Unified underpinning of human mobility in the real world and cyberspace

You may also be interested in:

Communication cliques in mobile phone calling networks
Ming-Xia Li, Wen-Jie Xie, Zhi-Qiang Jiang et al.

Modeling the mobility with memory
Jeehye Choi, Jang-Il Sohn, K.-I. Goh et al.

Statistically validated mobile communication networks: the evolution of motifs in European and
Chinese data
Ming-Xia Li, Vasyl Palchykov, Zhi-Qiang Jiang et al.

An empirical study on human mobility and its agent-based modeling
Tao Jia, Bin Jiang, Kenneth Carling et al.

Exact Solution of the Gyration Radius of an Individual's Trajectory for a Simplified Human Regular
Mobility Model
Yan Xiao-Yong, Han Xiao-Pu, Zhou Tao et al.

Two-state Markov-chain Poisson nature of individual cellphone call statistics
Zhi-Qiang Jiang, Wen-Jie Xie, Ming-Xia Li et al.

New Journal of Physics

The open access journal at the forefront of physics

CrossMark

PAPER

Unified underpinning of human mobility in the real world and cyberspace
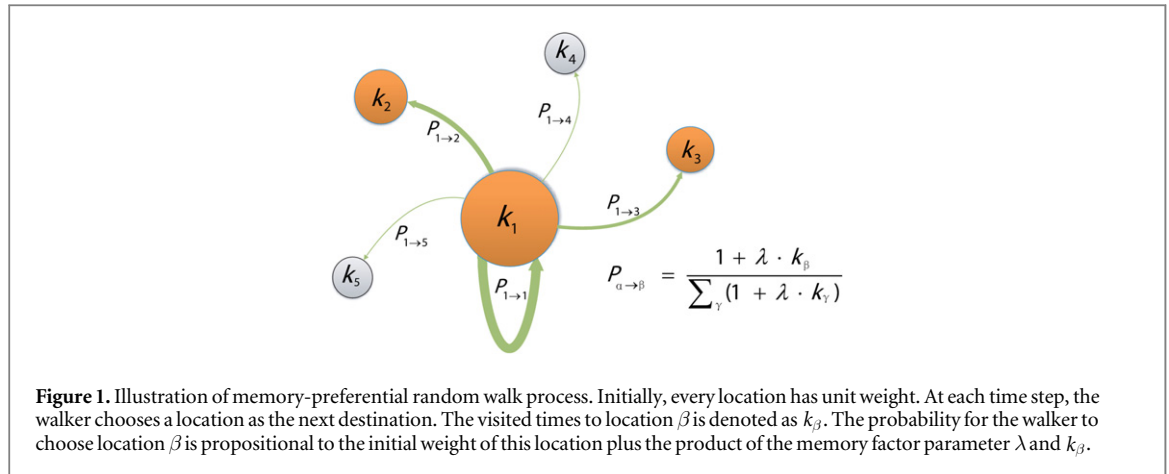
Yi-Ming Zhao[1], An Zeng[1], Xiao-Yong Yan[2,3], Wen-Xu Wang[1] and Ying-Cheng Lai[4]

1   School of Systems Science, Beijing Normal University, Beijing, 100875, People's Republic of China
2   Systems Science Institute, Beijing Jiaotong University, Beijing 100044, People's Republic of China
3   Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China
4   School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona 85287, USA

E-mail: anzeng@bnu.edu.cn and yanxy@bjtu.edu.cn

Abstract

Human movements in the real world and in cyberspace affect not only dynamical processes such as epidemic spreading and information diffusion but also social and economical activities such as urban planning and personalized recommendation in online shopping. Despite recent efforts in characterizing and modeling human behaviors in both the real and cyber worlds, the fundamental dynamics underlying human mobility have not been well understood. We develop a minimal, memory-based random walk model in limited space for reproducing, with a single parameter, the key statistical behaviors characterizing human movements in both cases. The model is validated using relatively big data from mobile phone and online commerce, suggesting memory-based random walk dynamics as the unified underpinning for human mobility, regardless of whether it occurs in the real world or in cyberspace.

1. Introduction

Human mobility in the real world and cyberspace plays an ever increasing role in the modern society and economy. Many important processes are affected by the patterns of human mobility, such as epidemic and information spreading [1–5], traffic congestion [6–8], and e-commerce. Modern research on human mobility dynamics began with the trajectory-based approach [9], e.g., by tracing the trajectories of dollar bills in the real world, which revealed a number of scaling relations such as a truncated power law in the distribution of the traveling distance. Analysis of mobile phone data demonstrated that individual travel patterns can be characterized by a spatial probability distribution, indicating the existence of universal patterns in human trajectories [10]. The question of whether human mobility patterns are predictable was addressed through an analysis of the limits of predictability in human dynamics [11]. More recently, human mobility in cyberspace and its relation to that in the real world were studied using big data analysis and phenomenological modeling [12].

Fundamental to the study of human mobility dynamics is the development of models to reproduce the phenomena and scaling relations obtained from empirical data [13]. A pioneering work in this field is the articulation of a statistical, self-consistent microscopic model [14]. Subsequent studies focused on predicting the mobility flow between two locations through, e.g., the classic gravity model [15]. A stochastic process capturing local mobility decisions, the so-called radiation model, was introduced [16], which yields better agreement with empirical data than the gravity model. Alternative mechanisms were introduced to model human trajectories [17–23]. The modeling effort has also been extended to cyberspace [12]. While many models were developed to reproduce the scaling laws obtained from various human mobility empirical data, a physical and first-principle based understanding of the underlying dynamics is still missing. In particular, the widely studied model of human microscopic trajectories [14] imposes the hypothesis that the probability of individuals visiting new locations in the real world has a power-law form: $P_{new} = \rho S^{-\gamma}$, where S is the number of distinct locations

Figure 1. Illustration of memory-preferential random walk process. Initially, every location has unit weight. At each time step, the walker chooses a location as the next destination. The visited times to location β is denoted as k_β. The probability for the walker to choose location β is propositional to the initial weight of this location plus the product of the memory factor parameter λ and k_β.

already visited, with ρ and γ being two parameters. However, the underlying mechanism accounting for the power-law probability of exploring new locations is yet elusive, prompting us to wonder if there is a unified, minimal model capable of predicting all known scaling laws for human mobility in both the real world and cyberspace.

In this paper, we show that all observed statistical features of human trajectories in the real world and cyberspace can be quantitatively predicted through a unified mechanism; memory based, preferential random walk process. The memory effect has been known to be important to human dynamics in general [24–28], which in limited space is a basic ingredient in our minimal model. The probability for an individual to visit a new location can then be obtained from *first principle considerations* without the need to hypothesize any particular mathematical form. The basic idea is intuitive (as most of us have experience with); if an individual visited a location in the past, the location would imprint a memory effect on the individual, enhancing the probability for him/her to visit the same location in the future. The striking finding is that, this simple rule, with only a single parameter, is capable of generating all the known statistical properties of human mobility (e.g., those predicted by the models of self-consistency [14] with more parameters and scaling assumption about the probability). Solving our minimal model analytically, we obtain scaling relations that agree well with the empirical ones from mobile phone check-ins and online shopping data sets that record human trajectories in the real world and cyberspace, respectively. Our minimal model thus establishes the unified underpinning of human mobility, representing a significant step forward in understanding modern human behaviors through statistical physics. This has the potential to advance a number of disciplines such as social sciences and online economics.

2. Methods

2.1. Memory-preferential random walk model

We consider a finite space of M locations, in which N individuals perform random walk with the probability of visiting a position proportional to its weight. For convenience, we use Latin and Greek letters to denote individuals and locations, respectively. The weight of a location α with respect to individual i, w_α^i, is updated during the process. An actual visit of i to α will increase the weight w_α^i through λ — the memory factor parameter. For $\lambda = 0$ and $\lambda > 0$, the random walk is unbiased and memory-preferential, respectively.

In our memory-preferential random walk (MPRW) model, for individual i the weight sequence of all M locations at time step t can be written as $\{1 + \lambda k_1^i(t), 1 + \lambda k_2^i(t), 1 + \lambda k_3^i(t), \ldots, 1 + \lambda k_M^i(t)\}$, where $k_\alpha^i(t)$ is the number of times that position α has been visited before time t. When i is about to move to a new location at $t + 1$, the probability to go to α is proportional to the weight of α, i.e., $w_\alpha^i \sim 1 + \lambda k_\alpha^i(t)$. We have

$$p_\alpha^i(t + 1) = \frac{1 + \lambda k_\alpha^i(t)}{\Sigma_\beta [1 + \lambda k_\beta^i(t)]}.$$

A typical step of the memory-preferential random walk process is schematically illustrated in figure 1.

The three statistical quantities [10, 11, 14, 16, 24–28] characterizing the human mobility dynamics are: (*i*) the total number $S(t)$ of distinct locations that an individual visited within time t, (*ii*) the probability $P(z, t)$ of returning to the zth discovered location at time t, and (*iii*) the fraction $P(k, t)$ of locations that have been visited k times at time t. As we mainly investigate $P(z, t)$ and $P(k, t)$ under certain t value, so we denote respectively these two distributions as $P(z)$ and $P(k)$ for simplicity in the rest of the paper. The quantity $P(z)$ can be used to infer whether previously visited locations are more likely to be visited than newly discovered locations, which we

will show possesses a more complex form than the well-known Zipf's law [29]. The quantity $P(k)$ is similar to the degree distribution in complex networks [30–32]. The three quantities can be used to validate our model through a detailed comparison between theoretical prediction and numerical results. In human mobility research, there are actually some other measures relying on geographic distance information, such as the mean square displacement (MSD). However, our model focuses on the abstract space without the distance information. This characteristic guarantees that we can validate our model in both online and offline systems (see the results section). Therefore, we mainly consider the measures independent of geographic distance information in this paper.

2.2. Analytical solutions

We aim to obtain the analytic expectation values of the three characterizing quantities. Since walkers are independent of each other, it suffices to analyze a single walker.

(i) *The number of distinct locations, $S(t)$.* $S(t)$ is defined as the total number of distinct positions visited by the person within time t. Inspired by the master equation method, we write down the probability of visiting a new position:

$$P_{\text{new}} = \frac{M - S}{M + \lambda t}.\tag{1}$$

By solving it, we have

$$S(t) = M - (M - 1)\left(\frac{M + \lambda}{M + \lambda t}\right)^{\frac{1}{\lambda}}.\tag{2}$$

In addition, we obtain the solution to P_{new} as

$$P_{\text{new}} = \frac{(M - S)^{\lambda+1}}{(M + \lambda)(M - 1)^{\lambda}}.\tag{3}$$

Even though the formula has a different form from $P_{\text{new}} \sim S^{-\gamma}$, one can see that the formula is indeed in a power-law form and P_{new} decreases with S, which are in general consistent with the one assumed in the work with the mobile phone data [14].

(ii) *The visit probability of positions discovered at different time, $P(z)$.* By using same method as above we have

$$P(z) = \frac{(M + \lambda t)(\lambda + 1)}{(\lambda + N)\lambda t}\left(\frac{M - z}{M - 1}\right)^{\lambda}.\tag{4}$$

(iii) *The visit probability of each position, $P(k)$.* To calculate $P(k)$, we note that the total number of the visited locations is S. Each location has its own ordinal z, which gives a relation between k and z. Suppose $k(z)$ is a monotonously decreasing function, we can obtain its inverse $z(k)$, also a monotonously decreasing function. The measure of $k = x$ is $|\Delta z| = |z'(k = x)\Delta k|$. We have

$$P(k = x) = \frac{|\Delta z|}{S} = \frac{|z'(x)\Delta k|}{S}.\tag{5}$$

As k is an integer and $\Delta k = 1$ in the system, we have
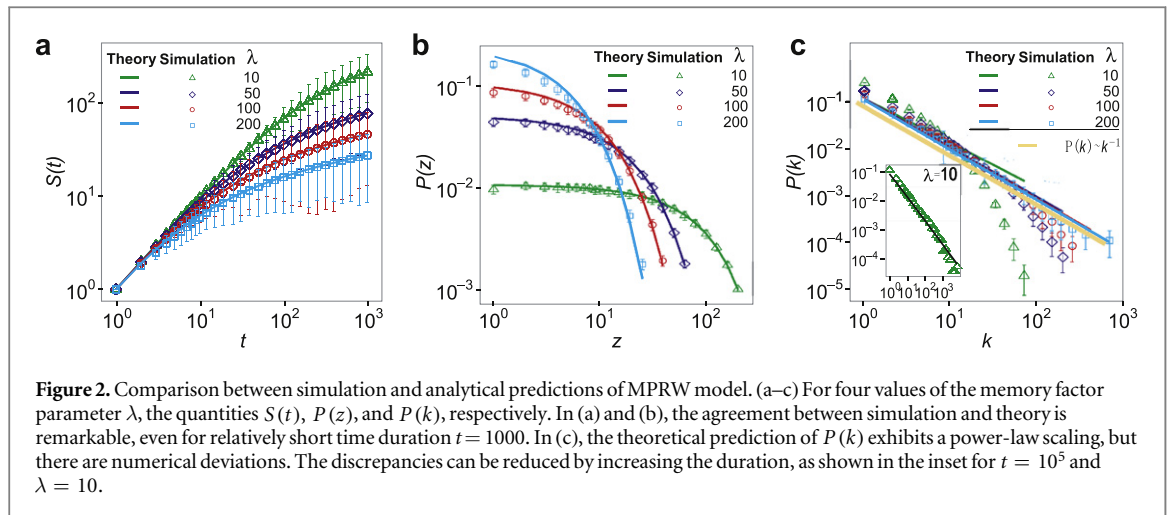
$$P(k) = \frac{|z'(k)|}{S(t)},\tag{6}$$

where

$$\begin{cases} |z'(k)| = -z'(k) = (M - 1)\left[\dfrac{\lambda + M}{(M + \lambda t)(\lambda + 1)}\right]^{\frac{1}{\lambda}}(k\lambda + 1)^{\frac{1}{\lambda}-1}, \\[2mm] S(t) = M - (M - 1)\left(\dfrac{M + \lambda}{M + \lambda t}\right)^{\frac{1}{\lambda}} \end{cases}\tag{7}$$

More details can be seen in the appendix section.

3. Results

3.1. Numerical validation on artificial systems

We conduct systematic numerical simulations of our MPRW model to obtain the scaling laws governing the three characterizing quantities, using the concrete setting where 100 walkers are distributed in a space of 1000 locations and perform 1000 walks, i.e., $N = 100$, $M = 1000$ and $t = 1000$. As $S(t)$, $P(z)$, $P(k)$ are defined for
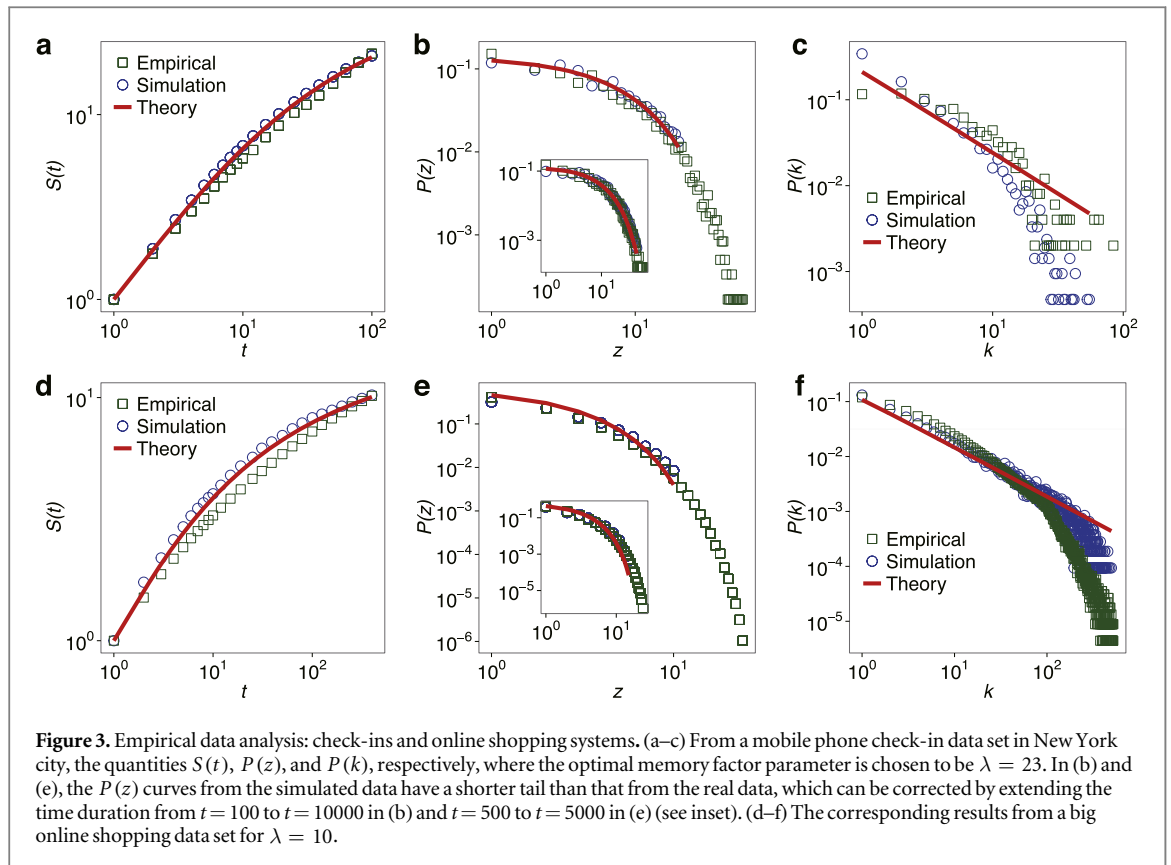
Figure 2. Comparison between simulation and analytical predictions of MPRW model. (a–c) For four values of the memory factor parameter λ, the quantities $S(t)$, $P(z)$, and $P(k)$, respectively. In (a) and (b), the agreement between simulation and theory is remarkable, even for relatively short time duration $t = 1000$. In (c), the theoretical prediction of $P(k)$ exhibits a power-law scaling, but there are numerical deviations. The discrepancies can be reduced by increasing the duration, as shown in the inset for $t = 10^5$ and $\lambda = 10$.

each walker. It is necessary to aggregate the results from all walkers to uncover the general features. Our approach is the following. (*i*) For each user i, we obtain the relation between $S^i(t)$ and t, where $S^i(t)$ is the total number of previously visited distinct locations within time t. We have $S(t) = (1/N)\sum_i S^i(t)$. (*ii*) To calculate $P(z)$, we let $P(z_\alpha^i)$ be the probability of i's visiting the location α. Say i has visited $z - 1$ distinct locations before walking into α. The quantity $P(z_\alpha^i)$ is then the fraction of times that walker i visited α, and we have $P(z_\alpha) = (1/N)\sum_i P(z_\alpha^i)$. (*iii*) For $P(k)$, we note that, each location α can be visited by different times for each walker. Let k_α^i be the number of times that walker i visited α. The aggregated frequency of visit to α is $k_\alpha = \sum_i k_\alpha^i$, and $P(k)$ can be obtained through the histogram of the sequence $\{k_1, k_2, k_3, ..., k_M\}$.

Figure 2(a) shows the function $S(t)$ for different values of λ, which is a sub-linear increasing function. We see that a stronger memory effect corresponds to a smaller rate of increase, which is natural due to individuals' resistance to explore new locations. Figure 2(b) shows the behavior of $P(z)$ at $t = 10^3$, where we see that the memory effect in general decreases the value z at which $P(z)$ begins to decrease rapidly, meaning that nostalgic individuals tend to discover few locations, a behavior that is consistent with that in figure 2(a). For both figures 2(a) and (b), the simulation results agree well with the analytical prediction by equations (2) and (4), respectively. Figure 2(c) shows the distribution function $P(k)$ at $t = 10^3$, which exhibits a general power-law scaling behavior. For large values of λ, the scaling exponent is about -1. However, small values of λ, $P(k)$ apparently deviate from the theoretically predicted power-law form. The deviation is a result of relatively short simulation duration. When we increase the duration to $t = 10^5$ in the case of $\lambda = 10$, the deviation diminishes, as shown in the inset of figure 2(c). In the solution of $P(k)$ (see equations (6) and (7)), both λ and t can be regarded as parameters. Given certain values of λ and t, $P(k)$ can be written as $P(k) \sim \left(k + \frac{1}{\lambda}\right)^{\frac{1}{\lambda} - 1}$. If the parameter λ is large enough, $P(k)$ can be rewritten as $P(k) \sim k^{-1}$. In figure 2, we added the curve of $P(k) \sim k^{-1}$ for comparison. One can see that the curves of $P(k)$ with different λ are indeed very close to $P(k) \sim k^{-1}$.

3.2. Validation with the location-based check-ins application
We validate our model with location-based check-in data [33]. Our data set recorded, in New York city, the positions of 42035 individuals as they use the check-in application, where the whole city is divided into 197 blocks. In order to obtain a long enough time series, we focus on the individuals who have at least 100 recorded locations and analyze their first 100 records. There are 60 individuals whose recorded data fulfill this requirement. The quantities $S(t)$, $P(z)$ and $P(k)$ are computed by aggregating the data from different individuals. The parameters that can be input to the MPRW model are thus $N = 60$, $M = 197$ and $t = 100$. A key to validating the model is the choice of some suitable value of the memory factor, λ. The optimal value, denoted as λ^*, can be estimated by comparing $S(t)$ from simulation and from real data. Specifically, limiting the choices of λ to integer values, we can calculate a set of square distance values, $d(\lambda)$, between the two $S(t)$ curves. The value of λ^* is one that minimizes $d(\lambda)$. For the mobile phone data, we have $\lambda^* = 23$, as shown in figure 3(a). We see that, for this choice of λ, the model predicted function $S(t)$ agrees well with that from the mobile phone check-ins data. With λ^* determined solely from $S(t)$, we also obtain a good agreement between the model and empirical results for the quantities $P(z)$ and $P(k)$, as shown in figures 3(b) and (c), respectively. which is remarkable. From figure 3(b), we note that the $P(z)$ curve from the model has a shorter tail than that from the

Figure 3. Empirical data analysis: check-ins and online shopping systems. (a–c) From a mobile phone check-in data set in New York city, the quantities $S(t)$, $P(z)$, and $P(k)$, respectively, where the optimal memory factor parameter is chosen to be $\lambda = 23$. In (b) and (e), the $P(z)$ curves from the simulated data have a shorter tail than that from the real data, which can be corrected by extending the time duration from $t = 100$ to $t = 10000$ in (b) and $t = 500$ to $t = 5000$ in (e) (see inset). (d–f) The corresponding results from a big online shopping data set for $\lambda = 10$.

real data. This is due to the difference in the time scale between the simulation and real data, e.g., $t = 100$ in the real data may correspond to a much longer time duration in the model. Extending the time duration to $t = 10000$ in the model gives a better agreement, as shown in the inset of figure 3(b). We find that $S(t)$ and $P(k)$ are immune to this effect, insofar as the time duration is not too small.

3.3. Online commercial data

The big data set is from *Taobao.com* [34]. As a main business branch of the Alibaba Group (a giant Chinese Internet company), Taobao is regarded as China's equivalent of eBay. The data set consists of the click records of Taobao users. When a user intends to make a purchase on Taobao, he/she clicks a sequence of links to obtain the relevant information (e.g., brand and price) of the product and then chooses one to buy. This process can be regarded as users surfing online web pages, i.e., movements in the cyberspace. Our data set consists of the records of 34330 individuals. After initial filtering to remove the individuals with abnormally long or very short click strings, we obtain a slightly smaller data set with 33462 users, for which the total number of web pages is 25. Each web page is corresponding to a category of products such as books and domestic appliances. The mobility of users in this online shopping web site indicates that users are browsing among different categories of products.

To cast the online shopping process in the framework of MPRW, we regard each web page as a location. The memory-preferential random walk can be interpreted as the tendency of a user to revisit the category of products that he/she is interested in (i.e. frequently visited). To be consistent with the data, we set $M = 25$ in the MPRW model. N and t, however, are two tunable parameters. N is the number of users considered in our model, corresponding to the number of independent realizations. We have found that the statistical properties of the model actually already reach the stable state when $N = 100$. In order to reduce the computational complexity, we use $N = 100$ instead of 33462. The parameter $t = 500$ means that each user jumps 500 steps in the space (i.e. each user clicks on 500 total web pages). Using the same method as for the mobile phone data, we determine the optimal value of the memory factor parameter to be $\lambda^* = 10$. The results of $S(t)$, $P(z)$, and $P(k)$ are shown in figures 3(d)–(f), respectively. Again, the results from MPRW model agree well with those from the data (for $P(z)$ a good agreement is achieved when an extended time duration, $t = 5000$, is used in the model, as shown in the inset in (b)), suggesting the model's wide applicability.

4. Discussion

To summarize, we develop a random walk model with a single parameter to reproduce the statistical scaling behaviors of the three quantities characterizing human mobility. The key element that makes our model distinct from previous ones is a memory-preferential mechanism in limited space. We demonstrate that, when this mechanism is incorporated into a standard random walk process, the analytically predicted behaviors agree, at a detailed and quantitative level, with those from two representative real data sets, one for real world and another for cyberspace movements. This is remarkable, considering that model is minimal with only a single adjustable parameter, the memory factor parameter. The main message is then that, while various mechanisms can be considered for human mobility, such as planted or social events [35] and gender difference [36–38], our findings provide strong evidence that random walk with memory is the unified underpinning of the human movement dynamics.

In fact, the model in this paper is similar to a preferential attachment model with a fixed number of nodes (short for PA_n) which has been studied in the literature [39]. However, we would like to emphasize that there are some key differences between our model and existing literature on PA_n. As our model and PA_n describe two very different processes (human mobility versus network evolution), these two models have many differences in their statistical properties. One straightforward example is the upper limit of node degrees. If the number of nodes in PA_n is n and e links will be added to the network, the possible maximum degree of node is $n - 1$. This is due to the fact that two nodes can only be linked once. In our model, if a walker walks e steps in a space with n places, the possible maximum degree (i.e. visits) a place can get is e instead of $n - 1$. Therefore, the analytic solution of the degree distribution is power-law in our model but exponential in PA_n. In addition, our model uncovers many statistical properties which never been studied with PA_n. As PA_n focuses on network evolution, most of related research studies the resultant degree distribution, i.e. $P(k)$. However, as our model is applied to study human mobility patterns, we investigate not only $P(k)$, but also $S(t)$ and $P(z)$ which are the statistical properties of human mobility and never been studied in network evolution. Taken together, though our model has some level of similarity to a PA with a fixed number of nodes, it has many novel properties as it is applied to model a completely different system.

While we assume in the present work that the walkers are homogeneous, the analysis can be extended to model incorporating memory heterogeneity [40]. Though introducing heterogeneity of user behavior could make a simple analytical solution intractable, it may lead to much richer statistical properties emerged from the model. Therefore, we believe that taking into account user heterogeneity in our model would be an important extension. In addition, we use two real data sets from totally different backgrounds to validate our model, one online and one offline. It would be interesting to test our model in many different data sets and look at more elaborate properties in future research. Given a data set from any generic behavior of human movements, the optimal memory factor parameter for the MPRW model can be estimated by comparing the behavior of an elementary statistical quantity from data and model. This feature has the additional benefit of assessing and quantifying the degree of intrinsic memory effect in the real system, which has potential applications to problems of significant current interest such as traffic optimization and online recommendation.

Acknowledgments

Appendix. The derivation details of the analytical solutions

The number of distinct locations, $S(t)$

We have the probability of visiting a new position equation (1). So we have the equation

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{M - S}{M + \lambda t}. \tag{8}$$

By solving it, we have

$$S(t) = M - c\left(\frac{1}{M + \lambda t}\right)^{\frac{1}{\lambda}},$$ (9)

where c is a constant. At the beginning of the random walk, we have the initial condition: $t = 1$ and $S(t) = 1$. Accordingly, we can obtain the value of c as

$$c = (M + \lambda)^{\frac{1}{\lambda}}(M - 1).$$ (10)

Thus we can have the solution of $S(t)$ as equation (2).

The visit probability of positions discovered at different time, $P(z)$

Let k_z denote the position z has been visited k times and t_z denote the first time when position z is visited. By using the same method, to k_z, we have

$$\frac{dk_z}{dt} = \frac{1 + \lambda k_z}{M + \lambda t}.$$ (11)

Solving it with the initial condition $t = t_z$ and $k_z = 1$, we have

$$k_z = \frac{(\lambda + 1)(\lambda t + M)}{\lambda(M + \lambda t_z)} - \frac{1}{\lambda}.$$ (12)

Actually, we have a hidden condition $S(t_z) = z$. By using it, we can get

$$k_z = \frac{(M + \lambda t)(\lambda + 1)}{(\lambda + M)\lambda}\left(\frac{M - z}{M - 1}\right)^{\lambda} - \frac{1}{\lambda}.$$ (13)

The visited frequency of each position $P(z)$ is proportional to k_z, so

$$P(z) \propto k_z.$$ (14)

As the sum of the k equals to evolving time t, so we have

$$P(z) = \frac{k_z}{t} \approx \frac{(M + \lambda t)(\lambda + 1)}{(\lambda + N)\lambda t}\left(\frac{M - z}{M - 1}\right)^{\lambda}.$$ (15)

The visit probability of each position, $P(k)$

We already have $S(t)$ as equation (2). And we can solve $k(z)$ inversely to obtain

$$z(k) = M - (M - 1)\left[\frac{(\lambda k + 1)(\lambda + M)}{(\lambda + 1)(M + \lambda t)}\right]^{\frac{1}{\lambda}}.$$ (16)

So we have equation (6) and $P(k)$ as equation (5).

References

[1] Lima A, De Domenico M, Pejovic V and Musolesi M 2015 Disease containment strategies based on mobility and information dissemination *Sci. Rep.* **5** 10650
[2] Frías-Martínez E, Williamson G and Frías-Martínez V 2011 An agent-based model of epidemic spread using human mobility and social network information *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Int. Conf. on Social Computing (SocialCom)* 57–64
[3] Belik V, Geisel T and Brockmann D 2011 Natural human mobility patterns and spatial spread of infectious diseases *Phys. Rev.* X **1** 011001
[4] Tizzoni M *et al* 2014 On the use of human mobility proxies for modeling epidemics *Plos ONE* **10** e1003716
[5] Wesolowski A *et al* 2012 C. O. quantifying the impact of human mobility on malaria *Science* **338** 267–70
[6] Deville P *et al* 2014 Dynamic population mapping using mobile phone data *Proc. Natl. Acad. Sci. USA* **111** 15888–93
[7] Jiang S *et al* 2013 A review of urban computing for mobile phone traces cuurent methods, challenges and opportunities *Proc. of the 2nd ACM SIGKDD Int. Workshop on Urban Computing, UrbComp '13*
[8] Berlingerio M *et al* 2013 AllAboard: a system for exploring urban mobility and optimizing public transport using cellphone data *Machine Learning and Knowledge Discovery in Databases Lecture Notes in Computer Science* **8190** 663–6
[9] Brockmann D, Hufnagel L and Geisel T 2006 The caling laws of human travel *Nature* **439** 462–5
[10] González M C, Hidalgo C A and Barabási A L 2008 Understanding individual human mobility patterns *Nature* **453** 779–82
[11] Song C, Qu Z, Blumm N and Barabási A L 2010 Limits of predictability in human mobility *Science* **327** 1018–21
[12] Zhao Z Z, Huang Z G, Huang L, Liu H and Lai Y C 2014 Scaling and correlation of human movements in cyberspace and physical space *Phys. Rev.* E **90** 050802(R)
[13] Blondel V D, Decuyper A and Krings G 2015 A survey of results on mobile phone datasets analysis *EPJ Data Science* **4** 1–55
[14] Song C, Koren T, Wang P and Barabási A L 2010 Modelling the scaling properties of human mobility *Nat. Phys.* **6** 818–23
[15] Zipf G K 1946 The P_1P_2/D hypothesis: on the intercity movement of persons *Am. Sociol. Rev.* **11** 677–86
[16] Simini F, González M C, Maritan A and Barabási A L 2012 A universal model for mobility and migration patterns *Nature* **484** 96–100
[17] Šćepanović S *et al* 2015 Mobile phone call data as a reginal socio-economic proxy indicator *Plos ONE* **10** e0124160

[18] Eagle N, Macy M and Claxton R 2010 Network diversity and economic development *Science* **328** 1029–31

[19] Yan X Y, Han X P, Wang B H and Zhou T 2013 Diversity of individual mobility patterns and emergence of aggregated scaling laws *Sci. Rep.* **3** 02678

[20] Yan X Y *et al* 2014 Universal predictability of mobility patterns in cities *J. R. Soc. Interface* **11** 0834

[21] Hou L, Pan X, Guo Q and Liu J G 2014 Memory effect of the online user preference *Sci. Rep.* **4** 06560

[22] Hu Y, Zhang J, Huan D and Di Z 2011 Toward a general understanding of the scaling laws in human and animal mobility *Europhys. Lett.* **96** 38006

[23] Saramäki J *et al* 2013 Persistence of social signatures in human communication *Proc. Natl Acad. Sci. USA* **111** 942–47

[24] Yamasaki K *et al* 2005 Scaling and memory in volatility return intervals in financial markets *Proc. Natl Acad. Sci. USA* **102** 9424–28

[25] Cattuto C, Loreto V and Servedio V D P 2006 A Yule-Simon process with memory *Europhys. Lett.* **76** 208–14

[26] Cattuto C 2006 Semiotic dynamics in online social communities *Eur. Phys. J* **46** 33–7

[27] Goh K I and Barabási A L 2008 Burstiness and memory in complex systems *Europhys. Lett.* **81** 48002

[28] Szell M *et al* 2012 Understanding mobility in a social petri dish *Sci. Rep.* **2** 00457

[29] Simon H A 1955 On a class of skew distribution functions *Biometrika* **42** 425–40

[30] Newman M 2010 *Networks: An Introduction* (Oxford: Oxford University Press)

[31] Albert R and Barabási A L 2002 Statistical mechanics of complex networks *Rev. Mod. Phys.* **47** 74

[32] Barabási A L and Albert R 1999 Emergence of scaling in random networks *Science* **286** 509–12

[33] Levandoski J J, Sarwat M and Eldawy A 2012 LARS: a location-aware recommender system *Proc. of 28th Int. Conf. on Data Engineering* 450–61

[34] Zhao Z D *et al* 2012 Emergence of scaling in human-interest dynamics *Sci. Rep.* **3** 3472

[35] Calabrese F *et al* 2010 The geography of taste: analyzing cell-phone mobility and social events *Pervasive Computing Lecture Notes in Computer Science* **6030** 22–37

[36] Smoreda Z and Licoppe C 2000 Gender-specific use of the domestic telephone *Soc. Psychol. Q.* **63** 238–52

[37] Kovanen L, Kaski K, Kertész J and Saramäki J 2013 Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences *Proc. Natl Acad. Sci. USA.* **110** 18070–5

[38] Blumenstock J E and Eagle N 2012 Divided we call: disparities in access and use of mobile phones in rwanda *Inform. Tech & Int. Develop.*

[39] Ben-Naim E and Krapivsky P L 2012 Popularity-driven networking *Europhys. Lett.* **97** 48003

[40] Pappalardo L *et al* 2015 Returners and explorers dichotomy in human mobility *Nat. Commun.* **6** 8166