Characterizing the human mobility pattern in a large street network

Bin Jiang,[1,*] Junjun Yin,[2,†] and Sijian Zhao[3,‡]
[1]*Department of Technology and Built Environment, Division of Geomatics, University of Gävle, Sweden*
[2]*Digital Media Centre, Dublin Institute of Technology, Ireland*
[3]*Academy of Disaster Reduction and Emergency Management, Beijing Normal University, China*
(Received 17 October 2008; revised manuscript received 26 January 2009; published 31 August 2009)

Previous studies demonstrated empirically that human mobility exhibits Lévy flight behavior. However, our knowledge of the mechanisms governing this Lévy flight behavior remains limited. Here we analyze over 72 000 people's moving trajectories, obtained from 50 taxicabs during a six-month period in a large street network, and illustrate that the human mobility pattern, or the Lévy flight behavior, is mainly attributed to the underlying street network. In other words, the goal-directed nature of human movement has little effect on the overall traffic distribution. We further simulate the mobility of a large number of random walkers and find that (1) the simulated random walkers can reproduce the same human mobility pattern, and (2) the simulated mobility rate of the random walkers correlates pretty well (an R square up to 0.87) with the observed human mobility rate.

DOI: 10.1103/PhysRevE.80.021136 PACS number(s): 05.40.Fb, 03.65.−w, 45.70.Vn, 89.40.−a

I. INTRODUCTION

Human mobility in a street network consists of numerous flights (or jumps) from one street to another. To be specific, a person at the origin (O) with street ***a*** will randomly (sometimes with a priority to the most connected ones) decide to choose a connected street ***b***; she walks along street ***a*** before she reaches the junction of ***a*** and ***b***. Then she will choose another connected street ***c*** and so on and so forth until she has reached her destination (D) (Fig. 1). Thus human mobility in a street network, a sort of network constrained movement, is like a frog jumping from one street to another at the topological level and like a turtle walking persistently along individual streets at the geometric level. This movement behavior might be over simplified because human mobility in reality tends to be purposive with destinations such as schools, offices, and homes. Despite the purposive nature of human mobility at the individual level, human movement tends to be predictable at the level of crowds in terms of how many people come to individual streets. This is the basic hypothesis we intend to verify in this study.

The fundamental idea behind this hypothesis can be further addressed by how intensively a street is used by vehicles. By street, we refer to it as a set of consecutive street segments that can be naturally perceived as a street with a good continuity—the natural street [1,2] or the self-organized natural street [3]. Suppose that every vehicle in a street network is equipped with a global positioning system (GPS) receiver, recording a signal every 10 s of the vehicle's position (x_i, y_i, t_i) along the street. How intensive a street is used by vehicles can be approximated by the number of recorded positions along the street. Note that the number of positions should be adjusted by the speed effect, i.e., the faster the speed, the fewer the number of positions. The number of recorded positions over a sufficiently long time period (e.g., one day for daily traffic) in a street network reflects the intensity of street use or traffic distribution on individual streets. The traffic intensity, or traffic distribution, constitutes a human mobility pattern in the street network. Now, let us assume something different. Suppose that we set up the same number of random walkers and let them move around arbitrarily (see the above behavior description) in the same street network. Would the random walkers be similar to human beings in forming some mobility patterns? Our primary answer to this question is yes. To put it differently, some mobility pattern formed by the random walkers is surprisingly the same as the one by human beings.

This paper aims to provide statistical evidence to support our answer to the above stated question and test the hypothesis of the similarity between human mobility pattern and that of random walkers. Although human mobility has been studied using massive trajectories of dollar bills [4] and mobile phones [5], as well as the more accurate GPS data capturing trajectories of human movement [6], all the studies without an exception adopt a Euclidean space, assuming human mobility is not constrained by the underlying street net-

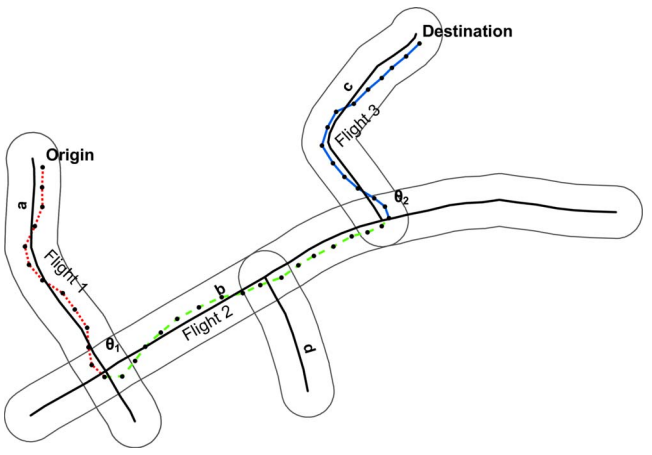


FIG. 1. (Color online) Illustration of a synthetic trail from the origin to the destination and its flights. (Note: flight 1—red dotted, flight 2—green dashed, and flight 3—blue solid.)

*bin.jiang@hig.se
†junjun.yin@student.dit.ie
‡scanzhao@hotmail.com

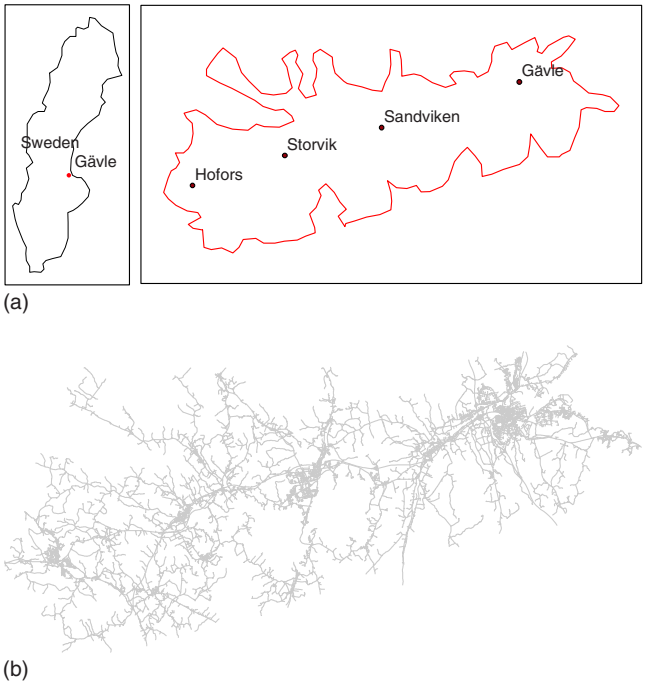


FIG. 2. (Color online) Maps (a) the taxi service area and (b) street network covering the four cities.

work. Furthermore, none of the studies attempted to illustrate a possible mechanism as to why human mobility demonstrates the Lévy flight behavior. In this regard, our study is an attempt, which adopts a network constrained space in illustrating human mobility patterns.

The remainder of this paper is organized as follows. Section II presents the data sources for the study. Section III introduces the methods for detecting a power law or Lévy flights using movement data. Section IV examines in detail the statistics of observed human mobility such as power law and Lévy flights. Section V further simulates random walkers' movement and illustrates their overall mobility properties or patterns. In Sec. VI, we try to uncover a possible mechanism behind the power-law distribution and Lévy flight behavior and further compare the observed human mobility and simulated random walkers' movement. Finally, Sec. VII summarizes the study and points to future work.

II. DATA

Three data sets are involved in this study: GPS data of 50 taxicabs' positions obtained automatically by GPS receivers, anonymized customer data as to when customers are picked up and dropped off, and the underlying street network. The data were collected from four cities or towns in the middle of Sweden: Gävle, Sandviken, Storvik, and Hofors [Fig. 2(a)]. The GPS data are massive; a GPS signal is captured every 10 s during a six-month period (October 2007, January to May 2008) for every one of the 50 cabs, capturing 59 983 958 positions. The customer data cover the same time period, recording anonymously a total of 166 679 customers' information as to when they were picked up and dropped off. The GPS data are represented by a quadruple (k_i, x_i, y_i, t_i), recording a cab (k_i)'s position (x_i, y_i) at time t_i, while the customers data are represented by a triple (k_i, t_p, t_d), representing pick-up time t_p and drop-off time t_d. Apparently, the association between a customer and a cab is known through the associated cab and the association between (t_p, t_d) and (x_i, y_i, t_i). The street network covers an area of 1253 km^2, with a total of 2445 km in length [Fig. 2(b)]. Based on the street network, we generate 4056 natural streets from a total of 10 439 street segments.

Both the GPS data and the customer data were used to extract trails. A trail is the GPS trajectory, representing a movement from where one is picked up to where one is dropped off. We derived 72 688 trails from the six-month data sets. Note the difference between the number of trails (72 688) and that of customers (166 679) as some invalid trails were not counted in our statistics. For example, trails lasting more than 2 h are considered to be invalid. This is because in a normal condition no one would take a taxi for such a long time. The extracted trails represent a fairly true picture of human mobility or an overall traffic distribution in the street network. Associated with trails are flights. As an example, Fig. 1 illustrates a synthetic trail passing over three different streets ***a***, ***b***, and ***c***, and thus it consists of three flights. Flights have two parameters: flight length and turning angle. The turning angle is in fact the deflection angle between the corresponding adjacent street segments. For example, the turning angle between flights 1 and 2 is the deflection angle between street ***a*** and ***b***. We extracted 578 585 flights from the 72 688 trails for the investigation of statistical properties.

III. METHODS FOR DETECTING LÉVY FLIGHTS: FITTING POWER LAWS TO HUMAN MOVEMENT DATA

Developed by the French mathematician Paul Lévy (1886–1971), a Lévy flight model is to extend the famous Brownian motion to be a more general random walk model [7]. A random walk model is a mathematical formalization of the intuitive idea of taking successive jumps or flights, each in a random direction. A Lévy flight is a particular random walk model that involves two different distributions: a uniform distribution for the turning angle (θ_i) on the one hand and a power-law distribution for the jump length (ℓ_i) on the other, i.e.,

$$P(\ell_i) \sim \ell_i^{-\alpha} \tag{1}$$

with $1 < \alpha < 3$.

Herein, the jump length lacks an average scale $(\bar{\ell})$, hence it is often called scale free [8]. It is important to point out that the Lévy stable distribution of the sums of jump length converges to a GAUSSIAN distribution when the Lévy exponent is set to $>=3$, indicating a Brownian motion. More significantly, Lévy flights differ from Brownian motions in movement patterns. The jump length of Brownian motions exhibits a GAUSSIAN distribution, while that of Lévy flights possesses a power-law distribution (Fig. 3).

To detect a Lévy flight is no more than to detect a power-law behavior of the jump length and further estimate the exponent α to see whether it is within the range $1 < \alpha < 3$.

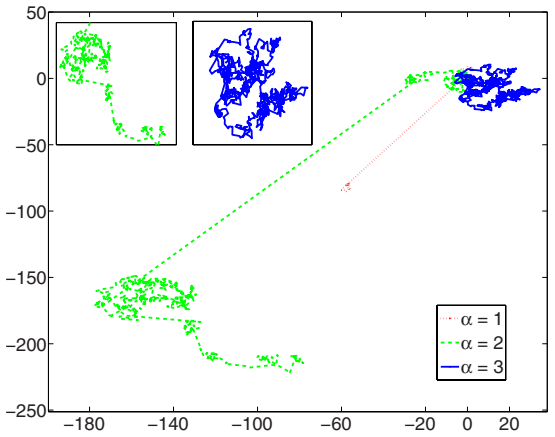


FIG. 3. (Color online) Illustration of Lévy flights with different exponent α. [Note: $\alpha \to 1$, straight lines red dotted; $\alpha \geq 3$, Brownian motion clusters (zoomed inset in blue solid); and $1 < \alpha < 3$, Lévy flights being scale invariant (zoomed inset in green dashed).]

This is commonly achieved by plotting the histogram of jump length in logarithmic scales. If the histogram closely follows a straight line, we can claim that a power law exists. This can be seen clearly from the logarithmic transformation of Eq. (1), i.e.,

$$\ln P(\ell_i) = -\alpha \ln(\ell_i) + b, \tag{2}$$

where α is the Lévy exponent and b a constant.

However, the above commonly used method suffers from errors in the logarithmic tail of the distribution (e.g., [9–11]). The end of the logarithmic tail looks messy because each bin only has very few samples in it. One possible solution is to use varying widths of bins in the histogram, with each bin m being increased by 2^m, to achieve a more homogeneous number of samples per bin. This can help us to reduce errors in the tail. This method can be further refined, i.e., by using the frequency per logarithmic bin normalized by dividing by bin width and total number of jumps to get the probability density (e.g., [10,12]). This first solution is actually based on the probability density function (PDF). Another solution, probably a much better one, is simply to use the cumulative density function (CDF), i.e., a plot of the probability that the jump length is greater than or equal to a certain value. This form of the power-law distribution is sometimes called Zipf's law or Pareto distribution.

The above-mentioned alternative solutions do not terminate here if our goal is to obtain the power-law exponent. Drawing a conclusion on Lévy flight relies on an accurate estimate of the power law or the Lévy exponent α. This is usually done by using the least-squares fit, but it is known to introduce systematic bias [9]. Because of the fact that a power-law distribution is likely to be confused with other heavy-tail distributions, such as the log-normal distribution [13] and the stretched exponential distribution [14], it is very tricky to draw a power-law hypothesis. Many researchers (e.g., [9,10,15]) have suggested more reliable methods, based on maximum likelihood methods and the Kolmogorov-Smirnov (KS) test, for identifying and quantifying power law distributions. Their methods can be used not only to fit a power law to data or part of the data but also for assessing how good the fit is, in comparison with other heavy tailed distributions. The estimated exponent is given by

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{\ell_i}{\ell_{\min}}\right]^{-1}, \tag{3}$$

where α denotes the estimated exponent and $\ell_{\min}$ is the smallest jump length for which the power law holds.

A modified KS test suggested by Clauset *et al.* [15] is adopted in this study to assess the goodness of fit, i.e., how good human movement data fit a power-law distribution. A fundamental idea is the maximum distance (δ) between the CDFs of the data and the fitted model,

$$\delta = \max_{x \geq x_{\min}} |f(\ell) - g(\ell)|, \tag{4}$$

where $f(\ell)$ is the CDF of the synthetic data with a value of at least $\ell_{\min}$ and $g(\ell)$ is the CDF for the power-law model that best fits the data while $\ell \geq \ell_{\min}$.

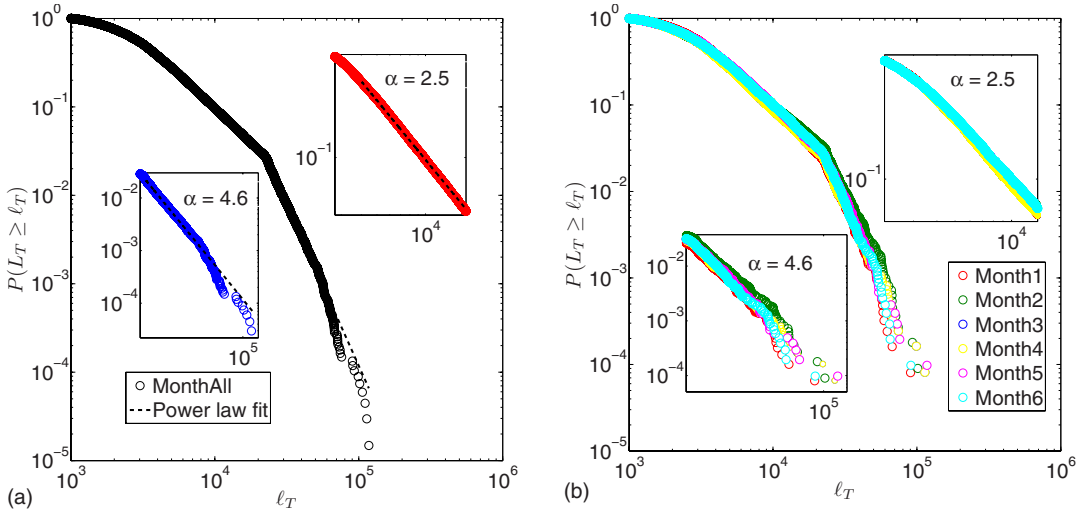


FIG. 4. (Color online) (a) Bipartite power-law distribution for the trails of 3–23 and >23 km for all six months together. (b) The same power-law distribution plotted month by month for checking consistency.

TABLE I. Exponent α and p value for both intra and intercity trails (support of power-law behavior).

	Intracity trails		Intercity trails	
	α	p	α	p
Month 1	2.69	0.02	4.73	0.20
Month 2	2.45	0.04	4.44	0.85
Month 3	2.48	0.49	4.63	0.03
Month 4	2.57	0.22	4.68	0.15
Month 5	2.54	0.16	4.72	0.44
Month 6	2.52	0.09	5.05	0.51
Month all	2.54	0.12	4.64	0.51

With the fitted model $g(\ell)$, we generate 1000 synthetic data sets that follow a perfect power law above $\ell_{\min}$ but have the same nonpower law distribution as the observed data below and recalculate the maximum distance between $f(\ell)$ and the fitted model, i.e., $\delta_i (i=1,2,\ldots,1000)$. A goodness of fit index p value is defined by

$$p = \frac{\text{No. of } \delta_i \text{ whose values are greater than } \delta}{1000}. \quad (5)$$

The p value indicates to what extent the data fit the model. The larger the p value, the more significant is the model, and p values greater than 0.05 are considered to be acceptable for a goodness of fit. However, some literature [5] suggests a smaller threshold around 0.01.

Although remaining controversial and requiring further and closer scientific scrutiny [16], Lévy flight behavior has been found in a variety of animal mobility and dispersals [12,17,18] and it becomes an appealing random walk model. Interested readers should refer to [17] and references therein for more details. The study by Viswanathan *et al.* [12] and many follow-up studies (e.g., [18]) have shown that Lévy flights can theoretically increase the success rate of a random search for a diverse variety of animals. Recently, human mobility has also been found to follow the random walk model [4,6], although a continuous Euclidean space is assumed in the studies. In what follows, we adopt a network constrained discrete space and aim to illustrate human mobility patterns from the perspectives of trails and flights.

IV. OBSERVATION OF HUMAN MOBILITY

A. Power-law fit for trails

Applying the above methods to the trail length, we found two very striking power laws, respectively, for the trails between 3 and 23 km and for the trails greater than 23 km, i.e., $P(\ell_T) \propto \ell_T^{-2.5}$ $(3 \leq \ell_T \leq 23)$ and $P(\ell_T) \propto \ell_T^{-4.6}$ $(\ell_T > 23)$, where ℓ_T denotes trail length (Fig. 4). This finding resembles the human behavior observed by Barabási [19] on human dynamics. It is also observed that 97% of trails are in the range of 3–23 km, while only 3% are greater than 23 km. Table I lists the exponent and p value for the goodness of fit. The two ranges (3–23 km and >23 km) of the trails correspond, respectively, to intracity and intercity movements, implying that for both intracity and intercity mobilities each shows a scaling property. We can note that the exponent for the intercity mobility trails is too big to make any sense, and it may be simply considered to be a cutoff.

B. Power-law fit for observed flights

A flight represents part of a movement along a persistent direction in a continuous space. In the discrete space of a street network, we define a flight as part of a movement along one particular street, which could be linear or curved (Fig. 1). With the extracted flights, we find that the flight length fits a power law with an exponential cutoff, i.e., $P(\ell_F) \propto \ell_F^{-2.3} e^{-\lambda \ell_F}$ $(\lambda = 0.000\,04)$, better than any alternatives such as log-normal, exponential, or stretched exponential cutoff (the details on the goodness of fit test can be found in Table II). The power-law exponent α is between 1 and 3, indicating a Lévy flight behavior. This finding conforms to previous studies on human mobility patterns [4]. On the other hand, the turning angle is found to exhibit a very sharply peaked bimodal distribution (Fig. 5). This distribution clearly deviates from the definition of Levy flights given in Sec. III. This deviation is probably due to the fact that human mobility occurred in a network constrained space, while the initial definition of Levy flights is given in a continuous Euclidean space. It should be noted that there is a

TABLE II. Tests of Lévy flight behavior with observed flights (support of power law with cutoff) (LR=likelihood ratio, for more details about it, refer to [15]).

	Log normal		Exponential		Stretched exp.		Power law with cutoff		Exponent	Rate
Flight	LR	p	LR	p	LR	p	LR	p	α	λ
Month 1	−33.2	0.0	1343.5	0.0	−35.3	0.0	−45.4	0.0	2.4	0.00005
Month 2	−12.4	0.0	1581.5	0.0	−13.6	0.0	−27.4	0.0	2.4	0.00003
Month 3	−16.7	0.0	1503.4	0.0	−17.9	0.0	−28.8	0.0	2.4	0.00003
Month 4	−34.2	0.0	1303.8	0.0	−36.8	0.0	−49.8	0.0	2.3	0.00005
Month 5	−51.0	0.0	1056.8	0.0	−54.1	0.0	−64.2	0.0	2.2	0.00006
Month 6	−51.0	0.0	1152.1	0.0	−53.3	0.0	−60.9	0.0	2.2	0.00006
Month all	−181.9	0.0	7882.7	0.0	−194.1	0.0	−260.5	0.0	2.3	0.00004

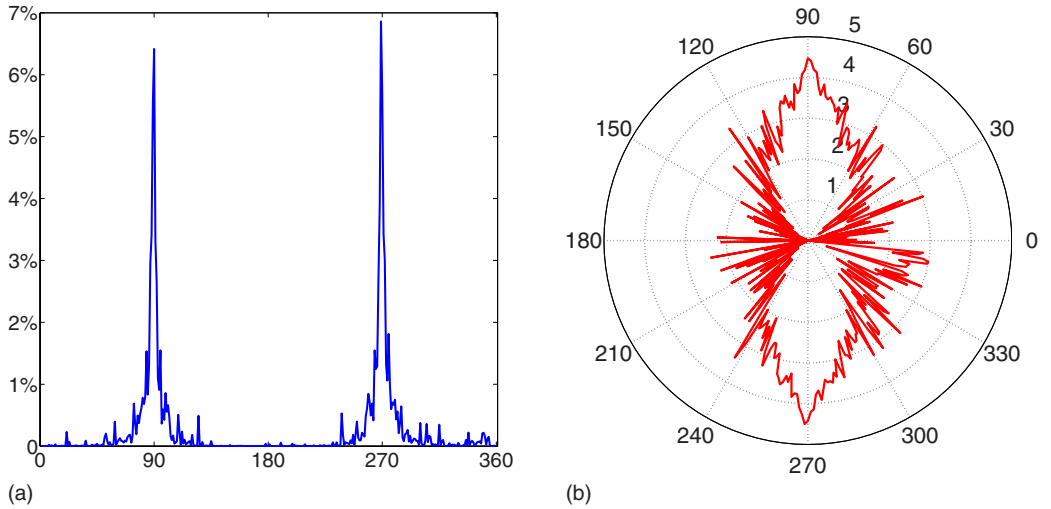


FIG. 5. (Color online) Distribution of the turning angle with the observed flights. (a) Histogram in the range of 0–360, each subrange (0–180 and 180–360) showing a very sharply peaked bimodal distribution. (b) The same histogram in a polar plot with logarithm for the frequency.

little bump in Fig. 6. This is due to the fact that the number of flights greater than 9 km makes up a very small portion. The same observation can be made regarding the simulated flights, but not as obvious as with the observed flights.

C. Power-law fit for observed traffic (90/10)

The number of positions recorded along the individual streets indicates to a certain degree the overall traffic distribution. However, this does not take into consideration of the speed effect, i.e., the faster the speed, the fewer the number of positions. In order to remove this speed effect, we made an adjustment using the equation

$$F = q\bar{\nu}, \tag{6}$$

where q is the number of positions recorded in an individual street and $\bar{\nu}$ is the average speed of a cab along the street.

For instance, three different cabs with speeds of 18, 7.2, and 3.6 km/h drive along a 100 m long street, and the numbers of positions recorded would be 2, 5, and 10, respectively, given that the positions are recorded every 10 s. After this speed adjustment using Eq. (6), every cab equally contributes to the traffic distribution, no matter what speed each holds.

We found a power-law behavior in the observed traffic. This implies that a minority of streets account for a majority of the traffic, i.e., the top 20% of streets account for 80% of traffic [20]. With the current study, it is found that the top 10% of streets account for over 90% of traffic (Fig. 7 and Table III).

V. SIMULATION OF RANDOM WALKERS' MOBILITY

A. Simulation settings

We simulate the mobility of numerous random walkers based on the idea of agent-based simulations in order to examine whether or not the random walkers can reproduce the observed human mobility patterns. The simulation is carried out from both topological and geometric perspectives, respectively (see Appendix). We adopt two types of random

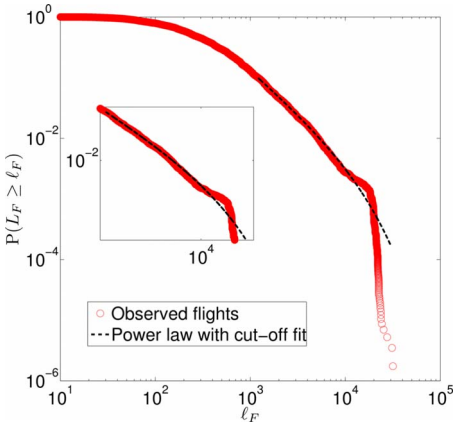


FIG. 6. (Color online) Observed flights showing a Lévy flight (length in km) behavior.

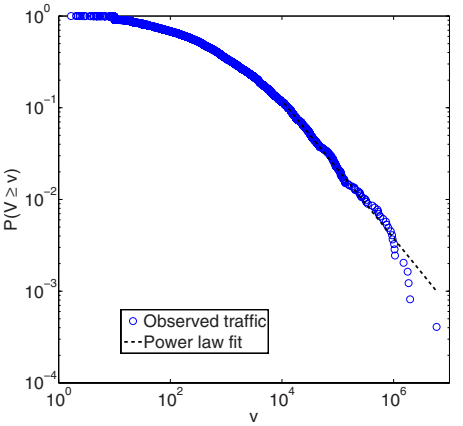


FIG. 7. (Color online) Power-law distribution of observed traffic (counted by the number of positions).

TABLE III. Tests of power law behavior of observed traffic assigned to individual streets (support of power-law behavior).

	α	p
Month 1	1.72	0.15
Month 2	1.76	0.39
Month 3	1.75	0.29
Month 4	1.75	0.29
Month 5	1.74	0.14
Month 6	1.78	0.28
Month all	1.75	0.52

walkers, inspired by the PageRank (PR) algorithm [21] and its modified version, the weighted PageRank (WPR) algorithm [22], hence they are referred to as PR walkers are WPR walkers. The only difference between the two types of random walkers is whether or not priority is given in choosing the next connected streets. In other words, the PR walkers are completely random, whereas the WPR walkers are less so due to a higher priority given to the most connected streets. Both the random walkers have the damping factor set to 1.0, according to the previous study [20].

The next question is how many (N_s) random walkers are used and how long (T_s) the simulation will last, as both variables have a significant effect on the number of simulated positions. For example, when T_s is too short or N_s is too few, most of the streets have no chance to be visited, so the simulation result would be biased. Therefore, the simulation must reach a saturated state before the simulated traffic can be compared against the observed human traffic. The saturated state we set is the time when the frequency of visiting individual streets is highly correlated with the calculated PR or WPR scores. In our simulation, the correlation coefficient R square value is 0.92. This is based on the fact that the probability of a random surfer visiting a web page can be accounted by its PageRank score. Note that the frequency of visiting individual streets and the traffic flow of individual streets are two different concepts, although they are closely related. The former is defined from the topological perspective, and it is counted only once when a connected street is chosen or visited, no matter how long the walkers move along it. However, the latter (traffic flow) is defined from the

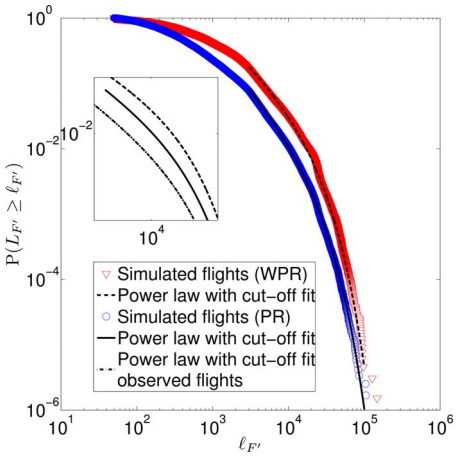


FIG. 8. (Color online) Simulated flights (length in km) approximately exhibit the same Lévy behavior, i.e., $P(\ell_{F'}) \propto \ell_{F'}^{-2.4} e^{-\lambda \ell_{F'}}$ ($\lambda = 0.000\ 03$ for WPR walkers and $\lambda = 0.000\ 04$ for PR walkers), in comparison with observed ones (inset).

TABLE IV. Tests of Lévy flight behavior with simulated flights (support of power law with cutoff) (LR=likelihood ratio, for more detail about it, refer to [15]).

	Log normal		Exponential		Stretched exponential		Power law with cutoff		Exponent	Rate
	LR	p	LR	p	LR	p	LR	p	α	λ
WPR walkers										
Month 1	−915.91	0.00	12967	0	−1018	0.00	−1328.7	0	2.0847	0.00003
Month 2	−838.44	0.00	13647	0	−932.9	0.00	−1242.7	0	2.12556	0.00003
Month 3	−872.80	0.00	13602	0	−976.8	0.00	−1317	0	2.096245	0.00003
Month 4	−894.62	0.00	13477	0	−997.3	0.00	−1319.5	0	2.094951	0.00003
Month 5	−896.77	0.00	13011	0	−997.8	0.00	−1310.3	0	2.089882	0.00003
Month 6	−990.32	0.00	13487	0	−1102	0.00	−1435.5	0	2.066448	0.00003
Month all	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
PR walkers										
Month 1	−548.45	0.00	7328	0	−600.8	0.00	−754.43	0	2.274887	0.00004
Month 2	−578.23	0.00	7274.9	0	−634	0.00	−791.21	0	2.252193	0.00004
Month 3	−665.81	0.00	6991.2	0	−726.6	0.00	−877.13	0	2.205745	0.00004
Month 4	−682.51	0.00	6903.2	0	−745.1	0.00	−905.56	0	2.204626	0.00004
Month 5	−604.54	0.00	6850.3	0	−666.1	0.00	−825.92	0	2.229495	0.00004
Month 6	−641.62	0.00	6891.3	0	−704.9	0.00	−870.96	0	2.208273	0.00004
Month all	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

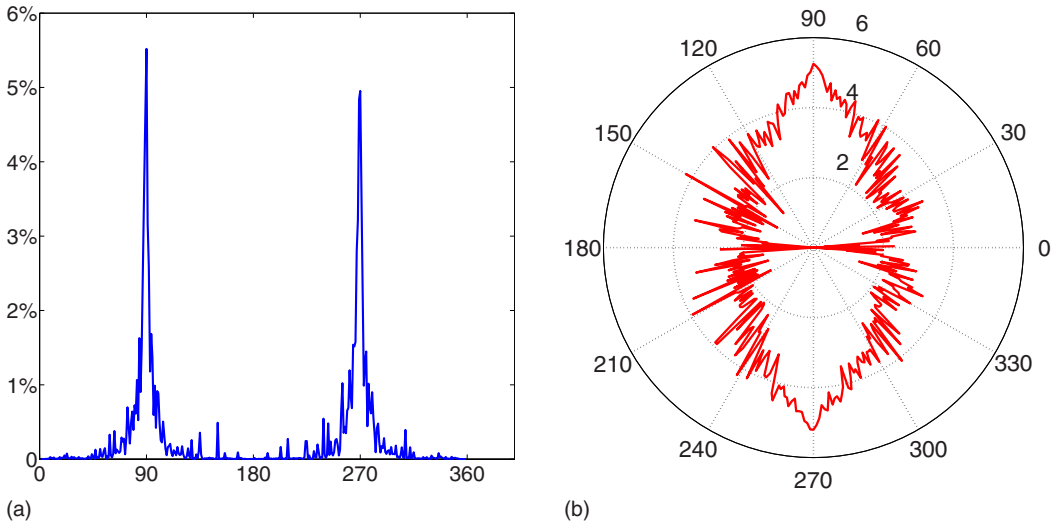


FIG. 9. (Color online) Distribution of the turning angle with the simulated flights (WPR walkers). (a) Histogram in the range of 0–360, each subrange (0–180 and 180–360) showing a very sharply peaked bimodal distribution. (b) The same histogram in polar plot with logarithm for the frequency.

geometric perspective, and it is accounted for by how long the walkers move along it. To reach the condition of the saturated state, we set either more walkers ($\wedge N_s$) with less time ($\vee T_s$), or fewer walks ($\vee N_s$) with more time ($\wedge T_s$), both having the same effect.

We set up 500 random walkers of each type to move at a persistent speed of 18 km/h for 5×10^6 s: (1) each random walker arbitrarily chooses one of the observed origins and destinations (O/D) as its origin, and then follows the random behavior, (2) how long each random walker shall walk for is determined by a random generator, which generates a series of travel times (T_m) for all the walkers. The travel time (T_m) follows a power-law distribution with an exponent of 2.5. The reason why we chose this particular exponent is based on the observation of the trail length, the majority of which (97%) exhibit a power law with an exponent of 2.5. The speed of 18 km/h reflects the average speed of the 50 taxicabs.

B. Power-law fit for simulated flights

Simulated flights demonstrate Levy flight behavior, but more precisely, a power law with cutoff (Fig. 8, Table IV). We are unable to do the test for all the six-month flights together because of the huge size, but the individual one-month flights do suggest a consistent result, i.e., a support of a power law with cutoff. On the other hand, the turning angle is found to exhibit a very sharply peaked bimodal distribution (Figs. 9 and 10). A slight difference can be observed in the distribution of turning angles with the polar plots, i.e., the one for PR walkers is a bit fatter than the one for WPR walkers. Also, two peaks are a bit different; in human walk-

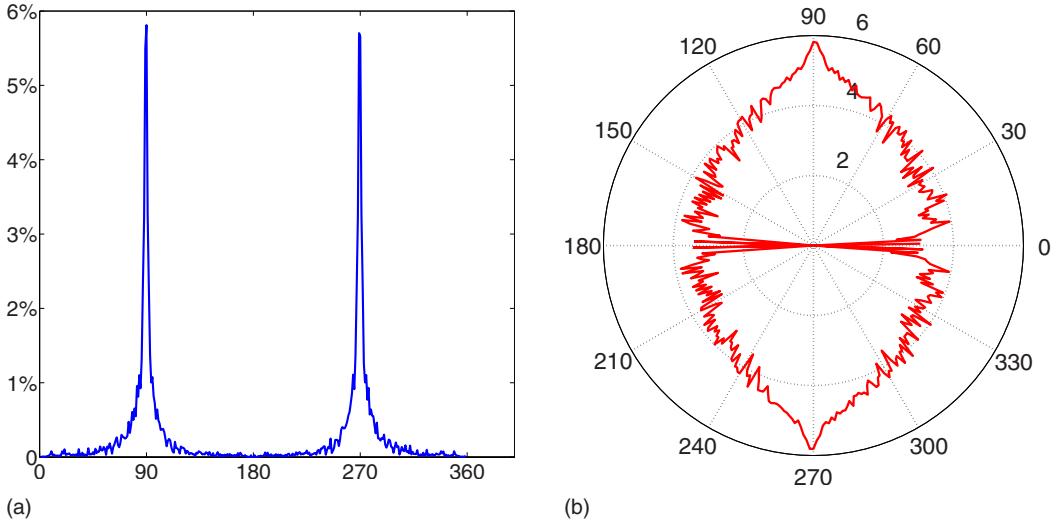


FIG. 10. (Color online) Distribution of the turning angle with the simulated flights (PR walkers). (a) Histogram in the range of 0–360, each subrange (0–180 and 180–360) showing a very sharply peaked bimodal distribution. (b) The same histogram in a polar plot with logarithm for the frequency.

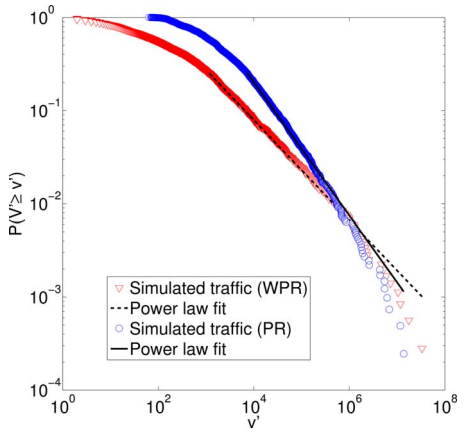


FIG. 11. (Color online) Power-law distribution of simulated traffic (counted by the number of positions).

ers, right turns are more common than left turns, while it is the opposite in random walkers. As to the reason, we suspect that it is to do with the randomness of random walkers and the purposiveness of human movement.

C. Power-law fit for simulated traffic (90/10)

Power-law behavior is again found in the simulated traffic (Fig. 11 and Table V), implying that a minority of streets account for a majority of traffic, i.e., the top 10% of streets account for over 90% of traffic.

VI. MECHANISM BEHIND THE LEVY FLIGHT BEHAVIOR AND OBSERVED TRAFFIC VERSUS SIMULATED TRAFFIC

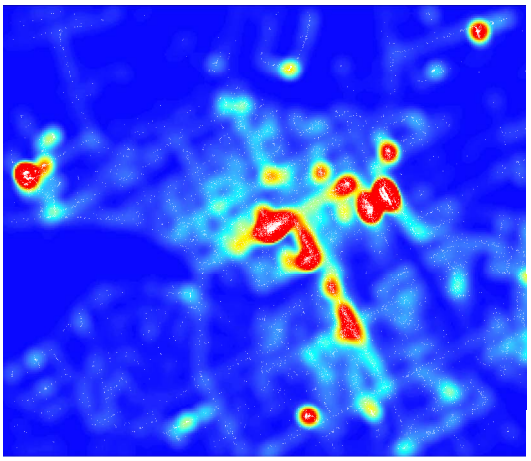
A. Possible mechanism behind power-law behavior of trail length

To uncover a possible mechanism behind the scaling property, we further examine the distribution of all possible

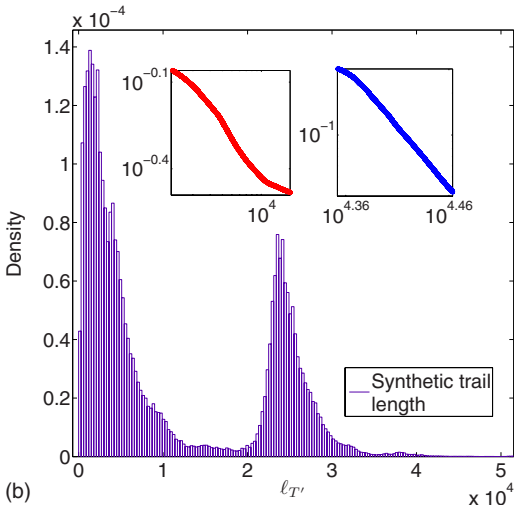
TABLE V. Tests of power-law behavior with simulated traffic assigned to individual streets (support of power-law behavior).

	WPR walkers		PR walkers	
	α	p	α	p
Month 1	1.53	0.65	1.66	0.19
Month 2	1.54	0.33	1.77	0.10
Month 3	1.53	0.59	1.79	0.19
Month 4	1.53	0.13	1.94	0.53
Month 5	1.55	0.53	1.77	0.20
Month 6	1.53	0.34	1.79	0.54
Month all	1.54	0.59	1.72	0.11

synthetic trails between O/D [Fig. 12(a) and Table VI]. The synthetic trails are assumed to be the shortest paths between the pairs of O/D. The number of O/D is massive, about 25 000 in one month, generating about 312 500 000 synthetic trails using the well-known Dijkstra shortest path algorithm. This number would make the power-law fit test too computationally intensive. Thanks to the clustered nature of O/D, and in order to reduce the computational burden, we first adopted the k means clustering method [23] to create 4000 clusters from 25 000 O/D in a one month period. The gravity centers of the 4000 clusters are considered to be new O/Ds, which are snapped into the street network to derive $(4000\times4000)/2-4000=7\,996\,000$ shortest paths or synthetic trails. The PDF of the synthetic trails are then plotted to assess a power-law fit. Although no striking power laws exist with the synthetic trails, the CDF of the synthetic trails in the ranges of 3–23 and >23 km does show a tendency for a power-law distribution (insets of Fig. 12). It is observed that about 80% of O/D are scattered, while 20% of them, indicated in red islands, are highly clustered, showing a clear



(a)



(b)

FIG. 12. (Color online) A possible mechanism behind the power-law behavior of trail length: (a) the tiny white spots are O/D during January 2008 at the center of Gävle city. The density of O/D is visualized in color from the highest (red islands) to the lowest (blue background). The largest red island or hotspot in the center is the downtown square, and other hotspots are the hospital, the central station, a district center, and the location of the taxi company. (b) The PDF of the 7 996 000 synthetic trails generated from the 4000 clusters, and the CDFs of the trails in the ranges of 3–23 km (inset to the left in red) and >23 km (inset to the right in blue) indicate two heavy-tail distributions.

TABLE VI. R square values between observed traffic and simulated traffic assigned to individual streets (support of a significant correlation).

	WPR walkers	PR walkers
Month 1	0.85	0.85
Month 2	0.87	0.83
Month 3	0.86	0.83
Month 4	0.85	0.85
Month 5	0.85	0.84
Month 6	0.86	0.83

scale invariance and fractal property. It is the scale free nature of the O/D distribution that underlies the power-law distribution of trail length.

B. Possible mechanism behind Levy flights behavior

We believe that it is the underlying street network that governs the Lévy flight behavior. To verify this belief, we further examine the topological structure of the street network in terms of the street-street intersections. First we derived 4056 natural streets based on the every-best-fit join principle and using degree 45 as the threshold. The reason why we adopted this particular principle is because the natural streets formed by this principle tend to be similar in morphological structure to the named streets [24]. The generated streets are transformed into a connectivity graph in terms of which street intersects which street, a sort of street topology based on street-street intersections. Both street length and connectivity are assessed to see whether or not they follow a power-law distribution. Not surprisingly, the underlying street topology demonstrates a power-law distribution with the exponent around 2.5, i.e., $P(c) \propto c^{-2.6}$ for street connectivity and $P(\ell_S) \propto \ell_S^{-2.4}$ for street length (Fig. 13). This statistical regularity can be expressed by the 80/20 principle: 80% of streets are less connected below the average and 20% of streets are well connected above the average [25]. The recorded taxicabs' positions every 10 s are overlapped with the street network to get an overall traffic distribution. We illustrate the fact that a minority of streets account for a majority of traffic [20], but with the current study it is adjusted to be that 10% of well connected streets account for 90% of traffic.

C. Observed traffic versus simulated traffic

To our surprise, the simulated traffic is highly correlated with the observed one, and the relevant R square value is between 0.83 and 0.87. This finding implies that a significant fraction of traffic can be accounted for by the underlying space (the street network topology and spatial distribution of O/D), and human free will or the purposive nature of mobility has little effect on the traffic distribution.

VII. CONCLUSION AND FUTURE WORK

In summary, in this study we investigated human mobility patterns from two different perspectives: trails between O/D and flights between streets. We found scaling properties with both trails and flights. The flights show a Lévy behavior in nature. More importantly, through simulation and experiments, it is illustrated that the scaling properties are attributed to the underlying street network and spatial distribution of O/D. We further conjecture that such a mechanism can be applied to human mobility patterns at a larger scale, e.g., at a country level. It is the fractal property of space itself that governs the scaling properties of human mobility patterns. Herein the space refers to manmade infrastructure such as cities, airports, train stations, and highways that all bear the scale-invariance property that governs the overall human mobility patterns. In this regard, the study by Barthélemy and Flammini [26] provides an interesting model. Our future work should concentrate on further investigation of geo-

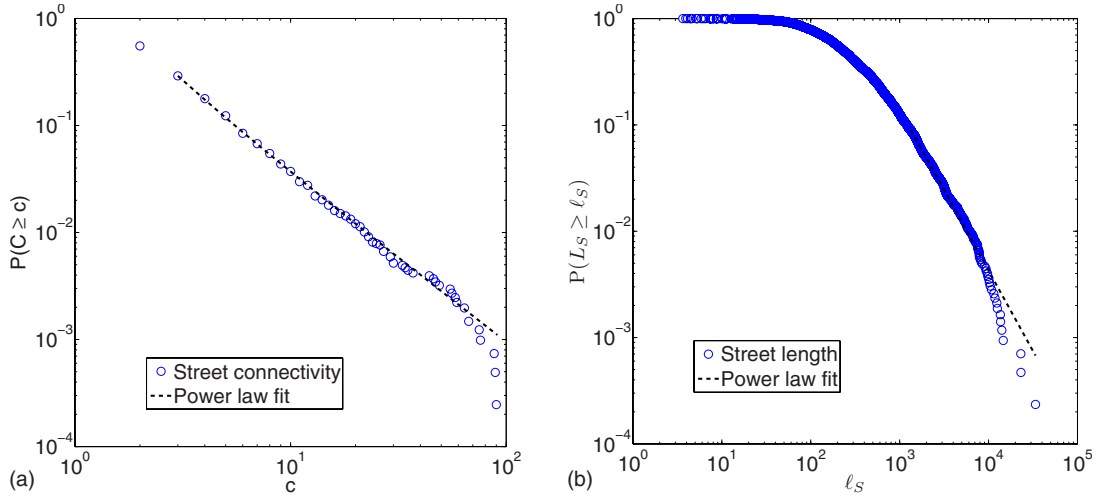


FIG. 13. (Color online) Scaling of streets (a) The power-law distribution of street connectivity, i.e., $P(c) \propto c^{-2.6}$, and (b) the similar power-law distribution of street length (in meters), i.e., $P(\ell_S) \propto \ell_S^{-2.4}$, where the exponent is slightly less than that of the street connectivity's distribution.

graphic space, i.e., how does space in general have an effect on human mobility patterns?

ACKNOWLEDGMENT

The research idea was conceived by B.J., data processing, analysis, and simulations were done by B.J., J.Y., and S.Z, and the paper was written by B.J. This study was financially supported by research grants from the Hong Kong Polytechnic University and the Swedish Research Council FORMAS. We would like to thank TAXI Stor och Liten for kindly providing the mobility data. The street network data are provided by the city of Gävle. In this connection, Anders Wahlgren and Annelie Hook deserve our special thanks for the data preparation and transferring. J.Y. would like to thank Aaron Clauset and César Hidalgo for their advice on the modified KS test.

APPENDIX: ALGORITHM FOR SIMULATING MOBILITY OF RANDOM WALKS

The algorithm is a detailed implementation of the random movement behavior given at the beginning of the paper. It is important to note a few points on the implementation: (1) each random walker arbitrarily chooses one of the observed O/D as its origin and then follows the random behavior, (2) how long each random walker shall walk for is determined by a random generator, which generates a series of travel time (T_m) for all the walkers. The time (T_m) follows a power-law distribution with an exponent of 2.5. The reason why we choose the exponent range is based on the observation of the trail length, the majority of which (97%) exhibit a power law with the exponent of 2.5.

Input: A street-based network shape file

Output: A txt file recording the number of positions in walker's trajectories for each street

Variables: D: damping factor value

T: simulation period

//The main function of simulation

Sub Simulating ( )

Transform the street-based network into a connectivity graph;

Create N turtles distributed randomly in the O/D-located streets;

Use a power law (α=2.5) function to randomly create a travel time T_m for each walker;

t=0;

While (t<=T) **do**

For Each walker **in** Walkers

If the walker's real travel time>T_m **Then**

Let walker jump randomly to an O/D-located street;

Use a power law (α=2.5) function to randomly recreate a travel time T_m for it;

End If

Call walker::MoveOn(dt) function to let the walker move to a new location;

Next

t=t+dt;

End while

//Output

Output the number of locations for each road to a txt file;

End sub

CLASS: Walker

// Walker walking behavior

Sub Walker::MoveOn(dt)

If (walker's next street doesn't exist) **Then**

Call walker::SearchNextStreet function to get a next street;

Get the junction between walker's current street and this street;

If (no such junction) **Then**

//Leap to another street

Locate the walker in the middle of this street;

Assign this street to the walker's current street;

Increment the walker-tracked number of current street by one;

Set the walker's next street to null;

Return;

Else

Assign this street to the walker's next street;

Compare the spatial relationship between the walker's current position and the junction to decide the walker's walking direction;

End If

End If

Use walker's speed, direction and time increment to calculate the next position;

If (walker reaches the junction between current street and next street) **Then**

Locate the walker in the junction;

Replace the walker's current street with next street;

Set walker's next street to null;

Increment the walker-tracked number of current street by one;

Else

Locate the walker in the new position;

Increment the walker-tracked number of current street by one;

End If

End sub

//Search a next street based on weighted PageRank (WPR) or standard PageRank (PR)

Function Walker::SearchNextStreet(walker's current street) as street

---

Get the connected street list A from walker's current street;
Generate a random number $0<d<1$;
If ($d<=D$) **Then**
If (PR) **Then**
Pick up a street randomly from List A;
Return this street;
Else if (WPR) **Then**
Pick up a street from List A according to the probability p;
// P is decided by the weight (w=connectivity/sum of connectivity)
// this pick up is in contrast to random pick up
Return this street;
End If
Else ($d>D$) Then
Select a street randomly;
Return this street;
End If
End Function

[1] R. C. Thomson, in *Bending the Axial Line: Smoothly Continuous Road Centre-line Segments as a Basis for Road Network Analysis*, Proceedings of the Fourth Space Syntax International Symposium, edited by J. Hanson (University College London, London, 2003).
[2] B. Jiang and C. Liu, e-print arXiv:0709.1981, Int. J. Geograph. Inf. Sci. (to be published).
[3] B. Jiang, S. Zhao, and J. Yin, J. Stat. Mech.: Theory Exp. (2006) P07008.
[4] D. Brockmann, L. Hufnage, and T. Geisel, Nature (London) **439**, 462 (2006).
[5] M. Gonzalez, C. A. Hidalgo, and A.-L. Barabási, Nature (London) **453**, 779 (2008).
[6] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, *On the Lévy-Walk Nature of Human Mobility: Do Humans Walk like Monkey?* (INFOCOM, Phoenix, AZ, 2008).
[7] J. Klafter, M. F. Shlesinger, and G. Zumofen, Phys. Today **49**(2), 33 (1996).
[8] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[9] M. L. Goldstein, S. A. Morris, and G. G. Yen, Eur. Phys. J. B **41**, 255 (2004).
[10] M. E. J. Newman, Contemp. Phys. **46**, 323 (2005).
[11] D. Sims, D. Righton, and J. W. Pitchford, J. Anim. Ecol. **76**, 222 (2007).
[12] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. G. E. da Luz, E. P. Raposo, and H. E. Stanley, Nature (London) **401**, 911 (1999).
[13] E. Limpert, W. A. Stahel, and M. Abbt, Bioscience **51**, 341 (2001).
[14] J. Laherrere and D. Sornette, Eur. Phys. J. B **2**, 525 (1998).
[15] A. Clauset, C. R. Shalizi, and M. E. J. Newman, SIAM Review (to be published), e-print arXiv:0706.1062 .
[16] A. M. Edwards *et al.*, Nature (London) **449**, 1044 (2007).
[17] G. M. Viswanathan, E. P. Raposo, and M. G. E. Luz, Phys. Life. Rev. **5**, 133 (2008).
[18] D. W. Sims *et al.*, Nature (London) **451**, 1098 (2008).
[19] A.-L. Barabási, Nature (London) **435**, 207 (2005).
[20] B. Jiang, Int. J. Geogr. Inf. Syst. **23**, 1033 (2009); e-print arXiv:0802.1284.
[21] L. Page and S. Brin, Proceedings of the Seventh International Conference on World Wide Web, 1998 (unpublished), Vol. 7, pp. 107–117.
[22] W. Xing and A. Ghorbani, Second Annual Conference on Communication Networks and Services Research CNSR'04, Fredericton, New Brunswick, Canada, 2004 (unpublished), pp. 305–314.
[23] J. A. Hartigan and M. A. Wong, Appl. Stat. **28**, 100 (1979).
[24] B. Jiang and C. Claramunt, Environ. Plann. B **31**, 151 (2004).
[25] B. Jiang, Physica A **384**, 647–655 (2007).
[26] M. Barthélemy and A. Flammini, J. Stat. Mech.: Theory Exp. (2006) L07002.