Evolving connectivity graphs in mobile phone data

Olivera Novović, Sanja Brdar, Vladimir Crnojević
BioSense Institute
University of Novi Sad, Serbia
onovovic@gmail.com, {brdars, crnojevic}@uns.ac.rs

1. Introduction

Connectivity graphs inferred from mobile phone data uncover pulse of human interaction. In the recent years many innovative applications based on this rich data emerged, such as urban sensing, transport planning, social analysis and monitoring epidemics of infectious diseases [1][2]. Anonymous mobile communication data from telecom operators can be utilized for sensing activities occurring within a city and can further fit into wider vision of smart cities that aims at monitoring and optimizing urban landscapes. Several studies explored mobile phone data in the context of urban sensing. Cici et al. analyzed aggregated cell phone activity per unit area that allowed them to detect seasonal patterns (weekday/weekend), anomalous activities and to segment a city into distinct clusters [3]. In another study, authors examined interactions among city inhabitants and visitors and identified the city's hotspots [4]. Mobile phone data can be also used to derive city land use information [5].

Here we utilized graph theory to study connectivity patterns on a city scale. We focused on the dominant backbone of networks - the most significant part of overall communication interaction. Pairwise communication was aggregated over spatial units of a city and one day time intervals and analysed throughout two months period. In our graphs nodes are spatial units and links were drawn if communication strength between units was significant. This allows us to study the backbone connectivity graphs as evolving structures and to examine temporal and spatial dynamics within a city. We measured global and local graph properties and here present a part of obtained results.

1. Data

Mobile phone service providers collect large amount of data for every user interaction. Every time a user makes interaction using mobile phone (SMS or call), one *Call Detail Record* (CDR) is created in Telecom operator database. CDRs used in our research are provided by the Semantics and Knowledge Innovation Lab (SKIL) of Telecom Italia [6]. The records refer to the communication inside city of Milan for a time period of two months (November and December 2013). Telecommunication interaction between mobile phone users is managed by Radio Base Stations (RBS) that are assigned by the operator. Every RBS has unique id, location and coverage map that provide approximate user's geographical location. CDRs contain the time of the interaction and the RBS which handled it. In available data collection CDRs are spatially aggregated on the grid containing 10 000 cells and temporally aggregated on time slots of ten minutes. We used telecommunication interactions set that comprises measured intensities between different cells. Only cells that spatially intersect with administrative area of Milan city were selected. The city is divided into 88 administrative zones [7]. The map with zones is presented in Fig 1, where each is represented by unique id placed in its center.

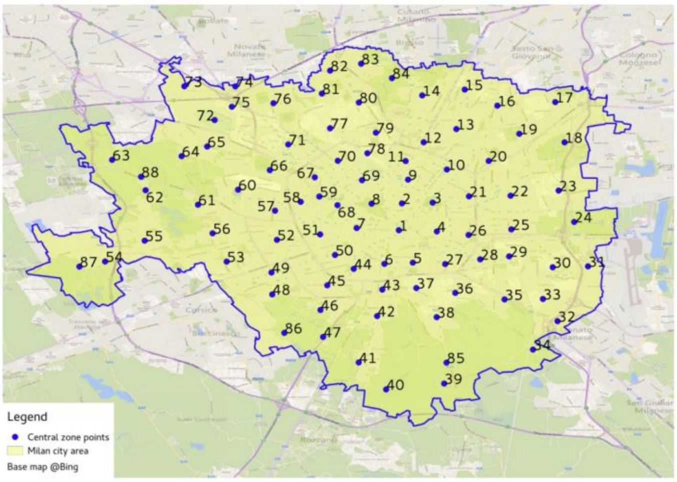

Fig. 1. Centres of 88 adnisistartive zones of Milan.

2. Connectivity graphs

To create connectivity graphs, we further aggregated communication from grid cells to 88 zones of Milan. Each grid cell is assigned to corresponding zone, thus our graphs refer to zones interactions. The connectivity matrices for 61 days were made in pairwise manner. Matrix element on the position [*i*, *j*] represents aggregated communication strength between zone *i* and zone *j*. After creating the connectivity matrices the filtering was performed to eliminate weak links. The strong and weak links distinction was made by equation 1 by calculating the significance of links α_{ij} [8]:

$$\alpha_{ij} = 1 - (k-1)\int_0^{p_{ij}} (1-x)^{k-2}\, dx < \alpha, \qquad (1)$$

where α denotes significance threshold, p_{ij} is the probability of having link between nodes *i* and *j*, and *k* is the number of nodes. In our case $k = 88$ and α was set to 0.05. After the weak links were eliminated, graph structure for each day was created from remaining links in connectivity matrix. The final graphs are sparse and suitable for visual analytics. We presented links with QGIS and selected four typical graphs (Fig 2.). The first is typical weekday, where the strongest communication links tend to appear and concentrate near city center and maximum communication strength is much higher than observed on the weekends. The second is the holyday (Friday 2013-11-01), where the strongest communication

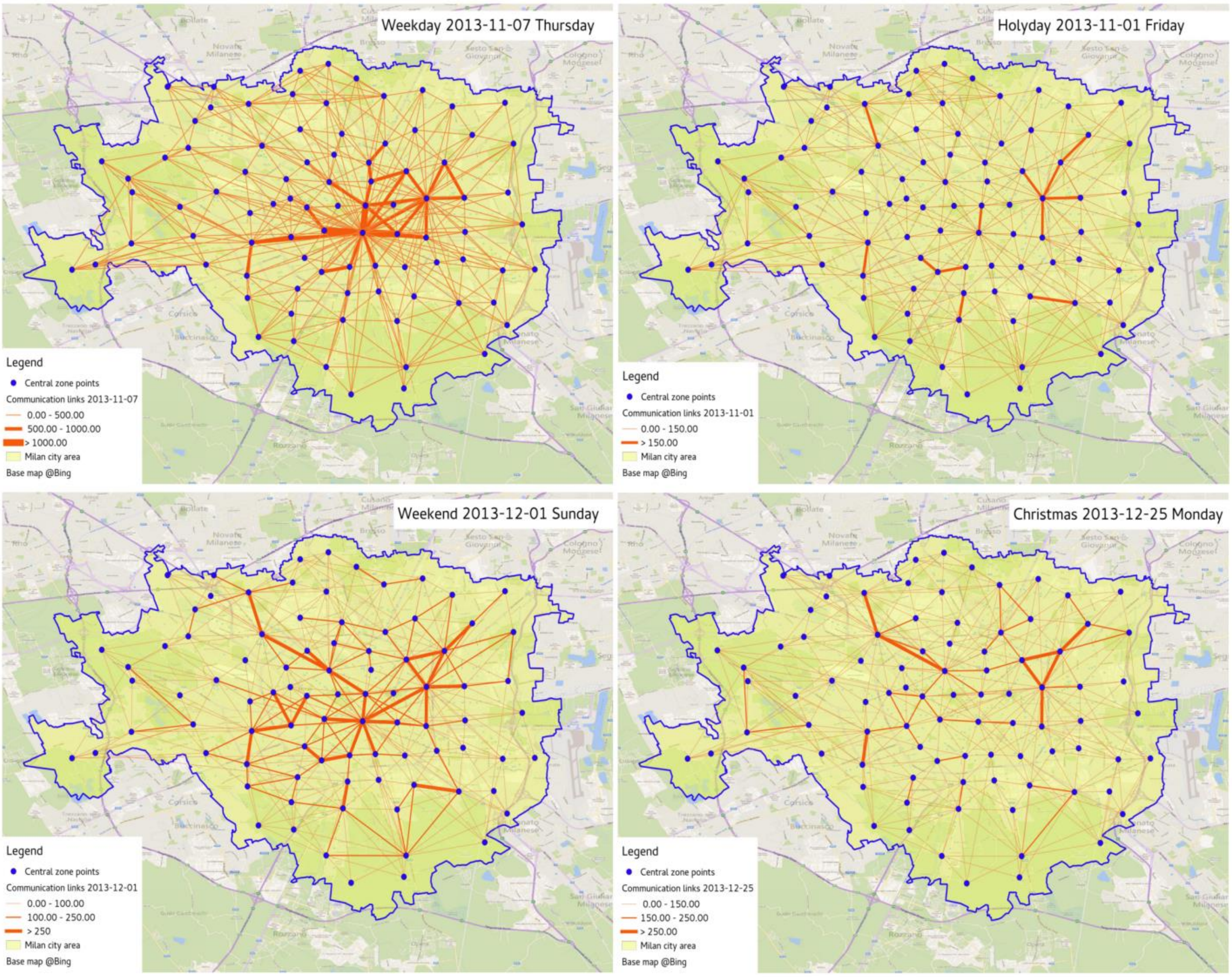

Fig. 2. Bacbone connectivity graphs between zones of Milan

links tend to disperse and the maximum communication strength is very low. The third graph presents typical weekend. The strongest communication links tend to disperse across city but the maximum communication strength is higher than on the holyday. Finally, the Christmas day is presented in the fourth graph. Its structure is similar as the one presented for another holiday. The strongest links are dispersed across residential parts of the city and the overall communication strength is low, which is typical for holydays.

3. Graphs properties

Along with visual inspection of graphs across two month period we quantified graphs properties and did deeper analysis of their changes during time and identified interesting weekday/weekend distinctions. We performed both, global and local, graphs analysis [9].

Global graph properties provide information on a global structure and further allow comparisons among graphs. We calculated the number of edges, maximum weight, radius, diameter, max clique size, average clustering for all inferred graphs. We compared global graph properties on different day types: weekdays and weekends and discovered differences. We observed that number of edges is higher on the weekdays than on weekends (Fig 3). Distributions of measured diameters unveil that weekday graphs have lower diameter compared to weekend, indicating faster information flow during work days and larger "connectivity distance" in weekends.

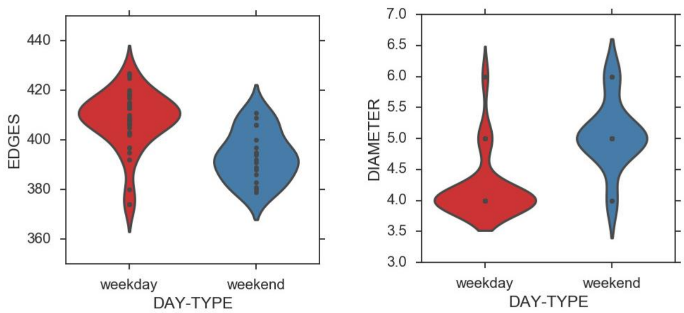

Fig. 3. Distribution of number of edges and diameter in different day types

Local graph properties uncover localized patterns in the graphs. We measured numerous local properties such as clustering coefficient, node degree, page rank, betwenness centrality, etc. Here we selected three nodes (Zone 1, 71 and 23, see Fig 1.) and presented changes in time of betwenness centrality (Fig 4.) and PageRank (Fig 5.). Betwenness centrality calculates the number of shortest paths that pass through examined node and PageRank determines the relevance or influence node in graph. Zone 1 that encompasses city center has the largest betwenness centrality and PageRank and we can notice clear weekday-weekend pattern. Zone 71, has opposite pattern, it increases on weekends and drops on workdays. The strong links involving Zone 71 are also visible in Fig 2. during weekends and holidays. Interestingly Zone 23, that covers part of Lambrate district has unusual jump in betweenes centrality at 15th December, that could be due to event *The Lambrate Bicycle Film Festival*. PageRank pattern for Zone 23 differs considerably form betweenes centrality implying that different graph properties can provide complementary information.

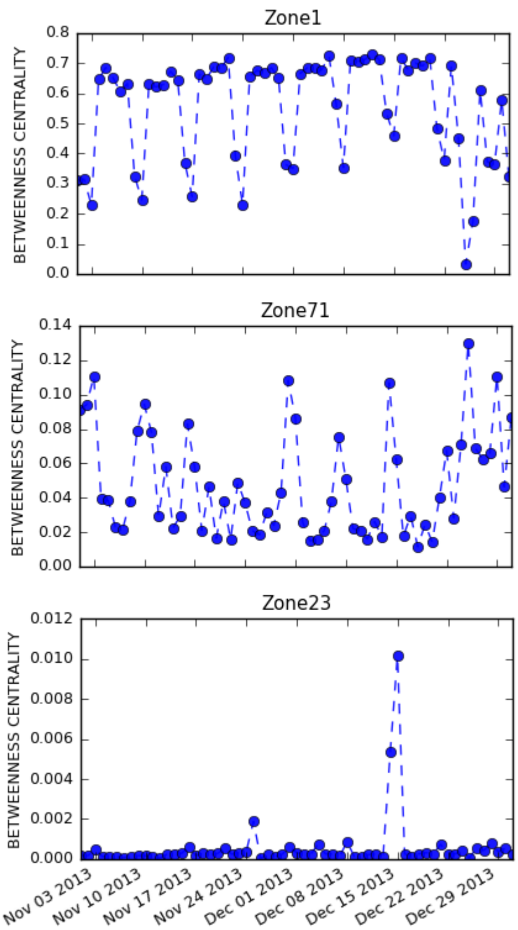

Fig. 4. Betwenness centrality of zones 1, 71 and 23

4. Conclusions and Future Work

Our analysis of the connectivity backbone networks in the city provided new insights into social interactions and their changes across the city zones in different day types. Through the lenses of graph theory we discovered properties that can serve for detecting the patterns and deviations from typical observations. Our future work will include more graph-based formalism in identifying strong temporally consistent links, patterns of change and evolving graph sequences [10] as well as unveiling underlying social pulse that is reflected in mobile phone data.

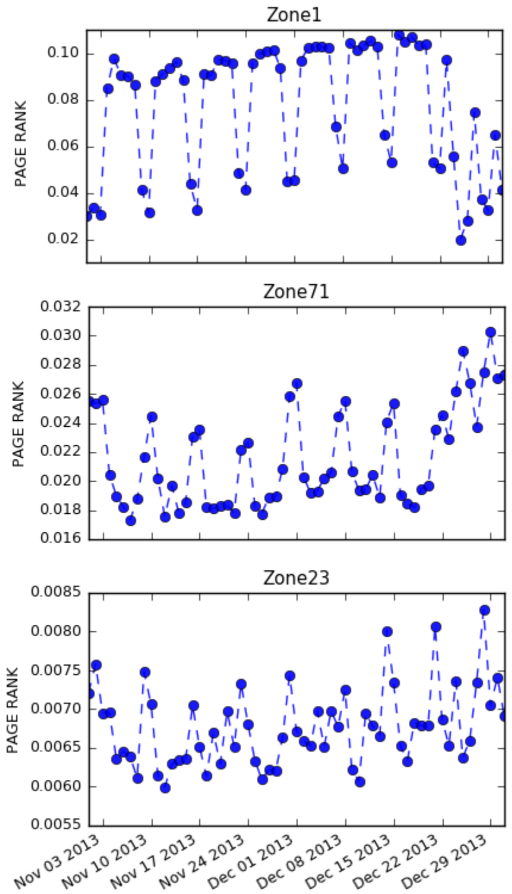

Fig. 5. Pagerank of zones 1, 71 and 23

References

[1] V.D. Blondel, A. Decuyper, and G. Krings. "A survey of results on mobile phone datasets analysis." EPJ Data Science vol. 4, no. 1, 2015.

[2] D. Naboulsi, M. Fiore, S. Ribot, S and R. Stanica,. Large-scale mobile traffic analysis: a survey. IEEE Communications Surveys & Tutorials, vol. 18, no.1, pp. 124-161, 2015.

[3] B. Cici, M. Gjoka, A. Markopoulou, and C. T. Butts, On the decomposition of cell phone activity patterns and their connection with urban ecology. In Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 317-326, June 2015.

[4] P. Bajardi, M. Delfino, A. Panisson, G. Petri & M. Tizzoni. Unveiling patterns of international communities in a global city using mobile phone data. EPJ Data Science, vol. 4, no. 1, 2015.

[5] T. Pei, S. Sobolevsky, C. Ratti, C., S. L. Shaw, T. Li, and C. Zhou. A new insight into land use classification based on aggregated mobile phone data. International Journal of Geographical Information Science, vol. 28, no. 9, pp. 1988-2007, 2014.

[6] G. Barlacchi et al. “A multi-source dataset of urban life in the city of Milan and the Province of Trentino,” Scientific data, vol. 2, 2015.

[7] Milan's public data. http://dati.comune.milano.it/, September 2016

[8] M. Á. Serrano, M. Boguná, & A. Vespignani, Extracting the multiscale backbone of complex weighted networks. Proceedings of the national academy of sciences, vol. 106, no. 16, pp. 6483-6488, 2009.

[9] L. D. F., Costa, F. A. Rodrigues, G. Travieso, and P. R Villas Boas.. Characterization of complex networks: A survey of measurements. Advances in physics, vol. 56, no. 1, pp. 167-242, 2007.

[10] A. Kosmatopoulos, K. Giannakopoulou, A.N. Papadopoulos, and K. Tsichlas, An Overview of Methods for Handling Evolving Graph Sequences. In Algorithmic Aspects of Cloud Computing, pp. 181-192, 2016.