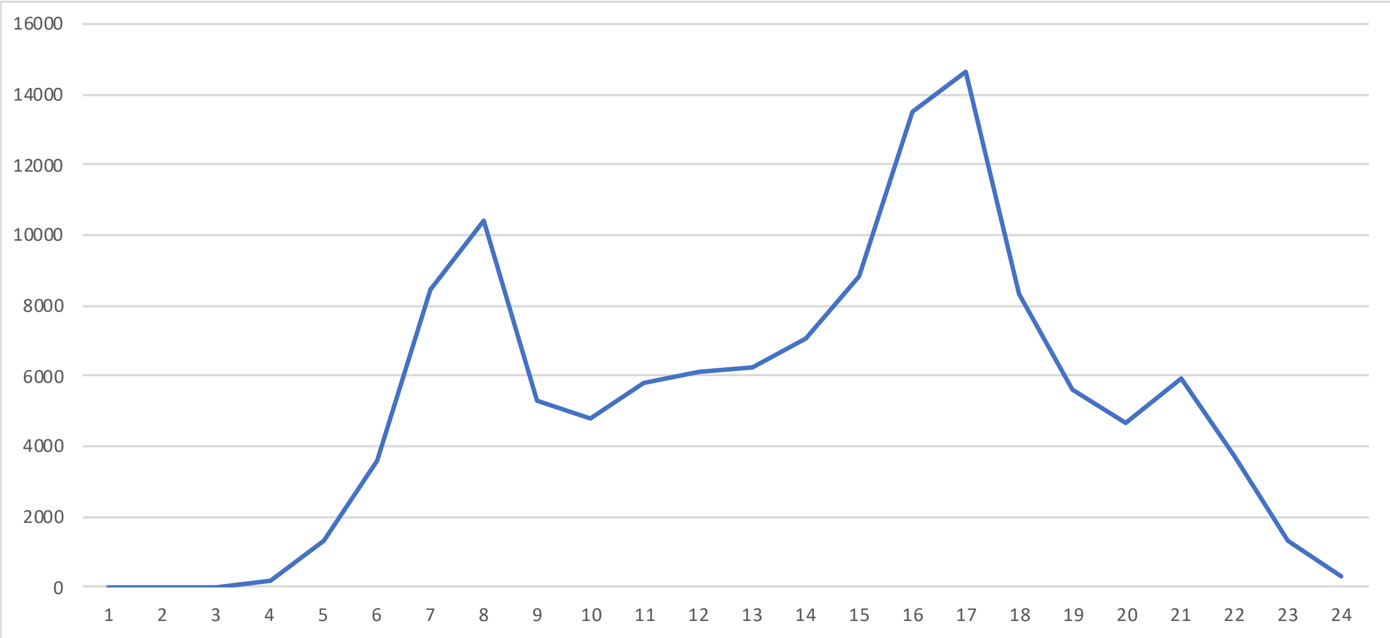
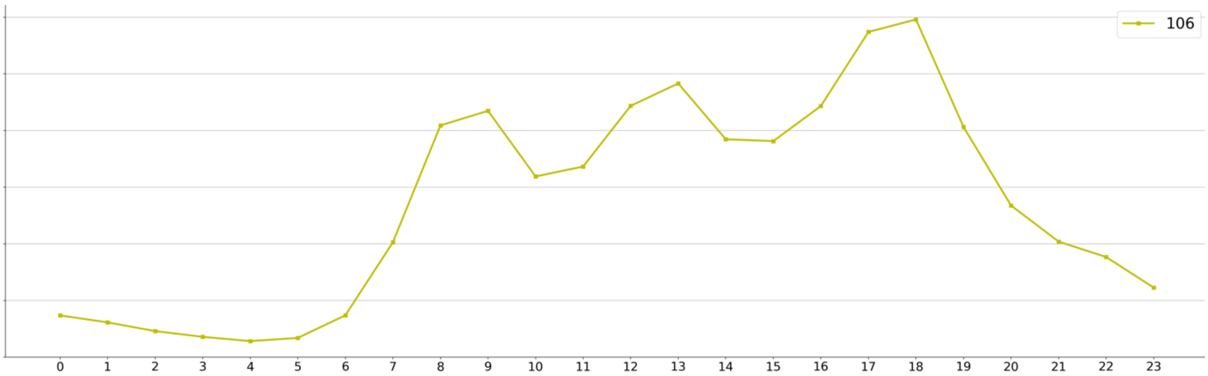
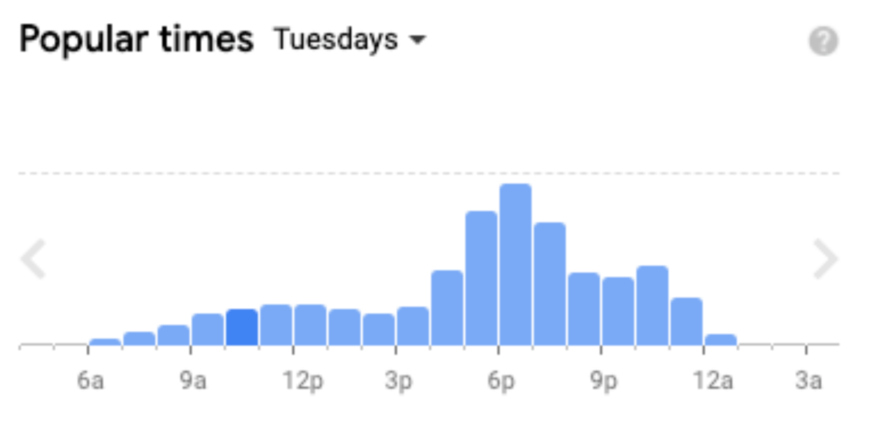
Tottenham court Road Tube station 15min aggregated counts 2017 weekday



Sensor at the corner of Oxford St and Tottenham Court Rd



Google Popular times Tottenham Court Rd Tube station



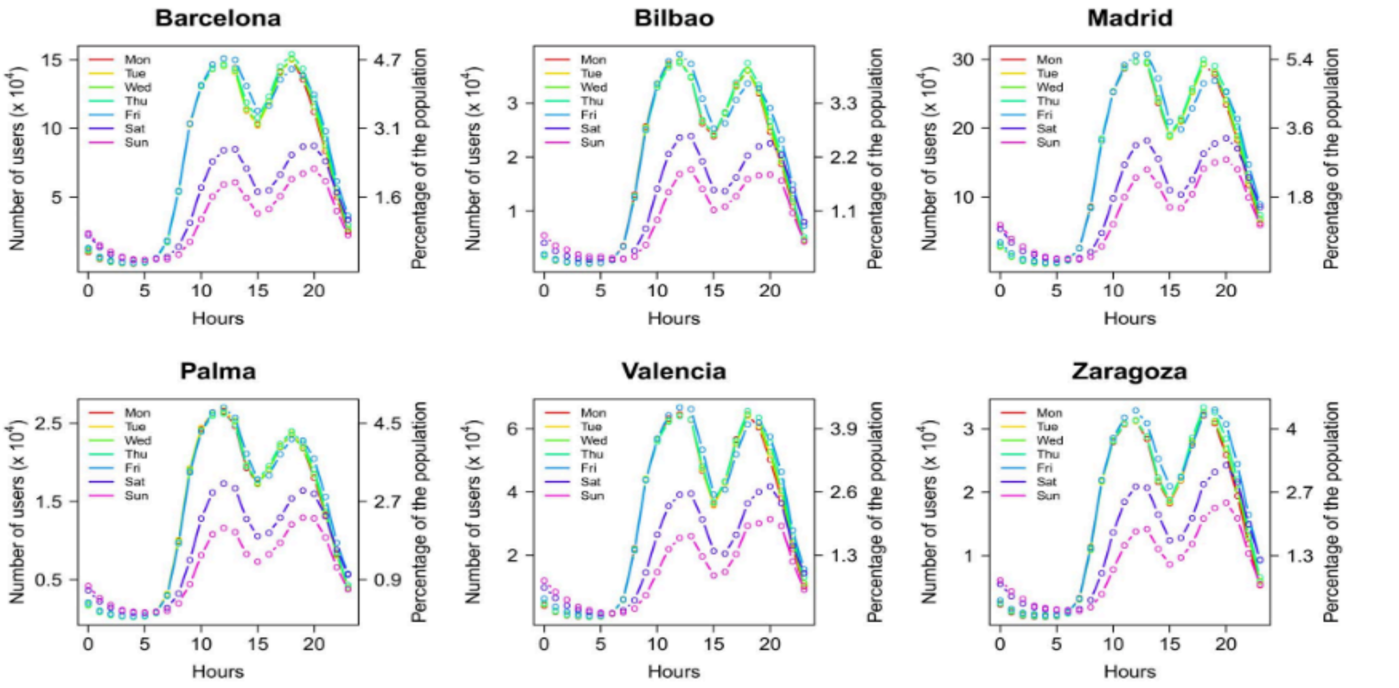
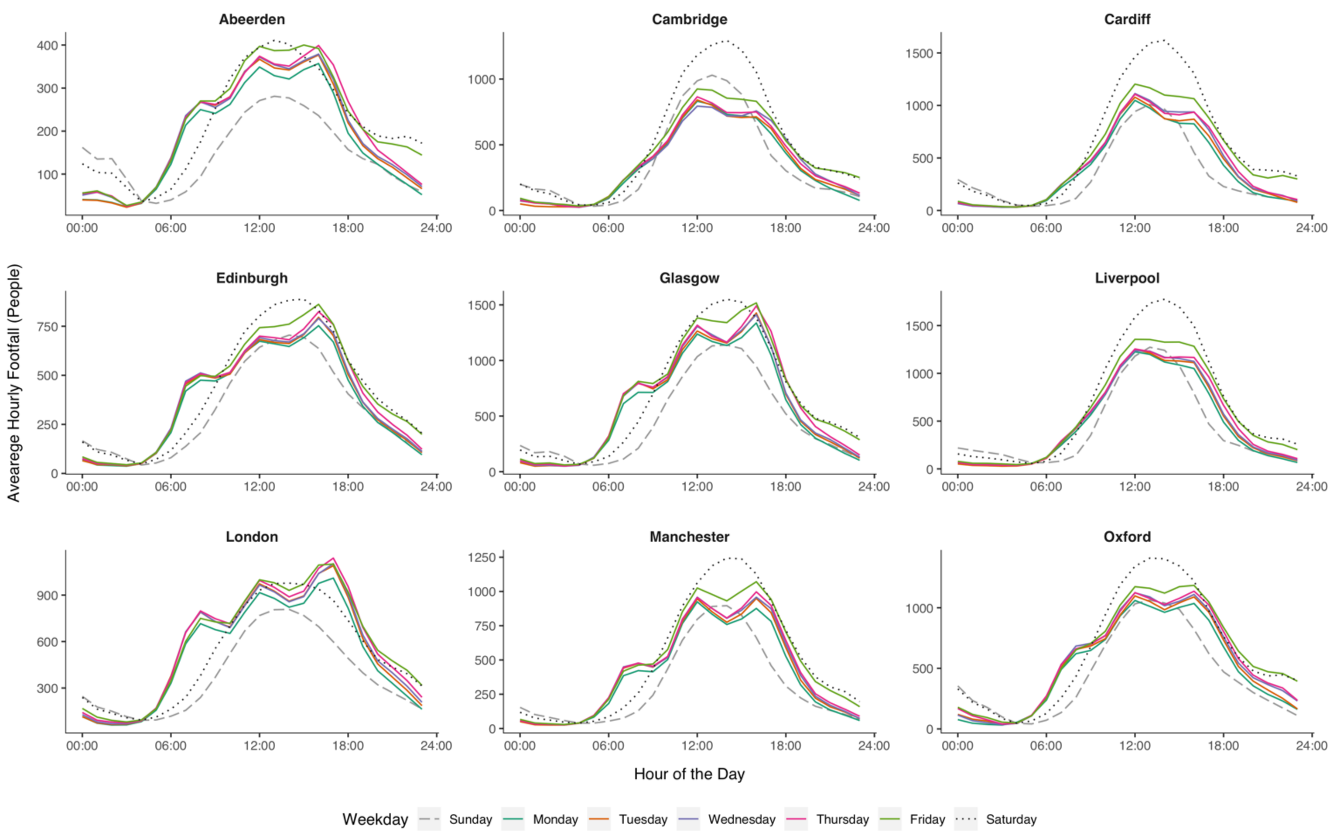


Figure 4 | Number of mobile phone users according to the hour of the day, for each day of the week, in six Spanish metropolitan areas. This figure was created with R.

From mobile phone data to the spatial structure of cities - Thomas Louail et al. DOI: 10.1038/srep05276

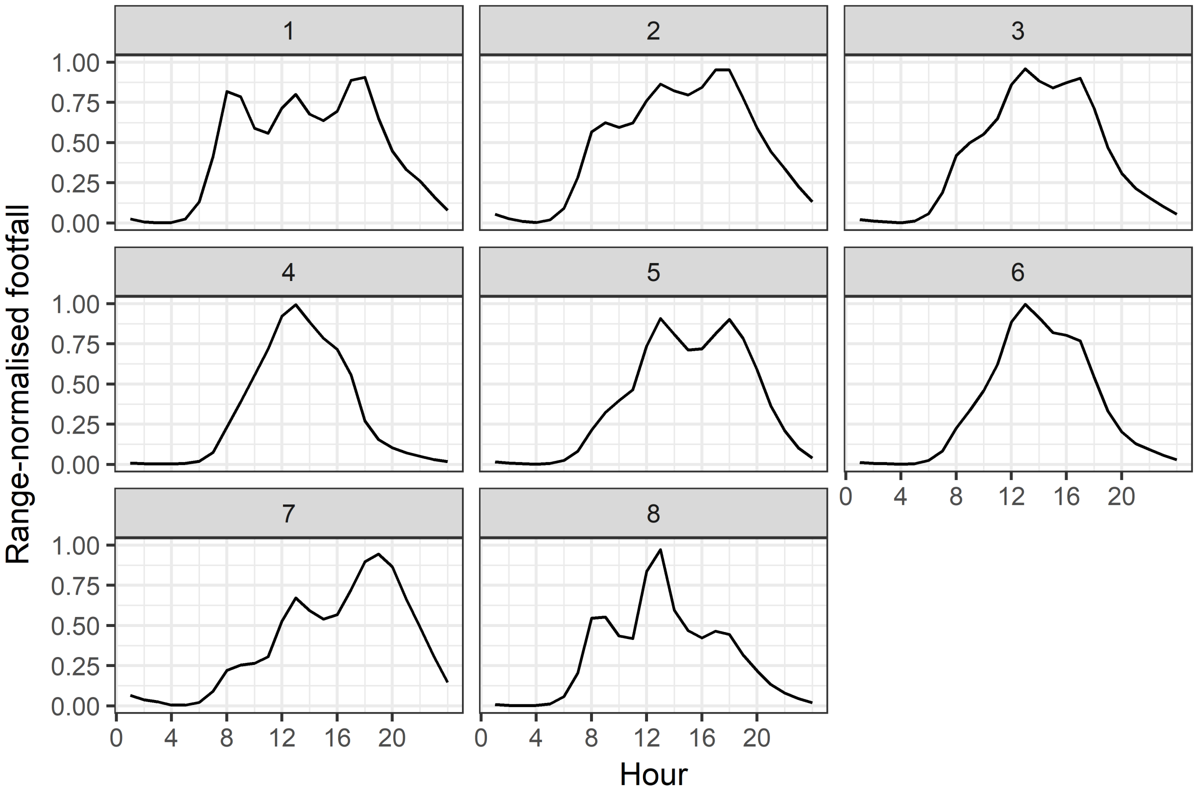


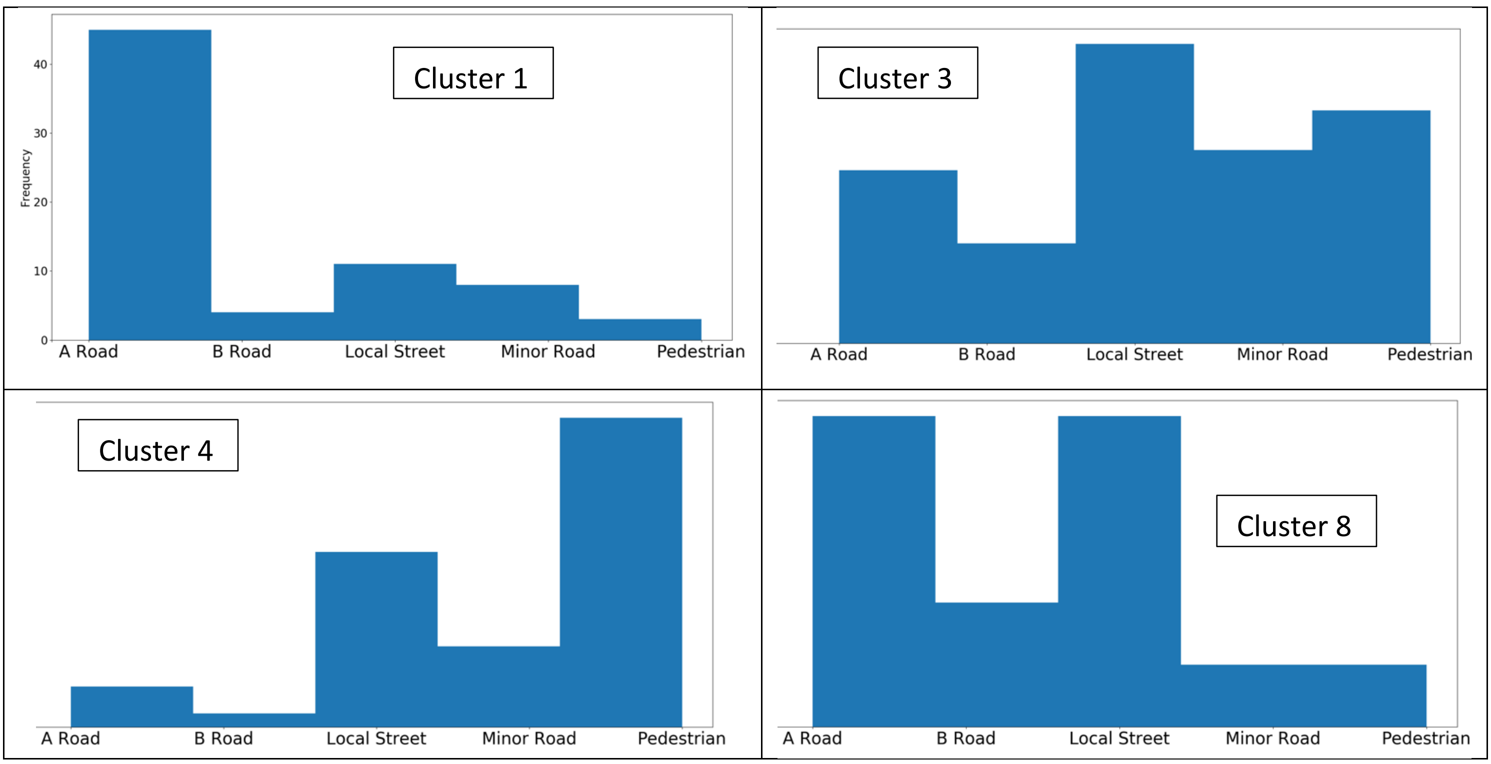
Main results

1. Is there a typical FF signal?

YES! But not a single one

DTW for all sensors 2017





Urban flows

(or coupling locations)

With these data, can we say something about the flow of people in the city?

Can a, let's say, time series analysis or another statistical approach throw some light in the μ-dynamics of the different locations?

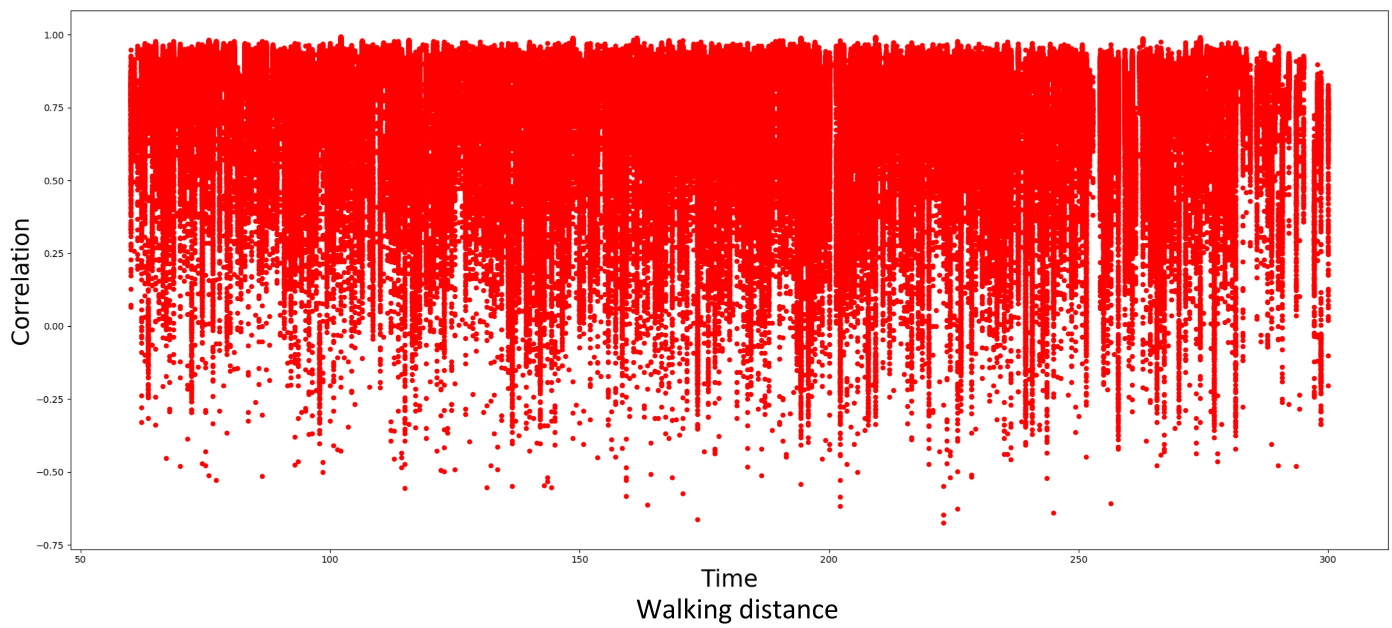
Two sets of sensor's pairs depending on the network walking distance between them:

$Wd > 5min$;

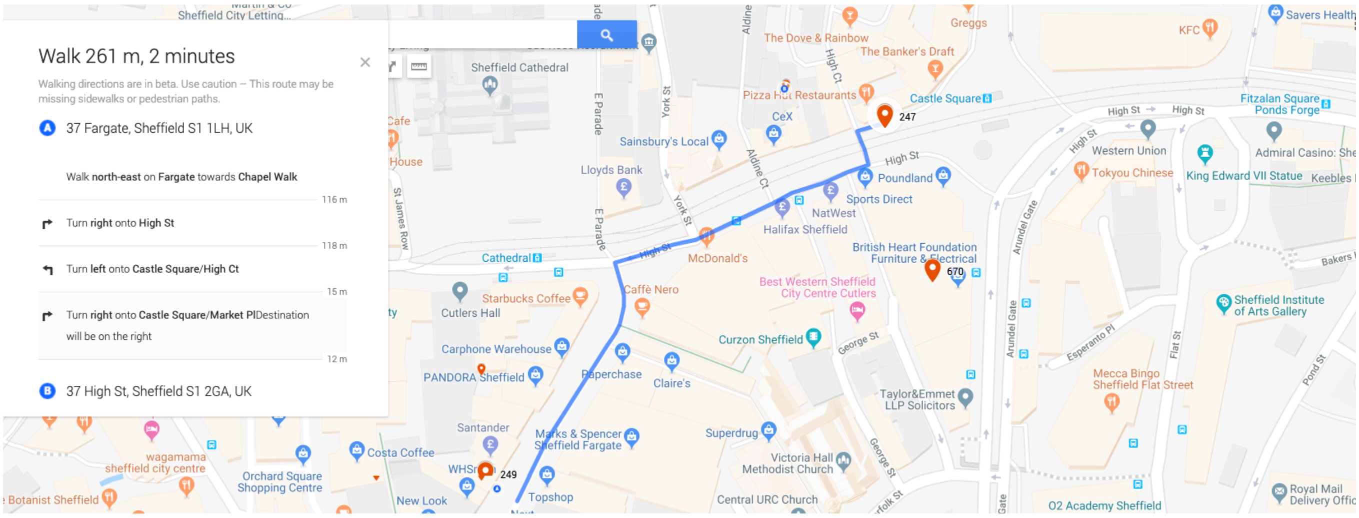
$1min \leq Wd \leq 5min$ (844 pairs)

In this research, we kept the second group and calculate the daily correlations between each pair of FF signals

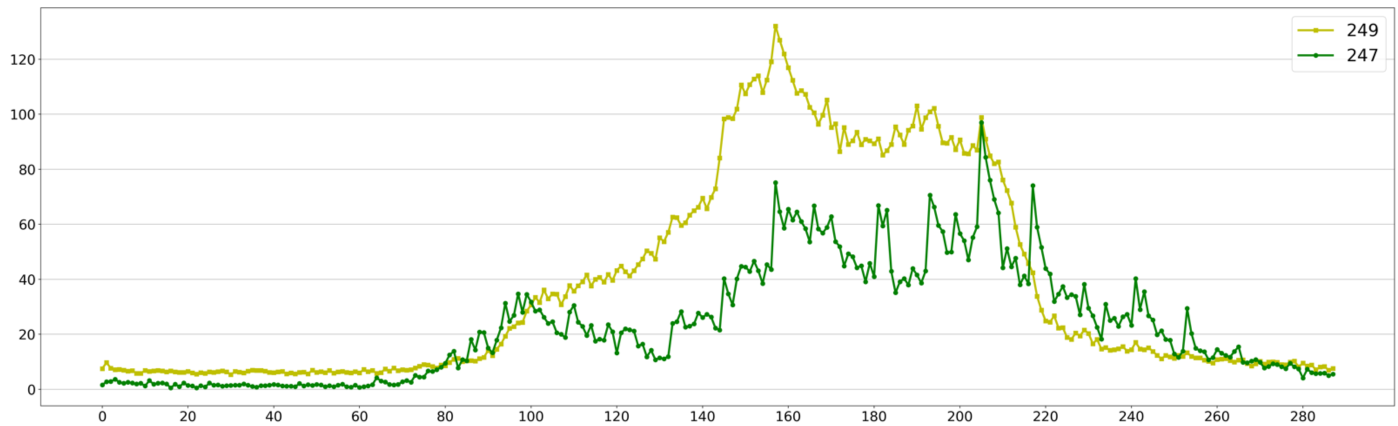
Weekday correlation 2017



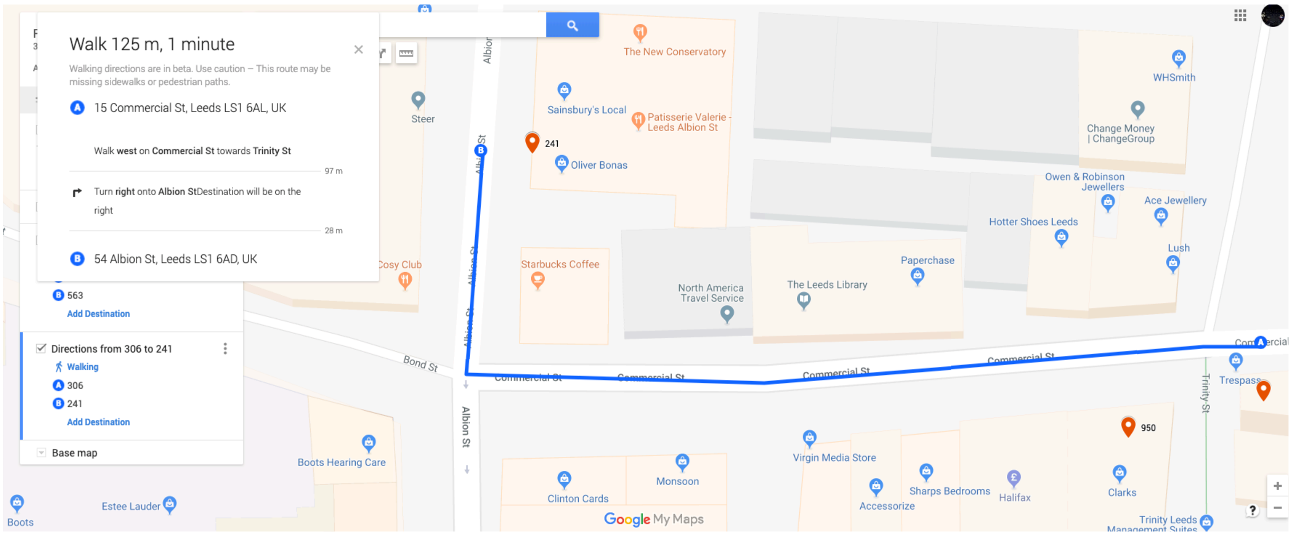
HCHT



This pair shares hours on 297 days in 2017
127 weekdays with HC
68 weekdays with LC



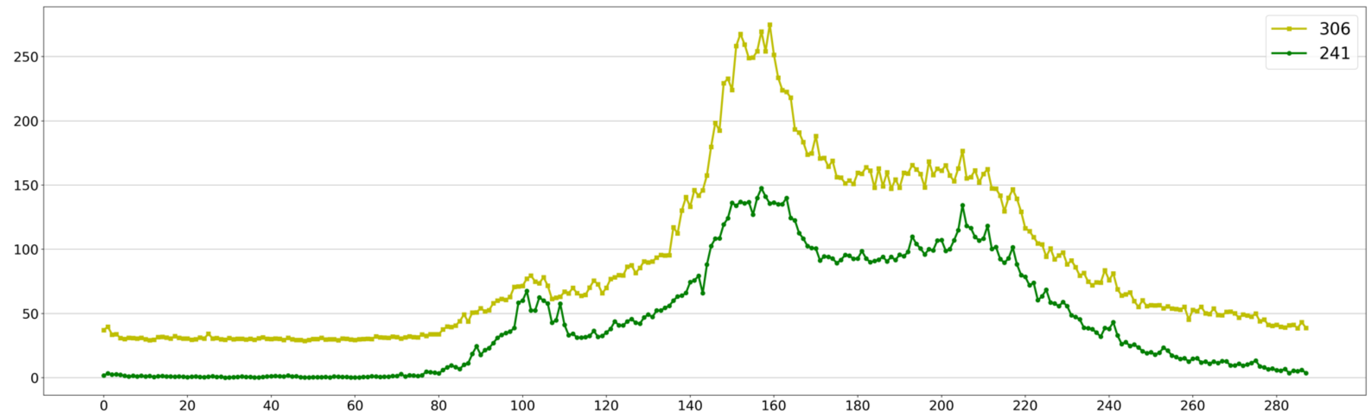
HCLT

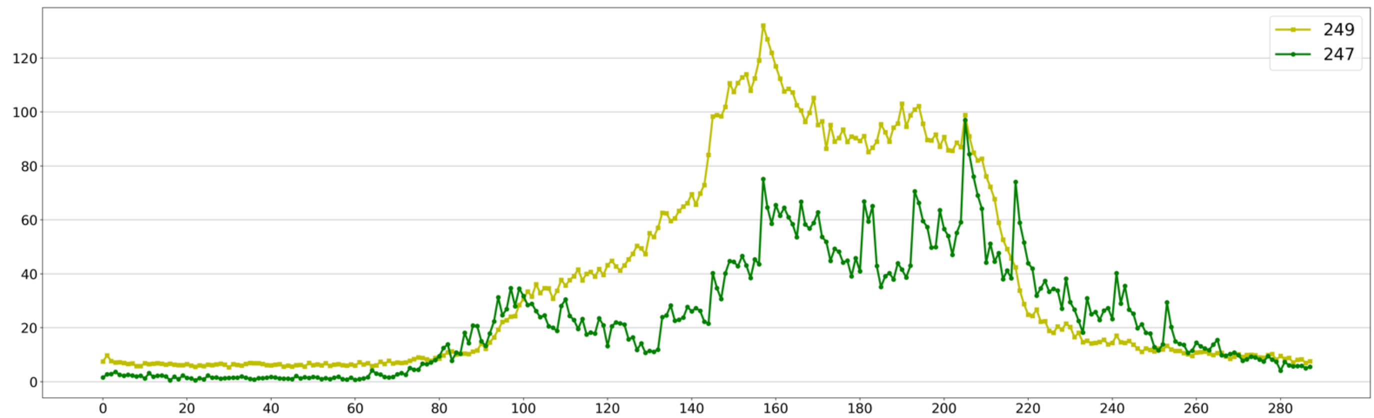


This pair shares hours on 305 days in 2017

211 weekdays with HC

1 weekday with LC





Information measures

Shannon entropy represents the basic measure of information is the preferred measure for detecting the reduction in uncertainty by any measurement x of a random variable whose probability is p(x)

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \qquad , \sum_{x} p(x) = 1$$

Extending Shannon entropy to measure the uncertainty between two interacting random variables X and Y – Mutual Information

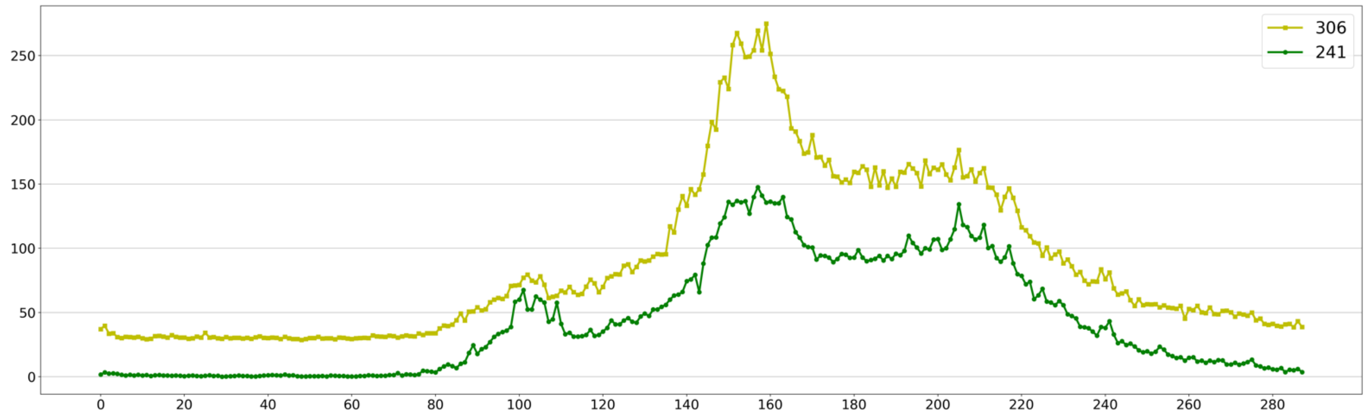
$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x).p(y)}$$

Transfer entropy, TE, was developed by Schreiber to overcome the time symmetric limitation of mutual information

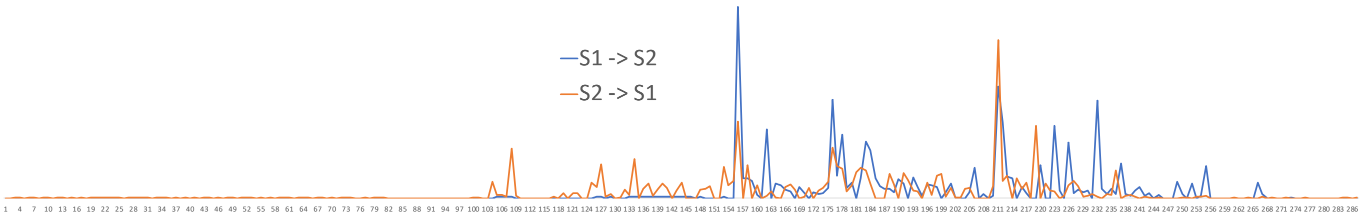
$$T(Y,X) = \sum_{t=1} p(x_{t+1}, x_t, y_t) \log \left(\frac{p(x_{t+1}, x_t, y_t).p(x_t)}{p(x_t, y_t).p(x_{t+1}, x_t)} \right)$$

$$T(X,Y) = \sum_{t=1} p(y_{t+1}, y_t, x_t) \log \left(\frac{p(y_{t+1}, y_t, x_t).p(y_t)}{p(y_t, x_t).p(y_{t+1}, y_t)} \right)$$

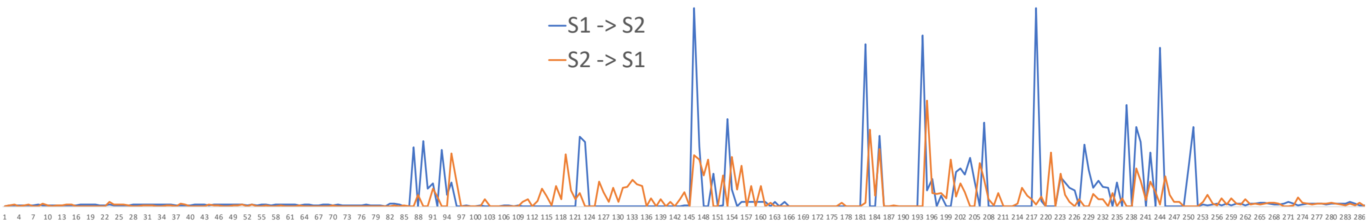
We are using a modified version of these formulation known as Local Transfer Entropy

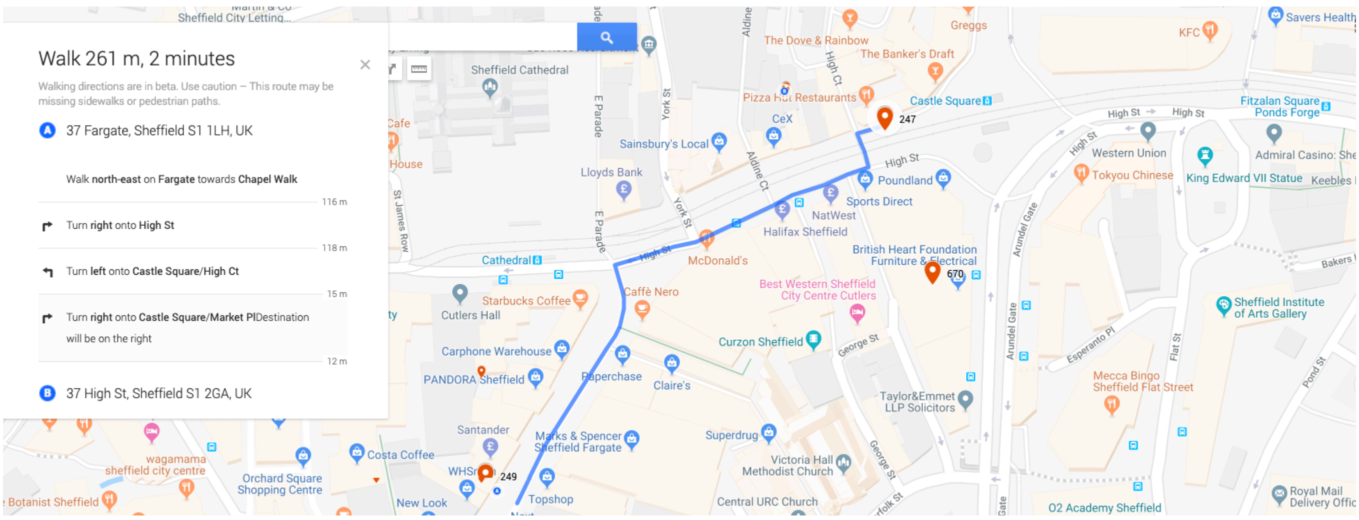


TE for a particular day on quad 1



TE for a particular day on quad 4





We found that we can explain, in ~90% of the cases, the different regimes in different weekdays for the same pair:

For HCHT the flow between locations should be equal

For LCHT the flow between locations should be greater from one to the other.

And at some extent we can apply the same line of thought for different pairs. However, we have an important number of pair locations that appear in different quadrants but share the same TE profile.

In general, the high / low correlation doesn't depend on the distance or the flows (or the counts at each location). So, it must be something else

When ask for walking directions, the Google API response with a series of instructions that share a set of five basic commands: **Walk, Turn, Slight, Continue and Take**

We extracted the walking directions for each pair and assign a weight to each one of those five words:

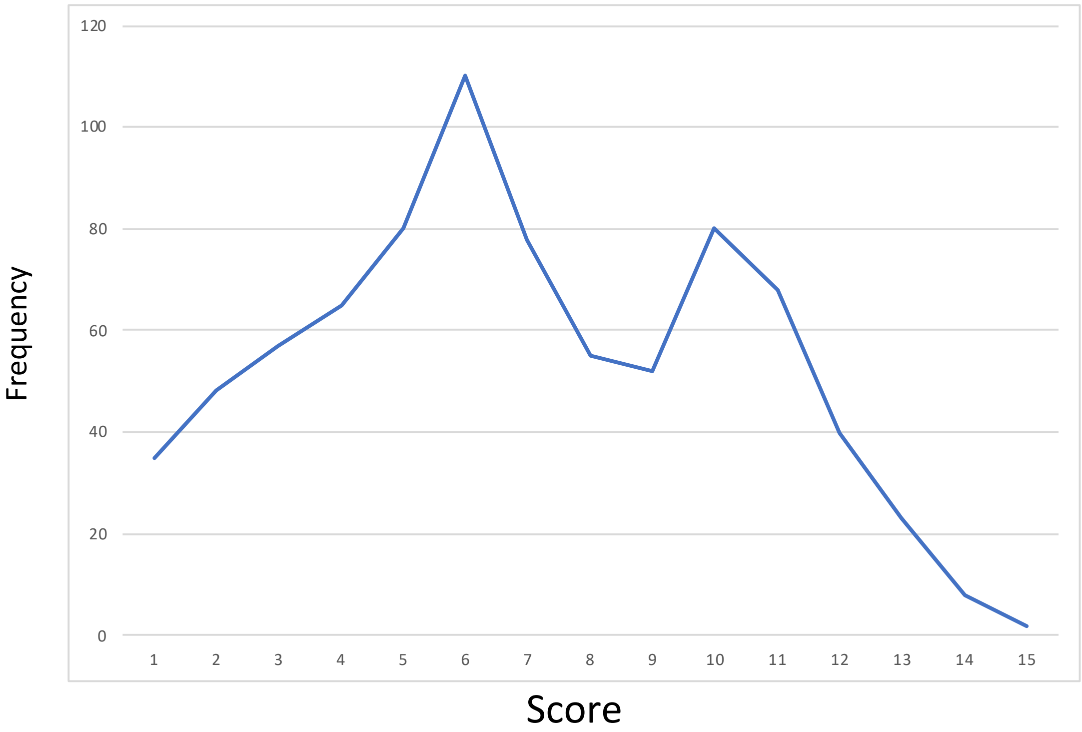
Instruction	Weight
Walk, Continue	1
Slight	2
Turn	3
Take	4

Then, we simply define a score for each route as:

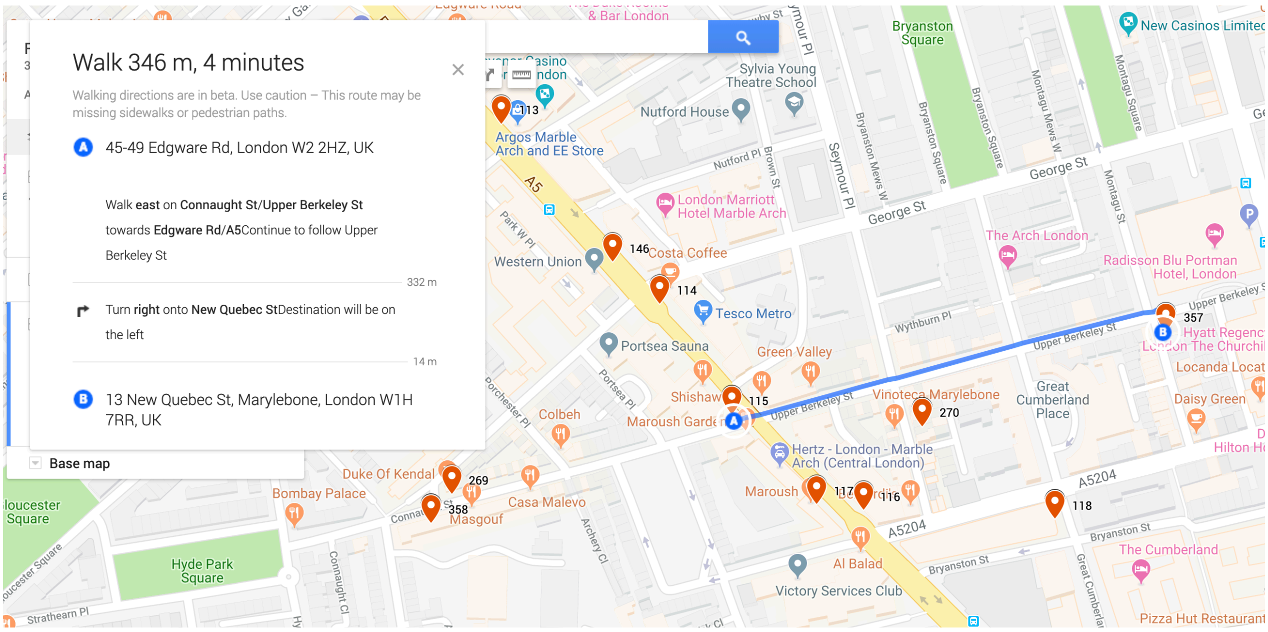
$$Score_r = \sum_{i=1\dots4} W_{ir}$$

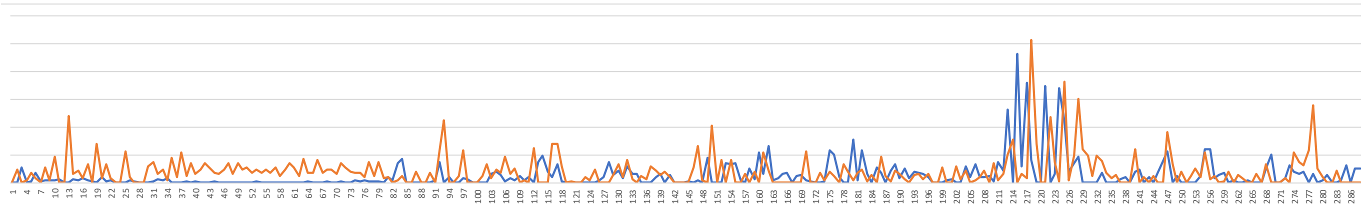
For example, the previous areas have a score of 10 and 4 respectively.

Our pairs exhibit a binomial distribution with respect to the score

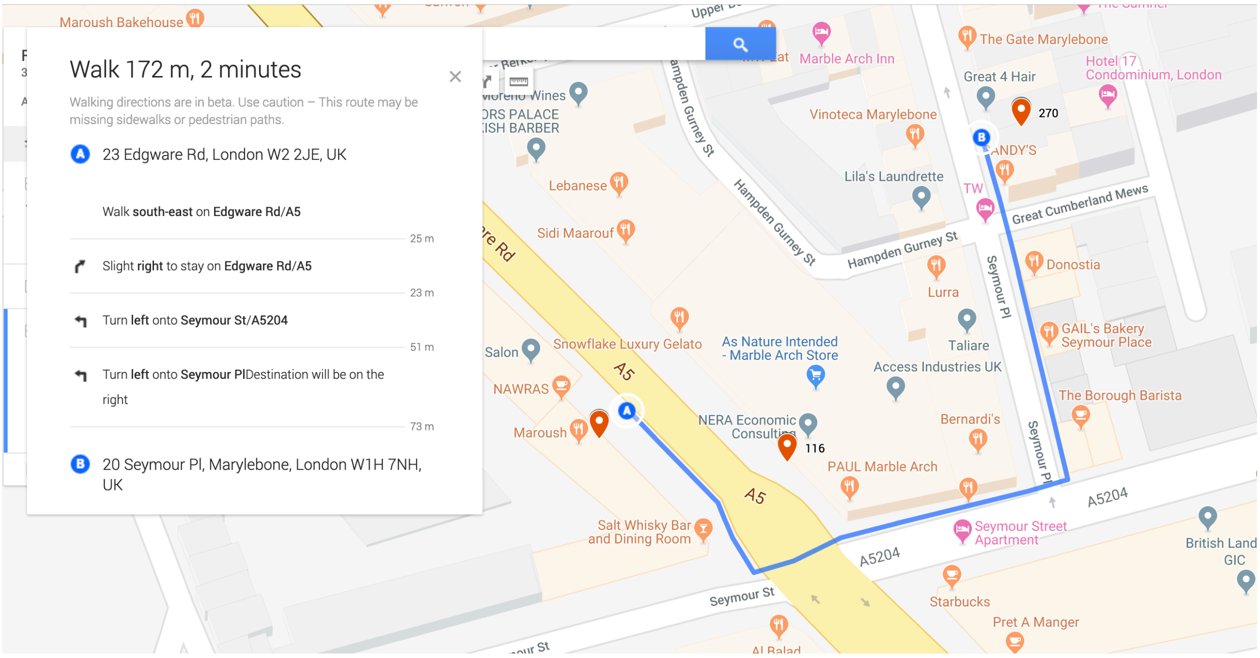


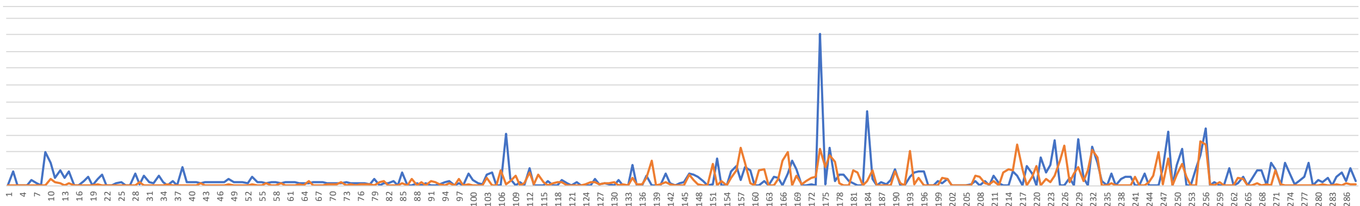
For example:
LCHT – Score 4





LCLT – Score 9





Routes with high score are less correlated in general on weekdays, BUT some routes with low scores are also low correlated → Intervention of oportunity

Conclusions

1. Can we use probe requests as a proxy to detect human activity? YES
2. Can we use FF signals to classify areas? YES
3. The relation between two signals is NOT a function of the walking distance between them.
4. We can have a feeling about the flow of people between two locations using LTE
5. Routes complexity could probably explain the correlations, but doesn't tell us anything about the flows.

TODO
Reweight using instruction + time
Hierarchy of locations → Renormalization group?

That's it!