Analysis of Mobility Patterns during a Large Social Event

Gergő Pintér, László Nádai and Imre Felde

August 3, 2018

Analysis of Mobility Patterns during a Large Social Event

Gergő Pintér*, László Nádai*, Imre Felde*
*Óbuda University/John von Neumann Faculty of Informatics, Budapest, Hungary
pinter.gergo@nik.uni-obuda.hu, nadai@uni-obuda.hu, felde@uni-obuda.hu

***Abstract*—The cellular networks including the millions of mobile devices generate a huge volume (mobile phone call records) of data by which the movements of people could be sensed. In this paper, a method (based on mobile phone call records) applied to characterize the patterns of urban population during a big social event is underlined. The proposed method is also to show the regions in the city where the similar trend of mobility takes place. This paper demonstrates the applicability of the approach on the data provided by a Hungarian mobile network operator during a public demonstration event. The results obtained can be helpful in understanding the structure of the city in the sense of human mobility related to large-scale events.**

I. Introduction

The recent cellular communication technologies, a.k.a the mobile phone network provides directly or indirectly the "sense" in some aspects of the human behavior. The digital fingerprints that the population leave behind during using the mobile devices reflect their behavior in a certain granularity.

In the last two decades, massive mobile phone location data have been studied and shown to have great potential to numerically characterize the human mobility [1] [2] [3]. In addition, studies on long-term mobility data showed how this kind of data could be a helping tool for improving the quality of urban life [4].

The ability to recognize and evaluate mobility patterns provides huge information for spatio-temporal data, that could be understood in aggregated urban mobility patterns [2], therefore it opens the door to a wide range of applications ranging from urban transportation to crime. Most data sources collected by the cellular network operators provide information about the mobile devices only if it shows activity like starting/receiving calls, sending/receiving text messages or transferring data. Consequently, the position of a device during most of the time is unknown. This fact leads to several problems in finding mobility patterns by means of widely used data mining techniques. The aim of the work is to test and present an approach for spotting quantitatively significant regularities in the way the population density deviates from the expected values on different areas of the urban area. In particular, discovering groups of regions that consistently behave in a coordinated way is aimed, suggesting the existence of some kind of connection among them. It is stressed that the objective is to provide to the domain expert hints about patterns in the life of the city/geographical area that emerge directly from data and that might be unknown or simply unexpected. That contrasts with other approaches to human mobility study that starts from a preconceived framework or model, and aim at positioning the city within the framework or at estimating some parameters to fit the model.

The used method of the analyses is similar to [5]. The human traffic flows are analyzed based on phone data during a big social event, which focuses on detecting the movements of the people among the areas.

The paper is organized as follows. First, the used Framework is introduced (Section II). In Section III, the data derived from the cellular network is presented. In Section IV, the approach suggested to characterize the mobility patterns is described. The method suggested is tested on a big social event, the findings are shown in Section V. Finally, the conclusions are summarized.

II. Framework applied

The proposed methodology aims at providing insights for better understanding the movement of the people in the urban environment during big social events. Two sets of data have been applied for the study as input data for the introduced methodology:

- The mobile device activities of anonymous users recorded by Vodafone Hungary (Call Detail Records, CDR)
- The map of GSM cells of the operator including all the geographical information (position of antennas, geometry of the cell)

A Big Data application framework (Fig 1) has been developed to process and present the activities of mobile phone users [6].

III. Cellular Network data

The data source is anonymous (CDR) from a mobile operator Vodafone in Hungary, focusing in Budapest, in April 2017. This operator has over 2.4 million active clients in Hungary. The case study area, of about 525.2 km^2, was identified inside the City of Budapest, the capital of Hungary.

Anonymous Call Detail Records (CDR) were received from Vodafone Hungary, which had about 25% market share in September, 2016 [7], recorded during the April of 2017. In addition to the Call Detail Records, information about the cells such as the centroid of the cell was also received. A Call Detail Record contains a timestamp, a device ID and a cell ID, thus devices (and via them people) can be connected to cells and rough geographic position. The precision of the positioning depends on the area of the cell. In downtown, cells

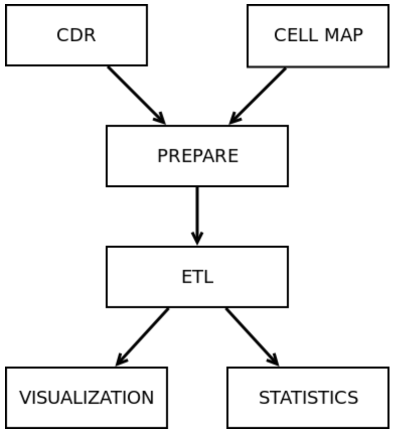


Figure 1. The scheme of the framework developed

are smaller and placed more densely than in the suburban areas and in the countryside. From another aspect the Call Detail Records can also be used to measure the load of the cell, so roughly determine how many people are present in an area in a given time.

The analysis is focused on the mobility of the people on 9th of April, 2017 in Budapest, when a demonstration was held for the Central European University (CEU) and against the modification of the higher education law.

This event motivated tens of thousands of people [8] on a Sunday afternoon to participate in the movement. The demonstration started at the Castle Garden (Fig. 2a) from where the demonstrators went to the Hungarian Parliament Building (Fig. 2d) through the Széchenyi Chain Bridge (Fig. 2b) and via the campus of the Central European University (Fig. 2c). Though the demonstration officially ended at the Parliament, a portion of the people kept on demonstrating and went to the Oktogon (Fig. 2e), a major intersection of Budapest and then to the City Park (Fig. 2f).

Roughly 11,000 unique device IDs were registered during the demonstration in those cells that can reach this path. Assuming that the choice of cell phone operator among the demonstrators corresponds with the nationwide trends then the total number of the crowd could be even fourfold of this number, but it has to be noted that it cannot be knonw for sure that if someone was actually participated in the demonstration. On the other hand it is also not realistic that every demonstrator used their phones during the demonstration.

The daily records before the demonstration are used as training or reference data and the records from the day of the demonstration as test data. For every cell which centroid is inside the administrative borders of Budapest, the daily activity change are calculated. Fig. 3 shows the activity changes of three cells. These cells cover some area around the main locations of the event, but they are not the only affected ones. Fig. 3a shows the activity of cell 4019, which is at the Castle Garden where the demonstration started. The peak is around 17:00 hour when the event has been officially started. The activity raise in the cell 4362 (Fig. 3b) is also significant.

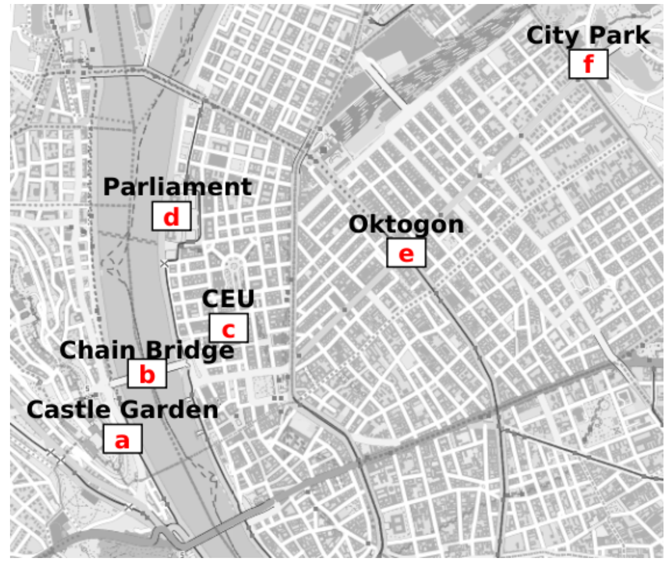


Figure 2. The main locations of the demonstration

This one is at the CEU building, and the peak is around 18:00 o'clock. Fig. 3c shows the activity of the cell 4434 which is at the neighborhood of the Hungarian Parliament Building and the peak is around the official end of the demonstration (19:00 hour).

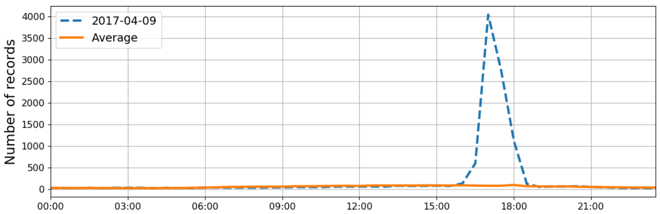
IV. The method applied

The CDR data is discretized on both dimensions. As for the time dimension, the records are aggregated into thirty-minute intervals for every cell and every day of the training data. The first eight days of the month (1-8 April, 2017) was used as a training data to determine the average usage of the cells for every thirty-minute intervals. The day of the demonstration (9th April 2017) was used as test data. Both the train (**T**) and the test (**S**) data can be treated as a matrix, where rows represent the cells and the columns represent the time intervals.

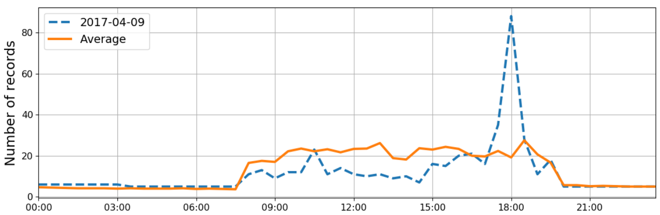
The ratio matrix (**R**) of the activity during the examined day or days and during the days used as reference are calculated according to the equation 1. The train matrix is subtracted from the test matrix and the result matrix is divided element-wisely by the test matrix to get the element-wise ratios. These ratios are then discretized into categories denoted with letters from A to U. The intervals for the discretization are chosen to handle small and big activity changes as well, therefore the intervals are wider at the edges of the spectrum (nearing to the letter A and U), see figure 4.

$$R = (S - T) \oslash T \tag{1}$$

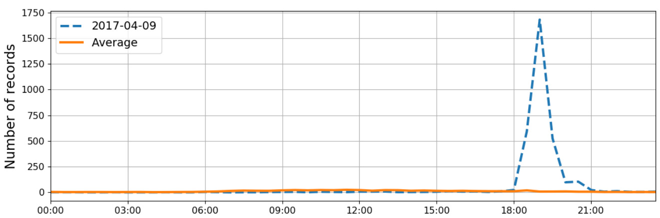
The activity rates denoted with letters are then handled as a string for every cell. From these exactly 48 letter long words (hence every letter represents the deviation of activity for 30 minutes), 2-grams and 3-grams are composed and used for pattern mining. this method is also used in Natural Language Processing [9] [10]. Two or three letter patterns are examined



(a) Activity of cell 4019 (at the Cartle Garden)



(b) Activity of cell 4362 (at CEU building)



(c) Activity of cell 4434 (at the Parliament)

Figure 3. The mobile phone activity changes (dashed lines) of cell 4019 (at the Castle Garden), cell 4362 (at CEU building) and cell 4434 (at the Parliament) in comparison with the average activity of the preceding days (solid line).

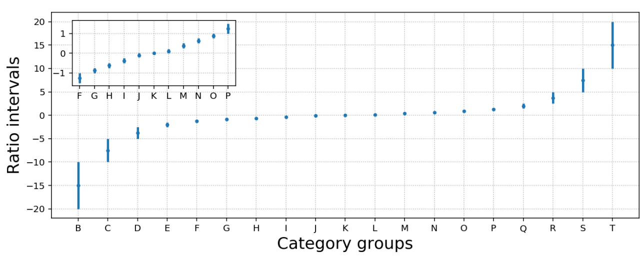


Figure 4. The intervals of the activity ratio categories from category B to T. The intervals are wider at the edges of the spectrum and the omitted categories (A and U) represent ratios smaller than -20 and larger than 20, respectively.

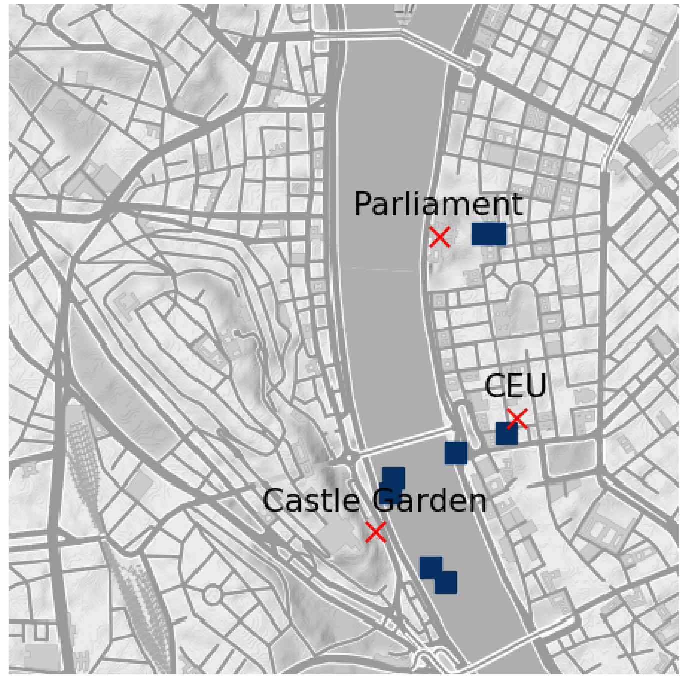


Figure 5. Cell centroids with at least 20 times more activity "UU" pattern) during the demonstration than in the previous days.

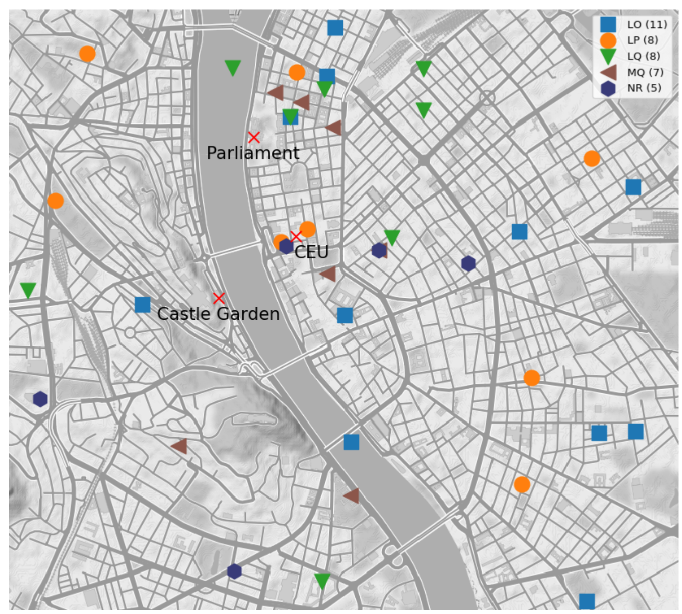


Figure 6. Cell centroids in Budapest downtown denoted with the most frequent activity patterns during the demonstration. The main locations of the demonstration marked with red crosses. The frequency of the patterns is provided in the legend.

as they represent 60 and 90 minutes of activity deviation changes and these time intervals fit for the demonstration. All of the bigrams and trigrams are generated for every cells and the ones are selected that represents a significant increase of activity. If a cell activity increases at least three categories in 60 and 90 minutes then a pattern considered interesting enough to examine. The cells with the pattern "UU", representing at least 20 times more activity during the demonstration than on average (Fig. IV). The main locations of the demonstration marked with red crosses. The frequency of the patterns is provided in the legend.

When the same pattern is shown in different cells in the same time range (60 or 90 minutes) then it means that those cells have a similar behavior in the given period. These cells are assumed equivalents concerning the same mobility patterns.

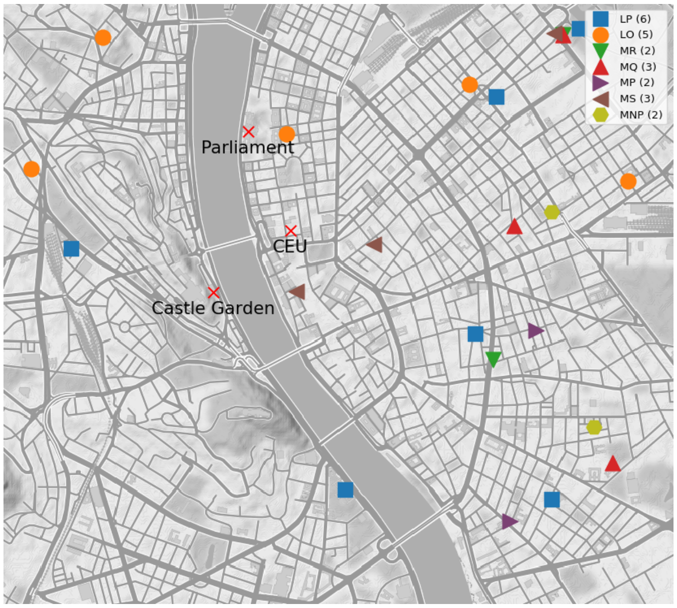


Figure 7. Cell centroids in Budapest downtown denoted with activity patterns after the demonstration. The main locations of the demonstration marked with red crosses.

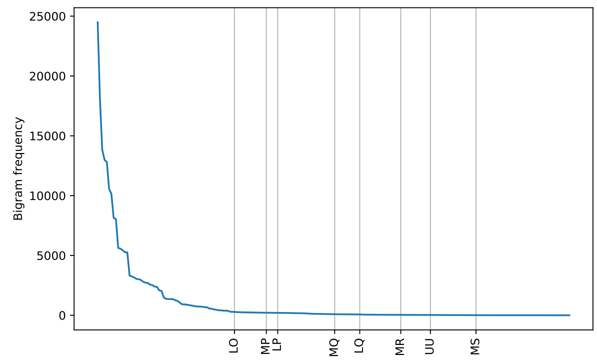


Figure 8. The frequency of the bigrams in the whole day data of Budapest, and the spatially relevant patterns are highlighted.

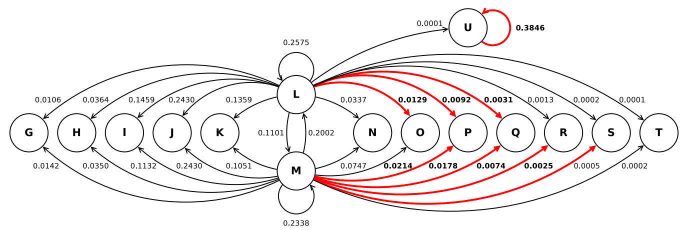


Figure 9. The transition probabilities for the most relevant starting categories (L, M) in respect of the demonstration.

V. Bigram Statistics

The bigrams representing significant activity change in a cell during an hour are filtered based on the relevant time intervals of the demonstration. Considering the spatial relevance of the demonstration (Fig. 6 and 7) the bigrams might seem frequent, especially the peaks “UU” pattern.

It has to be noted that during an ordinary Sunday these patterns are rare and patterns representing small changes are much more frequent. The most common patterns indicated no or very little change from the average, for example “II”, “JJ”, “GG”, “IJ”, “JI” and “KK”. Figure 8 shows the heavy-tail distribution of the ordered bigram frequencies where the spatially relevant patterns are marked.

If in a thirty-minute interval the activity level is “L” (which means at least 5% and less than 25% activity increase), then the next letter will most probably be “L” again ($P(L \mid L) = 0.2575$). The probability of “LO” ($P(O \mid L) = 0.0129$), “LP” ($P(P \mid L) = 0.0092$) or “LQ” ($P(Q \mid L) = 0.0031$) are significantly smaller. These patterns indicate significant increase in the mobile phone activity which is generally rare and can express a large social event like a demonstration. Figure 9 shows the transition probabilities of the most relevant starting categories (L, M) in respect of the demonstration. The transitions which indicates a significant activity increase are highlighted.

VI. Conclusions

A method has been developed and implemented to characterize the people’s mobility during a Big Social Event. The approach is based on the comparison of average and recent density of population in a selected area and time slot. The equivalency of regions in the city could be analyzed based on quantitative criteria. The approach is tested on CDR data collected during a demonstration obtained in Budapest in 2017. As future work, the density measures of the population is intended to enrich with origin-destination data in order to characterize the propagation of mobility as well.

Acknowledgment

The authors acknowledge the financial support of this work by the Hungarian State and the European Union under the EFOP-3.6.1-16-2016-00010 project.

References

[1] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, “Understanding individual human mobility patterns,” *nature*, vol. 453, no. 7196, p. 779, 2008.

[2] M. C. González and A.-L. Barabási, “Complex networks: From data to models,” *Nature Physics*, vol. 3, no. 4, p. 224, 2007.

[3] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, “Limits of predictability in human mobility,” *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[4] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, “Real-time urban monitoring using cell phones: A case study in rome,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 141–151, 2011.

[5] R. Trasarti, A.-M. Olteanu-Raimond, M. Nanni, T. Couronné, B. Furletti, F. Giannotti, Z. Smoreda, and C. Ziemlicki, “Discovering urban and country dynamics from mobile phone data with spatial correlation patterns,” *Telecommunications Policy*, vol. 39, no. 3-4, pp. 347–362, 2015.

[6] G. Pintér, L. Nádai, G. Bognár, and I. Felde, "Evaluation of mobile phone signals in urban environment during a large social event." IEEE 12th International Symposium on Applied Computational Intelligence and Informatics, 2018.

[7] N. Gyöngy-Kovács, "Hogyan érinti majd a telekomot a digi belépése a magyar mobilpiacra?" 20.04.2017, accessed: 2018-05-01. [Online]. Available: https://www.vg.hu/penzugy/hogyan-erinti-majd-a-telekomot-a-digi-belepese-a-magyar-mobilpiacra-487360/

[8] TheGuardian, "Thousands protest in Hungary over threat to Soros university," accessed: 2018-05-01. [Online]. Available: https://www.theguardian.com/world/2017/apr/09/thousands-protest-in-hungary-over-bill-threat-to-soros-university

[9] C. Y. Suen, "N-gram statistics for natural language understanding and text processing," *IEEE transactions on pattern analysis and machine intelligence*, no. 2, pp. 164–172, 1979.

[10] I. Markov, J. Baptista, and O. Pichardo-Lagunas, "Authorship attribution in portuguese using character n-grams," *Acta Polytechnica Hungarica*, vol. 14, no. 3, 2017.