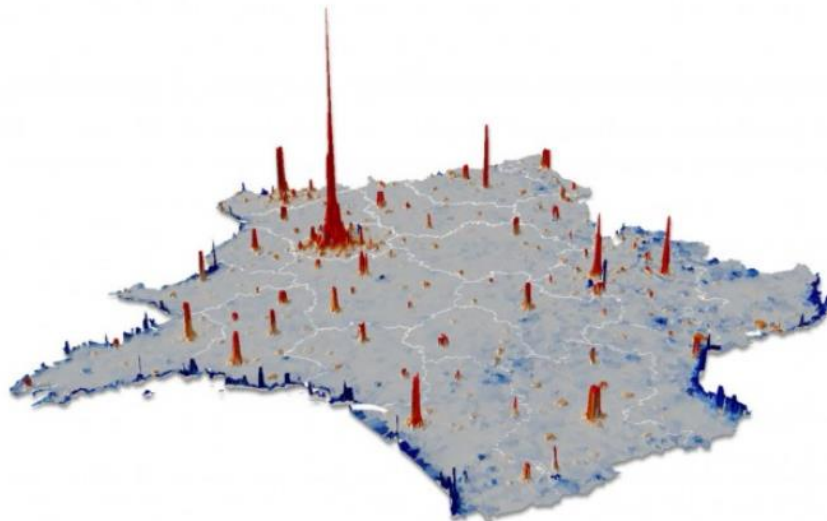
CATHERINE LINARD

A population density map of France derived from more than 1 billion cellphone call records shows that people congregate in urban areas during working periods (indicated by orange spikes), and head for coastlines during holidays (indicated by the blue spikes).

Geographies of sustainable commuting in the UK

Investigate large-scale commuting patterns and define which places have the highest potential to promote alternative, more sustainable commuting

Commuting patterns, sustainability, geographical analysis, visualisation

Concept

There is a lot of talk about how we need to promote a more sustainable human mobility by motivating people on public transport or by providing better public transport infrastructure. helping people consider public transport or by building new infrastructure.

However, the way commuting currently works in the UK is very badly understood. How many people commute from where to where? What mode do they use? Which distance do they cover? How long does it take them? Moreover, the geography of these patterns is little investigated. Where are the places where people use both bus and cars? Which locations are similar to each other and which aren't?

The intriguing bit is that there is quite some data at hand to answer these question. But the data is rather complex and so we will need to explore it in different ways: Can we find clusters of locations that have similar profiles of commuting patterns (for example, a similar distribution of bus, car, and rail shares)?

Can we detect the places where bus is really a competitive mode for cars, and where it is not (imagine a rural village where the bus takes 2 hours to get to the next town)? Can we find locations where public transport is available but people do not tend to use it (maybe some London neighbours have very competitive bus and rail options but people still take the car because they are rich?)

I'd propose we start from the available data on travel times and amount of people that commute and see if we can find data analytics, GIS techniques, or some small machine learning applications to unravel these patterns and answers some of these questions. Obviously, some GIS and visualisation skills will also be needed to make sure our findings can be properly communicated.

Input data

- Commuting figures, per mode, between all MSOAs in the UK, as captured by the Census
- Estimations of travel times between all MSOAs and for all different modes as created by the researchers in CASA working on the Quant project.

Reading materials

Have a look at this geodemographic classification:

https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/abouttheareaclassifications

Maybe we can do something similar but based on commuting profiles?

Mise en œuvre

- Explorative analysis of commuting patterns (GIS and Python/R)
- Clustering of MSOA profiles based on commuting patterns and travel times
- Visualisation of outcomes
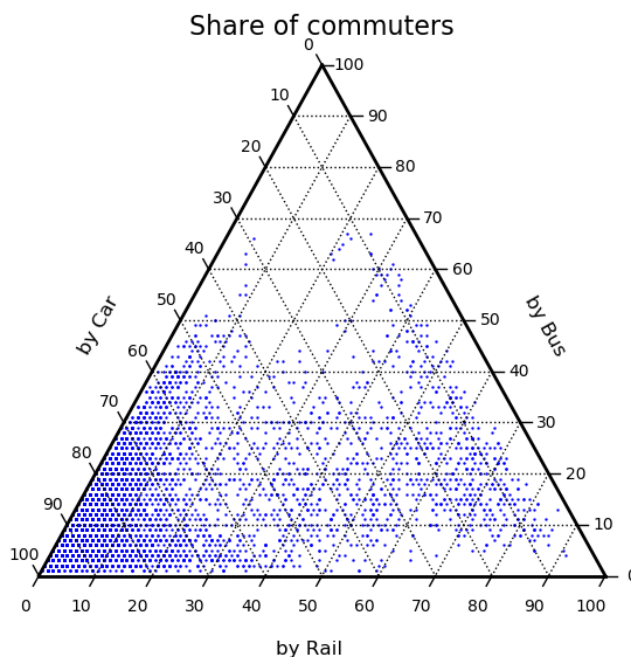
Ternary plot showing the share of mode transports in commuting in the UK. Each dot is one MSOA.

Your very own proposal

Come and have a chat with me to discuss your own proposal or a proposal proposed by our industry partners.

World domination?

Concept

You get that beautiful brain of yours going, you create some ideas that you are interested in, you come over (preferably during the surgery session, potentially on another moment) and we have a chat.

I would be very interested to work with people that want to explore visualisation work using d3.js or webGL frameworks on larger datasets, but of course, you can propose any idea to me!

If you need inspiration, you can always have a look at the proposal proposed by our industry partners. I'm sure you'll find something that interests you there and I would be happy to see if we can take this further.