Trends in urban flows: An information theory approach

Roberto Murcio[1], Balamurugan Soundararaj[2]andhttps://www.overleaf.com/project/5c030bdbc6fd9

January 17, 2019

1 INTRODUCTION

The accurate estimation of human activity in cities is one of the first steps towards understanding the structure of the urban environment [1]. Human activities are highly granular and dynamic in both the spatial and temporal dimensions [2] and estimating them with confidence is crucial for decision-making in numerous applications such as urban management, retail, transport planning and emergency management. Detecting general trends in flows of people between spatial locations is neither obvious nor an easy task due the high cost of capturing these movements without compromising the privacy of those involved. Online activity and general acknowledgement of social media have brought new data sources on human behaviour which were not accessible before and therefore allow explanation and further understanding of human drives and decisions. They also support individual users in decision making as more shared knowledge becomes available at people's fingertips everyday. Insighs into data mobile crowd sourced sensors demonstrate the present power of inter connectivity between people, space and time. The common -obvious- finding in all these studies is that people doesn't move randomly, i.e. following a Levy flight. Instead, there's a strong evidence for a conserved quantity in human mobility []. Ultimately, it's regularly and hindered by geographical distance.
There's a whole family of distributions
This work intends to address this problem by examining the movement of people in a network of street sensors at a fine spatial and temporal resolution. A novel methodology to the field of Big Data using mathematical models from information theory is introduced,

++ Indeed, the morphology of a route is shaped by the embedded spatial pattern of a city (land use and street topology) in association with dynamical factors such as congestion, accessibility, and travel demand, which relate to various attendant

*[1]Centre for Advanced Spatial Analysis, University College London [2] Department of Geography, University College London

socio-economic factors. Analyzing the morphology of routes, therefore, allows us to potentially uncover the complex interactions that are hidden within the coarse-grained spatial pattern of a city. Furthermore, the morphology also encodes the collective property of routes, including their long-range functional effects. For example, a single street, depending on its connectivity and location, can have influence that spans the dynamics across the whole city (Broadway in New York City for instance) ++ However, these datasets are hard to access an usually rely on very disclosing data (like the exact position of a person or its particular weekly routine from home to work). Here, we investigate the movement of people without explicit routes ==We need to recover here all the work done in pedestrians routes and make a strong case why these FF snapshots are worth it.

¿¿¿Dynamical properties of the spatial structure of cities: how much does the city shape change through the course of the day? Where are the citys hotspots located at different hours of the day? How are these hotspots spatially organized? Is the hierarchy and the spatial organization of hotspots robust through time?

In this work we addressed some of these questions using a dataset contains details of passive WiFi signal probing from a sensor network across Great Britain [CDRC cite], known as the SmartStreetSensors data. ==These data are used as a proxy for estimating footfall at retail locations... potentially identifiable information collected is hashed at sensor level, and the data is sent to the central server via an encrypted channel for storage. The temporal resolution of this network is sufficiently granular to detect general patterns across a day without compromising some individuals privacy. ===

2 Data

BALA The SmartStreetSensor project is one of the most comprehensive studies undertaken on consumer volume and characteristics in retail areas across United Kingdom. The data for the study is generated through sensors installed at around 1000 locations across UK. These sensors capture a series of public signals - known as probe request frames - generated by WiFi enable devices. From July 2015 to May 2017, there are around 652 have been installed and operational across UK. The number of probe requests captured is in the order of 2.6 billion records and growing at a rate of 6 million new requests per day. Each sensor generates a stationary time series representing footfall counts around a particular location through out the day in 5 minute intervals as shown in figure 1.
We analysed 53 million records corresponding to all the aggregated 5-minutes FF counts during 2017 at 889 locations (35 % of them in the Greater London area). The following map shows the spatial distribution of this network. The brightest areas correspond to cities with a higher number of sensors in operation (which does not necessarily implies a higher FF, although in London it is the case)

++Here we put the UK and central london heatmap ++ ++Repeat Blog++ //

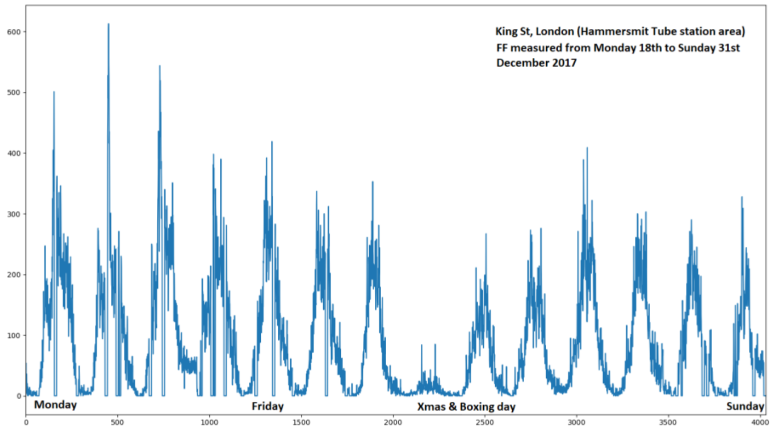

Figure 1: Five-minute footfall counts at Kings Cross, London on Monday 18th Sunday 31st December 2017

2.1 Time Series Analysis

TS Decomposition KARLO
Can we visualise trends for the same cities as BALA, an using plots like 14.2 and 14.10 from https://serialmentor.com/dataviz/visualizing-trends.html
++After this we briefly mention how the counts are constructed making sure that we spell out the randomisation issues. And we wrote all this in detail at an SI++

3 Results

3.1 FF signals

Each location has a particular signal FF counts as a function time that inform us about the local movements of pedestrians around that area, like the one at 1

==Next we argue about the circadian rhythms in different cities and how these repeat from one city to another showing Bala's plot and the collapsed one for the average FF by hour in a weekday / weekend . The difference between this two plots inform us about the structure of the city and **this is our first result**. If possible put an example of a weekend/weekday location==

Our second result is the type of signals and the correlation with type of street. With this, we signal that location is more important that type of business without compromising Karlo's paper.

Our third and main result is the flows of people. We do this using two tools: KARLO - Correlation analysis and my TE

This two measures are affected by distance and, more important by the street's

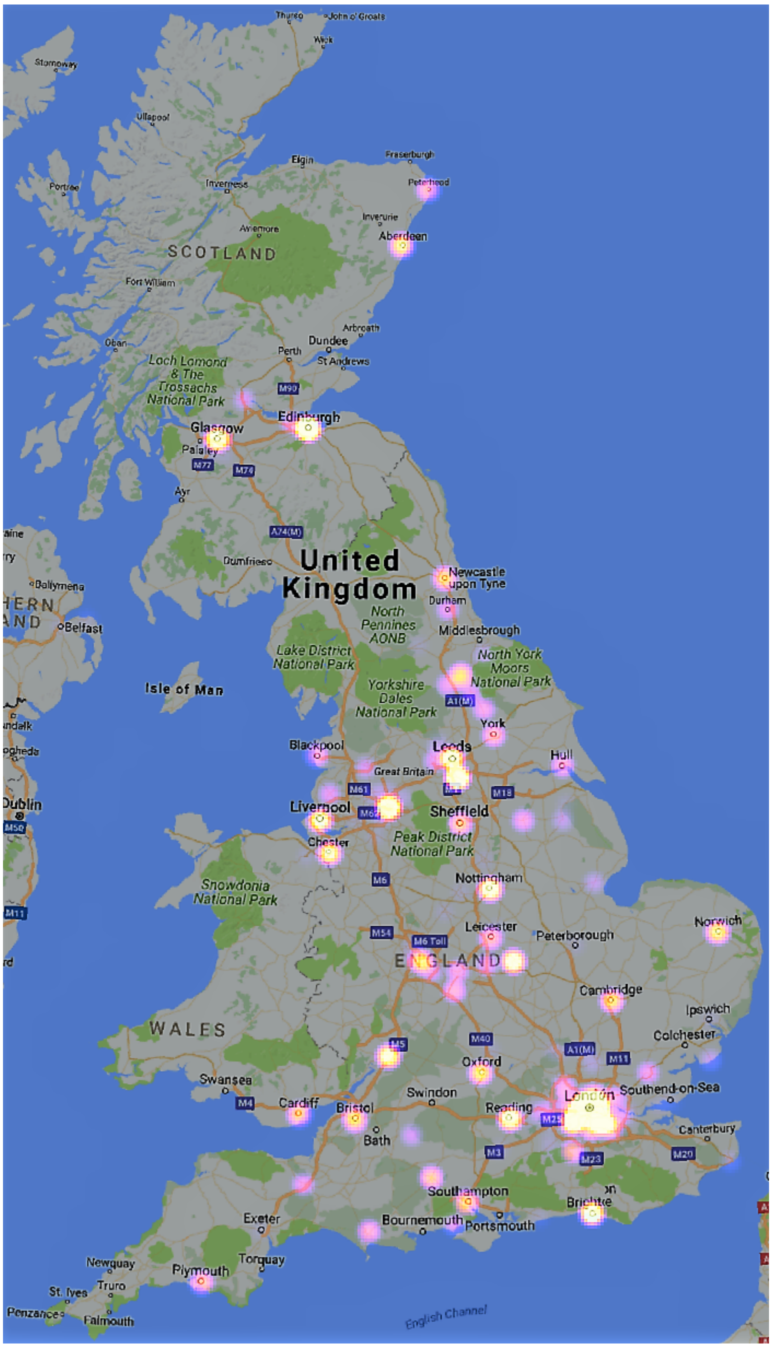

Figure 2: UK sensor distribution

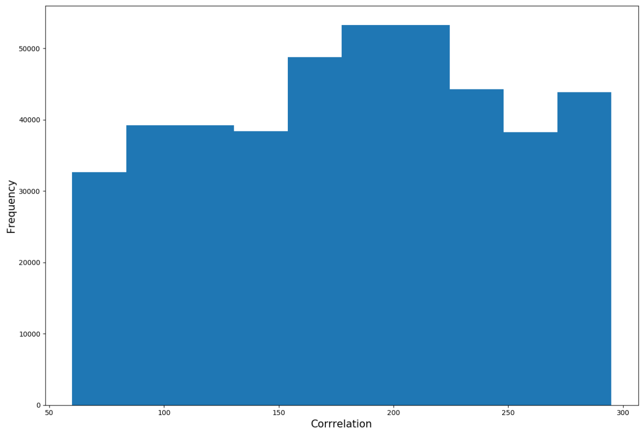

Figure 3: Time Distribution ¡5min

configuration. Consider TWO pairs of sensors (with strong correlation values), we can have the following scenarios:

Annual mean time less than 5 min = 182.07
Annual mean correlation less than 5min = 0.7108

Correlation	Distance	Configuration	Action
same	same	same	Areas are equal - look at volume
same	same	diff	Areas same function
same	diff	same	weight by distance effect
same	diff	diff	weight by configuration effect
diff	same	same	Areas diff function
diff	same	diff	weight by configuration effect
diff	diff	same	weight by distance effect
diff	diff	diff	Areas diff function

The TE will inform us about the micro-changes and flows of people Street network type. Compare the possible routes.
Consider the array of sensors shown at Figure . Let's assume that we have a flow of people walking past the location 116 and then diffusing towards the remaining sensors. Counts generated by the sensor are aggregated per five minute intervals, so if, for example, it takes one minute to walk from the location 116 to the location 117, the number of people detected at 117 from minutes 2 to 5, would correspond to the percentage of people detected at 116 from minutes 1 to 4. In other words, the similarity between the time series of counts at the locations under consideration are correlated. The aim is to, without actually tracking people, provide a measure for the size of the flow between each pair of sensors.

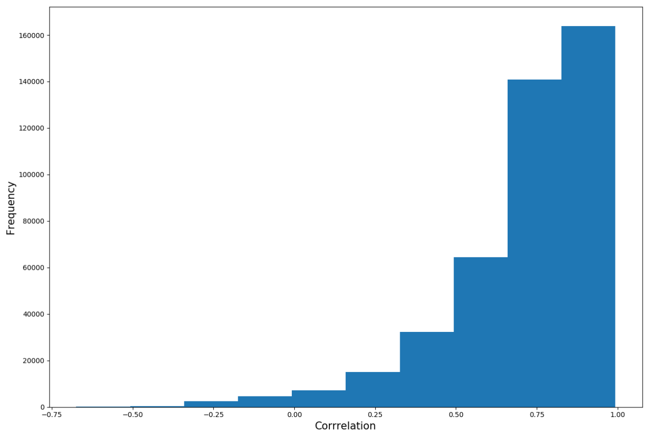

Figure 4: Correlation Distribution ¡ 5min

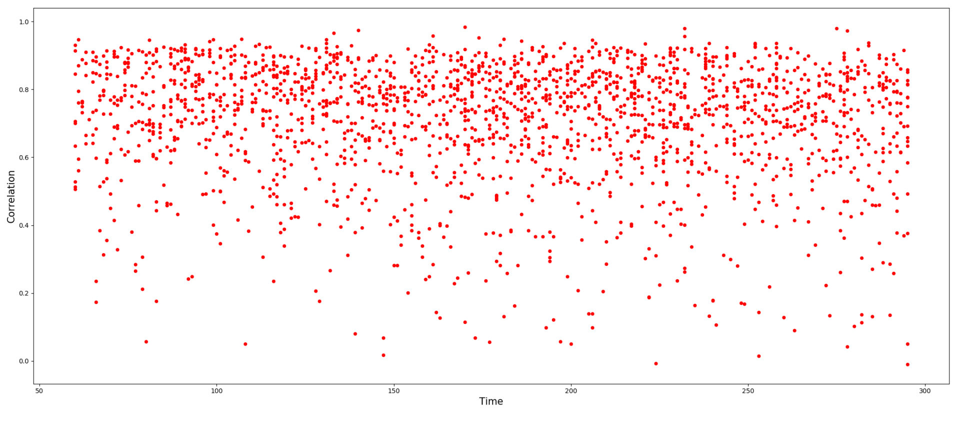

Figure 5: Time - Correlation

One way to accomplish this, is to think of this motion of people as flows of information among distinctive sources, so we can relate the number of people reaching one sensor from another by measuring the uncertainty between two interacting random variables. For this, we used an information theory concept known as Transfer Entropy *TE* [5] defined by:

$$TE(X,Y) = \sum_{t=1} p(y_{t+1}, y_t, x_t) log(\frac{p(y_{t+1}|y_t, x_t)}{p(y_{t+1}|y_t})) \tag{1}$$

Where t indicates a given point in time. Basically, Eq. 1 measures the reduction in uncertainty at y_t, given x_t and y_{t-1} in comparison with the case when only y_{t-1} is known. If this measure is applied directly to our people's movement problem and $X = location_i$, $Y = location_j$ and t runs for a whole day, the *TE* would represent an indicator of the direction of the flow, as the counts at y_{t+1} are more accurately estimated using the information of x_t.

4 Results

Taking again Figure as a reference, we measured the TE between sensor 116 and the rest of the sensors. The walking time is not constant and each sensor has counts at all times i, i.e., there are people passing by these sensors that came from locations outside the network. The numbers at each line represent the TE measured between each pair of sensor locations. The largest TE value found was between 117 to 115. The asymmetry of the TE is clear here, as the value in the opposite direction (115 to 117) is considerably lower. Another interesting value is the pair 116-117, where TE(116,117) $<<$ TE(117,116). This demonstrates that in this four-way crossing the predominant direction of flow is from location 117 to location 116 (from the bottom of the figure upwards, or from west to east in reality). These results suggest that, in general there is a larger flow of people from West side to East side of Edgware road and larger flow of people from South to North. The results are consistent with our intuition that there is a larger flow of people from South to North along this road towards the Edgware road underground station.

There is still a series of situations yet to be addressed by this model, such as the decay of probabilities with distance and the number of interventions of opportunity encountered by people while walking from one sensor to another. However, this first initial set of results is encouraging..

- It's really important to define the time frame for the calculations. It cannot be the whole day. It could be Morning, lunch, Afternoon and for a restricted number of places, nighttime. Check if we have sensors around night tube stations.
- Pedestrian dynamics is a term used to define the movement and interaction of people, both with one another and the built environment. Analysis of pedestrian dynamics in a robust and scientific manner helps professionals

from architects and transport planners to fire engineers and security advisors optimise the design and operation of a facility.

- The assessment of people movement is rapidly becoming a requisite in the design and operation of new and remodelled facilities. This is especially true where space is limited or the potential for increased demand is significant.
- In these circumstances, the possibility of uncomfortable and even dangerous conditions developing is significant. In order to build facilities that permit safe and efficient circulation, as well as effective emergency evacuation, it is necessary to understand and predict the way in which people move through the built environment.
- From mobile phone data to the spatial structure of cities - Reproduce Fig 4,5 and the idea of identify temporal hotspots and only over these make the TE analysis. We need to compare areas between cities and inner areas in London.
- Modelling Passenger Flows in Public Transport Facilities - This is a thesis , sections 3.3 and 4.7.1 is relevant to validate our results. Section 4.2 for general introduction is useful.
- A Tale of Many Cities: Universal Patterns in Human Urban Mobility - Lots of good references. This one is really important for us as propose a probability of transition from A to B

To get the distance Matrix
Road Graph
plugin1
plugin2

References

[1] Louail, T. et al., 2014. From mobile phone data to the spatial structure of cities. arXiv:1401.4540v1 [physics.soc-ph], 18, pp.114.

[2] Steenbruggen, J. et al., 2013. Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities. GeoJournal, 78(2), pp.223243.

[3] Brockmann, D. et al., 2006. The scaling laws of human travel. Nature, 439(7075), pp.462465.

[4] Song, C. et al., 2010. Limits of Predictability in Human Mobility. Science, 327(5968), pp.10181021.

[5] Schreiber T. (2000). Measuring Information Transfer Phys. Rev. Lett. , 85,461464