Territory analysis using cell-phone data

Vincent Aguiléra
University of Paris-Est

Chloe Milion
Orange Labs

Territory analysis using cell-phone data

Vincent Aguiléra[a,1], Chloé Milion[b], Sylvain Allio[b]

[a]*Laboratoire Ville Mobilité Transports, Université Paris Est, Champs-sur-Marne, France*
[b]*Orange Labs, Belfort, France*

Abstract

During the daily operations of a cellular telecommunication network, approximate location data are exchanged between the cell-phones and the network, in both directions. Those data include particular messages called signalling events. Signalling events record some very specific information required for the proper operations of the network. Among other things, a signalling event is generated every time a phone is switched on/off, every time a phone is entering a location area (i.e. a group of cells), every time a cell border is crossed while the phone is in communication, and so on. This paper elaborates on the idea that signalling events do capture a significant part of a given territory's activity. It shows, through illustrative examples, that some land-use information can be retrieved from cells' activity profiles, along with other inputs required by transportation and planning studies, such has flow maps or trip tables. Anonymized and aggregated data were provided by the Orange telecom operator.

Keywords: Cell phone data; land use.

Résumé

Au cours de son fonctionnement quotidien, un réseau de téléphonie mobile et les mobiles qui y sont connectés échangent des données approchées de localisation. Parmi ces données se trouvent des messages particuliers appelés les événements de signalisation. Ces événements contiennent des informations relatives à la position des mobiles. En particulier, un événement de signalisation est enregistré chaque fois qu'un mobile est éteint/allumé, chaque fois qu'un mobile change de zone de localisation (regroupement de plusieurs cellules), chaque fois qu'une frontière de cellule est franchie en cours de communication, etc. Cet article part de l'idée que l'ensemble des événements de signalisation reflète l'activité d'un territoire. Il montre, par des exemples illustratifs, que les profils d'activité des cellules téléphoniques permettent d'inférer le type d'activité au sol (e.g. zone résidentielle, zone commerciale), ainsi que d'autres quantités utilisées par les études transport/aménagement, telles que des cartes de flux ou des matrices origine-destination. Les données, anonymisées et agrégées, ont été fournies par l'opérateur de téléphonie Orange.

Mots-clé: Données de téléphonie cellulaire; usage du sol.

[1] Corresponding author. Tel.:+33-181-668-861
E-mail address: vincent.aguilera@enpc.fr

1. Introduction

Travel behaviour in metropolitan areas is usually studied by the mean of travel surveys, and large metropolitan areas typically only do such a survey once every decade. Those surveys are usually conducted for land-use and transportation planning purposes, and are mainly concerned with the inhabitants commuting behaviour. Hence, they do not capture important parts of the activity and mobility patterns. For instance, leisure and/or shopping motivated trips made either by non-residents—or by residents out of the daily routine—, are not easy to survey, especially during non-working days. Cell-phone data have recently emerged has a new source of data for studying people's mobility. So far, most research on the topic have used Call Detail Records (CDRs), i.e. time and location information available to the service provider when a communication (either voice, SMS or data) is emitted or received (see for instance Calabrese et al., 2013). But, in addition to CDRs, a cell-phone network also captures a variety of mobility related data. During the daily operations of a cell-phone network, location data are exchanged between the cell-phones and the network, in both directions. Those data include particular messages called signalling events. Signalling events convey some very specific information required for the proper operations of the network. Among other things, a signalling event is emitted every time a phone is switched on/off, every time a phone is entering a new location area (i.e. a group of cells), every time a cell border is crossed while the phone is in communication, and so on.

This paper elaborates on the idea that the flow of signalling events do capture a significant part of a given territory's activity. It shows how some land-use information can be retrieved from cell activity profiles (e.g. residential area, shopping area), along with other inputs required by transportation and planning studies, such has flow maps or trip tables, using anonymized and aggregated data provided by the Orange telecom operator. In contrast with previous works focused on CDRs, where only communication activity is dealt with, one important contribution is to show that inactivity is also reflected in signalling events. The paper proceeds as follows. Section 2 is an overview. It first provides a gentle introduction to signalling events and cell-phone location in a cell-phone network, and then presents the geographical area of interest. Section 3 is a literature review. Section 4 presents the dataset. Two illustrative examples of possible uses of such data are presented in Sections 5 and 6.

2. Overview

2.1. Signalling events and cell-phone location in a cell-phone network: the basis

A cell-phone sends and receives information (voice, messages, data) by radio communication. Radio signals are transmitted between the phone and the nearest antenna. Each antenna is connected to a Base Transceiver Station (BTS). In most cases the antennas and the BTS are installed at the top of towers. A tower covers a geographical area known as a cell. Typically, the beam-width of an antenna is 120°, and an ideal network may be seen as a mesh of hexagonal cells, each with a base station at its centre, and three antennas to ensure a 360° coverage. The cells overlap at the edges to ensure the mobile phone always remain within range of a BTS. From a logical (as opposed to physical) point of view, cells are grouped together into larger units called Location Area (LA). There are typically tens to hundreds of cells per LA. Signalling events transit between the cell-phones and the elements of the network architecture which are in charge of the mobility management or the communication management. Those events contains the identifier of the cell the phone was connected to at the time the event was triggered. For instance, when a mobile phone is switched on, it exchanges data with the nearest antenna to gather the LA it is located in. It then emits an attachment event (ATT1). The phone stores the current LA. Then, on a regular basis, the phone probes the network to check if the LA it is located in matches the one it has in memory. If so, nothing happens. If not, it emits a LA Update (LAU) event. Also, if no signalling event has been emitted for the past three hours, a LAU event is emitted in any case. Those two kinds of LAU events have distinguished acronyms: one is LAUN, that corresponds to a normal LAU (i.e. a true change in the Location Area), the other is LAUP, that corresponds to a periodic LAU (i.e. a LAU that happens if no signalling event has occurred for the past three hours). When the phone is switched off, it emits a detachment event (ATT0). One point that might be worth noticing is that, even if no communication (voice, SMS/MMS, data) has occurred between the attachment and the detachment LAUP events are triggered anyway. Additionally, as with CDRs, signalling events related to communication occur as well. Let mention the following events: when voice communication or a SMS/MMS is received (COM0 event) or initiated (COM1 event); during web browsing and periodic updates from the Internet server (GMM_SERVICE event); when a voice communication is processed while the mobile phone is moving a handover (HO event) happens.

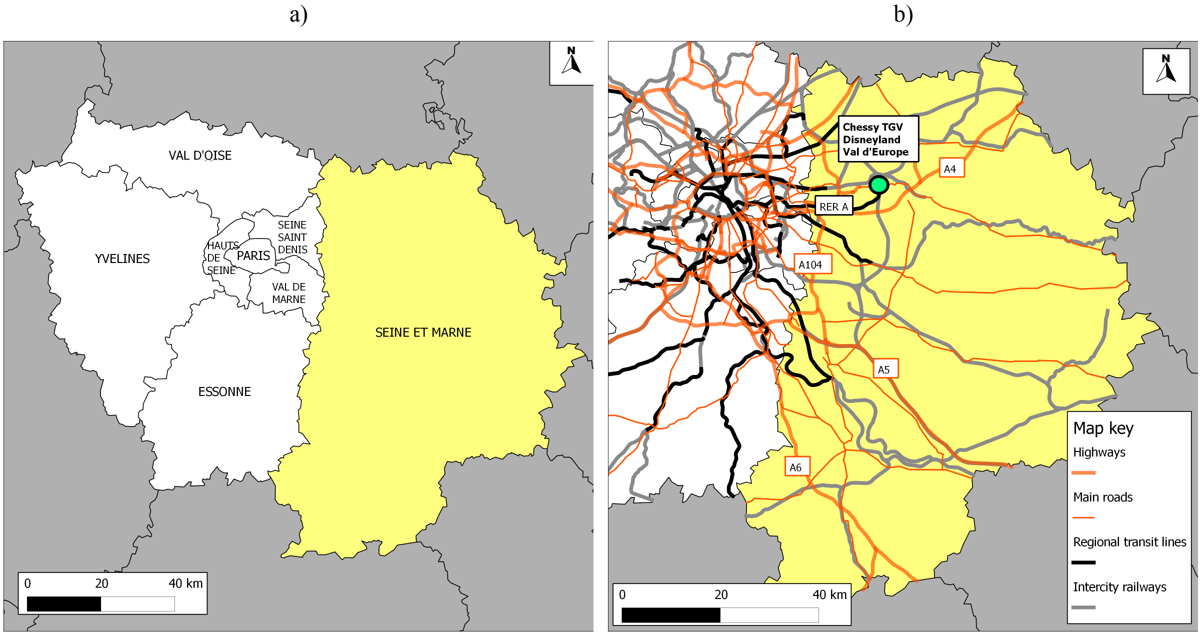

Fig. 1. (a) Location of the Seine-et-Marne department in the region surrounding Paris (Ile-de-France); (b) Main transportation infrastructures (road and railway) in Seine-et-Marne, and the geographical area of interest, around the Disneyland Resort Paris leisure park.

2.2. The area of interest

The geographical area of interest is Seine-et-Marne, one of the height departments in Ile-de-France, the region surrounding Paris (Fig. 1a). A particular attention is paid to the surroundings of Disneyland Resort Paris, with three points of interest in the vicinity of each other. Besides the leisure park, there is a huge shopping center called Val d'Europe, and a high-speed railway hub called Chessy TGV. The area is connected to Paris by the A4 highway and by the RER A, a regional transit line (Fig. 1b). The Chessy TGV hub allows connections with the european high-speed railway network, including Eurostar (London, Bruxelles) and Thalys (Bruxelles, Cologne, Amsterdam). The main other transportation facilities in Seine-et-Marne are the A104 ring road, and the A5 and A6 motorways. The population of Seine-et-Marne is around 1.33 million inhabitants, to be compared to 11.73 millions in Ile-de-France, according to the 2009 census. At the same time, the Seine-et-Marne department represents almost one half of the surface area of Ile-de-France. Hence, the population density is below the regional average. Also, inhabitants are evenly distributed. There is a huge concentration of population located close to the west boundary (Fig. 2a), with one cluster near Bussy Saint-Georges, along the RER A transit line and the A4 highway, and another cluster along the A6 motorway, starting from Melun. The spatial distribution of cell-phone towers closely follows that of the population. In the Voronoï diagram plotted Fig. 2b, the center of each Voronoï cell is a cell-phone tower. The comparison between Fig.2a and Fig.2b is instructive. The Seine-et-Marne department is divided into 544 administrative divisions called communes. The map on the left hand side (Fig. 2a) represents the spatial density of population of each commune. The map on the right hand side (Fig. 2b) represents the spatial density of the set of signalling events used in the sequel. This data set, to be described with more details in section 4, contains millions of signalling events extracted from 14 Locations Areas. The union of those Location Areas ensures a full coverage of the Seine-et-Marne department. There are more than 600 towers in the dataset, among which 395 are located in Seine-et-Marne. The two maps clearly show a correlation between the spatial coverage of a cell-phone tower and the population density: the denser the population, the narrower the Voronoï cell. Almost all Voronoï cells with more than 100,000 events per square kilometer overlap communes in which there exists a town with more than 15,000 inhabitants. One noticeable exception is the area around Disneyland, around 7 km east-north-east of Bussy-Saint-Georges, where there is no such a town but a very intense cell-phone activity.

a) b)

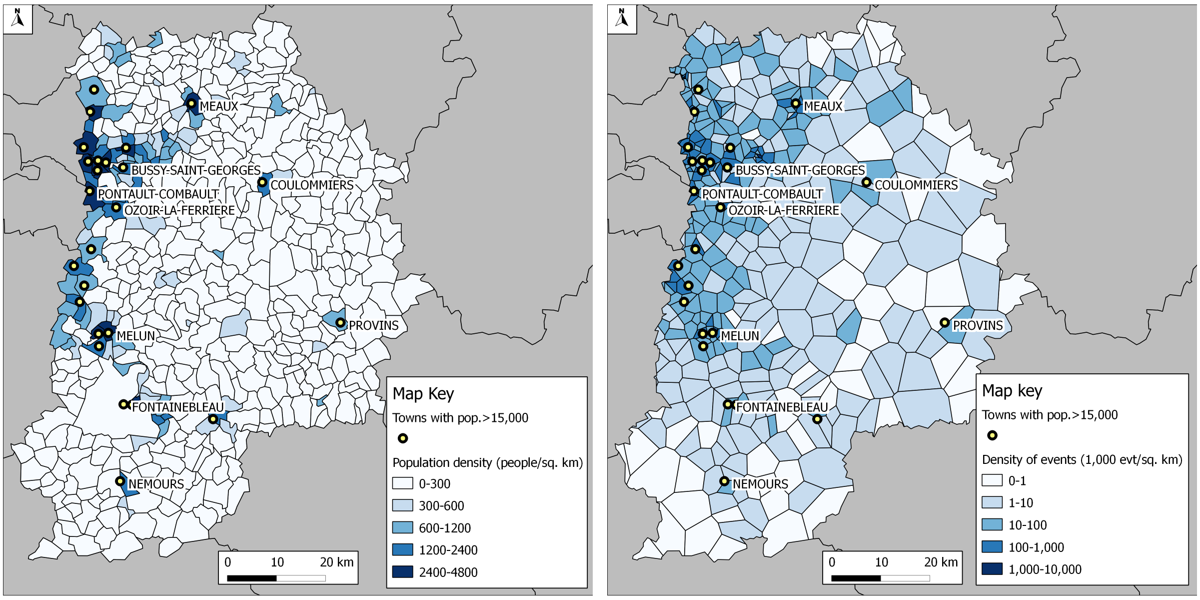

Fig. 2. (a) population density in Seine-et-Marne; (b) density of signalling events in Seine-et-Marne.

3. State of the art and literature review

One of the first attempt in using probe mobiles and radio communications for the purpose of monitoring a road network is reported by Linaartz et al. (1994), in the San Francisco Bay Area. In France, the use of cell-phone data to sense road traffic flow has been investigated since year 2000, first in a simulation framework (Ygnace et al. 2000), then during field operational tests along the French Rhone corridor (Ygnace 2001). From then, the topic has been regularly revisited by researchers (Bar-Gera, H., 2007; Caceres et al. 2010). At the same time major cell-phone operators have developed business solutions for the traffic data market. Most systems of this kind are based on handovers that occur during a communication while a cell phone performing is moving from on cell to another. Indeed, the proper management of a handover between two BSCs necessitate then to exchange very detailed signalling information (signal power, allocated frequency channel, …) that allow for a precise estimates of the location and speed of the mobile phone. Hence, locally, groups of cells can be used as virtual traffic loop detectors, with a level of accuracy that has been shown comparable to that of a standard traffic loop.

A second stream of research is based on the use of Calling Detail Records (CDRs). CDRs have been used: to observe dynamics of a city or a country through detection of dense area (Ratti et al., 2006; Rubio et al., 2013); for tourism surveys (Ahas et al., 2008); to model human mobility (Gonzales et al., 2008 ; Bagrow&Koren, 2009); to study the relations between human mobility and urban morphology (Kang et al., 2012); to study correlations betweens mobile phone usage and travel behavior Yang et al. (2012); to identify daily activity patterns (Phithakkitnukoon et al., 2010; Di Lorenzo& Calabrese, 2011). Meanwhile, CDRs have been shown to be noisy, sparse and incomplete (Farrahi&Gatica-Perez, 2010; Zheng et al., 2013). Spatial accuracy is an issue. The radius of a cell varies from hundreds of meters to tens of kilometers, depending on whether the cell is in a dense urban area or in the countryside. Some, like Calabrese et al. (2013), have used more accurate CDRs such as the ones provided by the AirSage company. Despite a greater spatial accuracy (the claim is an average uncertainty around 300 meters), a lot of short trips false detections remain, because of localisation uncertainty and fluctuations in the radio coverage with time. Also, users calling patterns are hard to predict. Several models have been proposed for the inter-call time distribution. Yang&Itoh (2009) have proposed an exponential low, while Gonzales et al. (2008) suggest a power law. Furthermore, calling activity patterns strongly depend not only on individual factors (age, gender), but also on other factors such as the time of the day or the day of the week. For privacy reasons, anonymised CDRs do not contain individual data, making links between the calling activity captured by CDRs and human activity hard to establish.

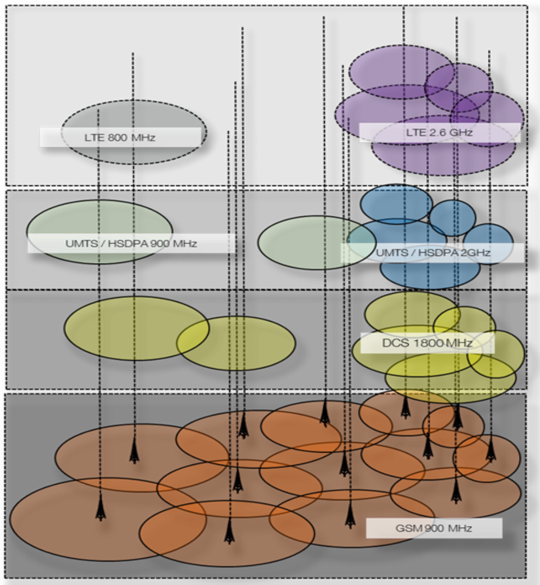

Fig. 3. Technology layers in a cell phone network.

4. The dataset

As explained in section 2.1 the infrastructure of a cell-phone network is based on towers carrying antennas. With time several technologies have emerged (Fig. 3). Each layer has its set of cells, and they may or not overlap. For each technology, the size of a cell depends on three factors. First, the local propagation conditions and terrain relief: radio signals are attenuated, absorbed or reflected by trees, hills and buildings. Second, the frequency band in which the network operates: the higher the radio frequency, the smaller the cell. Third, the capacity (the targeted number of simultaneous of calls) needed in any given area: base stations are typically spaced about hundreds of meters in urban areas, and tens of kilometers in the countryside. For all those reasons, the mobility of a cell-phone in the telecommunication network does not correspond straightforwardly to the geographic mobility. Orange provided us with aggregated statistics and point to point flow measurements built from the flow of signalling events coming from 14 Locations Areas that fully cover the Seine-et-Marne department. The data were processed on Saturday, December the 22th, 2012, from midnight to midnight. This day being close to Christmas celebration, one can expect observing a huge number of trips with shopping and/or leisure motivation. The data include not only those of Orange national subscribers, but also those of foreigners whose operator has a roaming agreement with Orange. The number of cell-phones observed in Seine-et-Marne may roughly correspond to 2 million people, to be compared to the total population of Seine-et-Marne (around 1.3 million). The average number of events per phone is close to 52.

Table 1. Number of events and of distinct cell-phones in the dataset.

	Whole dataset	Seine-et-Marne only
Number of events	$102.0\text{x}10^6$	$68{,}58\text{x}10^6$
Number of phones	$1.847\text{x}10^6$	$1{,}205\text{x}10^6$
Average number of events per phone	52.39	51.17

The four coming subsections present some statistics: the distribution of events per event type (Sec. 4.1); the links between calling activity and mobility (Sec. 4.2); the distribution of the number of events per cell-phone (Sec. 4.3); the distribution of delays between consecutive events (Sec. 4.4). Two illustrative examples of possible uses of such a dataset for mobility studies are given in Sections 5 and 6.

4.1. Distribution of events per event type

The distribution of the set of events per event type is plotted in the pie chart, Fig. 4. The meaning of most event types have been explained in section 2.1. For the others, prefixed by GMM_, it suffices to know that they correspond to 3G only signalling events. Three types of events, GMM_SERVICE, COM0 and LAUN represent almost 75% of the dataset, with an almost identical respective share, around 25% each. The GMM_SERVICE and COM0 event types correspond to incoming communications (data for GMM_SERVICE, voice/SMS for COM0). Then one can find COM1, GMM_RA and the other types of events at around 10% each. Within the other types of events, HO and LAUP represent the vast majority, at around 40% each.

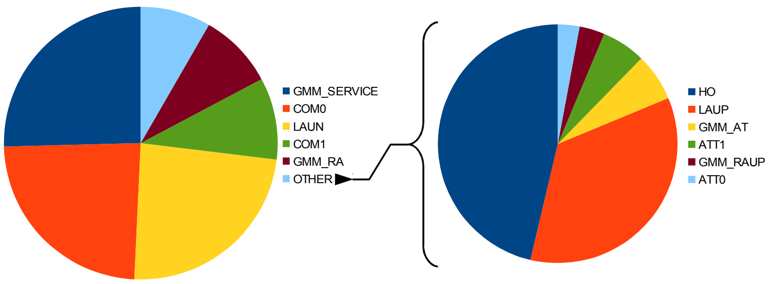

Fig. 4. Distribution of the events in the set of events, per event type.

4.2. Calling activity and mobility

Fig. 5 are plotted, for 10 bins of width 10 in the number of COM1 events per cell-phone, the number of cell-phones and a boxplot of the distribution of trajectories' lengths. All the bins from 10:19 to 90:99 have very similar distributions of trajectories' lengths. The distributions are asymmetric, with a median value around 12km, and the two median quartiles between 5 and 30 km. The bin 0:9 has slightly lower values. There is clearly no statistical evidence, in the dataset, of any correlation between the calling activity (voice+SMS) and mobility.

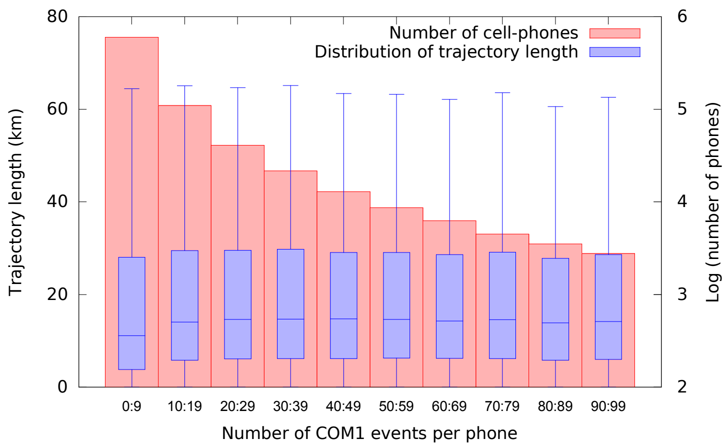

Fig. 5. Distributions of the number of cell-phones and of trajectory lengths, as a function of the number of events of COM1 events per phone.

4.3. Distribution of the number of events per cell-phone

The distribution of the number of events per cell-phone, for all observed cell-phones, is plotted Fig. 6. More than 90% of the observed cell-phones have emitted two or more events during the day;50% have emitted 14 or more events during the study period. 14% have emitted more than 100 events during the study period. The distribution looks almost regular, but presents a noticeable peak at 8 events per mobile: 5% of the observed mobile phones have emitted exactly this number of events during the day. The peak taken apart, the distribution should be easy to fit with a standard heavy-tailed distribution (e.g. a power law). The peak itself correspond in majority (more than 90%) to cell-phones that were switched on, but have not been used during the day: the 8 events are all of type LAUP (periodic Location Area Update).

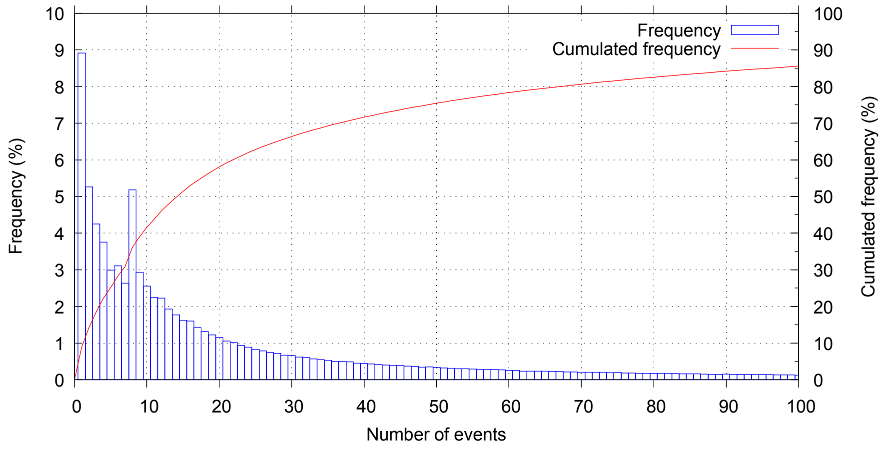

Fig. 6. Distribution of the number of cell-phones, as a function of the number of events per phone during the study period.

4.4. Distribution of time intervals between any two consecutive events per mobile

The distribution of time intervals between any two consecutive events per mobile is plotted Fig. 7. The shape of the distribution is almost regular in the range [1:170] minutes, with small peaks at 15, 30 60 and 120 minutes. Again, the regular part should be easy to model with a standard, super-exponential, distribution. The highest peak, at 180 minutes, corresponds obviously to the set of low activity cell-phones identified in section 4.3.

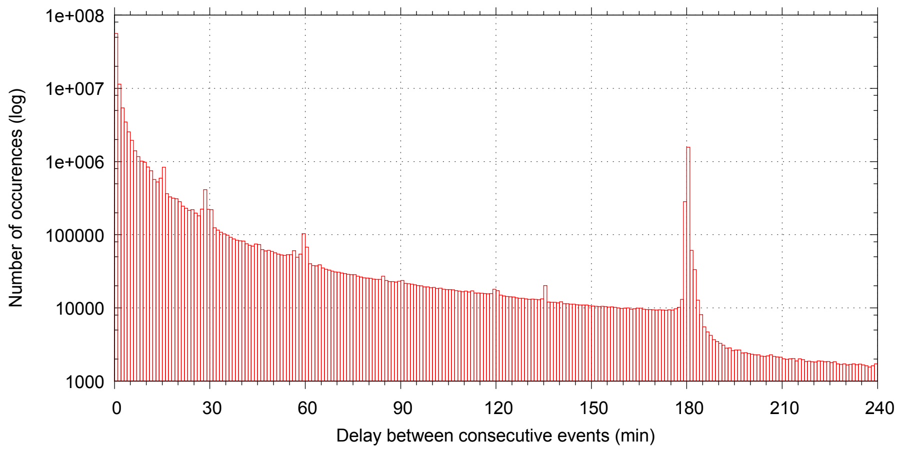

Fig. 7. Distribution of time intervals between any two consecutive events per mobile.

5. Immobility in Seine-et-Marne

Results from section 4.3 and 4.4 suggest that a non negligible fraction of cell-phones are not actively used—if used at all—while switched on. Let now have a closer look at the spatial inactivity. In the dataset, around 70% of the cell-phones have emitted two or more signalling events. Among them, almost 11% have not been detected as moving. We have defined a cell-phone as *immobile* if the time elapsed between its first and last event exceeds two hours, and if all events occurred in a single tower.

Now for each tower t let define the Tower Immobility Index (TII) as the ratio between the number of immobile cell-phones seen by t and the total number of cell-phones seen by t. The TII for all towers in Seine-et-Marne is mapped Fig. 8, along with the numbers of immobile cell-phones per tower. With no surprise, the towers overlapping circulated highways in the countryside (e.g. A5,A6, see Fig 1a) have a very low TII, below 1%. Towers with higher TII, between 8 and 16%, are located close to towns in the countryside (e.g. Coulommiers, Provins, Nemours, see Fig.1). The three towers with the highest number of immobile cell-phones are located in eastern Seine-et-Marne. Clearly, a cross-analysis with socio-demographic inputs is likely to reveal interesting correlations.

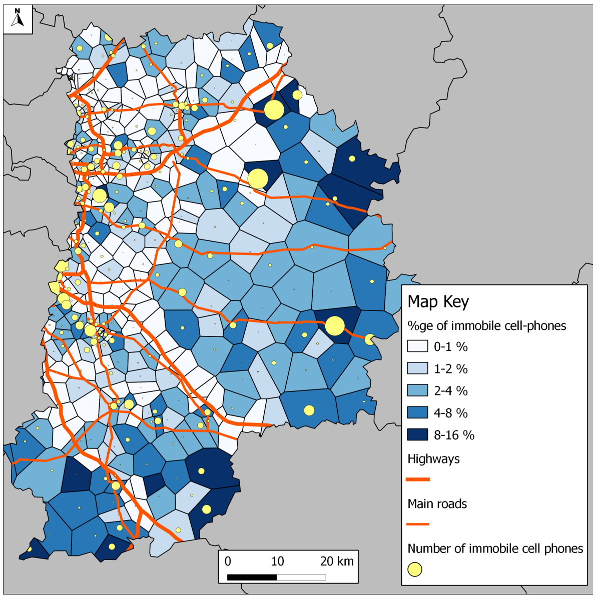

Fig. 8. Map of the Tower Immobility Index and of the number of immobile cell-phones per tower.

6. Flow analysis

Let now investigate a more classic case study. The purpose is to examine the flow of mobile phones in the Val d'Europe shopping center. Several quantities are likely to be of interest for various stakeholders in the retail business: the duration of stays in the shopping center; the shoppers traffic; the geographic origins of customers; the routes used to reach the shopping center; the places visited before/after the shopping center; and so on. Good estimates of such quantities are easy to derive from the dataset. The distribution of the estimated durations of stays, per 30 minutes bins, is plotted Fig. 9a. The estimates were computed from segments of point to point flows such that: the first and last points are located outside the shopping center; the intermediate point(s) are located in the shopping center; the time interval between the first point and the next point is less than 15 minutes; the time interval between the last point and the previous one is less than 15 minutes. It appears that most visitors have spent between half an hour and one hour and a half in the shopping center. The distribution within the day of the visitors traffic is illustrated Fig. 9b and Fig. 9c. The rush hours are between 15:00 and 18:00: not only the number of visitors reaches its maximum values, but also the ingress/egress flows. The spatial distribution of visitors is illustrated by the map Fig. 9d. To each tower is associated the number of distinct mobile phones having visited the shopping center during the day. The corresponding Voronoï diagram gives the spatial density of visitors. It allows for (i) the identification of residential areas from where visitors come from and (ii) the identification of the routes used by visitors.

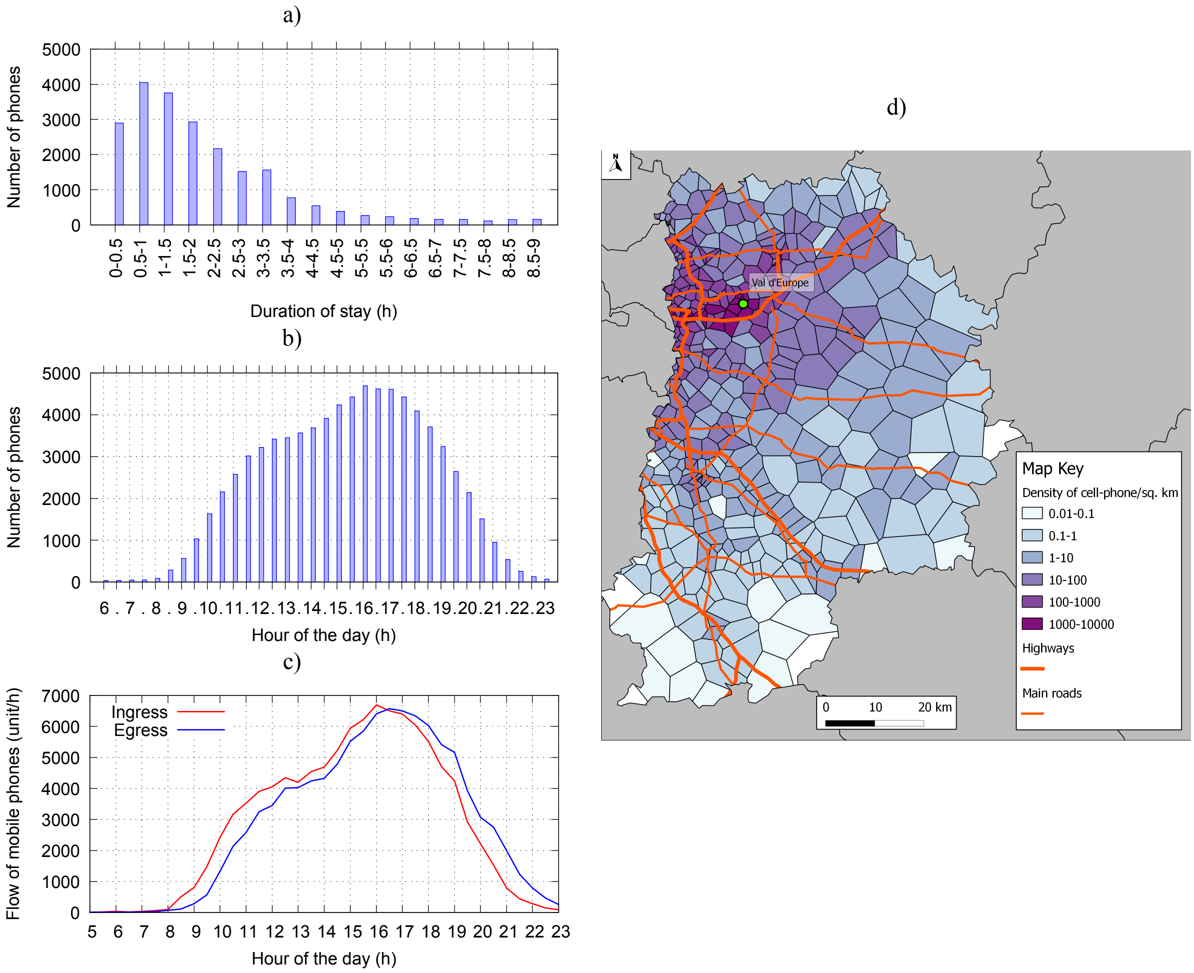

Fig. 9. Flow analysis at Val d'Europe shopping center. (a) Distribution of the estimated duration of stay. (b) Distribution of the cell-phones present at Val d'Europe, for every 30 minutes time slice. (c) Ingress and egress flows. (d) Density map of the cell-phones trajectories having visited Val d'Europe at least once during the day.

7. Conclusion

In this paper we have investigated possible uses of signalling events in a cell-phone network for the purpose of studying human mobility. Some points are worth remembering. First of all, the spatial resolution of the cell-phone network varies with people density, meaning that cells spatial extent is by design adjusted to the expected number of cell-phones that would be present in a given cell. Second, it appears that such a dataset contains appropriate spatio-temporal features to extract various aggregate quantities: in-flows, out-flows, route choice. Third, in comparison with traditionally used CDRs, where only communication activity is dealt with, it appears that precious information can be derived from communication inactivity and/or immobility.

References

Ahas, R., Aasa, A., Roose, A., Markb, U., Silma, S, 2008. Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, 29(3),.469–486.

Bagrow, J. P., Koren, T., 2009. Investigating bimodal clustering in human mobility. *Computational Science and Engineering, 2009. CSE'09. International Conference,* 4, 944-947.

Bar-Gera, H., 2007. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research Part C: Emerging Technologies*, 15(6), 380-391.

Caceres, N., Romero, L., M., Benitez, F., G., Del Castillo, J., M., 2010. Traffic Flow Estimates Inferred from Mobile Phone Networks. *12th WCTR*, Lisbon, Portugal.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr, J., Ratti, C., 2013. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26, 301-313.

Di Lorenzo, G., Calabrese, F., 2011. Identifying human spatio-temporal activity patterns from mobile-phone traces. *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference,* 1069-1074.

Farrahi, K., Gatica-Perez, D., 2010. Probabilistic mining of socio-geographic routines from mobile phone data. *Selected Topics in Signal Processing, IEEE* , 4(4), 746-755.

Gonzalez, M. C., Hidalgo, C. A., Barabasi, A. L., 2008. Understanding individual human mobility patterns. *Nature*, 453(7196), 779-782.

Kang, C., Ma, X., Tong, D., Liu, Y., 2012. Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and its Applications*, 391(4), 1702-1717.

Linnartz, J.P.M.G., Westerman M., Hamerslag, R. “Monitoring the Bay Area Freeway Network using Probe Vehicles and Random Access Radio Channels”. California PATH research report, U.C. Berkeley-ITS-PRR-94-23, October 1994.

Paul U., Subramanian A. P., Buddhikot M. M., Das S. R., 2011. Understanding traffic dynamics in cellular data networks. *INFOCOM, 2011 Proceedings IEEE,* 882-890.

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., Ratti, C., 2010. Activity-aware map: Identifying human daily activity pattern using mobile phone data. *Human Behavior Understanding, Springer Berlin Heidelberg,* 14-25.

Ratti, C., Williams, S., Frenchman, D., Pulselli, R. M., 2006. Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B Planning and Design,* 33(5), 727-748.

Rubio, A., Sanchez, A., Frias-Martinez, E., 2013. Adaptive non-parametric identification of dense areas using cell phone records for urban analysis. *Engineering Applications of Artificial Intelligence*, 26(1), 551-563.

Yang C., Itoh M., 2009. Empirical study on inter-call time distribution. *Consumer Electronics, 2009. ISCE'09. IEEE 13th International Symposium,* 52-53.

Ygnace, J., Drane, C., Yim, Y.B. de Lacvivier, R., 2000. Travel time estimation on the San-Francisco bay area network using cellular phones as probes. *University of California, Berkeley, PATH Working*, Paper UCB-ITS-PWP-2000-18.

Ygnace, J., 2001. Travel time/speed estimates on the French Rhone corridor network using cellular phones as probes. *SERTI V Program, STRIP (System for Traffic Information and Positioning) Project, INRESTS – LESCOT 0201,* Lyon, France.

Yuan Y., Raubal M., Liu Y., 2012. Correlating mobile phone usage and travel behavior–A case study of Harbin, China. *Computers, Environment and Urban Systems*, 36(2), 118-130.

Zheng, J., Liu, S., Ni, L. M., 2013. Effective Routine Behavior Pattern Discovery from Sparse Mobile Phone Data via Collaborative Filtering. *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 18, 29-37.