Spatiotemporal patterns of urban human mobility

Samiul Hasan[1], Christian M. Schneider[2], Satish V. Ukkusuri[1] and Marta C. González[2]

[1] School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907, USA

[2] Department of Civil and Environmental Engineering, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Abstract. The modeling of human mobility is adopting new directions due to the increasing availability of big data sources from human activity. These sources enclose digital information about daily visited locations of a large number of individuals. Examples of these data include: mobile phone calls, credit card transactions, bank notes dispersal, check-ins in internet applications, among several others. In this study, we consider the data obtained from smart subway fare card transactions to characterize and model urban mobility patterns. We present a simple mobility model for predicting peoples' visited locations using the popularity of places in the city as an interaction parameter between different individuals. This ingredient is sufficient to reproduce several characteristics of the observed travel behavior such as: the number of trips between different locations in the city, the exploration of new places and the frequency of individual visits of a particular location. Moreover, we indicate the limitations of the proposed model and discuss open questions in the current state of the art statistical models of human mobility.

PACS numbers: 64.60.aq, 89.65.Ef, 89.65.Lm, 89.75.Fb, 89.75.Kd

1. Introduction

Models of human mobility have a wide range of applications. Some can be direct approaches such as traffic forecasting [1], activity based modeling [2], and the design of transportation networks [3], while some others require more simplified models of human mobility such as the spreading of viruses and the evolution of epidemics [4, 5, 6, 7, 8], measuring human exposure to air pollutants [9], and simulating mobile networks of wireless devices[10]. In order to develop models of human mobility, conventionally, information is extracted from travel surveys [11, 12, 13, 14, 15]. These questionnaires provide us with the detailed itinerary of individual travels. They are designed to gather rich information about travel choices but are expensive to distribute over the entire population, covering typically, only a very limited time span and number of individuals. As an alternative, rich amount of information containing individual's coordinates is automatically recorded each time a person makes a call, sends an email, uses a credit card, or travels using a public transport smart card [16, 17, 18]. This traces of coordinates are recorded over months or even years. These data sources offer us the unique opportunity to understand and characterize the patterns of human travel behavior at a massive scale, unimaginable before. These big data sources contains incomplete information about individuals trips; which brings particular challenges to relate individual mobility to conventional models [11, 12, 13, 14, 15]. In general, these data sources lack the demographic information of individuals, the traces of their locations are not continuously collected and there is no detailed information about the specific choices of the individuals daily activities. In consequence, discrete choice models and transportation forecasting models [19, 20, 21] cannot be directly applied. In the light of this massive data availability, methods of analysis and modeling based in statistical physics are a suitable alternative. Our framework of analysis seeks to develop simple models and to gain understanding of the key features and difficulties involved in modeling complex socio-technical systems [22].

In this context, mobility studies, from a statistical physics perspective, started by analyzing animal trajectories, where the trajectories were approximated by scale-free random walks known as Lévy flights [23, 24]. Extending these results to human mobility, using data of half-a-million bank notes, the distribution of traveling distances on a high spatial scale (between $10 - 3200$km) decays as a power law suggesting that human trajectories are best modeled as a continuous-time random walk [25]. Later in Ref. [26], it is shown that when analyzed with the appropriate techniques mobile phone data offers the possibility of characterizing human trajectories over time. They found that human trajectory shows a high degree of spatial regularity; and its distribution follows a truncated power law form that is the result of differences in the characteristic distance traveled by individuals. An intrinsic limitation of the study was that the data collected over time depends on the bursty pattern of calling activity [27, 28], thus the coordinates of each user are recorded only when he or she initiates or terminates a call. In these previous studies there is no systematic characterization of individual's longitudinal mobility patterns from alternative big data sources. This lack of characterization limits the understanding of the duration of stays at each destination.This is a very important aspect for developing models of individual mobility. In this work, we consider the data set obtained from smart subway fare card transactions to characterize the urban human mobility patterns. We explicitly incorporate both the temporal and spatial variations of human mobility to develop a modeling framework that reproduces two non-trivial observations on the data: First, the frequency of visiting different locations by an individual decays with its rank of preference similar to a Zipf's law. Second, there is a slowly increasing probability with time that individuals explore new locations. Finally, we reproduce also the heterogeneous flow among stations measured in the entire subway network.

This paper is organized as follows. In Section 2, we describe the subway data source and present the observed spatial mobility patterns. In Section 3, we introduce a simple model reproducing some key observation in the data. In Section 4, we compare both model and data results as well as showing the limits of such kind of modeling approach. Finally, we conclude our findings and give an outlook for future work.

2. Empirical Observations

2.1. Data Source

For this study, we analyze the smart card transactions over a three-month period of 626 anonymous public transport users of London, UK [29]. Smart cards (called Oyster cards in London) are owned by the individual riders and generally record the time and place of every transaction that the card-owner makes on the public transportation system (e.g., subway station entry and exit). In London, more than seven million Oyster cards are regularly used, generating about 57 million weekly journeys (Transport for London, 2010). This constitutes records of about 80% of all the public transport services offered by Transport for London (TfL). Each record in the raw data collected from the Oyster smart card readers represents a transaction such as boarding on the bus or entering into or exiting from an underground subway station. Thus individual time resolved locations are found by tracking the smart card transactions in different stations in the public transport system network.

Initially, the transactions of 1000 users are collected for our analysis; however some of the users have incomplete journey information as some of their trips are made either by bus or rail. In these cases, we have only information where and when people get into the bus but no information about when and where they get off it. Hence, a lower threshold is selected: at least seventy percent of their recorded trips are made by subway. As shown in Fig. 1 we capture with this criterion the typical urban subway riders in London. A too small threshold would add users with limited information, while a too high threshold would include only unique users. In contrast to [18] we focus on a small group of regular users observed during three months, instead of analyzing the entire

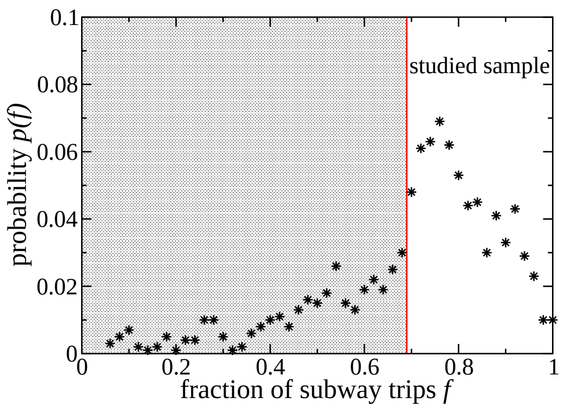

Figure 1. (Colors online) Identification of regular subway users. The distribution $p(f)$ of Oyster users which are using the subway for a given percentage f of their trips is shown. In our analysis only individuals with at least 70% of their total trips are made by subway are studied.

subway usage for fewer days.

2.2. Spatial Mobility Patterns: the Influence of Urban Contexts

In order to find the existence of the influence of urban geography on its peoples' mobility patterns, we first define for each individual the ranking of different places (i.e. subway station) based on the number of times the individual visits this place. For each individual we extract the different number of visited stations, and count the number of total visits to each of them. Then for each station we count the frequency that people had it as the most visited place and second most visited place during the entire period of observation. Figure 2a presents the distribution of most visited stations, it indicates that people's most visited places are scattered over the city with some subway stations having higher concentration.

However, if we look at the similar distribution for the second most visited place (see Fig. 2b) then an interesting pattern is revealed. The second most visited places have higher concentrations around the center of the city except one in the eastern part of the city which contains several office complexes. These two figures suggest that for most of the users the most visited place and second most visited place are actually their daily commuting options, likely capturing home and work respectively.

A third distribution can be derived by observing the frequency of visiting other ranked places in the city. Figure 2c presents the likelihood of visiting a place as 'other' place (not the most or second most visited place) for an individual. In general, these places represent those that people usually visit for recreational and social purposes; and the higher the likelihood the more popular a place is. Figure 2c shows that most of these popular places are actually located within the central part of the city in agreement with previous results [18].

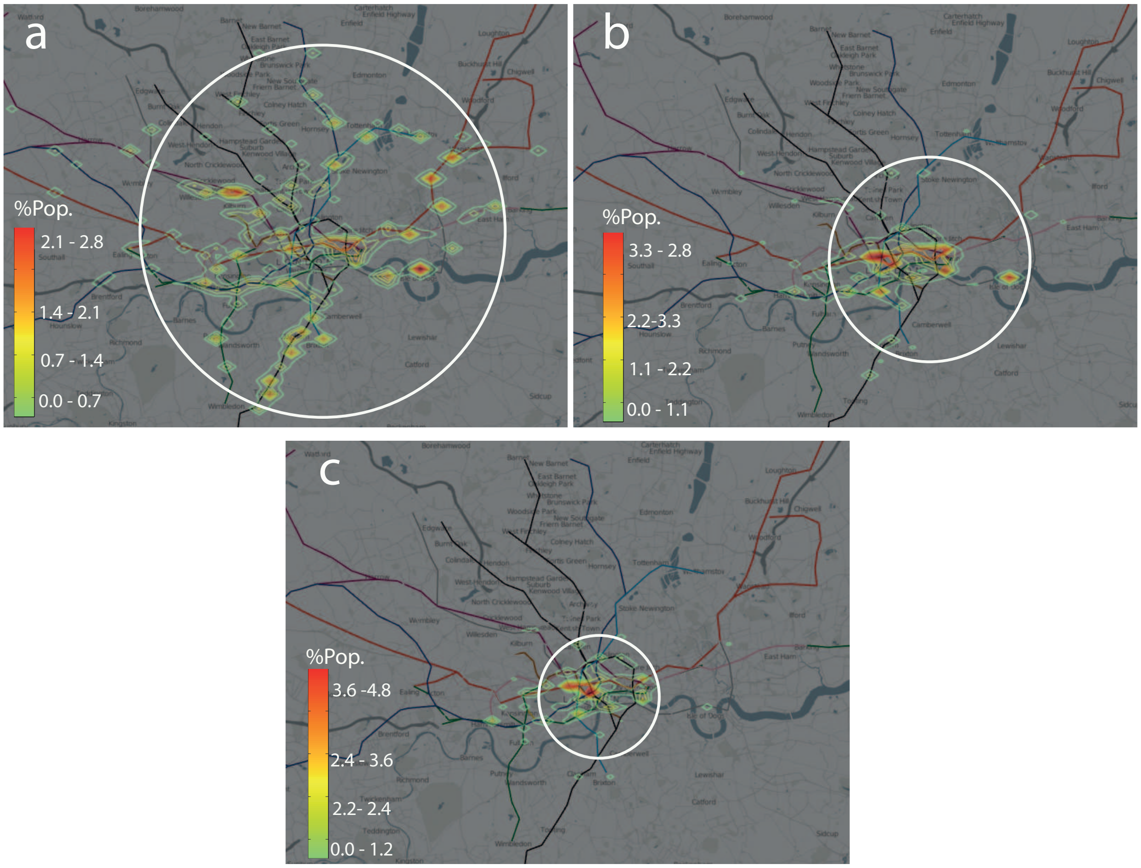

Figure 2. (Colors online) Spatial mobility patterns. (a) Distribution of peoples' most visited place in the subway station network. (b) Distribution of people' second most visited place in the network. (c) Distribution of peoples' other visited place (i.e. the places that people choose as their non-home or non-work destinations) in the network. The colors correspond to the likelihood of a station being the most visited place, second most visited place or other visited place among the users. The circles highlight the regions with a significant number of such stations.

2.3. Temporal Mobility Patterns-Trip Length Distribution

To uncover the temporal regularity of peoples' mobility we first investigate the distribution of displacements or trip length. We define a displacement in terms of the time spent to travel from one place to another. One advantage of this approach is that we can easily measure all the users' displacements into a common frame of reference. Moreover, it is realistically closer to urban subway users' perception of distance as most of the people consider the amount of time to be taken to move from one place rather than the actual distance between the two places. As an intrinsic advantage, this data provides us with individual's actual trip length. In [25], the observed displacements could be the combinations of two or more individual's displacements. On the other hand, in [26] although individual's displacements are captured, the phone activity limit the observation of trip lengths. Each displacement could be a combination of two or more displacements if no calls are made in the intermediate places.

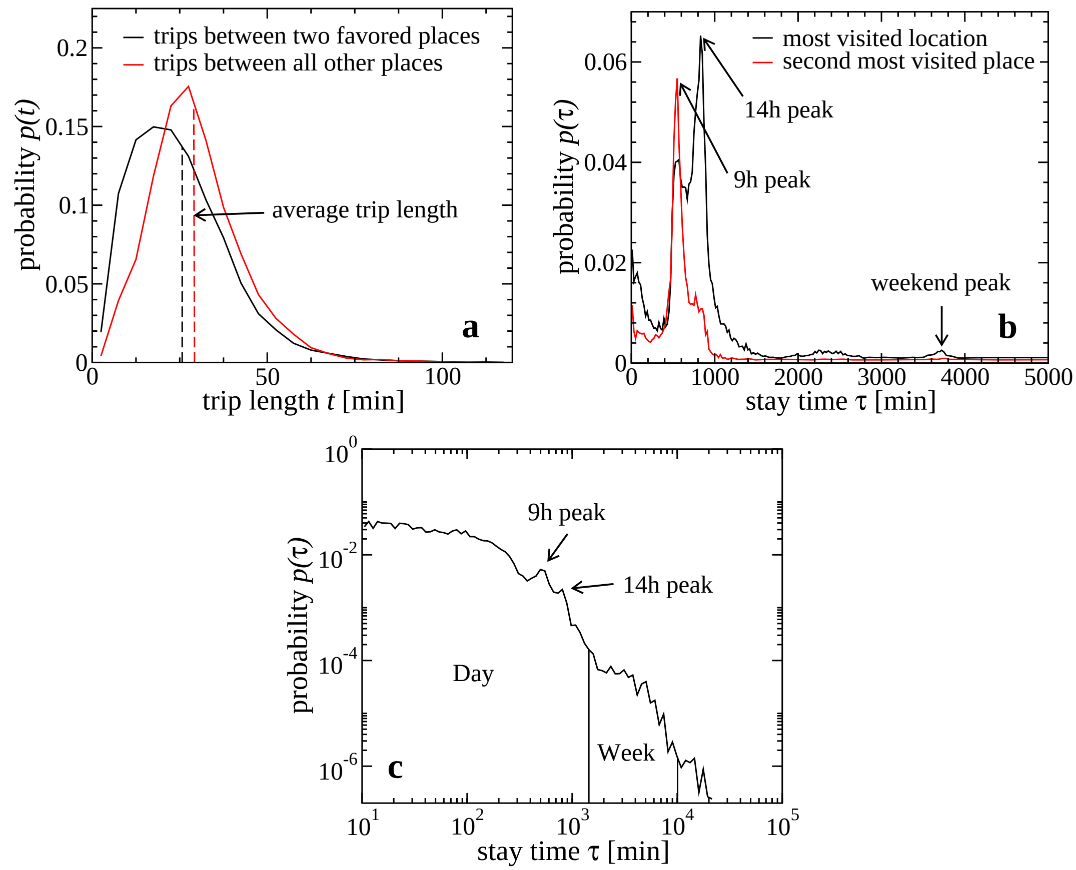

Figure 3. (Colors online) Temporal mobility patterns. (a) Distribution of peoples' displacements in minutes. (b) Distribution of stay times at the two most visited place (c) Distribution of stay times at places except the two most visited places (i.e. other places).

Figure 3a presents the displacement distributions for trips between the most visited places and trips from or to other places. We measure that, over the entire data, the typical trip time is about 25 minutes, while the average trip time between the most visited places is 15% shorter. The distributions of trip times show a central tendency around these values. In contrast, in previous studies [25, 26], power law distributions were observed for the trip length for larger scale of displacements. The deviation from power-law in our results might be due to two reasons: presence of a boundary on the subway network and the presence of a single transportation mode. The urban boundary restricts the observations to the intracity scale, while in previous studies the measurements are made on larger boundaries and without transportation mode restrictions.

2.4. Temporal Mobility Patterns-Stay Time Distribution

Another important characteristic regarding temporal patterns for urban human mobility is the stay time distribution. Stay time here is defined as the amount of time an

individual spends at a particular place. We observe in the dataset that for an individual there might be incomplete information of stay times due to the bus trips; therefore we restrict our analysis only for those stays for which we have the information of both the beginning and the end times of the stays. For measuring the distributions, we differentiate stay time in terms of staying at the most visited place, at the second most visited place and at all other places.
Figure 3b presents the stay time distribution at the two most visited places. For the most visited location two distinct peaks exist at approximately 9 hour and 14 hour. These two peaks captures two characteristic durations of stays at the most visited place. However, the stay time distribution at the second most visited place has a distinct peak at about 9 hour, which can be considered as the stay time for most of the people at their work place. A mixture of distributions might approximate both distributions. Observing the typical durations found in these two observations, it can be inferred that for most individuals the most visited location and the second most visited location represent individual's fixed destinations such as home and work place respectively. We also observe the probability distribution of stay time at other places (see Fig. 3c). In contrast, this distribution has no characteristic durations and the tail has a power law behavior representing the fact there exists a non-ignorable probability for people staying longer durations; such stays can represent vacations and international business trips. These findings deviate from that of the existing literature on waiting time distributions. Previous studies [25, 26] reported that waiting time distributions follow $P(\Delta t) \sim |\Delta t|^{-1-\beta}$ where $0 < \beta \leq 1$; we find that this waiting or stay time distribution is not generally true for all types of locations. Stay times at the most-visited place and the second-most visited places do not follow fat-tailed distributions.

2.5. Visitation Frequency and Zipf's law

To find the probability of visiting a place we rank (L) each individual's visited places based on the number of times one visits the places over the three-month period. For instance, rank 1 represents the most visited place; rank 2 the second most visited place and so on. Then we calculate the frequency of each of these ranked places. Individuals are grouped based on the total number of different places they visit (N).

Figure 4a shows the probability of visiting different places versus their corresponding ranks. This figure suggests two interesting features: a) Most of the times people pay visits only to a few locations (two most visited places) and the probability of visiting the most visited place and the second most visited place are close to each other in value; this indicates most individuals' regular routine pattern of movements between their home and work location. Note that less active users tend to have higher frequency in the first ranked locations, for example $p(1) + p(2) = 0.82$ for users with $N = 10$ locations while $p(1) + p(2) = 0.53$ for those with $N = 40$. b) People spend their remaining time at 3 to 43 different places that are visited with diminished regularity. We observe that the distributions in Fig. 4a follow a Zipf's law with an exponent that

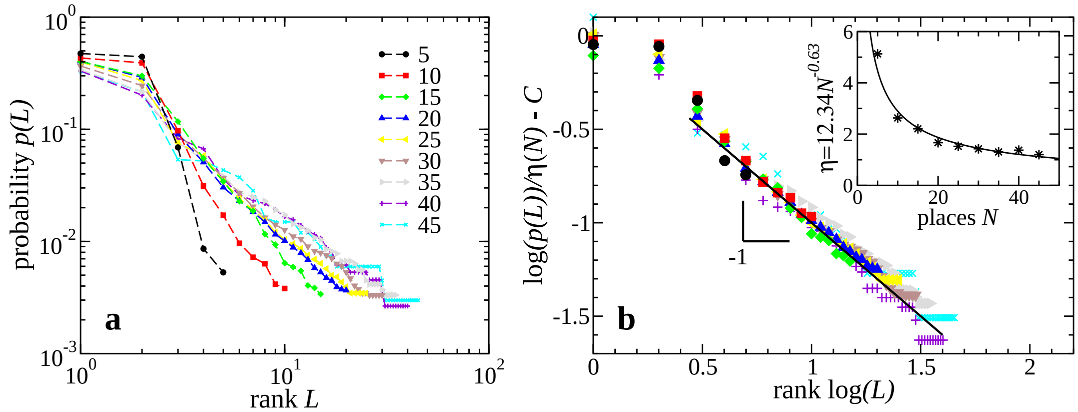

Figure 4. (Colors online) Probability of visiting different places. (a) Probability of visiting different places $p(L)$ vs. their corresponding ranking L in log-log scale. The numbers in the legend refers to N, the total number of different locations visited by individuals during the observation period. (b) The collapsed distribution of $p(L)$ vs L. The inset shows the observed relation between the η exponents of the power laws vs. N, the number of different places people visit.

depends on the total number of visited locations. Figure 4b (inset) shows the observed relation of the measured exponent and the number of visited locations: $\eta \sim N^{-0.63}$. By using this relationship, we observe that $p(L)$ vs. L collapses into a single distribution (see Fig. 4b). Our observations deviates from the classical Zipf's law in which the locations with rank 1 and 2 belong to the distribution, only locations with higher rank ($L > 2$) seem to follow this pattern.

3. Modeling Individual's Mobility Patterns

In previous sections, we have described temporal and spatial aspects of individual mobility patterns. In terms of spatial patterns we observe two important facts: first, people select their destinations in such a way that there are typically two fixed locations which have characteristic stay times. Second, the other places are selected with diminishing probability following a Zipf's law. The second observation tells us that people do not select the location of their non-home or non-work destinations randomly; rather they select these places based on the popularity of the corresponding place. From the modeling perspective, this means that if more people select a place it is more likely that other people will select it.

Inspired by the observations we propose a simple model of urban movement. We model two different choices that an individual makes when making mobility decisions: a) which places to visit, and b) how long to stay at each of these places. First, to incorporate the frequency of visits into the modeling framework we assume that there exists a fixed probability for a person to visit his home and work locations and these two locations are fixed for each agent and uniformly distributed within the city. At each time step each person selects home and work with fixed probabilities and other places only if these two

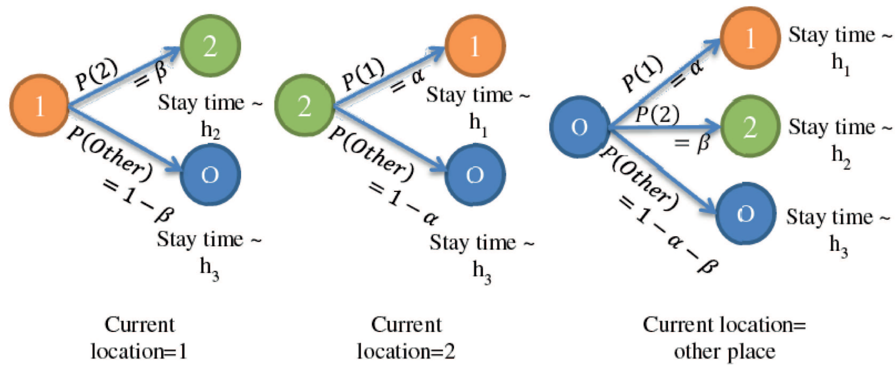

Figure 5. (Colors online) Schematic representation of the mobility model. The first (1) and second (2) most visited locations (could be thought as home and work) are selected with fixed probability α and β respectively while the other places are selected with probability $1-\alpha$, $1-\beta$ or $1-\alpha-\beta$ depending on the place of origin. Stay times are modeled by a duration function $h(t)$ that also depends on the current origin and uses the empirical observations.

locations are not selected. Figure 5 shows a schematic representation of the selection process in which the probabilities selecting home and work are α and β respectively.

A second important aspect is how to model the stay times at different locations. We adopt a hazard based duration modeling approach mainly to overcome the difficulties to sample from the distribution of stay time at different kind of places [30]. The hazard function, $h(t)$, is the instantaneous probability that a visit ends after time t, given that it has lasted until time t. If T is a non-negative random variable that represents the duration of a visit, $h(t)$ is a conditional probability that can be expressed in terms of the density function, $f(t)$, and the cumulative density function, $F(t)$, of T.

$$F(t) = Pr[T < t] = \int_0^t f(u)du \tag{1}$$

Then, $Pr[T \geq t] = 1 - F(t)$ is the survival distribution, or the probability to survive until time t, also known as endurance probability [30]. The duration function depends on the elapsed time of an activity t as:

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{2}$$

where $h(t)$ is the conditional probability that an individual will stay at a location between time t and $t + \mathrm{d}t$ given that the individual has not left the location until time t; and $f(t)$ and $F(t)$ are the corresponding density function and cumulative probability function of stay time respectively, which are measured from the data. We estimate three separate duration functions for the stay times at (1) home, (2) work place and (3) other visited places based on the corresponding stay times from the empirical observations.

Finally, we want to incorporate the non-uniform selection of the non-home and non-work destinations. To do so, we exploit the idea of preferential attachment into the location selection process. Basically, we find the probability $p(i)$ of selecting a particular place i as a function of the number of visits v_i that have already been made to that place by all individuals:

$$p(i) = \frac{v_i}{\sum_j v_j} \quad (3)$$

In order to run the simulations of the model, we first fix an arbitrary number of agents N_a and number of stations for the model N_s. We randomly assign to each agent two different stations as its most visited and second-most visited location with a fixed probability of $P(L)$ for $L = 1, 2$; specifically $\alpha = 0.4$ and $\beta = 0.3$. At each time step, based on its current location each agent makes a movement according to probabilities shown in Fig. 5 and the duration of stay has ended, by inspecting Eq. (2).

We rank all the station as its initial popular ranking ($v_i = i, i = 1, .., N_s$). Once an agent selects a station other than his most or second-most visited place we update the ranking of this place by counting an additional visit to this location. At each time step of the model each agent makes the location choice until the simulation time limit T is reached (for our simulation we select $T = 3$ months).

4. Model Results and Limitations

Notice that the proposed model is related to the model recently proposed by Song et al. [31], in terms of the preferential selection of locations depending on the number of visits, it has some subtle but important distinct characteristics. First, we do not restrict the jump size following a distribution (e.g. $P(\Delta r) \sim |\Delta r|^{-1-\alpha}$); instead we give each of the location an arbitrary contextual label. Our agents select their destinations according to their popularity, defined as the number of visits by all the agents, and not by the individual number of visits. Second, our stay time distributions do not follow $P(\Delta t) \sim |\Delta t|^{-1-\beta}$ where $0 < \beta \leq 1$ for all locations; instead we use the empirical distributions depending on the agent's individual rank of the destination. Third, Song et al. [31] introduced exploration and preferential return to find whether an agent will visit a new place or an already visited place as its next destination, with $P_{\text{new}} = \rho S^{-\gamma}$, and $P_{\text{ret}} = 1 - P_{\text{new}}$, where S is the number of previously visited locations by the agent and ρ as well as γ are parameters. Instead, we assign specific probabilities of moving to another place, depending on the type of the current location, leading to regular commuting between the two most visited locations. Note that new places can only be visited, if the agent move to another place and the agent did not visited that place yet, thus we have an upper limit N_s of new locations. The crucial difference is that the slow increase in number of locations by the agents is not based on P_{new} decaying with the individual S, but due to the tendency of the agents to select popular places in the network.

We generate the sequences of locations and times of visiting those locations for a particular number of agents following the described model. After generating these sequences we analyze the data using the same approach we followed for analyzing the

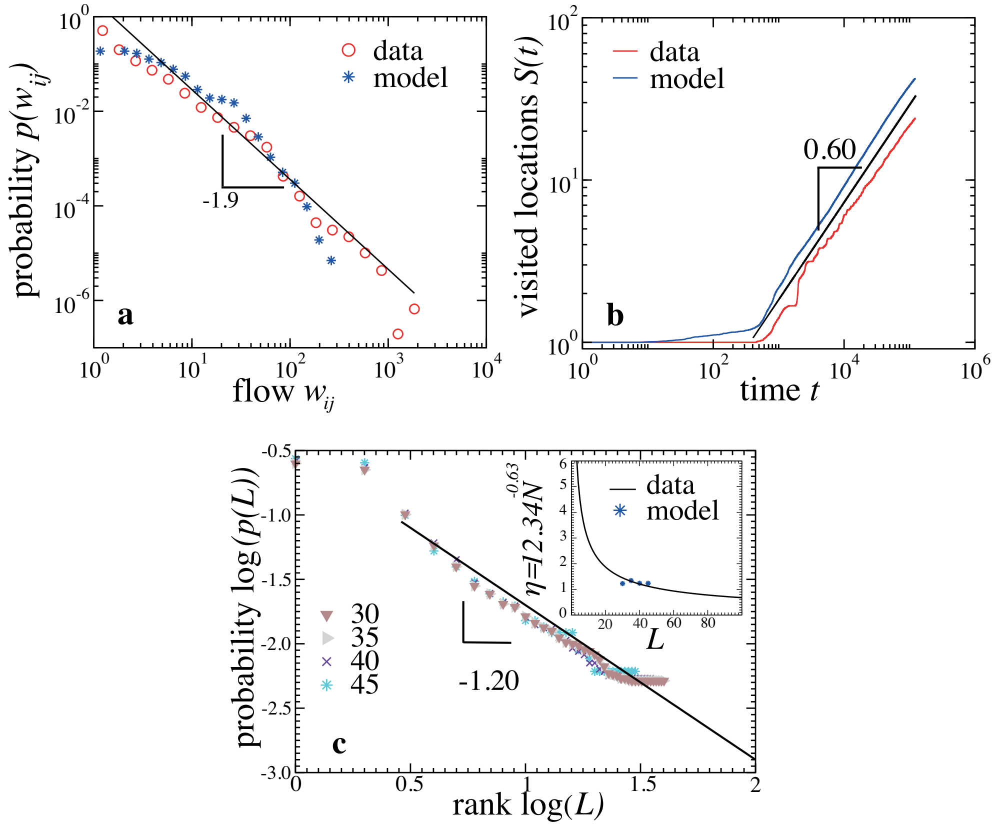

Figure 6. Comparison between model and empirical data. (a) Distribution of number of trips $p(w_{ij})$ between two locations i and j in the network of subway stations. The straight line is a power-law decay with an exponent -1.9. (b) Number of different visited locations $S(t)$ versus the observation time t. The straight line is a power-law with an exponent 0.6. (c) Frequency of visiting locations $p(L)$ with given ranks L. The legend refers to groups of agents with different total number of visited locations N. The straight line is a power-law decay with an exponent -1.2. All power-laws are shown as a guide to the eye. The simulations are made with $\alpha = 0.4$ and $\beta = 0.3$ and 100 stations.

data set of subway fare card transactions. Figure 6a shows the flow distribution between any two given locations, for both data and our model. This suggests that the preferential selection of destinations for non-flexible activities is sufficient to reproduce the broad heterogeneous flows among stations. More interestingly, this selection conditions the number of different locations S an agent will choose as a function of time as shown in Fig. 6b. This number increases sub linearly with time as $S(t) \sim t^{\mu}$, for both data and our model, presenting also an interesting agreement with previous observations with mobile phone data, with $\mu = 0.6$ [31]. These are surprising results for us as this model captures the aggregate mobility patterns and some aspects of the individual choices for the population despite the absence of many different factors for individual movements.

In consequence of the similarities in $S(t)$, the frequency of visitation of locations with different rankings obtained from our model and the data show very similar pattern for agents that have several visited locations during the three months, i.e. $N = 30$

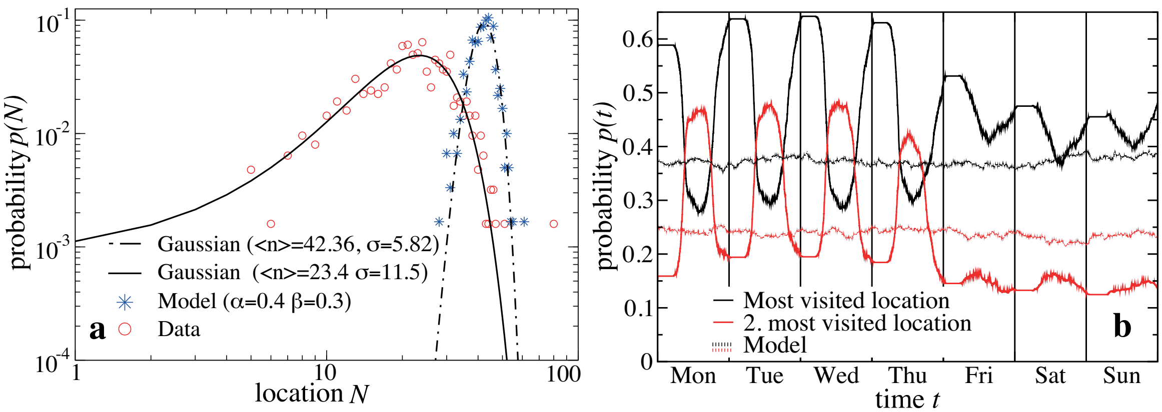

Figure 7. Limitation of the model. (a) Distribution of total number of visited locations. The smart card users have broader distribution than the modeled agents with fixed values of α and β. (b) The probability $p(t)$ to find an user in the most and second most visited location at a given time of the day t. While the model shows the introduced average value with few fluctuations, the real data is highly time-dependent with a characteristic circadian rhythm.

as shown in Fig. 6c. In this case the model generates an exponent $\eta = 1.2$ close to the saturation value of the data, and similar to the results from the mobile phone observations [31].

The simplicity of the models rises two main limitations that could be overcome in future extensions: first, it does not generate the heterogeneity of individuals in selecting the total number of locations N. Second, the selection of trips is independent of the time of the day. Figure 7a shows that the distribution of N is much broader for real users than in the model, nevertheless heterogeneity in the population could be introduced by selecting from a distribution of α and β instead of using the same fixed values for all agents. In Fig. 7b we show that the decisions of visiting a location for the users is independent on the time of the day, which could be corrected by introducing a temporal dependency on the selection of trips. In Fig. 8-10, we show individuals trajectories vs. time for both the data and the model. Figure 8 shows a user with 39 visited locations, which qualitatively compares well with a simulated agent in Fig. 9 with $N = 38$, except for the circadian restrictions to choose a particular time of the day when to travel. In contrast, a user with only $N = 7$ during the entire three months, as the one depicted in Fig. 10, cannot be trivially modeled with the proposed scheme.

5. Summary and Conclusions

This paper presents the spatial and temporal patterns of individual's mobility in a city using the data from smart subway fare card transactions. The empirical findings from this study give us some useful insights on human mobility patterns within a city. Regarding spatial mobility we observe two interesting facts: First, the two most frequent locations can be modeled with fixed probabilities; and second, when people select their

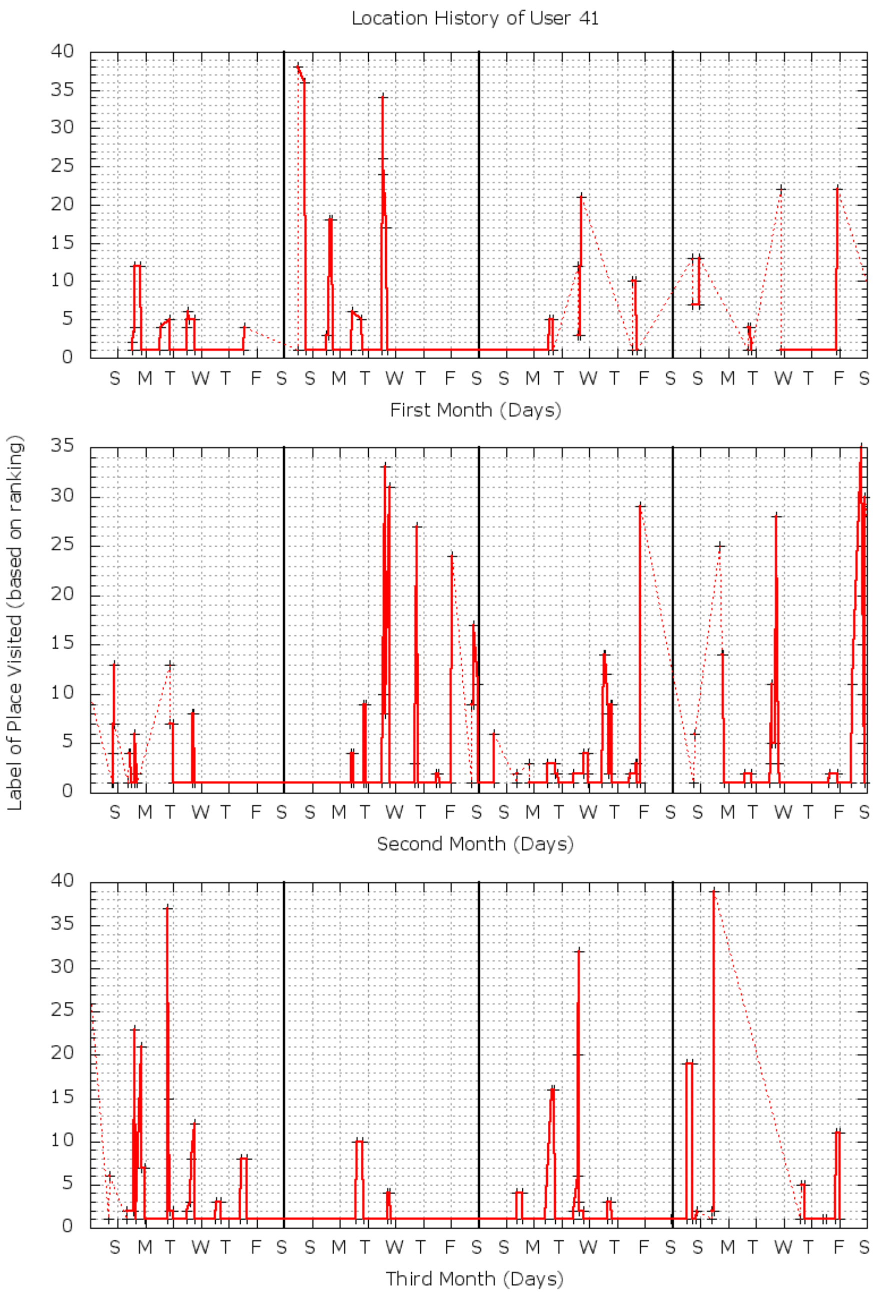

Figure 8. The trajectory of a subway user with N = 39 locations. The visited locations, ordered by the rank are shown over a three month period.

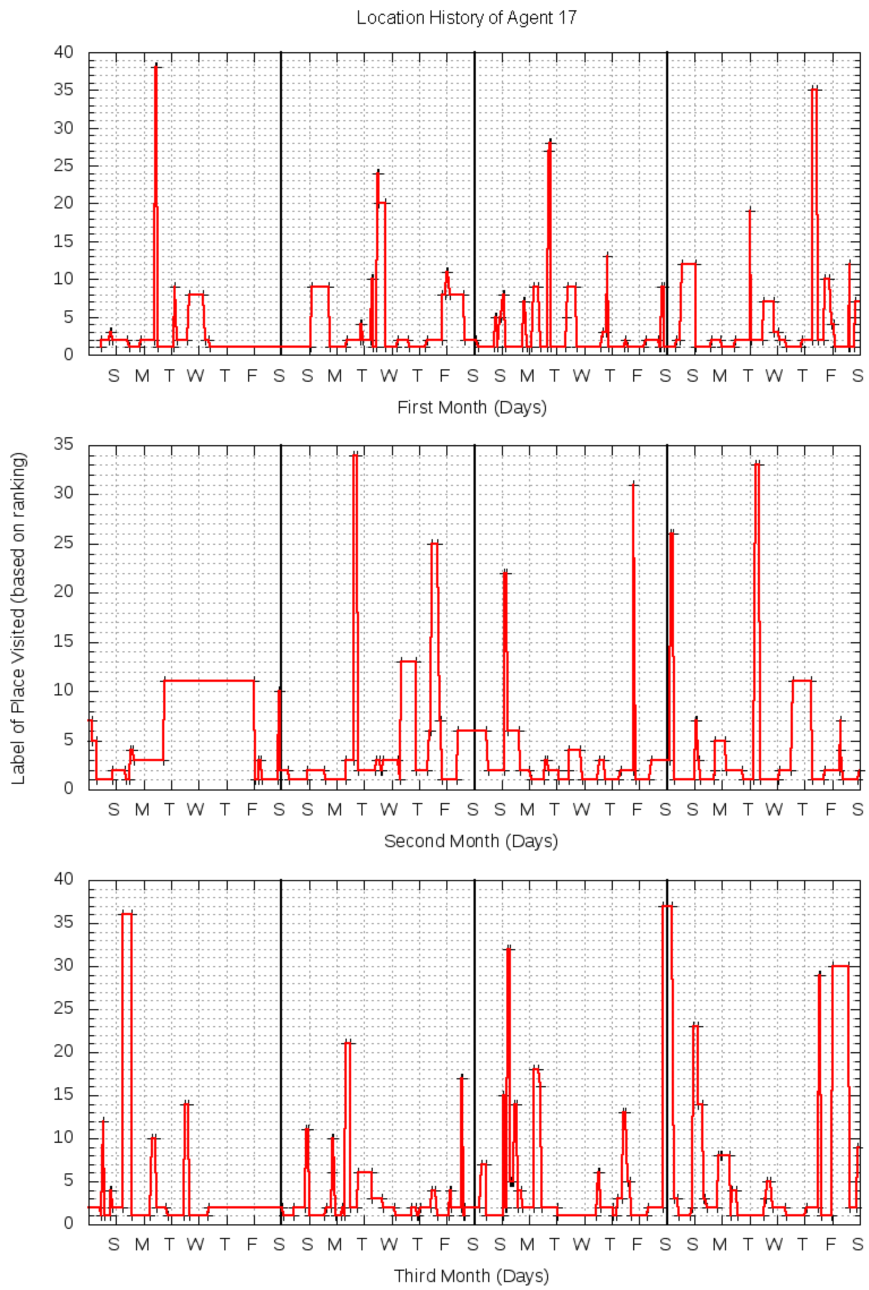

Figure 9. The trajectory of a modeled user with N = 38 locations. The simulated users visit the locations independent of the time of the day.

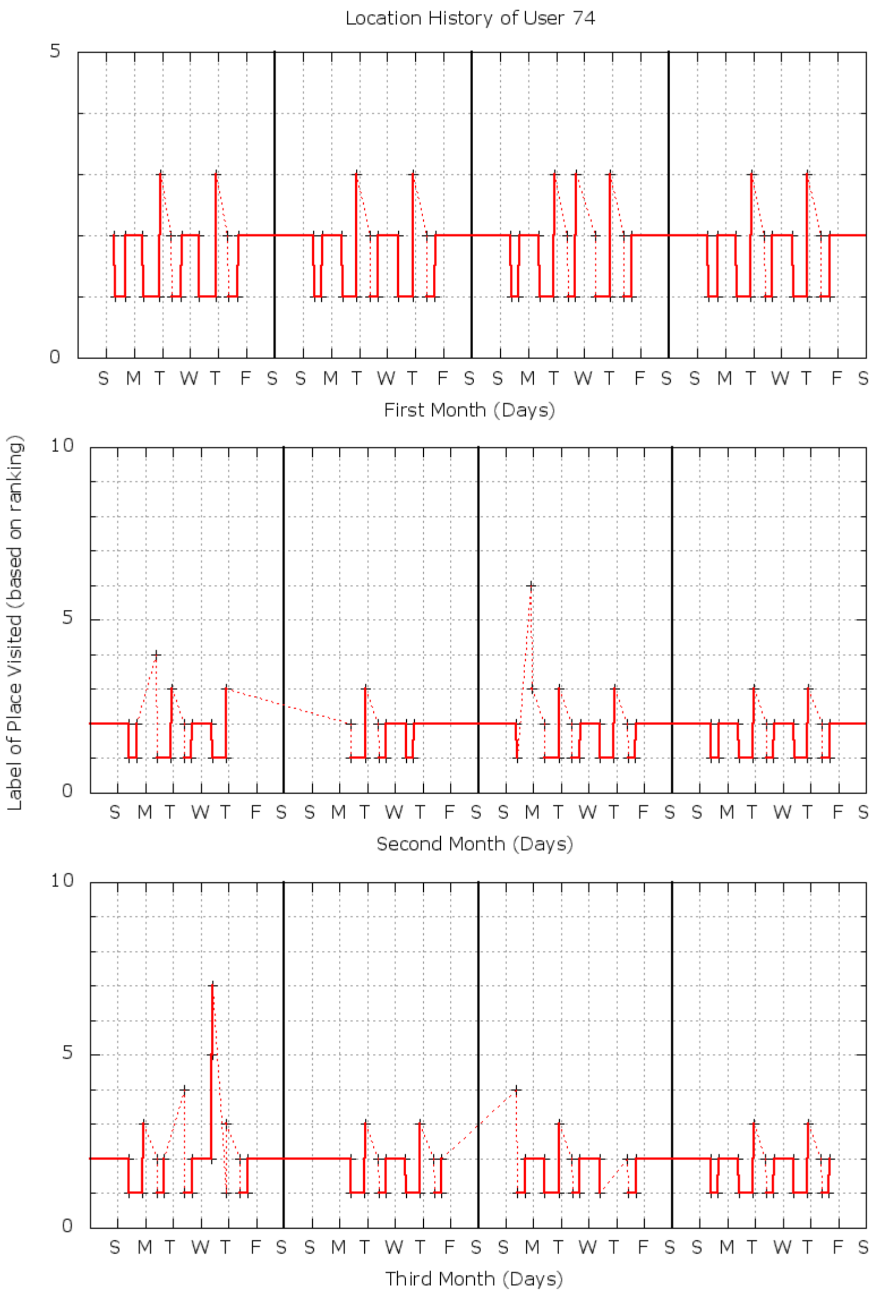

Figure 10. The trajectory of a subway user with N = 7 locations. These users show more regularity in their visits.

other destinations that are not either most-visited or second-most visited location they mostly select popular places in the city. In terms of temporal patterns we observe characteristic stay time distributions at the most visited place, second-most visited place and the other places.

We present a mobility model that utilizes the idea of preferential selection to popular places in the city for modeling the spatial distribution of individual's movements for selecting non-home and non-work destinations. It is able to generate the heterogeneous flows among locations at the aggregated level and it reproduces the frequency of visitation to other places as well as the sub lineal increase of number of different locations visited as a function of time at the individual level.

Although our simulation model has several simplistic assumptions, several contextual details of a particular city can be added to the model like a distribution of the residential areas or business places. This will capture the realistic distributions of home and work places for individuals in the city. In addition, the incorporation of trip selection depending on the time of the date and heterogeneity in the visits of locations must be incorporated. Our model is a good and simple approximation of urban mobility. It can be easily implemented in simulations for diseases spreading [4, 8, 32] within a city.

6. Acknowledgments

The Oyster card anonymous data was collected by Transport for London (TfL) for operational purposes, and we are grateful for their permission to use it in this paper. We also thank Prof. Nigel Wilson and Michael Frumin of MIT Transit Research Group for providing us the data; Prof. Chris Magee and Dr. Daniel Whitney for giving useful comments in the initial stage of this project.

[1] Kitamura R, Chen C, Pendyala R M and Narayaran R, 2000, *Transportation* **27**(1) 25-51
[2] Bhat C R and Koppelman F S, 1999, Activity-based modeling for travel demand (R. Hall)
[3] Ukkusuri S V, Tom V M and Waller S T, 2007, *Computer-Aided Civil and Infrastructure Engineering* **22**(1) 921
[4] Eubank S, Guclu H, Kumar V S A, Marathe M V, Srinivasan A, Toroczkai Z and Wang N, 2004, *Nature* **429** 180184
[5] Hufnagel L, Brockmann D and Geisel T, 2004, *Proc. Nat. Acad. Sci.* **101** 1512415129
[6] Colizza V, Barrat A, Barthélémy M, Valleron A-J and Vespignani A, 2007, *PLoS Medicine* **4** 95110
[7] Kleinberg J, 2007, *Nature* **449** 287288
[8] Nicolaides C, Cueto-Felgueroso L, González M C and Juanes R, 2012, *PLoS ONE* **7**(7):e40961
[9] Hanson S, 2005, *Proc. Nat. Acad. Sci.* **102** 1530115306
[10] Rhee I, Shin M, Hong S, Lee K, and Chong S, 2008, *Proceedings of INFOCOM*, Phoenix, USA
[11] Hanson S and Huff J, 1988, *Transportation* **15** 111-135
[12] Vilhelmson B, 1999, *GeoJournal* **48**(3) 177-185
[13] Ewing R and Cervero R, 2001, *Transportation Research Record* **1780** 87-113
[14] Schlich R and Axhausen K, 2003, *Transportation* **30**(1) 13-36
[15] Maat K, van Wee B and Stead D, 2005, *Environment and Planning B: Planning and Design* **32** 33-46
[16] Bagchi M and White P R, 2005, *Transport Policy* **12** 464474
[17] Seaborn C, Attanucci J and Wilson N H M, 2009, *Transportation Research Record* **2121** 55-62
[18] Roth C, Kang S M, Batty M and Barthélémy M, 2011, *PLoS ONE* **6**(1):e15923

[19] Ben-Akiva M and Bierlaire M, 1999, Discrete choice methods and their applications to short term travel decisions (R. Hall)

[20] Cetin N, Nagel K, Raney B and Voellmy A, 2002, *Computer Physics Communications* **147**(1-2) 559-564

[21] Axhausen K W, 2008, *Environment and Planning B: Planning and Design* **35**(6) 981-996

[22] Newman M E J, 2011, *Am. J. Phys.* **79** 800-810 (2011)

[23] Viswanathan G M, Afanasyev V, Buldyrev S V, Murphy E J, Prince P A and Stanley H E, 1996, *Nature* **381**, 413-415

[24] Edwards A M, Phillips R A, Watkins N W, Freeman M P, Murphy E J, Afanasyev V, Buldyrev S V, da Luz M G E, Raposo E P, Stanley H E and Viswanathan G M, 2007, *Nature* **449** 1044-1048

[25] Brockmann D, Hufnagel L and Geisel T, 2006, *Nature* **439** 462465

[26] González M C, Hidalgo C A and Barabási A-L, 2008, *Nature* **453** 779782

[27] Barabási A-L, 2005, *Nature* **435** 207-211

[28] Candia J, González M C, Wang P, Schoenharl T, Madey G and Barabási A-L, 2008, *Journal of Physics A: Mathematical and Theoretical* **41** 224015

[29] Oyster Factsheet, http://www.tfl.gov.uk/assets/downloads/corporate/oyster-factsheet.pdf, 2010, (Accessed November 1, 2011)

[30] Joly I, 2004, Travel time budget-decomposition of the worldwide mean *Conference of the International Association of Time-Use Research*, 27-29 October, Rome, Italy.

[31] Song C, Koren T, Wang P and Barabási A-L, 2010, *Nature Phys.* **6** 818-823

[32] Balcan D, Colizza V, Goncalves B, Hu H, Ramasco J J and Vespignani A, 2009, *Proc. Nat. Acad. Sci.* **106** 2148421489