

# Interpreting Google metrics by socio-economic background [working title]

Pavol Mulinka(1,2), Souneil Park(2) and Diego Perino(2)

(1) Czech Technical University in Prague, (2) Telefonica Research, Spain

pavol.mulinka@fel.cvut.cz, {souneil.park, diego.perino}@telefonica.com

**Abstract**—[TBD]

**Keywords**—[TBD]

## I. INTRODUCTION

[TBD]

Contributions of this paper are as follows:

- feasibility analysis of using network measurements for network planning based on socio-economic status of areas
- analysis and definition of user network behavior by correlation of their socio-economic background and Google network measurements
- analysis and evaluation of network service in the context of socio-economic background of the users
- analysis and interpretation of network measurements outliers based on the socio-economic knowledge of the area

## II. RELATED WORK

[TBD]

## III. RESULTS

We used Google Mobile Network Insights network performance data with a coordinate aggregation type. We merged it with an Index of Multiple Deprivation (IMD) Decile (where 1 is most deprived 10% of LSOAs). We focused on the area of London, UK.

### A. Google Mobile Network insights dataset description

Dataset consists of seven types of metrics: (i)Share of Observations, (ii)Share of Devices, (iii)Radio Access Technology, (iv)LTE Band Metrics, (v)Domestic Roaming Metrics, (vi)Signal Metrics, (vii)YouTube Metrics. We chose to aggregate the metrics by coordinates aggregation dimensions type of highest granularity of 280m for area of London, UK. This provides us highest granularity on LSOA user behavior insights. We chose the highest, five weeks period, time granularity to obtain dataset with highest density of coordinates.

[HOW TO GET DATASET/SOMETHING ABOUT API TBD by Diego]

*1) Metric analysis and filtering:* We filtered metrics based on their redundancy and usability by analyzing their definition, inter-correlation, correlation with LSOA IMD decile and boxplots for each IMD decile value. First, we filtered out types of metrics that are not related to network performance, i.e. : *Share of Observations, Share of Devices, LTE Band Metrics, Domestic Roaming Metrics*. Than we filtered out on types of *Radio Access Technology* not present in London, i.e. : *GSM %* and *CDMA %*. Than, we filtered percentile metrics from *Signal Metrics* as they provide us redundant information to original signal metrics.

[TBD - table/heatmap/something to show that percentile features have redundant information]

For boxplots of remaining metrics see Fig. 3.

RSRP/RSRQ interectional metrics "display the joint distribution of these two values for a given location". We kept all metrics to retain ability to put their trends into context. Their boxplots can be found in Fig. 4. Description of the values in diagrams:

- RSRQ[dB]: Poor  $\leq -13$ ; OK  $\langle -12, -8 \rangle$ ; Good  $\geq -7$
- RSRP[dB]: Poor  $\leq -110$ ; OK  $\langle -109, -90 \rangle$ ; Good  $\geq -89$

From YouTube metrics we filtered out *Percentage of observations and Users above threshold* for every resolution (i.e. 360p, 480p, 720p, 1080p, 4k). These metrics indicate "percentage of YouTube users who recorded goodput greater than or equal to the bitrate recommended for the specified resolution for 80% or more of their observations"[from Mobile Network Insights Manual]. We marked the recommended goodput bitrates in boxplot, see Fig. 5d, to include the notion of the filtered out metrics.

*2) Metric generalization value:* [TBD by Pavol] - Statistics that would show how well the London area is covered by aggregated coordinates from Google. Things that come to my mind : 1. put grid with 280x280m rectangles over a London area, see how many points could fit there vs how many points do we have from Google 2. put in contrast the size of the LSOAs Fig. 1 vs 280x280m grid vs number of NA values in YouTube metrics that were not filtered out in "Metric analysis and filtering" subsection.

### B. LSOA IMD dataset

[TBD by Souneil, short description from paper that Souneil submitted before Xmas]

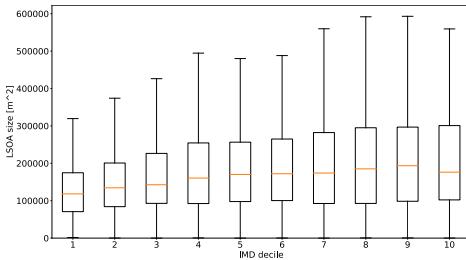


Fig. 1: LSOA size per IMD

### C. Merged of dataset

We joined Google Mobile Network insights dataset and LSOA IMD dataset by their geographical location. Resulting dataset consists of Google Mobile Network insights metrics subset with coordinates and LSOA IMD Decile values of LSOAs in which the coordinates reside.

### D. Merged dataset analysis

We computed the correlation coefficient between Google metrics and IMD values to see their relation, Fig. 2. We focused on metrics exhibiting correlation with IMD decile values and analysed box plots of all filtered Google metrics to put subset of metrics into perspective. We divided them into three categories : (i) Signal metrics Fig. 3, (ii) RSRP/RSRQ intersectional metrics Fig. 4, (iii) YouTube metrics Fig. 5.

#### 1) Google signal metrics:

- there is a noticeable correlation between IMD decile and LTE/WCDMA technologies, showing the higher class LSOAs tend to use more WCDMA technology compared to LTE
- LTE signal power [dBm] and LTE strong signal % is less present in higher class LSOAs. We are current guess is that this might be due to historical buildings and issues with introducing new technologies and hardware in these areas compared to new apartment buildings

#### 2) RSRQ/RSRP intersectional metrics:

- RSRQ is stable despite the signal power going down which means quality is assured, but higher class is less likely to use quality signal. At the same time YouTube goodput is equal among all IMD classes
- YouTube usage of 4k quality is ~10%; if the situation changes in upcoming years then the higher class citizen might experience quality issues as their YouTube goodput will not be sufficient

#### 3) Youtube metrics:

- despite lower LTE usage/coverage in high IMD areas, the areas have comparable YouTube goodput, start time, observed and reported video quality. This means that lower YouTube usage is the choice of users and not an outcome of worse bitrate

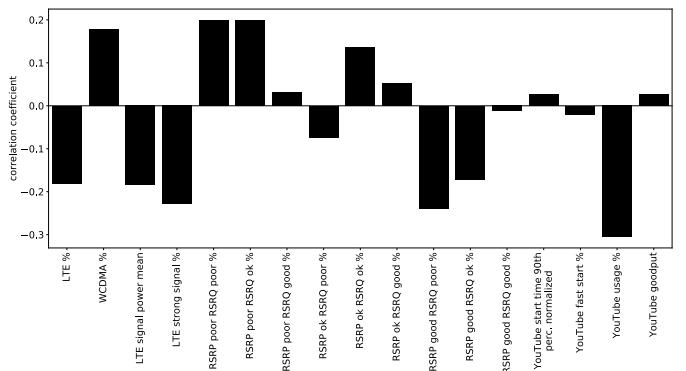


Fig. 2: Google metrics correlation with IMD decile.

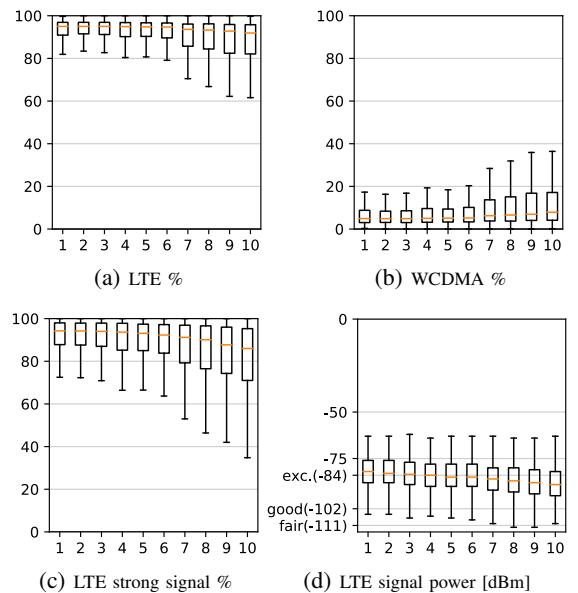


Fig. 3: Google signal metrics.

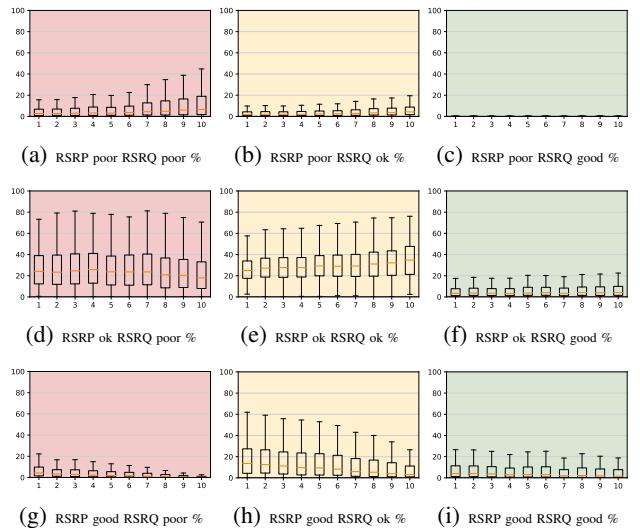


Fig. 4: Google RSRP/RSRQ intersectional metrics.

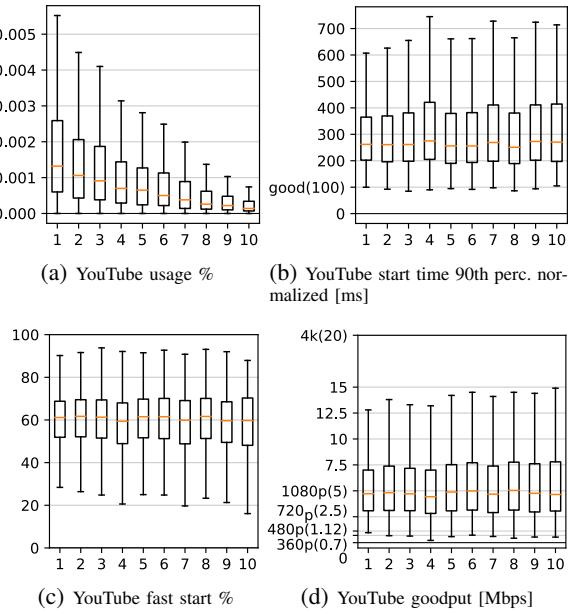


Fig. 5: *Google YouTube metrics.*

#### E. Outliers analysis

We analysed the outliers, i.e. fliers in the boxplots, by visualizing them on a map and putting them into context of their neighborhood and IMD value for give LSOAs, Fig. 6. [TBD by Pavol - maybe change the visualization or the description under the maps to make it easier to understand and search for explanation of the outliers]

#### F. Appendix

This subsection is just informational and is not usable for the purpose of the conference paper - for the moment at least. We created a two datasets from LSOA polygon dataset containing IMD decile values. First dataset contains LSOA name and its IMD decile value and seconda contains all neighbors of each LSOA. We fed these dataset to spatial machine learning algorithms in ELKI data mining software to see if there are any LSOA outliers Fig. 7. By LSOA outlier we understand an LSOA that has IMD value that is significantly locally distinct from its neighbors. The idea is that LSOAs with Google metric outliers might be the ones that are also spatial IMD decile outliers, eg. LSOA with IMD 10 has outliers because it is surrounded by LSOA that have IMD 1.

## IV. CONCLUSION AND FUTURE WORK

[TBD]

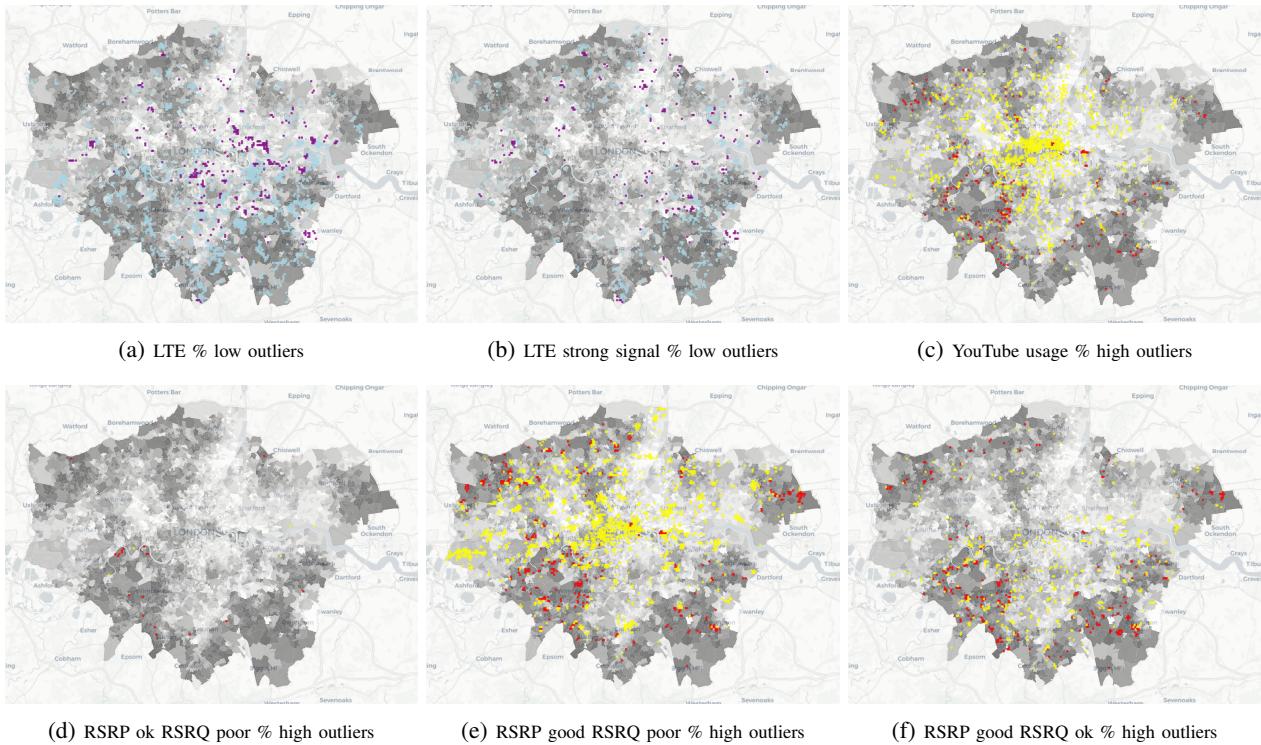


Fig. 6: *Google metrics outliers*. We differentiate between high(yellow-red) and low(blue-purple) outliers, i.e. outliers above and below whiskers in the boxplots. Red and purple dots show outliers in the IMD classes where the value of given metric is atypical. Red marks the high outliers in 9-10 IMD Decile LSOAs whereas purple marks low outliers in 1-2 IMD Decile LSOAs.

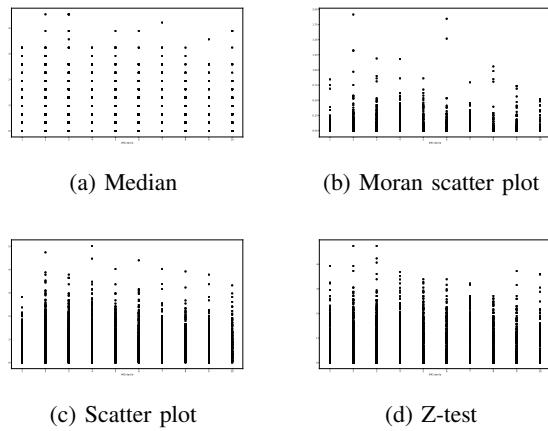


Fig. 7: *IMD spatial outliers*