# Trends in urban flows: An information theory approach

Balamurugan Soundararaj[1], Karlo Lugomer[1]and Roberto Murcio[1]

## I. 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 flow 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. This research intends to address this problem by examining the movement of people in a SmartStreetSensors network at a fine spatial and temporal resolution. A novel methodology to the field of Big Data using mathematical models from information theory is introduced, taking an area in central London as a case study.

## II. DATA

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.

## III. METHODS

Consider the array of sensors shown at Figure 2. 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,

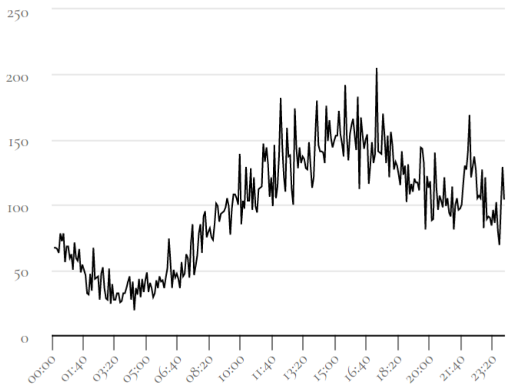


Fig. 1. Five-minute footfall counts at Edgware rd., London (location 146) on 02-July-2017

provide a measure for the size of the flow between each pair of sensors.

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:

85

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

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 $x_t$.

## IV. RESULTS

Taking again Figure 2 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

[1]Department of Geography, University College London

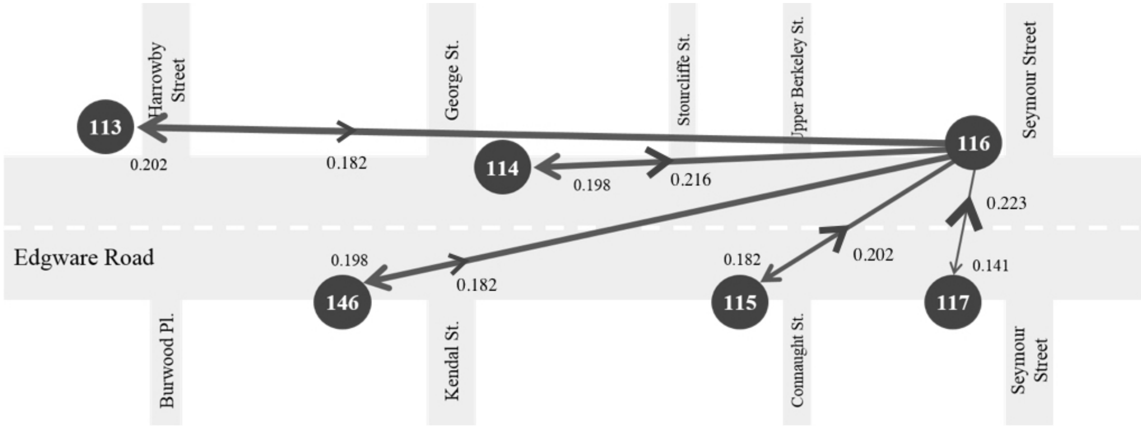


Fig. 2. Relationship of the transfer entropy between sensor to the distance between them

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.

## 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