Barbosa et al. *EPJ Data Science* (2015) 4:21
DOI 10.1140/epjds/s13688-015-0059-8

EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE Open Access

The effect of recency to human mobility

Hugo Barbosa[1*], Fernando B de Lima-Neto[2], Alexandre Evsukoff[3] and Ronaldo Menezes[1]

*Correspondence: hbarbosa@biocomplexlab.org
[1]BioComplex Lab, Florida Institute of Technology, 150 W University Blvd, Melbourne, FL 32901, USA
Full list of author information is available at the end of the article

Abstract

In recent years, we have seen scientists attempt to model and explain human dynamics and in particular human movement. Many aspects of our complex life are affected by human movement such as disease spread and epidemics modeling, city planning, wireless network development, and disaster relief, to name a few. Given the myriad of applications, it is clear that a complete understanding of how people move in space can lead to considerable benefits to our society. In most of the recent works, scientists have focused on the idea that people movements are biased towards frequently-visited locations. According to them, human movement is based on a exploration/exploitation dichotomy in which individuals choose new locations (exploration) or return to frequently-visited locations (exploitation). In this work we focus on the concept of *recency*. We propose a model in which exploitation in human movement also considers recently-visited locations and not solely frequently-visited locations. We test our hypothesis against different empirical data of human mobility and show that our proposed model replicates the characteristic patterns of the recency bias.

Keywords: human mobility; regularities in human dynamics; mobility data analysis

1 Introduction

The understanding of the fundamental mechanisms governing human mobility is of importance for many research fields such as epidemic modeling [1–3], urban planning [4, 5], and traffic engineering [6–8]. Although individual human trajectories can seem unpredictable and intricate to an external observer, in fact they exhibit many spatiotemporal regularities [9–17]. One of these patterns, largely observed in empirical data, is the strong tendency we have to spend most of the time in just a few locations [15, 18, 19]. More precisely, the distribution of visitations frequencies have been observed to be heavy tailed, being better approximated by a power law distribution [13, 18].

However, the fundamental mechanisms responsible for shaping our visitation preferences are still not fully understood. The *preferential return* (PR) mechanism, proposed by Song *et al.* [18], offered an elegant and robust model for the visitation frequency distribution. It defines the probability Π_i for returning to a location i as $\Pi_i \propto f_i$, where f_i is the visitation frequency of the location i. It implies that the more visits a location receives, the more visits it is going to receive in the future, which in different fields goes by the names of *Matthew effect* [20], *cumulative advantage* [21], or *preferential attachment* [22].

Although the focus of the PR mechanism - as part of the Exploration and Preferential Return (EPR) individual mobility model - was to replicate the scaling properties of human mobility, its robustness and modularity, combined to analytical formalism the authors employed in deriving its mechanisms, has turned it into a modeling platform itself, where authors can test their hypotheses by easily replacing or adding specific mechanisms to it [23]. For instance, Toole *et al.* [24] incorporated a social mechanism to the mobility dynamics.

However, the Preferential Return assumption as a property of human motion leads to two discrepancies. First, the earlier a location is discovered, the more visits it is going to receive. It implies that a early-discovered location will most likely be one of the most visited ones throughout the entire lifespan of the individual. Second, if the cumulative advantage indeed holds true for human movements, people would not change their preferences, which is clearly not true.

In this work we investigate the existence of a *recency* bias - a stronger influence of recent events - in human mobility, a phenomenon known to play an important role to other decision-making-related behaviors [25–27]. Our objective is to investigate the influences of accumulated mobility trajectories (*i.e.* visitation frequencies) and recent mobility context (*i.e.* recency) to human traveling behavior.

Notice that we are not implying a dichotomy between them but rather that *recency* and *frequency* are complementary mechanisms that ultimately share some level of dependency. From an individual's trajectory standpoint, it is obvious that frequently-visited locations are recurrent in one's trajectory and therefore the interval between two consecutive visits tend to be short. On the other hand, a recently-visited location does not depend on the number of previous visits to it.

In order to extract these two traits from individual human displacements, one needs to look at the evolution of visitation patterns over a large period of time. In this work, we propose a novel rank-based framework for human mobility characterization beyond the spatiotemporal dimensions, where each point in a trajectory can be decomposed into its frequency and recency ranks.

In our analyses, we used two human mobility datasets: the first one ($D1$) corresponding to 6 months of anonymized mobile-phone traces of 30K users from a large metropolitan area in Brazil. The second dataset ($D2$) is composed of more than 23M *check-ins* produced by more than 51K Brightkite users around the world.[a]

It is worth noting that the data we analyzed is subject to a sample bias. One way to reduce the influence of such bias is by analyzing multiple datasets representing differences in the populations across multiple dimensions. In our analyses, the datasets have important differences in terms of the population they represent. The data of $D1$ has a noticeable socio-economic bias due to the fact that approximately 75% of mobile phones in Brazil correspond to pre-paid lines, mostly used by lower-middle and working classes. Additionally, it is plausible to assume that the data in $D2$ have an age bias, with younger people being over-represented in it. See the Materials and Methods section for more information on the datasets.

Nevertheless, the generality of our approach and the patterns we observed across the different datasets suggest that the recency bias we uncovered is a true universal mechanism of human traveling behaviors. Also, our results show a strong tendency of individuals to return to recently-visited locations that are not conditioned to the number of previous visits. Last, we incorporate the recency bias into a human mobility model and show that it is an important mechanism of the human traveling behavior. In the next section we contextualize our work within the current human mobility literature.

2 Related works

Traditionally, quantitative investigation of human movements was largely based on survey data. Over the last decade the field has witnessed a paradigm shift, mostly due to the increasing availability of high-resolution time-resolved digital human traces. This was made possible thanks to the development and popularization of many information and communication technologies such as GPS devices [28–30], location-based social networks [31–33] and mobile phone communications [15, 34–36] to name but a few.

In 2006, Brockmann *et al.* [16] analyzed more than 460K dollar bills traces concluding that both the jump length and waiting-time distributions in human traveling behavior can be mathematically described by a two parameter *continuous time random walk*. In 2008, González *et al.* [15] empirically found two important regularities in human traveling behavior: first, humans tend to spend most their time in very few highly-frequented locations, and second, individuals trajectories can be described by a time-independent characteristic length scale. Later on, Song *et al.* further explored the fundamental scaling properties of human travels, and proposed a general model of individual mobility - namely Exploration and Preferential Return (EPR) - capable of reproducing not only the spatiotemporal properties of mobility but also the heavy-tailed visitation frequency distribution.

In the EPR model, the probability of returning to a given location does not take into account the current individual's location, nor the time elapsed since the previous visit to that place. However, when it comes to the predictability of individual's trajectories, the performance of Markovian predictors based on recent past history suggests the existence of a visitation bias toward recently-visited locations on a short time scale [37–39].

Szell *et al.* [23] analyzed the virtual trajectories of more than 1,400 players within the virtual world of the MMORPG Pardus, pointing to the fact that the EPR model could not capture sub-diffusive evolution of the mean squared displacement (MSD) exhibited by the users within the Pardus virtual world. It was partially due to the lack of a mechanism capable of reproducing a tendency of the players to return to recently-visited sites in the game [23].

Schneider *et al.* [19] applied a motif approach - brought from network science - to the investigation of the underlying mechanisms of daily human mobility patterns. In that study, individual daily trajectories were represented by directed networks, in which nodes and edges represent visited locations and the trips between them respectively. Since it aims at capturing the individual *daily* mobility graphs, a recency bias at this time scale would be indistinguishable from the small number of locations an individual typically visits on a day. For instance, in Ref. [19] the average number of locations visited on a single day was $\langle N \rangle \approx 3$.

In this study we explore the visitation patterns that emerge from the individual microlevel traveling behavior, under a time-scale-agnostic approach.

3 A rank-based analysis of human visitation patterns

In this section, we propose a rank-based approach to the analysis of human trajectories. For such, we defined two rank variables K_f and K_s characterizing respectively the *frequency* and *recency* of a given location in the context of a individual trajectory. Both ranks were measured in a expanding basis from the accumulated sub-trajectories. To illustrate, consider a particular user x with a trajectory $T = [(l_1, l_2, \dots, l_n), l_i \in [1, \dots, N]]$ composed of n steps to $S \leq N$ locations. For each step $j > 0$, we have the partial trajectory $\mathcal{T} = [l_1, l_2, \dots, l_{j-1}]$ composed of all the previous steps, with l_{j-1} being the immediate preceding step. From the sub-trajectory $\mathcal{T}$ we compute the frequency-based ranks K_f of all locations visited so far. If the step j is a return (*i.e.*, $l_j \in \mathcal{T}$) we say that the frequency rank of the location l_j is the rank $K_f(l_j)$.

As we mentioned, the PR mechanism suggests that the visitation probability of a particular location is proportional to the number of previous visits to it. Our claim is that the Zipf's Law observed in visitation frequencies distribution is influenced by a recency bias expressed as a *tendency to return to recently-visited locations*, represented here as K_s.

In other words, we can describe the two rank variables as:

- K_s is the recency-based rank. A location with $K_s = 1$ at time t means that it was the *previous* visited location. $K_s = 2$ means that such location was the second-most-recent location visited up to time t, and so on.
- K_f is the frequency-based rank. A location with $K_f = 1$ at time t means that it was the *most* visited location up to that point in time. Similarly, a location with $K_f = 2$ is the second-most-visited location up to time t, and so on.

Given the definitions above, we first analyzed the frequency of returns as a function of K_s. This analysis shows that such probability decays vary rapidly with K_s (Figure 1). More precisely, for $D1$, the probability $p(K_s)$ follows a truncated power-law distribution, defined as

$$p(x) = Cx^{-\alpha} e^{-x/\kappa}$$

with exponent $\alpha_{K_s} \approx 1.644 \pm 0.001$ and exponential cut-off $\kappa_{K_s} \approx 40.9 \pm 0.3$ whereas the best fit for the frequency-based rank distribution is achieved when $\alpha_{K_f} \approx 1.560 \pm 0.0009$ and $\kappa_{K_f} \approx 23.6 \pm 0.2$. For $D2$, the best fit for the return ranks distribution is obtained with parameters $\alpha_{K_s} \approx 1.699 \pm 0.001$ and $\kappa_{K_s} \approx 206.6 \pm 7.6$ for the recency rank, whereas the

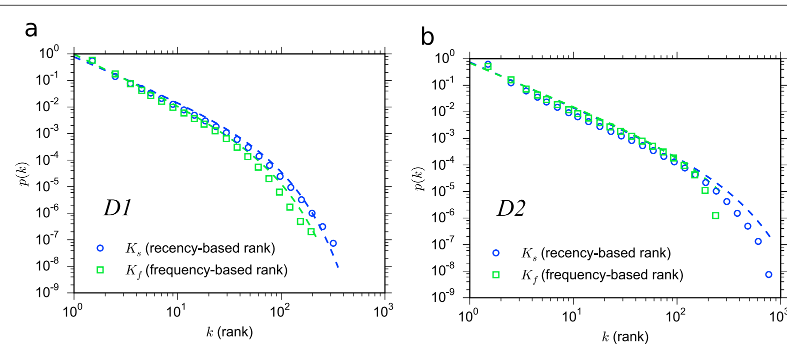


Figure 1 Comparison between the probability of return by recency and frequency ranks. The distributions of both ranks can be better approximated by truncated power laws (dashed lines). **(a)** The recency-based rank of $D1$ has exponents $\alpha_{K_s} \approx 1.644$ and exponential cut-off $\kappa_{K_s} \approx 40.94$, whereas the frequency-based rank distribution has a better fit for $\alpha_{K_f} \approx 1.56$ with $\kappa_{K_f} \approx 23.6$. **(b)** The best fit for the return ranks distribution in $D2$ is achieved with parameters $\alpha_{K_s} \approx 1.699$ and $\kappa_{K_s} \approx 206.6$ for the recency rank whereas the frequency rank has parameters $\alpha_{K_f} \approx 1.521$ and $\kappa_{K_f} \approx 64.3$.

frequency rank has the exponent $\alpha_{K_f} \approx 1.521 \pm 0.001$ and cut-off $\kappa_{K_f} \approx 64.3 \pm 1.3$ (see the Supporting Information (Additional file 1) for details on the curve fitting methods and results).

Notice that the exponents for the rank distributions were very similar for both datasets, regardless of their significant differences in terms of spatial coverage, number of users and time scale, suggesting that the distribution of the rank variables might be capturing a common underlying mechanism.

However, one can notice that the recency rank is a convolution of both frequency *and* recency biases, since highly-visited locations implies short intervals between visits. In order to quantify and decompose the recency bias from the recency rank we explore the intuition that even though low K_f implies low K_s, the opposite is not true. The recency dimension is memoryless in the sense that the K_s value of a location at time $t+1$ does not depend on the K_s at t and therefore, even recently-discovered locations can have a low K_s. The following analyses exploit this property of the recency rank by testing whether infrequently-visited locations can help us identify - and measure - the recency bias.

3.1 Recency over frequency: the role of recent events in human mobility

From the joint distribution of the rank variables we investigated the conditional frequencies of $P(K_s|K_f)$. If users have a bias for recently-visited locations we should observe:

1. lower values of K_s must be frequently observed over a wider range of K_f. It would suggest that we tend to return to recently-visited locations even if it was just discovered (*i.e.*, lower K_f rank);
2. higher values of K_f must deviate from lower K_f values, suggesting that the probability of return to a location decays with time, especially if it was a highly-visited location.

To test these hypotheses, we analyzed $P(K_s|K_f)$ for all K_f and K_s values. For example, a visit to a location with ranks $(10, 3)$ means a return to the 10th most visited site after visiting 3 other locations. The conditional frequencies are here represented as two-dimensional histograms (shown as heatmaps) (Figure 2).

The first pattern we can observe is that for both datasets the conditional probability distributions (Figure 2(a) and (b)) are highly right-skewed and asymmetric. The right-skewness results not only from a combination of the heavy tails of $p(K_f)$ and $p(K_s)$ individually, but also from the convolution of them.

From the asymmetries in the distribution we can extract important insights regarding the dynamics of the recency bias in human mobility. The first one is the fact that recency bias is more pronounced up to $K_s \approx 40$ visits, beyond which the return probability vanishes. One possible explanation for such upper bound to the recency effect is due to the maximum long-term temporal regularities observable in $D1$ and $D2$ (*i.e.* monthly and yearly respectively). In $D1$, the average number of visits per *month* a user made is 46.4 whereas in $D2$, the average number of visits per *year* was 46.7. Since it is difficult to determine the recency bias in such long-term regularities, from here on we will focus our attention on the short-term returns.

When it comes to our most-visited locations, we tend to return to them after visiting very few locations. It can be seen by the rapid decrease in the returns frequencies when K_s grows. For instance, in $D1$, more than 91% of the returns to the most-visited place occurred after visiting fewer than five other locations, while for $D2$, it was more than 86% (see Figure 3).

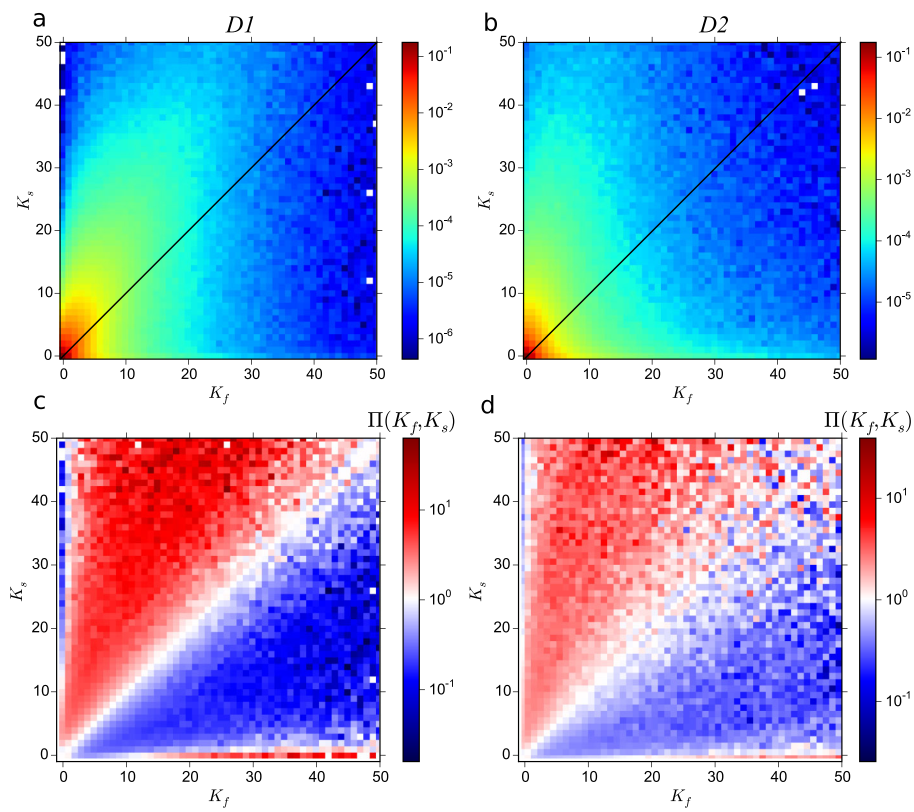


Figure 2 Return probabilities. Each point represents a return, whereas the color encodes the density of points. The top panels correspond to the rank-based recency distribution. The ranks here were shifted to have the highest-ranked locations at (0, 0) and a point (x, y) in the histogram represents a return to the $(x + 1)$th most-visited location after $y + 1$ steps. **(a)** Looking at the return ranks distribution for *D*1 we can observe that the recency influence is less pronounced in *D*1 in comparison with *D*2. **(b)** On the other hand, the finer-grained data of *D*2 shows a strong influence of recency. Return probability ratio $\Pi(K_f, K_s)$ for *D*1 **(c)** and *D*2 **(d)**. In particular, signatures of the dominance of recency should manifest themselves in the plot as red for $x > y$.

4 The recency bias to recently-discovered locations

As we mentioned before, one way to decompose the recency from the frequency bias is by looking at the returns to recently-discovered or infrequently-visited locations, characterized by a $K_f > C_f$, where C_f is a K_f value above which the recency bias stands out from the frequency bias in a given dataset. In fact, what we really want to measure is the likelihood of returning to a location whose frequency rank is $K_f = x$ after having visited $K_s = y$ locations such as $p(K_f = x|K_s = y) > p(K_f = y|K_s = x)$ and $x \gg y$. Thus, we define the probability ratio $\Pi(x, y)$ as

$$\Pi(x,y) = \frac{p(K_f = x|K_s = y)}{p(K_f = y|K_s = x)},$$

where for $p(x, y) > p(y, x)$, the ratio $\Pi(x, y) > 1$. For instance, $\Pi(20, 2)$ quantifies the proportion between: the number of visits to the 20th most visited location after visiting 2 other locations and the number of visits to the 2nd most-visited location after visiting 20 other locations. Figure 2 (bottom panels) shows the distribution of $\Pi(x, y)$. Hence, we defined

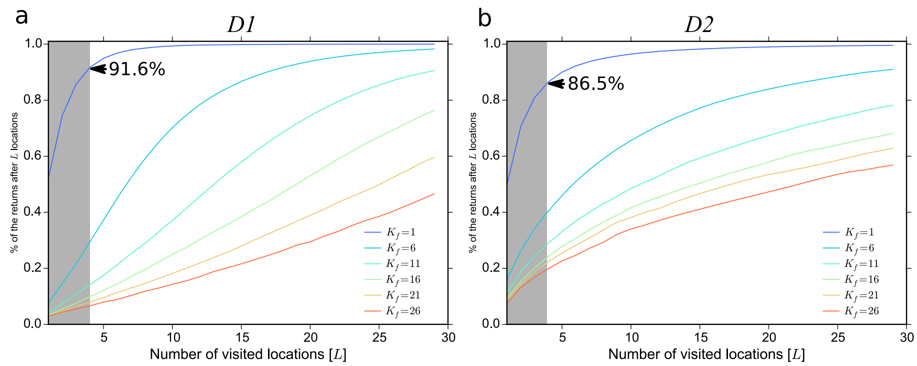


Figure 3 Fraction of returns to the K_f most-visited location occurring after the visitation of L different locations. Another way to look at the recency effect is by analyzing the correlation between the number of different visited locations between two visits to a location. We can see that people tend to return to their most-visited locations after visiting very few places. **(a)** In *D*1, more than 91% of the returns to the most-visited location occurred after visiting four or fewer locations while for *D*2 **(b)** it was about 86%.

C_f simply as

$$C_f = \min_x \left\{ \Pi(x,y) | \forall y : \Pi(x,y) > 1; x > y \right\}.$$

From Figures 2(c) and 2(d), we can visually estimate $C_f \approx 12$ and $C_f \approx 20$ for $D1$ and $D2$ approximately. Again, as expected, we can observe that the recency bias evident indeed becomes more and more prominent for larger K_f.

Based on what we described as the transient nature of the recency effect, it is clear that if a location is recurrently visited within short intervals for a reasonable time, it can climb up positions in the K_f rank. Moreover, since the recency information is entirely encoded within the order in which the places were visited. One simple but very useful implication of this property is that if we randomly shuffle a trajectory, the visitation frequencies are preserved whereas the recency bias is lost.

The first feature we can observe is that when we shuffle the trajectories in $D1$ (Figure 4(a)), the ranks distribution exhibit a similar pattern as observed on the original data. However, it supports our claim that the predominance of the preferential return, as captured by the aggregated mobile phone data of $D1$, is hindering the micro-level dynamics characteristic of the recency effect. A closer look at the bottom rows of Figure 4(a) does not show any increased probability due to recency. When we artificially destroy the power-law distribution of the visitation frequencies (Figure 4(b)) we can observe a dramatic change in the ranks distribution. It suggests that a significant part of the ranks distribution of $D1$ is indeed rooted on the visitation frequencies, as predicted by the PR mechanism.

When we analyze the randomized versions of $D2$ the influence of the recency becomes even more evident. As before, shuffling the individuals trajectories (Figure 4(d)) removes the features we described in Figure 2 (as before, the evidence in the bottom rows is not there). Moreover, by removing the temporal information from visitation sequences in $D2$, the rank distributions acquire the same form as the one of $D1$.

In summary, when we look at the recency rank distributions for the randomized data in both datasets, we see that the recency rank on the shuffled trajectories deviate from the empirical data. showing that the recency effect is indeed present in both datasets. More

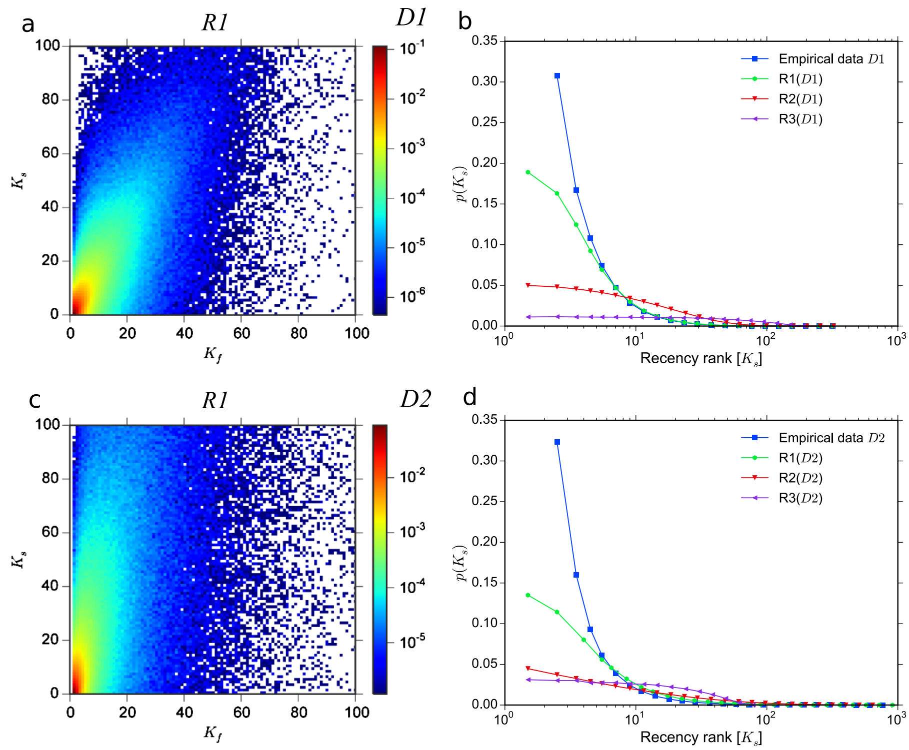


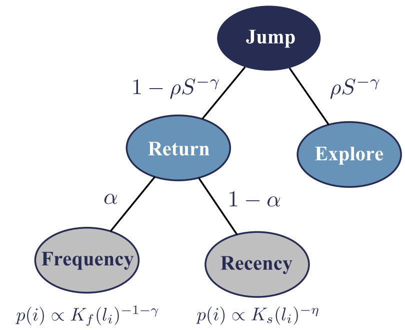
Figure 4 The rank-based analyses of randomized versions of the empirical datasets. (a) and **(c)** Conditional probabilities distribution of the randomized version *R*1 of *D*1 and *D*2 respectively (additional plots for the other randomization methods are on the Supporting Information). Rank variables were extracted from randomized versions of the datasets. Overall, the conditional probabilities have similar patterns as observed on the original data. However, when we look at the K_S distribution (in log-linear scale) (**(b)** and **(d)**), we see that the shuffled data deviates from the empirical data for $K_S \leq 4$. It is interesting to observe that when $K_S > 4$ the distributions for *R*1 and the original data converge again into a single curve.

striking, however, is the fact that this analysis not only shows that the recency effect is bounded to the most recently-visited locations but also suggests a possible existence of an upper limit to the effect. For instance, the recency effect could be observed more strongly when returns occur after visiting two locations in *D*1 and three locations in *D*2. It means that if an individual returns to a recently-discovered location before having visited 3 other locations, it is likely that this location will be visited again soon.

5 The recency-based model

Based on the empirical evidence of the recency bias in human mobility, the next natural step is to test the generative mechanisms of the features described on the previous section. For such, we propose a recency-based variation to the EPR model where the recency bias is incorporated. Also, we disregarded the CTRW component of the model. The noninclusion of CTRW let us better capture the recency visitation bias; in our analyses only the individuals' displacements (*i.e.*, successive observations in different locations) were considered. Therefore, waiting times would have absolutely no effect in our analyses since they would be removed in the pre-processing phase. A high-level representation of the model is depicted in Figure 5. Notice that in our definition we used uppercase K for the rank variables whereas in Ref. [18] the authors used lowercase k.

Figure 5 Recency-based individual mobility model. Notice that the exploration mechanism is kept the same as in the EPR model. In addition to the PR mechanism, the proposed model incorporates the recency effect, where recently-visited locations have also a high visitation probability.



The model can be described as follows: first, a population of N agents is initialized and scattered randomly over a discrete lattice with $M \times M$ cells, each one representing a possible location. The initial position of each agent is accounted as its first visit. At each time step agents can visit a new location if probability $p_{\text{new}} = \rho S^{-\gamma}$, where S corresponds to the number of distinct locations visited thus far. The parameters values were estimated from the empirical data (see Supporting Information for details) as $\gamma_{D1} = 0.73 \pm 0.03$ and $\rho_{D1} = 0.83 \pm 0.03$. For $D2$, the estimated parameters were $\gamma_{D2} = 0.50 \pm 0.08$ and $\rho_{D2} = 0.75 \pm 0.03$.

With complementary probability $1 - p_{\text{new}}$ an agent returns to a previously visited location. If the movement is selected to be a return, with probability $1 - \alpha$ the ith last visited location is selected from a Zipfian distribution (Zipf's law) with probability

$$p(i) \propto K_s(l_i)^{-\eta},$$

where $K_s(l_i)$ is the recency-based rank of the location l_i. The parameter η controls the number of previously visited locations a user would *consider* when deciding to visit a location. With probability α the destination is selected based on the visitation frequencies with probability

$$\Pi_i \propto K_f(l_i)^{-1-\gamma},$$

where $K_f(l_i)$ is the frequency rank of location l_i. Notice that when $\alpha = 1$ we recover the original preferential return behavior of the EPR model while when $\alpha = 0$, visitation returns will be based solely on the recency. We experimentally tested different parameters configuration for the model. Our analyses have shown that when $\alpha = 0$, the heavy tail of the visitation frequency disappears while for $\alpha = 1$ the power law of the recency distribution vanishes. *It suggests that both mechanisms must be present in order to reproduce those two features.*

The synthetic data produced by the EPR model seems to have a good approximation with the empirical data (see Figure 6(a)). However, when we compare the bottom-most rows of the histogram, it deviates from the empirical evidence, by not capturing the broader distribution of $p(K_f, K_s)$ for recently-visited locations. On the other hand, the recency-based mechanism (RM) reproduced the recency influence as observed in the empirical data (Figure 6(b)).

When we look at the K_f distribution, the EPR model recovers its heavy tail, as one would expect (inset of Figure 6(d)). On the other hand, when we look at each variable individually

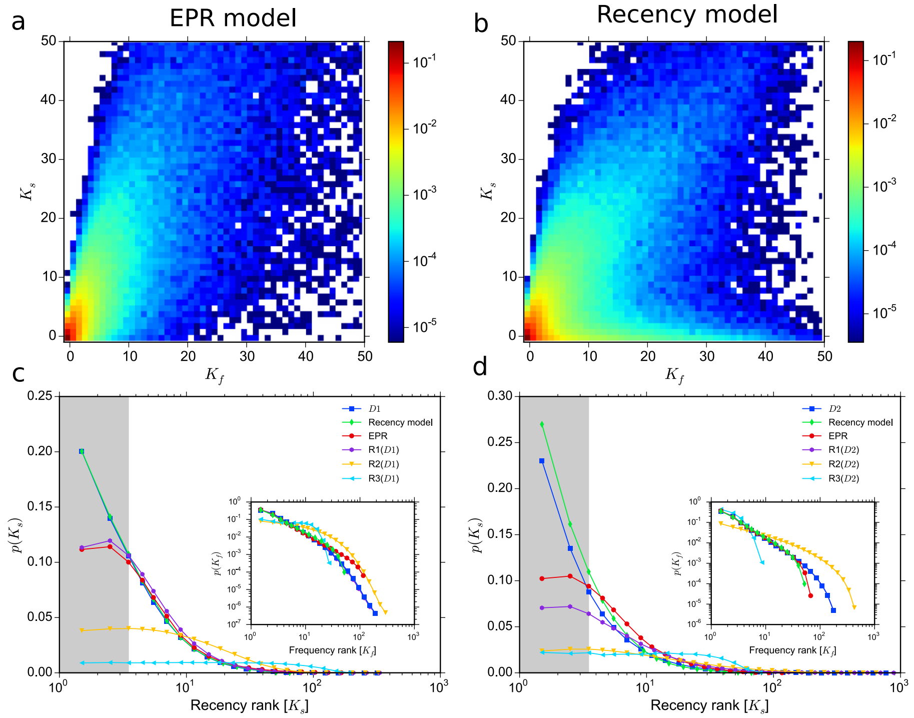


Figure 6 Comparison between the EPR model and the recency-based (RM) model. (a) The analysis of the return ranks generated by the EPR model shows that it reproduces a pattern similar to the one observed from the empirical analysis, especially of *D*1. **(b)** On the other hand, on the presence of the recency mechanism, we can observe the same high probability of return to recently-visited locations (*i.e.*, low K_s) as observed on the empirical data. **(c)** When we look at the distribution of the frequency ranks, the preferencial return mechanisms (labelled *EPR*) successfully exhibited a power-law distribution, in agreement with the empirical observations. Since the *R*1 data maintains the visitation frequencies, the K_f distribution of both variables are identical and hence their curves overlap. The activation of the recency mechanism does not affect the frequency rank distribution. **(d)** However, when we look at the K_s distribution, the EPR mechanism does not capture the power-law behavior observed on the empirical data.

we notice that the K_s distribution, as produced by the EPR model deviates from a power law. In fact, it is better approximated by an exponential distribution whereas recency-model maintains its power-law behavior. The differences in the K_s distribution as produced by both models become more evident in log-linear scale, where we can clearly see that the EPR model does not capture the preference for recently-visited locations (see main plot of Figure 6(c) and Figure 6(d)).

The validity of our approach in reproducing the recency bias was tested using a two-sample Kolmogorov-Smirnov (KS) test. As previously discussed, one way to observe the recency bias is by looking at the distribution of K_f for small K_s. Hence we tested the same-distribution hypothesis of K_f by comparing the empirical distributions from the data against those produced by the simulation models. In other words, we want to compare the visitation frequencies of the locations being visited after visits to at least K_s locations (Figure 7). To serve as a reference we applied the same approach comparing the K_f distributions of $D1$ against $D2$.

We can clearly see that the Recency model was the only one to reproduce the K_f distribution for small K_s values (*i.e.*, the recently-visited locations). Although the full K_f distribution produced by the *EPR* has strong agreement with the empirical data, it could not

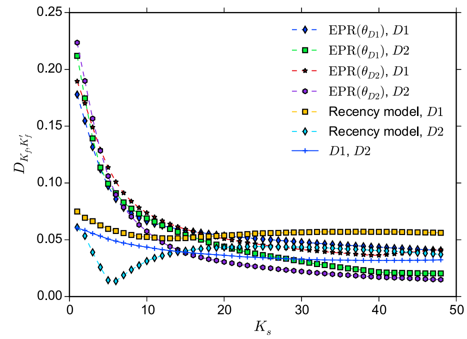


Figure 7 Two-sample KS statistic. Here we compare the goodness of fit offered by both the EPR and our Recency model with both the empirical datasets. Our analyses suggest that the recency effect is more noticeable in specific regions of the rank space. For this reason, we tested the same K_f distribution hypothesis for increasingly larger K_s ranges. In other words, this test evaluates the distance between the empirical and synthetic distribution of the K_f ranks of the visited locations up to a given K_f. θ_{D1} and θ_{D2} correspond to the EPR parameters vector as empirically estimated from $D1$ and $D2$ respectively, whereas $\mathrm{EPR}(\theta)$ represents the synthetic data produced by the EPR model using the parameters vector θ. Additionally, we applied the same approach to both empirical datasets to serve as a baseline for comparison.

reproduce recency effect as captured by conditional frequencies. For larger K_s values (*e.g.*, greater than 15), the *EPR* approximates again to the data, showing a fit even better than our approach, showing that the recency effect is indeed bounded.

Another interesting pattern observed in Figure 7 is that the goodness-of-fit test not only confirmed our findings that the importance of the Recency bias decays as we visit more locations between consecutive visits, but also it supports the evidence that such influence is bounded to approximately five locations.

6 Discussion

When it comes to visitation patterns, humans are extremely regular and predictable, where recurrent travels respond for most of our movements. An external observer can identify from ones's trajectories locations such as home and work, even after a very short period of observation. On the long term, however, these visitation patterns are not expected to remain the same. New locations are discovered. New social ties are established. New opportunities arise.

Akin to other human behaviors, traveling patterns evolve from the convolution between internal and external factors. A better understanding on the mechanisms responsible for transforming and incorporating individual events into regular patterns is of fundamental importance. In this work, we revealed that the recency bias - as observed in other human behaviors - also plays a role in human traveling patterns. Our results show that a single visit to a place strongly affects its likelihood of the further visits. More surprisingly, the recency influence is highly bounded to a few recently-visited locations. Our findings were drawn from a novel bivariate rank-based approach from which we could decompose the recency and frequency dimensions in determining individual visitation patterns.

Finally, we extended the EPR model to include a recency mechanism, which managed to successfully replicate some of the recency and frequency visitation patterns we described here. The importance of our results go beyond its scientific value for the human mobility community and their traditionally related areas such as urban planning and public health. The recency bias can be of great interest for areas such as public security (*e.g.*, detection of anomalies in individual trajectories) and strategic management (*e.g.*, offering a better understanding of customer visitation patterns) to name but a few. In a broader sense, our results add a small but important piece to our understanding of the human traveling behavior.

7 Materials and methods

7.1 The empirical datasets

In this work, we used two mobility datasets: the first one ($D1$) corresponds to 6 months of anonymized mobile-phone traces from a large metropolitan area in Brazil. This dataset is composed of 8,898,108 records from 30,000 users between January 1-June 30, 2014. The second dataset ($D2$) is composed of 23,736,435 *check-ins* from 51,406 Brightkite users in 772,966 different locations. Unlike the mobile phone data, locations in the Brightkite dataset correspond to the actual places where the users checked in - phone data locations correspond to the antenna tower the phone communicates with and hence are approximations of the user's actual location.

Since our interest here is on the individuals' *trajectories*, in this analysis we considered only the data that provides information relating to the users' displacements. Hence, we filtered out multiple repeated observations on the same place, resulting in a time series for each individual, representing their trajectories over the observed period. The rationale for removing the successive points in a same location is because in the context of this work, *recency* is defined in terms of visits to recent past destinations. Hence, successive observations within the same location cannot be considered as being influenced by a recency bias. Thus, since human displacements are interspersed by longer periods with no jumps, the bursty behavior, observed in many human activities (including mobile phone communications) [40, 41] would otherwise wrongfully boost the measurements of a recency preference.

To illustrate how the filtering works, if we assume that A, B and C are locations, and the data shows a user in the locations (in this order) $[A, B, B, B, C, C, A, A, A, B]$, the multiple consecutive observations at the same locations are filtered out. Hence, the trajectory to be analyzed would be $[A, B, C, A, B]$. Furthermore, to reduce the influence of co-located antennas (common in densely-populated sites), we merged those within less than 10 meters apart under the just one id.

7.2 The randomized datasets

Additionally, in order to verify whether the power law observed in the recency rank distribution is rooted on the temporal semantics of individuals' trajectories, we applied our rank-based approach to randomized versions of both empirical datasets ($D1$ and $D2$). The first randomized dataset we analyzed ($R1$) was obtained from uniformly shuffling each individual trajectory. This way, we artificially remove any temporal information possibly encoded within the individual trajectories, while maintaining the visitation frequencies intact. On the second randomization method ($R2$), we also remove the visitation frequencies by generating for each user a new random trajectory with the same number of displacements, and the same number of distinct visited locations. To serve as the baseline for the analyses, the data of the third randomization approach ($R3$) produces a new dataset with the same size as the original one, but keeping only the total number of users and locations. More precisely, for each of the datasets, we generated a randomized version of them with M random points

$$v_m = [u_m, l_m, m], \quad m \in [1, \dots, M],$$

where each u_m, l_m is uniformly sampled from U users and N locations respectively, with M, U and L the same as in $D1$ and $D2$.

Additional material

Additional file 1: Supporting Information: Statistical analysis and parameters estimation (pdf)

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
Developed the ideas, methods and analyses: HB and RM. Empirical data analysis: HB and AE. Wrote the manuscript: HB, FBLN and RM.

Author details
[1]BioComplex Lab, Florida Institute of Technology, 150 W University Blvd, Melbourne, FL 32901, USA. [2]Computational Intelligence Research Group, Polytechnic School, University of Pernambuco, Rua Benfica, 455, Recife, PE, 50720-001, Brazil. [3]COPPE, Federal University of Rio de Janeiro, Rua Horácio Macedo, Bloco G, 2030-101, Rio de Janeiro, RJ, 21941-450, Brazil.

Endnote
[a] Brightkite was a location-based social networking service launched in 2007 and closed in 2011 [33, 42].

Received: 24 July 2015 Accepted: 11 November 2015 Published online: 17 December 2015

References
1. Belik V, Geisel T, Brockmann D (2011) Natural human mobility patterns and spatial spread of infectious diseases. Phys Rev X 1(1):011001. doi:10.1103/PhysRevX.1.011001
2. Colizza V, Barrat A, Barthelemy M, Valleron A-J, Vespignani A (2007) Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. PLoS Med 4(1):13. doi:10.1371/journal.pmed.0040013
3. Balcan D, Vespignani A (2011) Phase transitions in contagion processes mediated by recurrent mobility patterns. Nat Phys 7:581-586. doi:10.1038/nphys1944
4. Toole JL, Ulm M, González MC, Bauer D (2012) Inferring land use from mobile phone activity. In: Proceedings of the ACM SIGKDD international workshop on urban computing. ACM, New York, pp 1-8. arXiv:1207.1115v1
5. Lenormand M, Gonçalves B, Tugores A, Ramasco JJ (2015) Human diffusion and city influence. J R Soc Interface 12:20150473. arXiv:1501.07788
6. Kitamura R, Chen C, Pendyala R, Narayanan R (2000) Micro-simulation of daily activity-travel patterns for travel demand forecasting. Transportation 27(1):25-51
7. Jung W, Wang F, Stanley H (2008) Gravity model in the Korean highway. Europhys Lett 81(4):48005. arXiv:0710.1274v1
8. Krajzewicz D, Hertkorn G, Wagner P, Rössel C (2011) SUMO (Simulation of Urban MObility): an open-source traffic simulation car-driver model
9. Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. Science 327(5968):1018-1021. doi:10.1126/science.1177170
10. Wang D, Pedreschi D, Song C, Giannotti F, Barabási A-L (2011) Human mobility, social ties, and link prediction. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining - KDD '11. ACM, New York, p 1100. doi:10.1145/2020408.2020581
11. Yang Y, Herrera C, Eagle N, González MC (2014) Limits of predictability in commuting flows in the absence of data for calibration. Sci Rep 4:5662. doi:10.1038/srep05662
12. Sadilek A, Krumm J (2012) Far out: predicting long-term human mobility. In: Twenty-sixth AAAI conference on artificial intelligence, pp 814-820
13. Krumme C, Llorente A, Cebrian M, Pentland AS, Moro E (2013) The predictability of consumer visitation patterns. Sci Rep 3:1645. doi:10.1038/srep01645
14. Lu X, Wetter E, Bharti N, Tatem AAJ, Bengtsson L (2013) Approaching the limit of predictability in human mobility. Sci Rep 3:2923. doi:10.1038/srep02923
15. González MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):479-482. doi:10.1038/nature06958
16. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. Nature 439(7075):462-465. doi:10.1038/nature04292
17. Hasan S, Schneider CM, Ukkusuri SV, González MC (2012) Spatiotemporal patterns of urban human mobility. J Stat Phys 151:304-318. doi:10.1007/s10955-012-0645-0
18. Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. Nat Phys 6(10):818-823. doi:10.1038/nphys1760
19. Schneider CM, Belik V, Couronné T, Smoreda Z, González MC, Couronne T (2013) Unravelling daily human mobility motifs. J R Soc Interface 10:20130246. doi:10.1098/rsif.2013.0246
20. Merton RK (1968) The Matthew effect in science: the reward and communication systems of science are considered. Science 159(3810):56-63. doi:10.1126/science.159.3810.56
21. Price D (1976) A general theory of bibliometric and other cumulative advantage processes. J Am Soc Inf Sci 27(5):292-306
22. Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509-512. doi:10.1126/science.286.5439.509
23. Szell M, Sinatra R, Petri G, Thurner S, Latora V (2012) Understanding mobility in a social petri dish. Sci Rep 2:457. doi:10.1038/srep00457

24. Toole JL, Herrera-Yaque C, Schneider CM, González MC (2015) Coupling human mobility and social ties. J R Soc Interface 12(105):20141128. doi:10.1098/rsif.2014.1128
25. Weber EU, Johnson EJ (2006) Constructing preferences from memory. doi:10.2139/ssrn.1301075
26. Hoch SJ (1984) Availability and interference in predictive judgment. J Exp Psychol Learn Mem Cogn 10(4):649-662. doi:10.1037/0278-7393.10.4.649
27. Hoch SJ (1985) Counterfactual reasoning and accuracy in predicting personal events. J Exp Psychol Learn Mem Cogn 11(4):719-731. doi:10.1037/0278-7393.11.1-4.719
28. Huang W, Li S, Liu X, Ban Y (2015) Predicting human mobility with activity changes. Int J Geogr Inf Sci 29(9):1569-1587. doi:10.1080/13658816.2015.1033421
29. Zheng Y, Zhang L, Xie X, Ma W (2009) Mining correlation between locations using human location history. In: Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems, pp 2-5. doi:10.1145/1653771.1653847
30. Zheng Y, Zhang L, Xie X, Ma W-Y (2009) Mining interesting locations and travel sequences from GPS trajectories. In: Proceedings of the 18th international conference on world wide web - WWW '09, pp 791-800. doi:10.1145/1526709.1526816
31. Lichman M, Smyth P (2014) Modeling human location data with mixtures of kernel densities. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining - KDD '14, pp 35-44. doi:10.1145/2623330.2623681
32. Yan X, Zhao C, Fan Y, Di Z, Wang W (2013) Universal predictability of mobility patterns in cities. J R Soc Interface 11:20140834. doi:10.1098/rsif.2014.0834
33. Cho E, Myers SA, Leskovec J (2011) Friendship and mobility. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining - KDD '11. ACM, New York, p 1082. doi:10.1145/2020408.2020579
34. Zhang D, Xiong H, Yang L, Gauither V (2013) NextCell: predicting location using social interplay from cell phone traces. IEEE Trans Comput. doi:10.1109/TC.2013.223
35. Smoreda Z, Olteanu-Raimond AM, Couronné T (2013) Spatiotemporal data from mobile phones for personal mobility assessment. In: Zmud J et al (eds) Transport survey methods: best practice for decision making. Emerald, Bingley, pp 745-768
36. Wang P, Hunter T, Bayen AM, Schechtner K, González MC (2012) Understanding road usage patterns in urban areas. Sci Rep 2:1001. doi:10.1038/srep01001
37. Gambs S, Killijian M-O, Del Prado Cortez MNN (2012) Next place prediction using mobility Markov chains. In: Proceedings of the first workshop on measurement privacy and mobility MPM 2012, pp 1-6. doi:10.1145/2181196.2181199
38. Herder E, Siehndel P, Kawase R (2014) Predicting user locations and trajectories. In: User modeling, adaptation, and personalization. Springer, Cham, pp 86-97
39. Song L, Kotz D, Jain R, He X (2006) Evaluating next-cell predictors with extensive Wi-Fi mobility data. IEEE Trans Mob Comput 5(12):1633-1648. doi:10.1109/TMC.2006.185
40. Karsai M, Kaski K, Barabási A-L, Kertész J (2012) Universal features of correlated bursty behaviour. Sci Rep 2:397. doi:10.1038/srep00397
41. Jo HH, Karsai M, Kertesz J, Kaski K (2012) Circadian pattern and burstiness in mobile phone communication. New J Phys 14:013055. doi:10.1088/1367-2630/14/1/013055
42. Grabowicz P, Ramasco J, Gonçalves B, Eguíluz V (2014) Entangling mobility and interactions in social media. PLoS ONE 9(3):e92196. arXiv:1307.5304v1