Measuring the impact of epidemic alerts on human mobility using cell-phone network data

Vanessa Frias-Martinez, Alberto Rubio, Enrique Frias-Martinez

Telefonica Research, Madrid, Spain
{vanessa, efm}@tid.es

Abstract. Accumulating evidence reveals a strong link between human mobility and the spread of epidemics. In order to control the spread of an epidemic, governments can implement mobility restrictions to its citizens. The effect of such restrictions on the mobility of the population has not been adequately studied at a large scale mainly due to the lack of relevant data. Nevertheless, the recent adoption of ubiquitous computing technologies enables the design of such studies. In this paper we measure the impact that the alerts issued by the Mexican government had on the mobility of the Mexican population during the H1N1 flu outbreak in April and May of 2009. The mobility of individuals was characterized using anonymized Call Detail Records (CDRs) traces. The results indicate a statistically significant reduction, of up to 80% in some cases, in the diameter of mobility of individuals.

Keywords: human mobility, call detail records (CDR), computational epidemiology, epidemic alerts.

1 Introduction

Human mobility plays a central role in the spatial spreading of infectious diseases [1-2]. Understanding its actual effect on pandemic propagation is a key issue in order to design adequate epidemiological models that might allow us to predict the impact of future epidemics and control its spread. The recent outbreaks of pandemics, such as H1N1, have caused a surge in the number of papers successfully combining epidemic spreading models and mobility models to optimize the strategies for epidemic containment [3-6]. In general, these studies show that mobility restrictions can delay the spread of epidemics but are not sufficient to contain them.

In case of a pandemic, the World Health Organization (WHO) recommends to authoritative bodies the assessment of the suspension of activities in educational, government and business units as a plausible measure to reduce the transmission of a disease [7]. Following these recommendations, governments have usually instituted policies that aim to reduce individual mobility in order to control an epidemic. The preventive actions implemented by the Mexican government to control the H1N1 flu outbreak of April 2009 constitute an illustrative example.

Although the role of mobility in epidemic spreading has been studied [3-6], research into the impact that government preventive actions have on the mobility of the general population is limited. Understanding the impact of such mandates is critical for the design of policies aimed at reducing human mobility and control the spread of future outbreaks. The deficiency of analytical results on the impact of such mandates is mostly due to the lack of large scale quantitative data about human motion. Nevertheless, the recent adoption of cell phones by very large portions of the population enables to capture large scale quantitative data about human mobility.

In this paper we measure the impact that the actions taken by the Mexican government during April and May of 2009 had on human mobility using phone Call Detail Records (CDRs). The actions consisted of alerts and/or mandates aimed at reducing mobility, and were issued in three stages: (a) a *medical alert [8] or stage 1*, issued on Thursday, April 16th, which was triggered by the diagnosis of H1N1 flu cases, followed by (b) the *closing of schools and universities [9] or stage 2*, enacted from Monday, April 27th through Thursday, April 30th, and (c) the *suspension of all non essential activities [10] or stage 3*, implemented from Friday, May 1st to Tuesday, May 5th. We evaluate the impact of the alerts using two approaches: (1) a *Population Mobility Analysis* that computes the aggregated mobility of all individuals and analyzes its change, and (2) a *Geographic Mobility Analysis*, that evaluates changes in human mobility at specific geographical locations. The aim of this paper is to evaluate the capabilities of CDR data as a new way of measuring the impact of epidemic alerts in order to complement traditional surveillance techniques.

2 Capturing Mobility Information

Cell phone networks are built using a set of towers or base transceiver stations (BTS) that are in charge of communicating cell phones with the network. Each BTS tower has a geographical location expressed by its latitude and longitude. The area covered by a BTS is called a sector. For simplicity, we assume that the area of coverage of each BTS tower can be approximated with a 2-dimensional non-overlapping polygon and use Voronoi tessellation to define it. Call Detail Records (CDRs) are generated when a cell phone connected to the network makes or receives a phone call or uses a service (e.g., SMS, MMS, etc.). In the process, and for invoice purposes, the information regarding the call is logged, including the BTS used, which gives an indication of the geographical position of the user. The location of each individual is known at a BTS level, no information of the location within the cell is known. Depending on the population density, the area covered by a cell ranges from less than $1Km^2$ in dense urban areas to more than $3Km^2$ in rural areas.

In this study, cell phone CDRs for 1,000,000 anonymized customers from one of the most affected Mexican states were obtained for a period of 5 months from January 2009 to May 2009. From all the information contained in a CDR, only the originating encrypted number, the destination encrypted number, the time and date of the call and the BTS used for the communication were considered in our study. Song et al. [12], sowed that mobility models computed from CDRs can accurately predict the real locations of users with 93% accuracy for users with an average call frequency > .5 calls per hour. In our case, we relaxed the requirement and only considered users with an average of two daily calls or more.

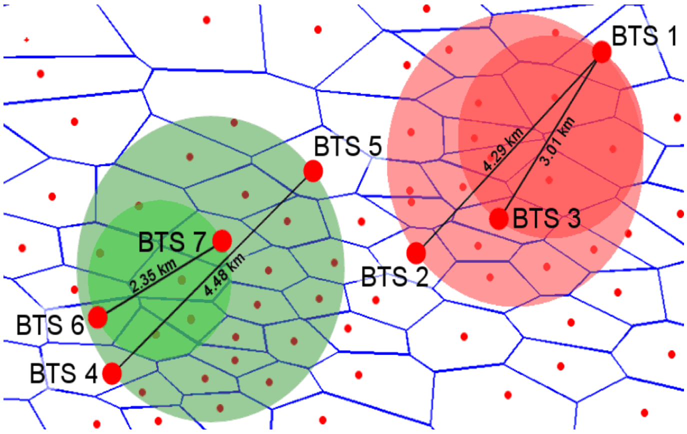

Figure 1. Diameter and area of influence of two individuals (one marked in green and the other in red) for the stage 2 alert and its baseline. Red dots represent BTS towers with their coverage approximated using Voronoi. In the case of the individual marked with red, the baseline diameter of mobility is 4.29km (defined by BTS1 and BTS2), and during the alert period the diameter is reduced to 3.01km (the distance between BTS1 and BTS3). For the individual marked in green, the baseline period has a diameter of 4.48km (defined by BTS4 and BTS5), and for the alert period the diameter is reduced to 2.35Km (defined by BTS6 and BTS7).

3 Population Mobility Analysis

The Population Mobility analysis focuses on comparing the aggregated mobility of the population during the different alert stages with a baseline intended to characterize *typical* mobility behavior. For that purpose, the mobility of each individual during each alert period was characterized by the *diameter* of his or her area of influence, where the area of influence is defined as the geographical region where the daily activities of that individual take place. The diameter of the area of influence is defined as the maximum distance between all the BTS towers used by an individual during the temporal period of study.

On the other hand, the baseline was defined as the average diameter for each individual during a set of *normal* time periods, and is computed differently for each alert stage: (a) *baseline 1* is used to quantify the changes in mobility that took place during stage 1 alert period (from Thursday 16th to Wednesday 22nd). This baseline was computed using mobility data from four 7-day time periods – from Thursday to Wednesday – prior to April 16th considered to represent the typical weekly mobility behavior under *normal* circumstances, specifically: January 15th-21st, January 22nd-28th, February 12th-18th and March 5th-11th; (b) *baseline 2*, defined to evaluate changes in mobility during the stage 2 alert (from Monday 27th to Thursday 30th), was computed using data from four 4-day time periods – from Monday to Thursday – prior to April 16th, specifically January 19th-22nd, January 26th-29th, February 16th-19th and March 9th-12th ; and (c) *baseline 3*, defined to identify mobility changes during the stage 3 alert (Friday May 1st to Tuesday May 5th), was computed using data from Easter holidays in order to represent the *typical* behavior during a holiday period. Note that May 1st and May 5th were bank holidays in Mexico (Labor Day and Cinco de Mayo), hence the choice of Easter holidays to define *baseline 3*: Friday April 10th to Sunday April 12th, corresponding to Friday May 1st to Friday May 3rd of the alert, and Monday April 6th and Tuesday April 7th, corresponding to Monday May 4th and Tuesday May 5th of the alert. The limitations imposed by the data available imply that for this particular case the baseline can only be defined by one time period. Fig. 1 presents an example of the diameter of mobility of two subscribers for the stage 2 alert period and its baseline, as well as the change in mobility experimented by them.

In order to quantify the impact of the government calls on mobility, a one-sided t-test was used to compare the distribution of diameters between each stage and its corresponding baseline. All baseline and alert period distributions were previously checked against the Lilliefors test to guarantee that they followed a normal distribution. The validity of the baselines was also assessed by comparing them with the distribution of mobility diameters obtained during different control periods prior to April 16th. No statistically significant differences were found, thus indicating that baselines represent *typical* mobility behavior.

In the case of the stage 1 alert period, no significant change in human mobility was detected. However, the t-test revealed statistically significant differences for the stage 2 and stage 3 alert periods with $p<0.01$: the distribution of the diameter of mobility was reduced during these two stages compared to their baselines. The same statistical analysis was carried out on a daily basis by comparing the distribution of diameters for each day of stages 2 and 3 with its corresponding daily distribution in the baseline. The one-sided t-test was statistically significant in all cases except for Saturday, May 2nd and Sunday, May 3rd.

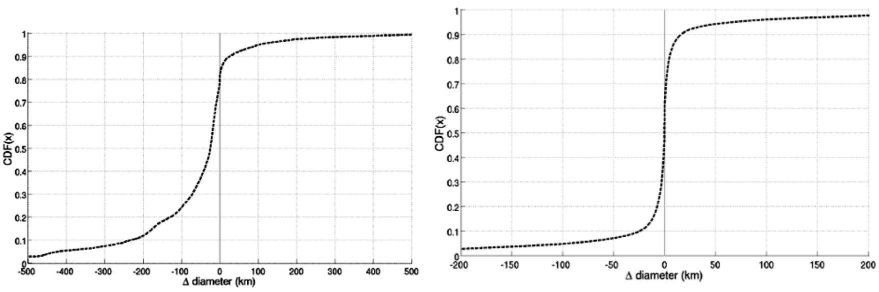

Figure 2. (Left) CDF of the change in diameter of individual mobility on April 27th, where the X-axis presents the variation in Km (a positive variation implies that the individual had a larger diameter during the alert period than during the baseline and viceversa); and (Right) CDF of the change in diameter of individual mobility on May 1st.

In order to quantify the changes in mobility during alert periods 2 and 3, we subtracted the diameter of each subject in each day of those stages from his/her diameter in the corresponding day of the baseline. Figure 2 (left) depicts the Cumulative Distribution Function (CDF) of diameter change for April 27th relative to its baseline (similar graphs were obtained for the remaining days of stage 2), where we observe that 80% of the population reduced its diameter, with around 50% of them reducing it by 20*km* or more. Similarly, Figure 2 (right) presents the diameter change for May 1st (similar graphs were obtained for the remaining days of stage 3, except for May 2nd and May 3rd), where 55% of the population reduced its diameter, and approximately 20% of them by more than 10*km*. We observe larger values in the reduction of the diameter of mobility during stage 2 alert when compared to stage 3 alert, probably due to the fact that stage 3 alert was already a holiday.

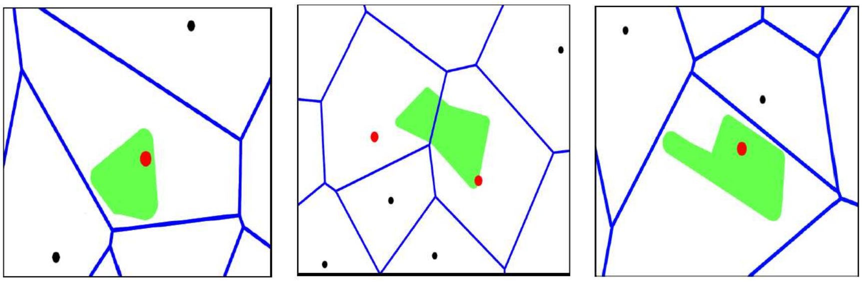

Figure 4. Area of coverage of each BTS and the space occupied by the corresponding infrastructure: (left) the university campus, (center) the hospital and (right) the airport.

4 Geographic Mobility Analysis

The Geographic Mobility analysis evaluates the impact that the alerts had on specific geographic areas that contain critical infrastructures. Such analysis aims to understand whether the number of individuals that visited these infrastructures varied as a result of the government mandates. Representing the coverage of each BTS tower using Voronoi tessellation allows us to identify the BTS towers that handle the calls of individuals located at specific infrastructures. It is important to note that the geographical coverage of a specific BTS might not only include the infrastructure under study but also other residential or business areas. We carry out the Geographic Mobility analysis on three different infrastructures: the main university campus, the main hospital complex and the airport at the capital of the state under study, as shown in Figures 4. The red dots represent the set of BTS towers that give service to the infrastructure, the black dots represent neighboring BTS towers, the blue lines the coverage of each BTS and the green area the actual geographic location of the infrastructure. In the case of the airport terminal, the whole infrastructure is covered by one BTS. Similarly, the university campus is covered by a unique BTS, but in this case its coverage includes residential areas. Finally, the hospital complex is covered by two BTS towers whose coverage also includes densely populated residential areas.

In order to measure the impact of the mandates in these infrastructures, each one was characterized by the number of unique individuals that visited them, daily and hourly, during each alert period and those signals were compared to its baseline. The daily and hourly baselines were defined for each infrastructure as the average number of unique individuals whose calls were handled by the BTS tower that gives coverage to the infrastructure for each day/hour during the time periods previously defined.

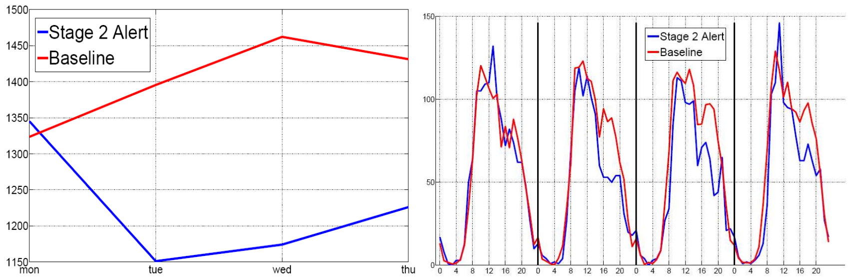

Figure 5. (Left) Number of individuals that visited the university campus during the second alert period (in blue) and its baseline (in red) aggregated daily; and (Right) aggregated hourly.

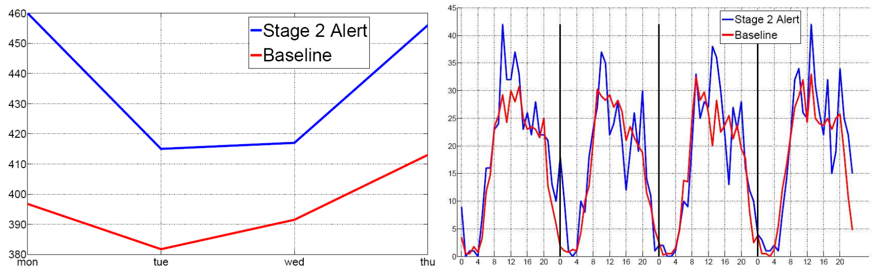

Figure 6. (Left) Number of individuals that visited the airport during the second alert period (in blue) and its baseline (in red) aggregated daily. (Right) The same data aggregated hourly.

As mentioned earlier, the coverage of the BTSs that gives service to an infrastructure might include residential areas. Since we are interested in computing the impact of the government mandates on specific infrastructures, each signal was corrected by subtracting all individuals whose residence is located in the area of coverage of the corresponding BTS. By eliminating all the individuals from the residential areas, we can focus our analysis on the infrastructure itself. The residence location of the individuals was computed using a residential detection algorithm that associates a *home BTS* to an individual based on its cell phone usage [11]. Fig. 5 presents the number of unique users that visited the university campus during the second alert period and its corresponding baseline aggregated daily (left) and hourly (right). Similarly, Fig. 6 shows the number of unique users that visited the airport during the second alert period and its baseline daily (left) and hourly (right).

	Stage 1 Alert	Stage 2 Alert	Stage Alert 3
University Campus	h=0	h=1 (left) p=0.007	h=1 (left) p=0.0008
Hospital Complex	h=0	h=0	h=0
Airport Terminal	h=0	h=1 (right) p=0.012	h=0

Table1. Statistical significance of the change in number of visitors for each infrastructure selected and each stage of alert considering the daily representation.

Focusing on the daily representation, the Lilliefors test did not reject the null hypothesis (p>0.01) for any of the baseline and alert signals, indicating that they follow a normal distribution. In order to check the deviation during the alerts from the baseline in each infrastructure, a pair wise analysis of variance (ANOVA) between each pair of baseline and alert signal was performed. Table 1 presents for each alert period and each infrastructure the result of the test when using the daily representation, where h=1 rejects the null hypothesis and indicates that they originate from different distributions. If that is the case, the second parameter indicates if the number of individuals of the baseline is higher than the alert signal (noted as left) or if the number of individuals of the alert signal is higher than the baseline (noted as right). The third parameter indicates the significance value obtained. Table 1 shows no statistically significant difference in the number of visitors to any of the infrastructures under study during the stage 1 alert. This result is aligned with the evaluation obtained in the Population Mobility analysis. On the other hand, we observe that there is a statistically significant reduction in the number of visitors observed at the university campus during the stage 2 alert period (see Fig. 5 (left) for details). We also observe a statistically significant increase in the number of visitors to the airport during the same stage 2 alert period (see Figure 6(left) for details). This is possibly caused by, among others, the fact that after stage 2 finishes, a long holiday weekend follows, which combined with the enactment of suspension of all non essential activities of stage 3, would motivate people to leave the city or take a vacation. Table 1 also shows that during the stage 3 alert period, only the university campus experienced a statistically significant decrease in the number of individuals that visited the infrastructure.

When studying the hourly changes in the number of visitors to each infrastructure, the Lilliefors test did reject the null hypothesis (p<0.01) for all baselines and alert periods indicating that they do not follow a normal distribution. As a result, when comparing each hourly alert period distribution with its baseline, instead of using a t-test, a Kolgomorov-Smirnof test and a Wilcoxon rank-sum test were used. Both tests did not reject the null hypothesis (p>0.01) indicating that there is no statistically significant change in the number of hourly visitors during the alert periods and its baselines in any infrastructure. As can be seen in Fig. 5(right) and Fig. 6(right), there is a difference between the stage 2 alert distribution and its baseline for both the airport and the university, nevertheless such difference is scattered throughout the day and thus not statistically significant.

5 Conclusions

This paper represents a first step towards the use of CDR data for evaluating the impact of government mandates using as example the alerts issued by the Mexican government during the H1N1 outbreak of April 2009. The population mobility analysis provided evidence that: (a) medical alerts (stage 1) do not seem to significantly change human mobility, whereas (b) interventional actions (stage 2 and 3) significantly change the diameter of mobility, particularly if the intervention takes place during regular working days. The reduction in mobility is higher when schools and universities are closed during regular days (stage 2) than when all non-essential activities are closed (stage 3) during a period that already was a holiday. A direct consequence for the design of epidemic alerts is that the enactment of a total closure of activities during a holiday period is not as effective for slowing down the spread of the epidemic as the partial closing of some activities (typically schools) during regular working days.

The geographic mobility analysis indicated: (a) that the increase in number of visitors that the airport received during the stage 2 alert implies that mandates such as the total closing of infrastructures (stage 3 alert) might provoke an increase in the number of individuals that visit transport hubs before its enactment, thus limiting the containment and possibly causing an undesired increase in the spread of the epidemic and (2) that no statistical significant change in the number of visitors to the Hospital Complex was found throughout all the alert periods, indicating that medical alerts did not seem to push the population towards physically seeking medical advice. Also, these results not only qualitatively provide an answer of the impact of the alerts, but also provide a quantitative measure of the change in mobility.

References

1. Riley S. Large-Scale Spatial-Transmission Models of Infectious Disease. Science 2007; 316(5829):1298-1301.
2. Epstein JM, Goedecke DM, Yu F, Morris RJ, Wagener DK, Bobashev GV. Controlling pandemic flu: the value of international air travel restrictions. PLoS One 2007; 2(5).
3. Balcan D, Colizza V, Gonsalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc. Nat. Aca. Sci. 2009; 106(51): 21484-21489.
4. Cauchemez S, Valleron AJ, Boelle PY, Flahault A, Ferguson NM. Estimating the impact of school closure on influenza transmission from Sentinel data. Nature 2008; 452:750-754.
5. Wallinga J, van Boven M, Lipsitch M. Optimizing infectious disease interventions during an emerging epidemic. Proc. Nat. Aca. Sci. 2010; 107(2):923-928.
6. Colizza V, Barrat A, Barthelemy M, Vespignani A. The role of the airline transportation network in the prediction and predictability of global epidemics. Proc. Nat. Aca. Sci. 2006; 103(7):2015-2025.
7. World Health Organization - Guidance for National Authorities (2009) (www.who.int/csr/disease/swineflu/guidance/national authorities/en/index.html).
8. PAHO Office, World Health Organization (WHO). Announcement, Latin America and Caribbean on alert for influenza A April 29th, 2009). (new.paho.org/hq/index.php?option=com_content\&task=view\&id=1261%\&Itemid=1).
9. DOF Secretaria de Gobernación de México (SEGOB), Decreto ordenación diversas acciones para prevenir, controlar y combatir virus de influenza (April 25th, 2009). (dof.gob.mx/nota_to_imagen_fs.php?codnota=5088366&fecha=25/04/200%9).
10. DOF Secretaria de Gobernación de México (SEGOB), Acuerdo de suspensión de labores en la administración publica federal (April 30th, 2009). (http://dof.gob.mx/nota_detalle.php?codigo=5089064&fecha=30/04/2009).
11. Frias-Martinez V, Virseda J, Rubio A, Frias-Martinez E. Towards Large Scale Technology Impact Analyses: Automatic Residential Localization from Mobile Phone- Call Data. ICTD 2010, London, UK.
12. Song C, Qu Z, Blumm, Barabasi AL. Limits of predictability in human mobility. Science, vol. 327, no. 5968, pp. 1018–21, 2010.