Human Mobility in Advanced and Developing Economies: A Comparative Analysis

Alberto Rubio, Vanessa Frias-Martinez, Enrique Frias-Martinez and Nuria Oliver

Data Mining and User Modeling Group
Telefonica Research, Madrid, Spain
{arm, vanessa, efm, nuriao}@tid.es

Abstract

The deployment of ubiquitous computing technologies in the real-world has enabled the capture of large-scale quantitative data related to human behavior, including geographical information. This type of data creates an opportunity to characterize human mobility, with potential applications ranging from modeling the spread of viruses to transportation planning. This paper presents an initial study focused on understanding the similarities and differences in mobility patterns across countries with different economic levels. In particular, we analyze and compare human mobility in a developing and an advanced economy [1] by means of their cell phone traces. We characterize mobility in terms of (1) average distance traveled, (2) area of influence of each individual, and (3) geographic sparsity of the social network. Our results indicate that there are statistically significant differences in human mobility across countries with different economic levels. Specifically, individuals in the developing economy show smaller mobility and smaller geographical sparsity of their social network when compared to individuals in the advanced economy.

1. Introduction

The study and characterization of human mobility has a wide range of applications that include modelling the spread of viruses (Huerta and Tsimring 2002), improvements to cellular network efficiency (Zang and Bolot 2007), and the detection of urban hotspots (Djordjevic et al. 2008). While the mobility of animals has already been quantitatively studied (Viswanathan 1996; Sims 2008), our understanding of human mobility has been somewhat limited mostly due to the lack of large scale quantitative mobility data over wide geographical areas.

The recent adoption of ubiquitous computing technologies by a very large fraction of the population in both advanced and developing countries enables to capture –for the first time in human history– large scale quantitative data about human motion. In this context, mobile phones play a key role as sensors of human behavior as these are typically owned by individuals that carry them at –almost– all times. Hence, it comes at no surprise that most of the quantitative data about human motion has been gathered via GPS or Call Detail Records (CDRs hereafter) from cell phone networks.

Seminal work by Gonzalez, Hidalgo, and Barabasi (2008) analyzed the CDRs of $100,000$ individuals showing that the majority of people mostly travel between two nearby locations (i.e., work and home) and only occasionally travel larger distances. Other authors have carried out rather original human mobility studies using other sources of human mobility data, for example, bill notes (Brockmann, Hufnagel, and Geisel 2006) or Second Life traces by (La and Michiardi 2008).

In this paper we compare human mobility patterns (computed from CDRs) in a developing and an advanced economy, in order to shed light on which aspects of human mobility may be common *vs* different across countries with different economic levels. This analysis constitutes an initial effort towards understanding whether there exist universal human mobility patterns or if, on the contrary, mobility trends differ across economies. The broader goal is to understand human mobility in order to guide the implementation of public policies in areas such as disease spreading, transportation management, and crime prevention.

In particular, we focus on characterizing: (1) general mobility or the average distance traveled by an individual; (2) the area of influence of a person, which describes the size of the geographical area where a person spends most of his/her time; and (3) the geographic sparsity of the social network of an individual, which models the geographic spread of the contacts of an individual.

To the best of our knowledge, this is the first study of its kind to: (1) compare human mobility in a developing and an advanced economy; and (2) model the geographical relationship between an individual and his/her social network (geographic sparsity).

2. Related Work

The analysis of CDR data is at the forefront of human mobility modeling because of its availability for a large and diverse number of individuals. Other data sources (mainly GPS, cell-based location, LAN/WAN) will introduce a bias in the sample *e.g.,* data from wireless LAN/WAN only considers users with access to a wireless network. By compari-

[1]Based on the 2009 International Monetary Fund (IMF) World Economic Outlook Country Classification www.imf.org/external/pubs/ft/weo/2009/02/weodata/groups.htm

son, a sample derived from CDR data is typically less biased due to the pervasiveness of cell phones worldwide, both in developing and advanced economies.

As noted previously, the findings of Gonzalez, Hidalgo, and Barabasi (2008) indicate that human displacements of cell phone users typically concentrate on two nearby locations following a power-law distribution with an exponential cut-off. Additional uses of CDRs include the work by Hung, Peng, and Huang (2005) who described a procedure to mine similar user moving patterns computed from CDRs. Similarly, Seshadri et al. (2008) modeled specific user calling features including number of calls and duration of the calls. Although CDRs have been previously used to analyze human mobility, to the best of our knowledge this paper represents the first comparison of human mobility in developing and advanced economies.

Ultimately, differences or similarities in human mobility across countries might be useful when tackling issues such as epidemic spreading or transportation planning. For example, Anderson et al. (2004) studied the transmission dynamic and control of the SARS epidemic in Hong Kong from information on epidemiological, demographic and clinical variables on 1425 cases. As the authors state, one of the greatest difficulties lies in modeling diversity in the mobility of the population. We believe that mobility studies like the one presented in this paper could greatly enhance existing epidemic models.

In the area of transportation planning, Liu, Biderman, and Ratti (2009) evaluated real-time urban mobility dynamics in Shenzhen, China. The study used two different real-time data sources: GPS data from 5,000 taxis and data from 5 million smart cards from buses and subway. In order to improve transportation systems in rural and urban environments, it is crucial to also model the mobility of people who may lack means of public transportation. We believe that mobility models obtained from CDRs will provide insight of new routes that should be considered for future transportation developments.

3. Data Acquisition and Filtering

Cell phone networks are built using a set of base transceiver stations (BTS) that are in charge of communicating cell phone devices with the network. Each BTS is identified by the latitude and longitude of its geographical location. Call Detail Records (CDRs) are generated whenever a cell phone connected to the network makes or receives a phone call or uses a service (e.g., SMS, MMS). In the process, and for invoice purposes, the information regarding the BTS is logged, which gives an indication of the geographical position of the user at the time of the call. From all the information contained in a CDR, our study only considers the originating encrypted number, the destination encrypted number, the time and date of the call, the duration of the call, and the BTS that the cell phone was connected to when the call was placed.

In total, we obtained cell phone CDRs for nearly 300,000 anonymized customers in an advanced economy ($sample1$) as well as 700,000 anonymized customers in a developing economy ($sample2$). The CDRs were collected over the

same period of four months in each country. These subscribers have either a contract with the cell phone company, or use pre-paid cards for their calls. In the case of the developing economy, 5% of the subscribers have a contract while the remaining 95% used pre-paid cards. In contrast, subscribers with a contract in the advanced economy accounted for approximately 70% of the sample.

In order to make sure that we obtain robust mobility models, we only want to consider subscribers that use cell phones on a regular basis. For both samples, we computed the average number of calls per day per individual. Figure 1 displays the Cumulative Distribution Function (CDF) for the average number of calls per day *i.e.,* the percentage of subscribers y in the total population that have a specific average number of calls x per day. One can see that individuals in the advanced economy tend to use cell phones much more than in the developing economy. In fact, almost 75% of individuals in the advanced economy make/receive on average at least 2 calls per day, while only 27% of the population in the developing economy averages the same number of calls per day. This result is probably related to the different billing models between contract and pre-paid customers. Recall that most of the individuals in *sample1* had a contract whereas most of the individuals in *sample2* had a pre-paid card. Cell phone contracts typically entail a flat monthly rate or a fix cost per X number of minutes. Hence, contract customers have a tendency to consume their allowance of minutes each month, while pre-paid customers are more cautious about their cell phone use.
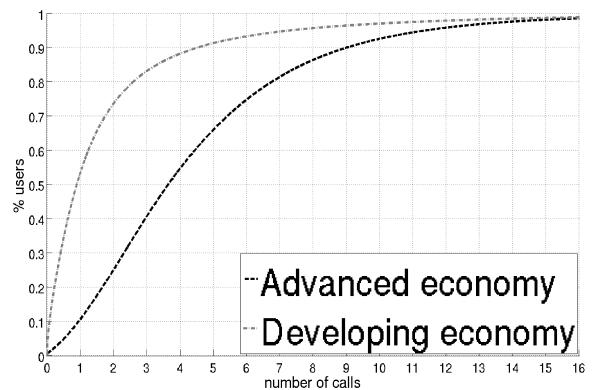

Figure 1: CDF of the average number of calls per day for a developing and an advanced economy.

Based on this analysis, for the rest of the paper we only consider subscribers with an average of at least two calls per day. Such restriction, guarantees a minimum amount of cell phone activity per subscriber. After applying this filter, we are left with 220,000 individuals in the advanced economy and 190,000 individuals in the developing economy.
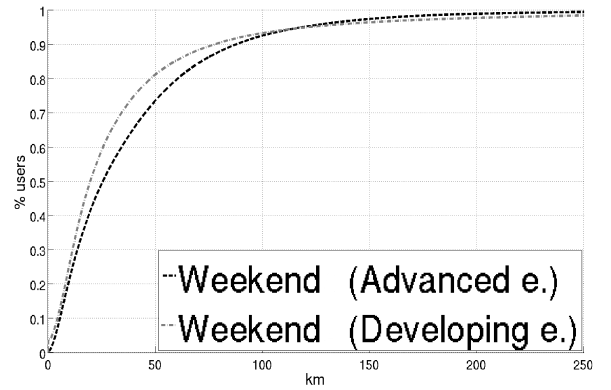
4. Description of Mobility Variables

In order to analyze and compare the human mobility in both samples, we consider the following three variables:

- *Average Distance Traveled*: Defined as the average distance traveled by a user during the period of time of study,
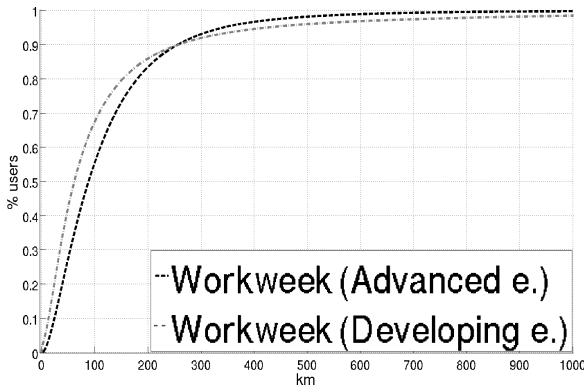
(a) Workweeks

(b) Weekends

Figure 2: CDF of the average distance traveled during (a) workweeks and during (b) weekends.

where each individual distance is computed as the geographical distance between the BTSs used by two consecutive calls. This variable provides an approximation of the average mobility of a user.

- *Area of Influence*: Defined as the diameter of the enclosing circle where a user moves. Formally, it is computed as the maximum distance (in kilometers) between the set of BTSs used to make/receive all calls during the period of time of study.

- *Geographical Sparsity of the Social Network*: Defined as the average distance between the user and all the members of his/her social network. For a specific period of time, it is calculated by averaging all the individual distances between the position of the user and each member of his/her social network every time a call is made/received. To the best of our knowledge, this is a novel parameter in the literature and introduces a measure of the geographical sparsity of an individual's social network by combining mobility and social networks measurements.

5. Comparative Mobility Analysis

In order to measure and compare these variables in our samples, CDRs for individual users were divided in two sets: the *workweek set* consisted of all the activity accumulated from Monday to Friday, and the *weekend set* consisted of all activity accumulated from Saturday to Sunday.

Average Distance Traveled

Figures 2(a) and 2(b) show the CDF of the average distance traveled per individual during the workweeks and weekends respectively. Throughout, the distance traveled is computed as the average of the workweek and weekend distances for each week in the period of four months. One can observe that during workweeks 90% of the population in the advanced economy has a higher average distance traveled than individuals in the developing economy. For example, while 46% of the population in the advanced economy travels an average of 100 kilometers during the workweek, only 34% of the population in the developing economy travels the same average distance. Interestingly, there exists a small

group of users in the population of the developing economy (approximately 10%) for whom the average distance traveled exceeds that of the population in the advanced economy. In fact, long distance travelers (users traveling more than 250 kilometers per workweek) cover larger distances in the developing economy than in the advanced economy.
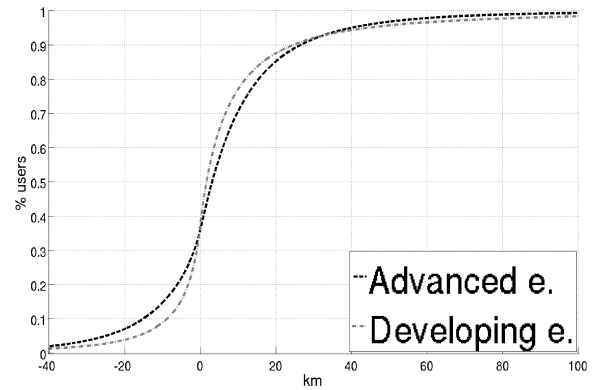

Figure 3: CDF of the distance traveled during an average weekend day subtracted from an average workweek day for each user.

A similar analysis follows for Figure 2(b) that represents the average distance traveled during weekends. The bulk of users in the developing economy (approximately 95%) travels less distance in comparison to users in the advanced economy. This trend is inverted for a small fraction of the population in the developing economy. For each pair of distributions, we run t-tests to determine whether the differences observed in the analysis were statistically significant. Both t-tests rejected the null hypothesis ($p < 0.01$) thus confirming the different nature of the distributions.

In order to understand the difference between workweek and weekend mobility, for each user we subtracted the average distance traveled during weekends from the distance traveled on workweeks. Figure 3 shows the CDF for the distances obtained from the subtractions. These distances have been normalized according to the number of days in each
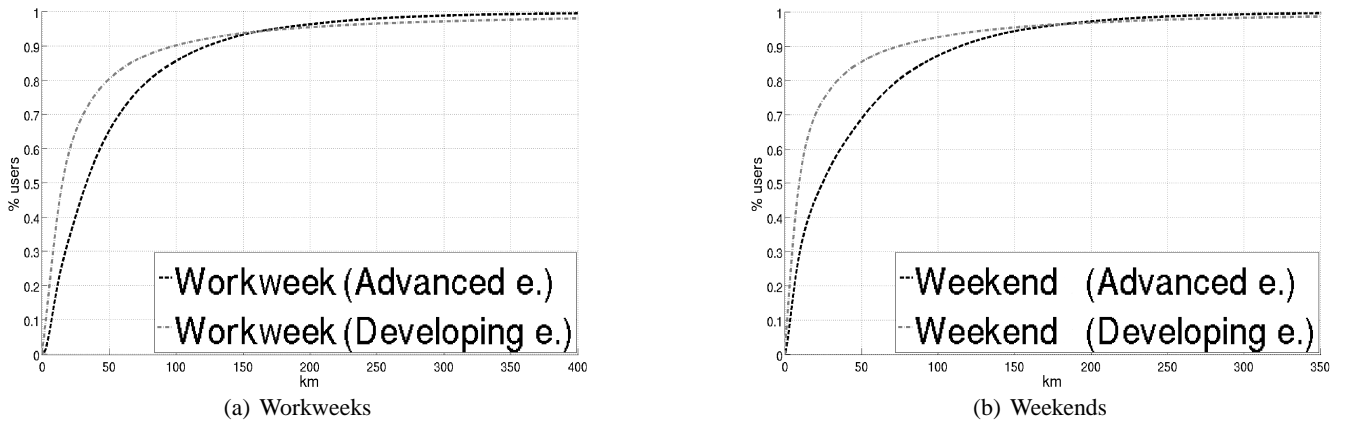
(a) Workweeks

(b) Weekends

Figure 4: CDF of the average diameter of the area of influence during (a) workweeks and during (b) weekends.

set. If the distance x resulting from the subtraction is less than zero, it indicates that the user travels more on weekends than on workweeks. If the distance is greater than zero, the user travels more on workweeks than during weekends.

In this case, one can see more similarities between the developing and the advanced samples. In fact, approximately 60% of users in both samples travel larger average distances during workweeks than during weekends, whereas the remaining 40% tends to travel more during weekends than workweeks.

One potential caveat in this analysis is the size of the country under study. In our case, the area of the advanced economy is five times that of the developing economy. However, the actual size of the country does not seem to impact the value of the average distance traveled since the ratio of distance values for both samples is almost one.

In summary, users in the advanced economy travel longer distances than users in the developing economy with the exception of a small percentage of long distance travelers. Potentially, this analysis of mobility could help scientists to model epidemic spreadings as well as to evaluate its impact in different countries. We conjecture that users with higher average distances traveled represent a higher risk of epidemic spread.

Area of Influence

The area of influence defines the diameter of the largest circle where the user has been for the period of time under study. Note that it is not necessarily related to the distance traveled. In fact, an individual could potentially travel frequent short displacements in the same geographical area, which would account for a high average distance traveled but a small area of influence. Throughout this analysis, the diameter of the area of influence is computed as the average of the workweek and weekend diameters for each week in the four month period.

Figures 4(a) and 4(b) depict the CDF of the average workweek and weekend diameter of the area of influence. In particular, Figure 4(a) shows that more than 90% of the total population in the developing economy have an area of influence smaller than that of users in the advanced economy.

In other words, individuals in the developing economy move in a geographical area which is smaller and more self-contained. For instance, 65% of individuals in the advanced economy have an area of influence with a diameter larger than 20 kilometers, whereas only 40% of individuals in the developing economy satisfy this condition. However, from a small percentage of users (approximately 6%) whose areas of influence have a diameter larger than 160 kilometers, the individuals in the developing economy have larger diameters than those in the advanced economy.
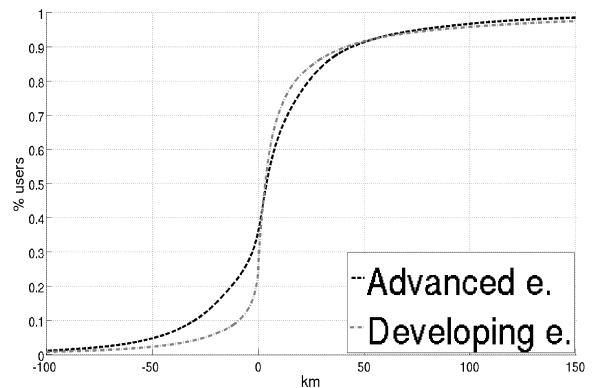

Figure 5: CDF of the diameter during weekends subtracted from the diameter during workweeks for each user.

Turning to the areas of influence during weekends, one observes an increase in the differences between the advanced and the developing economies (see Figure 4(b)). In fact, 50% of the subscribers in the developing economy have an area of influence with a diameter of approximately 10 kilometers, whereas the same percentage of individuals in the advanced economy share an average diameter of 25 kilometers.

Throughout the workweeks and weekends, individuals in the developing economy tend to have smaller areas of influence than subscribers in the advanced economy, with the exception of individuals with very large areas of influence that tend to have larger diameters among users in the developing
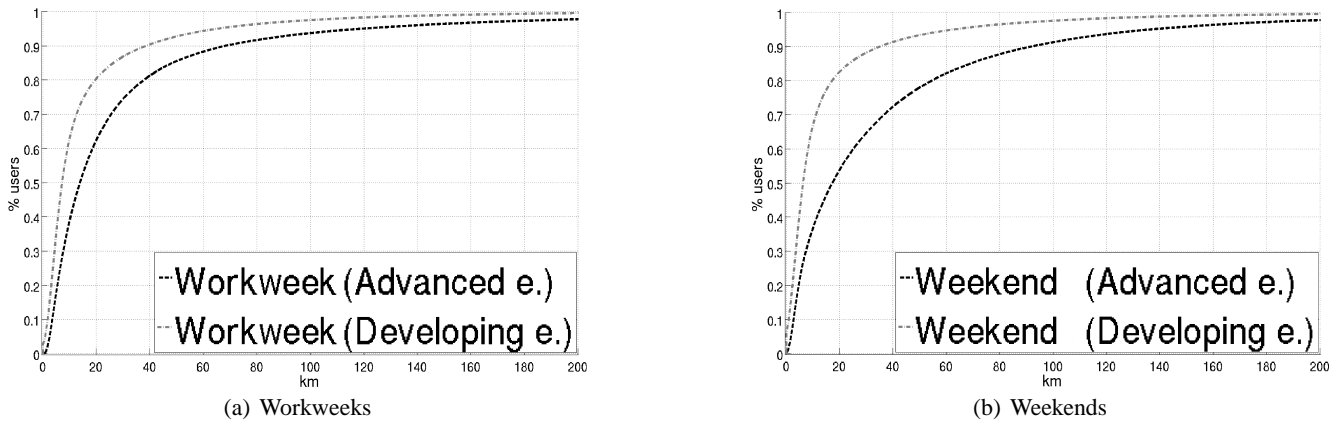
(a) Workweeks

(b) Weekends

Figure 6: CDF of the geographic sparsity of the social network during (a) workweeks and (b) weekends.

economy. For completeness, we run t-tests and checked that the pairs of distributions presented in Figures 4(a) and 4(b) were statistically significantly different with $p < 0.01$.

As previously done, Figure 5 depicts the difference (in kilometers) between the weekend diameters and the workweek diameters for each user. Approximately 60% of users in both developing and advanced economies increase their areas of influence during workweeks.

Eventually, this analysis could be used by transportation analysts to determine where to place future public transportation systems, focusing on geographical areas that are not yet covered but that show high levels of human mobility based on CDR analysis. For example, approximately 90% of the population in the developing economy has an area of influence with an average workday diameter of 20 kilometers (see Figure 4(a)). This scale could be used to determine the geographical extent of public transport in either urban or rural environments.

Geographic Sparsity of the Social Network

Figures 6(a) and 6(b) depict the CDF for the geographic sparsity of the social network for both the developing and the advanced samples. The numbers shown correspond to the average of workweek and weekend geographic sparsity values for each week over the four-month period.

From the Figures, one can gather that subscribers in the developing economy tend to have smaller geographic sparsity than users in the advanced sample during workweeks as well as weekends. Approximately 90% of the users in the developing economy have a social network with a geographic sparsity between $30 \approx 40$ kilometers whereas subscribers in the advanced economy keep larger social networks (between $60 \approx 75$ kilometers). This means that users in the developing economy tend to have their contacts geographically closer to them than users in the advanced economy. The distributions for the developing and advanced economies have been compared using a t-test ($p < 0.01$), and proven to be statistically different.

Interestingly, during weekends individuals in the developing economy slightly *reduce* the geographic sparsity of their social networks, *i.e.,* only a small fraction of users tend to communicate with people who are closer to them than those contacted during workweeks. On the contrary, users in the advanced economy *increase* the sparsity of their social networks during weekends *i.e.,* they tend to contact people who are farther away from them than those contacted during workweeks.

To see this more clearly, Figure 7 shows the CDF of the change in the social sparsity between workweeks and weekends for each subscriber. Note that a negative value indicates that an individual's social network has a larger sparsity on weekends than on workweeks. For example, 60% of the users in the advanced economy *increase* the geographic sparsity of their social network during weekends, probably indicating the existence of two differentiated social networks per user: one for the workweeks and another one for the weekends.

These results might be relevant in social and behavioral sciences, particularly in studies focusing on the analysis of personal relationships. Over time, one could model the geographic sparsity of the social networks across cultures and detect potential social changes or evolution in social relationships.
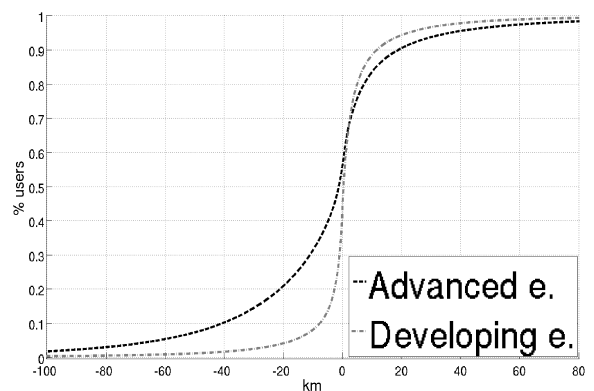

Figure 7: CDF of the weekend geographic sparsity subtracted from the workweek geographic sparsity for each user.

6. Conclusions and Future Work

In this paper we have presented a detailed comparative study of human mobility in a developing and an advanced economy using CDR data for $220,000$ cell phone users in the advanced economy and $190,000$ in the developing economy for a period of four months. We have focused our mobility analysis on three distinct variables: (1) average distance traveled, which provides an approximation of the average mobility of a user, (2) area of influence of each individual, that approximates the average diameter of the enclosing circle where a user moves, and (3) geographic sparsity of the social network, which introduces a measure of the geographical sparsity of a individual's social network by combining mobility and social networks measurements.

In our analysis, we have found that users in the developing economy tend to show smaller *average distances traveled* as well as smaller *areas of influence*. Intriguingly, we detect a fraction of the population (approximately 6%) in the developing economy for whom these variables exceed those of the population in the advanced economy. In particular, these individuals display average distances larger than 250 kilometers and areas of influence greater than 160 kilometers.

In terms of the geographical sparsity of the social networks, we have observed that subscribers in the advanced economy tend to have their social contacts at larger geographical distances in comparison to users in the developing economy. In addition, subscribers in the developing economy do not show significant changes in the geographical sparsity of their social network from workweeks to weekends. Conversely, subscribers in the advanced economy display a shift in the geographical coverage of their social networks.

We have also argued that a better understanding of human mobility across different economies could be useful in the creation of public policies to enhance human development in areas such as epidemic spread modeling, transportation alternatives and social sciences. Future work will focus on working together with epidemiologists and social scientists in order to understand the usability of our analysis, its impact and its possible improvements. In addition, we plan to continue this comparative work in more countries.

References

Anderson, R.; Fraser, C.; Ghani, A.; Donnelly, C.; Riley, S.; Ferguson, N.; Leung, G.; Lam, T.; and Hedley, A. 2004. Epidemiology, transmission dynamics and control of sars: the 2002–2003 epidemic. *Phil. Trans. R. Soc. Lond. B* (359):1091–1105.

Brockmann, D.; Hufnagel, L.; and Geisel, T. 2006. The scaling laws of human travel. *Nature* (439).

Djordjevic, B.; Gudmunsson, J.; Pham, A.; and Wolfe, T. 2008. Detecting regular visit patterns. 244–255.

Gonzalez, M.; Hidalgo, C.; and Barabasi, A. 2008. Understanding individual human mobility patterns. *Nature* (453):779–782.

Huerta, R., and Tsimring, L. 2002. Contact tracing and epidemics control in social networks. *Physical Review E 66*.

Hung, C.; Peng, W.; and Huang, J. 2005. Exploring regrassion for mining user moving patterns in a mobile computing system. In *Proceedings of the First International Conference on High performance computing and communications, HPCC*, 878–887. Lecture Notes in Computer Science, LNCS.

La, C., and Michiardi, P. 2008. Characterizing user mobility in second life. In *WOSP '08: Proceedings of the first workshop on Online social networks*, 79–84. New York, NY, USA: ACM.

Liu, L.; Biderman, A.; and Ratti, C. 2009. Urban mobility landscape: Real time monitoring of urban mobility patterns. In *Proceedings of the 11th International Conference on Computers in Urban Planning and Urban Management*.

Seshadri, M.; Machiraju, S.; Sridharan, A.; Bolot, J.; Faloutsos, C.; and Leskove, J. 2008. Mobile call graphs: beyond power-law and lognormal distributions. In *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 596–604. New York, NY, USA: ACM.

Sims, D. 2008. Scaling laws of marine predator search behavior. *Nature* (451):1098–1102.

Viswanathan, G. 1996. Levy flight search patterns of wandering albatrosses. *Nature* (381):413–415.

Zang, H., and Bolot, J. 2007. Mining call and mobility data to improve paging efficiency in cellular networks. In *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, 123–134. New York, NY, USA: ACM.