Late for Good

Vsevolod Salnikov and Renaud Lambiotte
naXys, University of Namur, Belgium

In this work, we perform a large-scale experiment on human mobility by using smartphones as sensors. To do so, we have developed an app continuously tracking significant displacements of phones and providing in return to their user an attractive service. Contrary to standard tracking procedures based on Call Detail Records, our methods offers the advantage of collecting data in real-time, in an anonymous way, and without the constraints usually imposed by mobile phone operators.

The last few years have witnessed the increasing use of mobile phone devices as a way to collect social data at a large-scale [1, 2]. Each of us carries a mobile phone, almost continuously, and its sensors and online services provide a more and more complete view on our life and our environment. Our phone knows our whereabouts, our friends, where we meet, our taste in music, etc. The integration and monetization of these data by corporations raises alarming privacy issues. Yet, when treated ethically and used for the common good, they also have the potential to provide innovative ways to solve problems in areas such as public health, ecology or urban planning [3, 4]. Examples include the automatic detection of dysfunctions in the transport infrastructure, and the participatory report of ecological or epidemiological data.

In the particular case of mobility tracking, two different types of approach have been developed to extract individual trajectories from mobile phone data. The first approach exploits data already routinely collected by cellular network or by online services. Examples include Foursquare check-in data [5], or Call Detail Records of a phone connecting to a cell tower [6, 7]. These data have the advantage of being acquired for free and of being analyzed at virtually no cost, but they have the disadvantage of being proprietary, often with strict confidentiality agreements, and of variable quality. For instance, CDRs are known to provide a sparse and heterogeneous sampling of the trajectories, with a fairly poor spatial resolution.

The second approach aims at tracking mobility in a more controlled way, either by handing mobile devices to a limited number of users, or by distributing software, typically downloadable apps, tracking motion. However, this approach either relies on ad-hoc infrastructure or on the active participation of users, and it has, as a consequence, been limited to small sample sizes so far [8]. The aim of our work is to address this limitation and to explore the possibility to attract a large number of individuals in a mobility tracking experiment. Because the deployment of infrastructure is prohibitively expensive at this scale, we focus on the development of apps and their adoption by the broader public [9, 10]. To be installed, an app needs to compete against thousands of others and, typically, to be run continuously in the background of the phone OS despite its energy cost. To be successful, the deployment of a tracking experiment thus requires proper incentives for a user to participate. For this reason, a majority of experiments have failed to reach a large part of the population, and they have typically been limited to circles of *geeks*, students and researchers.

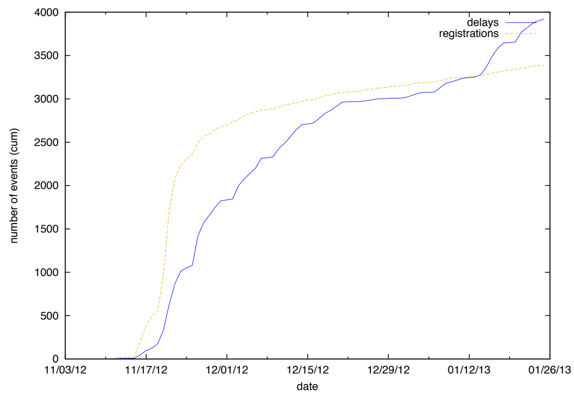

FIG. 1: Time evolution of the total number of users and submitted delays. After an initial rapid growth, we are in a steady regime, where the number of new users and delays is roughly constant.

In this work, we have developed an app for iPhones and Android devices, called *lateforgood* [13], where we offer an attractive service to the user, and not only the prospect of participating in a research experiment. The app is dedicated to Belgian train users, and helps them keep track of their train delays and submit a form for financial compensation to the national train company [14]. Incentives for the user are thus: money (paid by the train company for delays), time and convenience (the otherwise tedious form is automatically filled on our servers) but also the important feeling of being heard by the train company as a collective voice (the punctuality of train delays has been the subject of animated debates in Belgium in the last years).

In practice, the app works as follows. When it is first opened, it starts tracking the motion of the user, in order to detect closeby train stations. The app runs in

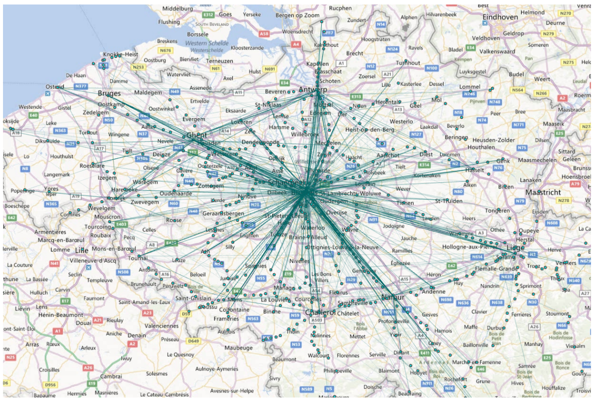

FIG. 2: Geographical representation of train delays. Two train stations are connected when a delay has been notified. Darker lines correspond to more delays.

the background on the mobile device, using a technology similar to the one used by *openpath* [11] to minimize impact on battery life: the app does not record the position continuously via GPS, but tracks only significant changes in position determined by the device API, via GPS, WI-FI access and triangulation of cell phone towers. When the user re-opens the app to declare a train delay, a list of likely official train schedules is proposed, based on recent mobility patterns, and a delay can be declared and uploaded to the user profile. When a sufficient number of delays has been accumulated, a compensation form is pre-filled and ready for download. Let us also stress that the app uploads data to our servers only when the app is used by its user. Moreover, the issue of privacy has been considered carefully as no personal information is hold: on our side, it is as if mobility patterns of anonymous mobile phones were collected.

The experiment is still in its acquisition phase. Since the launch of our service in November 2012, around 3500 users have downloaded it, among which 40% use it actively (Fig. 1). During this time period around 4000 delays have been submitted, leading to 600000 data points for the position of the users, and providing us with information on the main train lines in Belgium (Fig. 2). We are currently developing algorithms to distinguish between car and train mobility, to properly assign sampled trajectories to train routes [12], including connections between train lines, and to uncover a list of train lines and/or stations particularly affected by delays.

Acknowledgment

We acknowledge financial support from FNRS.This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office.

[1] R. Kwok, “Phoning in data”, Nature **458**, 959-961 (2009)
[2] J. Reades, F. Calabrese, A. Sevtsuk and C. Ratti, “Cellular Census: Explorations in Urban Data Collection”, IEEE Pervasive Computing **6**, 30–38 (2007)
[3] A. Pentland, D. Lazer, D. Brewer and T. Heibeck, “Using reality mining to improve public health and medicine”, Stud. Health Technol. Inform. **149**, 93–102 (2009)
[4] V.D. Blondel, M. Esch, C. Chan, F. Clerot, P. Deville, E. Huens, F. Morlot, Z. Smoreda and C. Ziemlicki, “Data for development: the D4D challenge on mobile phone data”, arXiv:1210.0137 (2012)
[5] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil and C. Mascolo, “A Tale of Many Cities: Universal Patterns in Human Urban Mobility”, PLoS ONE **7**, e37027 (2012)
[6] M.C. González, C.A. Hidalgo, A.-L. Barabási, “Understanding individual human mobility patterns”, Nature **453**, 779-782 (2008)
[7] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, J. Rowland and A. Varshavsky, “A tale of two cities”, Proceedings of the Eleventh Workshop on Mobile Computing Systems (HotMobile ’10), 19–24 (2010)
[8] N. Eagle and A. Pentland, “Reality Mining: Sensing Complex Social Systems”, Personal and Ubiquitous Computing **10**, 255–268 (2006)
[9] http://www.funf.org/
[10] M. Wirz, T. Franke, D. Roggen, E. Mitleton-Kelly, P. Lukowicz and G. Troster, “Inferring crowd conditions from pedestrians’ location traces for real-time crowd Monitoring during city-scale mass gatherings”, Proceedings of the IEEE 21st international workshop on enabling technologies, 367–372 (2012)
[11] https://openpaths.cc/
[12] R.A. Becker, R. Cáceres, K. Hanson, J. Meng Loh, S. Urbanek, A. Varshavsky and C. Volinsky, “Route classification using cellular handoff patterns”, Proceeding of Ubiquitous Computing (UbiComp 2011)
[13] More info can be found on http://www.sci-app.com
[14] The compensation is equal to 100% of the value of the ticket if the delay is of at least 60 minutes. In case of multiple delays within a 6 month period, a compensation of 25% (50%) can be claimed if, over a period of 6 months, at least 20 (10) delays of at least 15 (30) minutes are suffered. The compensation is received in the form of vouchers.