SCIENTIFIC REPORTS
nature research

OPEN

A universal opportunity model for human mobility

Er-Jian Liu[1,2] & Xiao-Yong Yan[2,3*]

Predicting human mobility between locations has practical applications in transportation science, spatial economics, sociology and many other fields. For more than 100 years, many human mobility prediction models have been proposed, among which the gravity model analogous to Newton's law of gravitation is widely used. Another classical model is the intervening opportunity (IO) model, which indicates that an individual selecting a destination is related to both the destination's opportunities and the intervening opportunities between the origin and the destination. The IO model established from the perspective of individual selection behavior has recently triggered the establishment of many new IO class models. Although these IO class models can achieve accurate prediction at specific spatiotemporal scales, an IO class model that can describe an individual's destination selection behavior at different spatiotemporal scales is still lacking. Here, we develop a universal opportunity model that considers two human behavioral tendencies: one is the exploratory tendency, and the other is the cautious tendency. Our model establishes a new framework in IO class models and covers the classical radiation model and opportunity priority selection model. Furthermore, we use various mobility data to demonstrate our model's predictive ability. The results show that our model can better predict human mobility than previous IO class models. Moreover, this model can help us better understand the underlying mechanism of the individual's destination selection behavior in different types of human mobility.

Predicting human, goods and information mobility between locations is an important topic in complex human behavior[1,2], transportation science[3,4], sociology[5], economic geography[6] and regional economics[7–9], and it also has practical applications in urban planning[10,11], population migration[12], cargo transportation[13], traffic engineering[14], infectious disease epidemiology[15–18] and emergency management[19–21]. For more than 100 years, researchers have proposed a variety of models for predicting the mobility of people between locations. The most influential model is the gravity model, which is analogous to Newton's law of gravitation, i.e., the flow between two places is proportional to their population and decays as the power of their distance. The gravity model is simple in form and has been successfully used to predict railway freight volume[22], subway passengers[23], highway traffic flow[24], air travel[25], commuting[26] and population migration[12]. Hereafter, researchers derived the gravity model from the perspective of destination selection behavior using the theory of determining utility[27], stochastic utility[28] and game theory[29]. Another classic model that is also established from the perspective of destination selection behavior is the intervening opportunity (IO) model[30]. Different from the gravity model, the IO model takes the total number of opportunities (often proportional to population) between the origin and the destination (named intervening opportunities), instead of the actual distance between the two places, as a key factor in determining human mobility. The concept of intervening opportunities provides a new direction for constructing the human mobility prediction model[31].

Inspired by the IO model, Simini *et al.* establish a parameter-free human mobility model named the radiation model[32]. The radiation model assumes that when seeking job offers, the commuter will choose the closest workplace to his/her home, whose benefit is higher than the best offer available in his/her home county, i.e., the benefit of home is higher than the benefits of the intervening opportunities and lower than the benefit of the workplace. The radiation model can better predict the commuting behavior between counties. Some researchers improve the radiation model and propose various commuting prediction models, such as the radiation model with selection[33], generalized radiation model[34], the flow and jump model[35], travel cost optimized radiation model[36] and a cost-based radiation model[37]. Yan *et al.* propose a population-weighted opportunities (PWO) model[38] by mining human daily travel data from several cities,

[1]Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing, 100044, China. [2]Institute of Transportation System Science and Engineering, Beijing Jiaotong University, Beijing, 100044, China. [3]Complex Lab, University of Electronic Science and Technology of China, Chengdu, 611731, China. *email: yanxy@bjtu.edu.cn

such as the GPS trajectories from vehicles and call detail records from mobile phones. The PWO model assumes that the probability of an individual selecting a destination is proportional to the number of opportunities at the destination and inversely proportional to the total population at the locations whose distances to the destination are shorter than or equal to the distance from the individual's origin to the destination, which can better predict intracity trips. Yan *et al.* further combine the PWO model with the continuous-time random walks model[39] to obtain a universal model of individual and population[40], which realizes the prediction of intracity and intercity mobility patterns at both the individual and population levels. Huang *et al.* propose a novel human mobility model that can capture real-time human mobility in a sustainable and economical manner, which broadens our view.[4] Sim *et al.* establish a deliberate social tie (DST) model[41] from the perspective of social interactions. The DST model assumes that an individual seeks out social ties only with other individuals whose attribute values are higher than the attribute value of the individual and the attribute values of the intervening opportunities. Motivated by the DST model, Liu and Yan propose an opportunity priority selection (OPS) model that assumes that the destination selected by the individual is the location that presents a higher benefit than the benefit of the origin and the benefits of the intervening opportunities[42]. In general, all of the IO class models[32–38,40–42] share two common assumptions: (i) using an agent to represent all of the individuals; (ii) when selecting a destination, the agent will compare the benefits of different locations. The difference between these IO class models is that the rules for comparing benefits of different locations are different. Although the radiation class models[32–37] can accurately predict commuting behavior and other IO class models[38,40–42] can accurately predict intracity and/or intercity mobility, an IO class model that can simultaneously describe the individual's destination selection behavior an different spatiotemporal scales is still lacking.

In this paper, we propose a universal opportunity (UO) model to characterize an individual's destination selection behavior. The basic idea of the model is that when an individual selects a destination, she/he will comprehensively compare the benefits of the origin, the destination and the intervening opportunities. Furthermore, we use various mobility data sets to demonstrate the predictive power of our model. The results show that the model can accurately predict different spatiotemporal scale movements such as intracity trips, intercity travels, intercity freight, commuting, job hunting and migration. Moreover, our model can also cover the classical radiation model and OPS model, presenting a new universal framework for predicting human mobility in different scenarios.

Results

Model. We assume that when an individual chooses a destination, like the radiation model[32] and the OPS model[42], she/he first evaluates the benefit of the location's opportunities[43] where the benefit is randomly chosen from a distribution $p(z)$. After that, the individual comprehensively compares the benefits of the origin, the destination and the intervening opportunities and selects a location as the destination. To characterize the behavior of an individual comprehensive comparison of the benefits of the locations, we use two parameters α and β. Parameter α reflects the behavior of the individual's tendency to choose the destination whose benefit is higher than the benefits of the origin and the intervening opportunities. Parameter β reflects the behavior of the individual's tendency to choose the destination whose benefit is higher than the benefit of the origin, and the benefit of the origin is higher than the benefits of the intervening opportunities. According to the above assumption, the probability that location j is selected by the individual at location i is

$$Q_{ij} = \int_0^\infty \mathrm{Pr}_{m_i+\alpha\cdot s_{ij}}(z)\mathrm{Pr}_{\beta\cdot s_{ij}}(<z)\mathrm{Pr}_{m_j}(>z)\mathrm{d}z, \tag{1}$$

where m_i is the number of opportunities at location i, m_j is the number of opportunities at location j, s_{ij} is the number of intervening opportunities[30] (i.e., the sum of the number of opportunities at all locations whose distances from i are shorter than the distance from i to j), $\mathrm{Pr}_{m_i+\alpha\cdot s_{ij}}(z)$ is the probability that the maximum benefit obtained after $m_i + \alpha \cdot s_{ij}$ samplings is exactly z, $\mathrm{Pr}_{\beta\cdot s_{ij}}(<z)$ is the probability that the maximum benefit obtained after $\beta \cdot s_{ij}$ samplings is less than z, $Pr_{m_j}(>z)$ is the probability that the maximum benefit obtained after m_j samplings is greater than z, α and β are both non-negative and $\alpha + \beta \leq 1$.

Since $\mathrm{Pr}_x(<z) = p(<z)^x$, we obtain

$$\mathrm{Pr}_x(z) = \frac{\mathrm{dPr}_x(<z)}{\mathrm{d}z} = xp(<z)^{x-1}\frac{\mathrm{d}p(<z)}{\mathrm{d}z}. \tag{2}$$

Equation (1) can be written as

$$\begin{aligned} Q_{ij} &= \int_0^\infty \mathrm{Pr}_{m_i+\alpha\cdot s_{ij}}(z)\mathrm{Pr}_{\beta\cdot s_{ij}}(<z)\mathrm{Pr}_{m_j}(>z)\mathrm{d}z \\ &= \left(m_i + \alpha s_{ij}\right)\int_0^1 \left(p(<z)^{m_i+(\alpha+\beta)s_{ij}-1} - p(<z)^{m_i+(\alpha+\beta)s_{ij}+m_j-1}\right)\mathrm{d}p(<z) \\ &= \left(m_i + \alpha s_{ij}\right)\left(\frac{p(<z)^{m_i+(\alpha+\beta)s_{ij}}}{m_i + (\alpha+\beta)s_{ij}}\Bigg|_0^1 - \frac{p(<z)^{m_i+(\alpha+\beta)s_{ij}+m_j}}{m_i + (\alpha+\beta)s_{ij} + m_j}\Bigg|_0^1\right) \\ &= \frac{\left(m_i + \alpha s_{ij}\right)m_j}{\left[m_i + (\alpha+\beta)s_{ij}\right]\left[m_i + (\alpha+\beta)s_{ij} + m_j\right]}. \end{aligned} \tag{3}$$

Then, the probability of the individual at location i choosing location j is

$$P_{ij} = \frac{Q_{ij}}{\sum_j Q_{ij}} \propto \frac{\left(m_i + \alpha s_{ij}\right)m_j}{\left[m_i + (\alpha + \beta)s_{ij}\right]\left[m_i + (\alpha + \beta)s_{ij} + m_j\right]}. \tag{4}$$

Further, if we know the total number of individuals O_i who travel from location i, the flux T_{ij} from location i to location j can be calculated as

$$T_{ij} = O_i P_{ij} = O_i \frac{\left(m_i + \alpha s_{ij}\right)m_j \Big/ \left(\left[m_i + (\alpha + \beta)s_{ij}\right]\left[m_i + (\alpha + \beta)s_{ij} + m_j\right]\right)}{\sum_j \left(m_i + \alpha s_{ij}\right)m_j \Big/ \left(\left[m_i + (\alpha + \beta)s_{ij}\right]\left[m_i + (\alpha + \beta)s_{ij} + m_j\right]\right)}. \tag{5}$$

This is the final form of the model and we name it the universal opportunity (UO) model.

The α and β parameters in the UO model reflect the two behavioral tendencies of the individual when choosing potential destinations (where the opportunity benefit is higher than the benefit of the origin). From Eq. (3), we can see that the larger the value of parameter α, the greater the probability that distant potential destinations will be selected by the individual. We name this behavioral tendency the exploratory tendency. On the other hand, the larger the value of parameter β, the greater the probability that near potential destinations will be selected by the individual. We name this behavioral tendency the cautious tendency. We choose average travel distance and normalized entropy as two fundamental metrics to discuss the influence of two parameters α and β on individual destination selection behavior. The average travel distance reflects the bulk density of individual destination selection[44–47], and normalized entropy reflects the heterogeneity of individual destination selection[48]. As shown in Fig. 1, the two fundamental metrics have the same regularities with a change in two parameters, whether the number of destination opportunities is a uniform or random distribution. When $\alpha = 0$, $\beta = 1$, the average travel distance is the shortest, and the normalized entropy value is the smallest; when $\alpha = 0$, $\beta = 0$, the average travel distance is the longest, and the normalized entropy value is the largest. From the definitions of the two parameters, we can easily explain the reasons for the regularities. When α is closer to 0, β is closer to 1, the individual is more cautious, and the probability of choosing near potential destinations is higher, so the shorter the average travel distance and the stronger the heterogeneity. When α is closer to 1, β is closer to 0, the individual is more exploratory, and the probability of choosing distant potential destinations is higher, so the average distance is increased while the heterogeneity is decreased. When α and β are both closer to 0, the individual attaches more importance to the benefit that the location brings to him/her and does not care about the order of locations, so the longer the average travel distance and the stronger the homogeneity.

Moreover, when α and β take extreme values (i.e., the three vertices of the triangle in Fig. 1), we can derive three special human mobility models. When $\alpha = 0$, $\beta = 0$, we name this model the opportunity only (OO) model (see details in Supplementary Information, The derivation of the OO model). In this model, the individual chooses the location whose benefit is higher than the benefit of the origin. Then, the probability of the individual at location i choosing location j as the destination is

$$P_{ij} = \frac{m_j \Big/ \left(m_i + m_j\right)}{\sum_j m_j \Big/ \left(m_i + m_j\right)} \propto \frac{m_j}{m_i + m_j}. \tag{6}$$

When $\alpha = 1$, $\beta = 0$, our model can be simplified to the OPS model, in which the individual chooses the location whose benefit is higher than the benefit of the origin and the benefits of the intervening opportunities (see details in Supplementary Information, The derivation of the OPS model). Then, the probability of the individual at location i choosing location j as the destination is

$$P_{ij} = \frac{m_j \Big/ \left(m_i + s_{ij} + m_j\right)}{\sum_j m_j \Big/ \left(m_i + s_{ij} + m_j\right)} \propto \frac{m_j}{m_i + s_{ij} + m_j}. \tag{7}$$

When $\alpha = 0$, $\beta = 1$, our model can be simplified to the radiation model, in which the individual chooses the location whose benefit is higher than the benefit of the origin and the benefits of the intervening opportunities are lower than the benefit of the origin (see details in Supplementary Information, The derivation of the radiation model). Then, the probability of the individual at location i choosing location j as the destination is

$$P_{ij} = \frac{m_i m_j \Big/ \left[\left(m_i + s_{ij}\right)\left(m_i + s_{ij} + m_j\right)\right]}{\sum_j m_i m_j \Big/ \left[\left(m_i + s_{ij}\right)\left(m_i + s_{ij} + m_j\right)\right]} \propto \frac{m_i m_j}{\left(m_i + s_{ij}\right)\left(m_i + s_{ij} + m_j\right)}. \tag{8}$$

From Eqs. (6)–(8), we can see that the OO model, the OPS model and the radiation model are all special cases of our UO model.

Prediction. We use fourteen empirical data sets, including commuting trips between United States' counties (USC), commuting trips between the provinces of Italy (ITC), commuting trips between the subregions of Hungary(HUC), freight between Chinese cities (CNF), internal job hunting in China (CNJ), internal migrations in the US (USM), intercity travels in China (CNT), intercity travels in the US (UST), intercity travels in Belgium (BLT), intracity trips in Suzhou (SZT), intracity trips in Beijing(BJT), intracity trips in Shenzhen (SHT), intracity

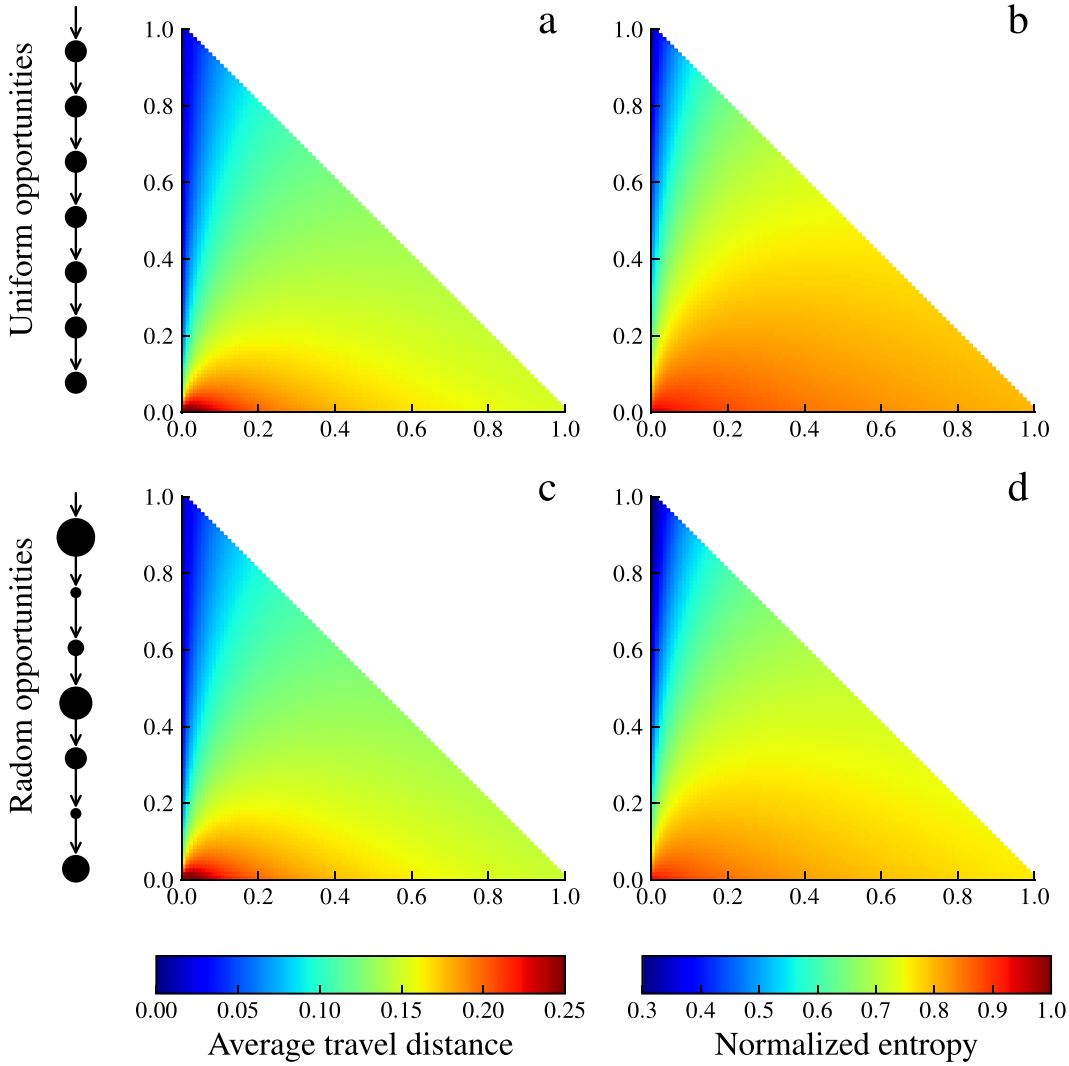

Figure 1. Average travel distance and normalized entropy versus different parameter combinations. (**a**,**b**) Average travel distance and normalized entropy values corresponding to different parameter combinations. Here, the number of destination opportunities is a uniform distribution. (**c**,**d**) Same average travel distance and normalized entropy values as in (**a**,**b**), but the number of destination opportunities is a random distribution.

trips in London (LOT) and intracity trips in Berlin (BET) (see **Methods**), to validate the predictive ability of the UO model. We first extract the flux T_{ij} from location i to location j from the data set and obtain the real mobility matrix. Then, we exploit the Sørensen similarity index[38] (SSI, see **Methods**) to calculate the similarity between the real mobility matrix and the mobility matrix predicted by the UO model under different parameter combinations. The results are shown in Fig. 2. Figure 2o shows the optimal values of the parameter α and β corresponding to the highest SSI for the fourteen data sets.

It can be seen from Fig. 2a–d that for USC, ITC, HUC and CNF, when α is close to 0 and β is close to 1, the SSI is relatively large. The reason is that for commuting data sets (USC, ITC and HUC), the commuting distance or time is very important for commuters. As a result, most people tend to choose near potential destinations when finding a job based on their place of residence or adjusting their place of residence after finding a job. This cautious destination selection tendency also exists in freight. Freight to far destinations will lead to an increase in transportation costs and a decrease in the freight frequency, which will have a negative impact on freight revenue. Thus, unless the destination opportunity benefit is very high, the individual tends to choose a near destination rather than a far destination for freight. For the migration and job hunting data sets (USM and CNJ), when α is close to 1 and β is close to 0, the SSI is relatively large, as shown in Fig. 2e,f. The reason is that both job seekers and migrants pay more attention to the destination opportunity benefit rather than the distance to the destination. In other words, they are more exploratory but less cautious. Even if a high benefit destination is far away, it will still be selected by individuals with a relatively high probability. The reason is that the distance to the destination has a smaller impact on long temporal scale mobility behaviors, such as migration and job hunting, than on daily commuting behaviors. For intercity travel data sets (CNT, UST and BLT), when α and β are both near the middle of

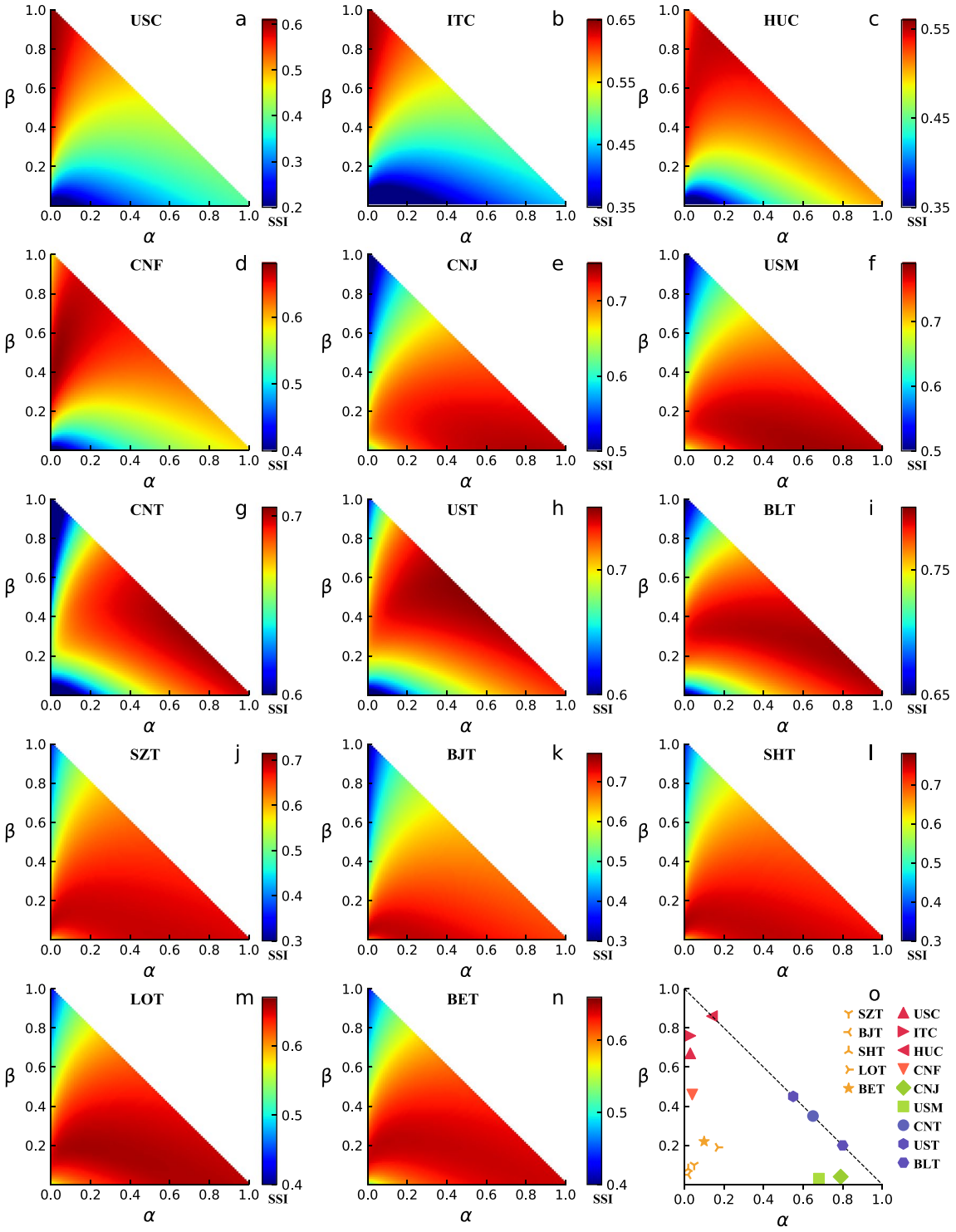

Figure 2. Results for empirical data sets. (**a**–**n**) We exploit SSI to calculate the similarity between the real mobility matrix and the predicted mobility matrix under different parameter combinations for the fourteen data sets. Here, the color bar represents the SSI, where a dark red (blue) dot indicates a higher (lower) SSI. (**o**) The optimal values of the parameters α and β correspond to the highest SSI for the fourteen data sets.

the diagonal line of the triangle, the SSI is relatively large, as shown in Fig. 2g–i. For most people, intercity travel is occasional and not as frequent as commuting. Travelers are less inclined than commuters to choose near potential destinations but they tend to explore distant potential destinations. Thus, the exploratory tendency parameter α of intercity travels is much larger than that of commuting. On the other hand, the importance of the travel cost of intercity travels is higher than that of the cost of migration. Thus, the cautious tendency parameter β of intercity travels is larger than that of migration. For intracity trips data sets (SZT, BJT, SHT, LOT and BET), when α and β

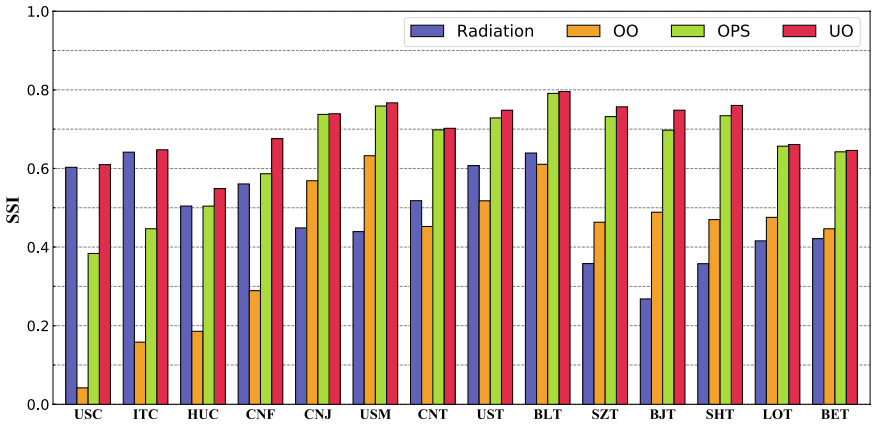

Figure 3. Comparing predicting accuracy of the UO model, the radiation model, the OPS model and the OO model in terms of SSI.

Data set	SSI-UO	SSI-RM	SSI-OPS	SSI-OO	RMSE-UO	RMSE-RM	RMSE-OPS	RMSE-OO
USC	0.610	0.603	0.384	0.042	2158.766	2308.054	2948.205	3654.402
ITC	0.648	0.641	0.447	0.158	1600.862	1696.488	2132.627	3033.019
HUC	0.549	0.504	0.504	0.186	477.254	546.878	429.377	612.904
CNF	0.676	0.561	0.587	0.289	111.201	184.724	128.789	183.655
CNJ	0.739	0.449	0.738	0.567	185.709	481.072	189.816	297.379
USM	0.767	0.434	0.759	0.632	1126.110	3275.661	1218.255	1521.585
CNT	0.702	0.518	0.698	0.452	441.063	829.869	438.463	731.153
UST	0.748	0.607	0.729	0.518	55.851	95.013	65.513	115.795
BLT	0.796	0.639	0.791	0.611	26.236	58.641	26.339	48.080
SZT	0.757	0.358	0.732	0.463	7.871	47.801	9.133	12.553
BJT	0.748	0.268	0.697	0.489	6.567	68.039	12.291	12.040
SHT	0.760	0.358	0.734	0.470	48.196	368.901	71.152	91.000
LOT	0.661	0.416	0.657	0.476	4.309	20.031	4.603	8.104
BET	0.646	0.421	0.642	0.447	3.288	11.271	3.356	5.323

Table 1. Comparison of models prediction accuracy. SSI is the Sørensen similarity index between the real mobility matrix and the mobility matrix predicted by different models. RMSE is the root mean square error of predicted mobility matrix. UO, RM, OPS, and OO stand for the universal opportunity model, the radiation model, the opportunity priority selection model and the opportunity only model, respectively.

are both close to 0, the SSI is relatively large, as shown in Fig. 2j–n. The reason is that compared with the intercity mobility behavior on a large spatial scale, the spatial scale of intracity mobility behavior is small. In this scenario, the individual is not necessarily concerned about the travel distance and focuses more on the benefit that the location will directly bring to him/her. Thus, the optimal values of α and β are both close to 0, as shown in Fig. 2o.

We next compare the predictive accuracy of the mobility fluxes of the UO model with the radiation model, the OPS model and the OO model. In terms of SSI, as shown in Fig. 3 and Table 1, the UO model performs best. However, the radiation model and the OPS model can provide only relatively accurate predictions for some data sets. For example, the radiation model can predict commuting and freight trips relatively accurately but cannot accurately predict other types of mobility. The reason is that the individual tends to choose near potential destinations rather than distant potential destinations in commuting and freight, where travel costs are more important. From Fig. 2o, we can see that for commuting and freight data sets, the optimal parameter β (which reflects the individual's cautious tendency) of the UO model is close to 1, and the optimal parameter α (which reflects the individual's exploratory tendency) is close to 0. Therefore, the prediction accuracy of the radiation model in which the individual only chooses the closest potential destination (i.e., $\alpha = 0$, $\beta = 1$) is close to that of the UO model in commuting and freight data sets. However, the prediction accuracy of the radiation model is considerably lower than that of the UO model in job hunting, migration and noncommuting travel data sets. The reason is that the individual is more likely to choose distant potential destinations in these data sets. In these cases, the prediction accuracy of the OPS model, in which the individual tends to choose distant potential destinations, is closer to that of the UO model. We further measure the fluxes predicted by different models compared with the real fluxes and find that the average fluxes predicted by our model are more in agreement with real observations than the other three models (see details in Supplementary Information, Comparison among different models). We also use a frequently used statistical index, named the root mean square error (RMSE), to measure the prediction errors of

the UO model and the other three models, and Table 1 lists the results. From the table, we can see that in most cases, the RMSE of the UO model is smaller than that of the other benchmark models, although the RMSE is not the parameter optimization objective of the UO model. These results prove that the three models only capture the individual's destination selection behavior at a specific spatiotemporal scale. Yet our UO model can accurately describe the individual's destination selection behavior at different spatiotemporal scales.

Discussion

Although previous IO class models are widely used to predict the mobility of people between locations[32–38,40–42], these models can only achieve accurate prediction at specific spatiotemporal scales. In this paper, we developed a UO model to predict human mobility at different spatiotemporal scales. Our model establishes a new framework in IO class models and covers the classical radiation model[32] and the OPS model[42]. Although the UO model has two parameters, they are different from the parameters in some regression analysis models or machine learning models in the sense that they simply improve the prediction accuracy of the model. These two parameters essentially describe the two tendencies, i.e., exploratory tendency and cautious tendency, of an individual's destination selection behavior. They not only enable the UO model to better predict human mobility at different spatiotemporal scales than the parameter-free models but also help us better understand the underlying mechanism of the individual's destination selection behavior in different types of human mobility.

Many phenomena in complex system field are strongly related to human mobility[31]. For example, the spread of disease is directly affected by human travel distance between locations and the population size of locations[15–17,49–52]. The UO model can accurately describe the individual's destination selection behavior at different spatiotemporal scales, which has potential applications for understanding the spread of disease within humans. Not only that, but the IO model can also describe an individual's selection behavior in social networks such as friend networks and scientific collaboration networks. In friend networks, the individual tends to choose friends who are close to him/her and have a high sense of identity[41,53]. In scientific collaboration networks, the individual tends to choose nearby scholars who have high scientific influence[54]. These phenomena indicate that when one seeks to build beneficial ties, she/he will take into account both the distance and the benefits of the opportunities. The UO model can describe the individual's interactive object selection behavior, providing a new perspective for social network analysis.

Despite its fine performance in predicting human mobility, the UO model has room for further improvements. For example, most existing IO class models use an agent to represent all of the individuals and neglect the diversity of individual selection behavior[46,55–59]. Building mobility prediction model for each individual may reflect the diversity in detail. However, it is extremely cumbersome and cannot grasp the commonality among individuals' mobility patterns. One possible approach is first clustering individuals according to their mobility behavior characteristics[60–62], then expanding our UO model for different classes of individuals, which may more accurately predict human mobility.

Methods

Material and methods. *Data sets.*

(1) Commuting trips. The commuting trips data sets include the commuting trips between United States' counties[32] (USC), the commuting trips between the provinces of Italy[35] (ITC) and the commuting trips between the subregions of Hungary[35] (HUC), which were downloaded from http://www.census.gov/population/www/cen2000/com-muting/index.html, http://www.stat.it/storage/cartografia/matrici_pendolarismo/matrici_pendolarismo_2011.zip and http://www.ksh.hu, respectively. Since we focus on mobility among zones(counties, provinces or subregions), all the residences/workplaces within a zone are regarded as the same with an identical zone label. Then, we can accumulate the total number T_{ij} of trips from zone i to zone j, which is also carried out in the following data sets.

(2) Freight between Chinese cities (CNF). The CNF data set is extracted from the travel records of freight between Chinese cities from 19 May 2015 to 23 May 2015. When freight is loaded or unloaded, the coordinates and time are recorded automatically by a GPS-based device installed in the truck. All the loading/unloading locations within a city are regarded as the same with an identical zone label.

(3) Internal job hunting in China (CNJ). The CNJ data set is extracted from more than 160 million job hunters' resumes from 2006 to 2016 and was downloaded from https://www.zhaopin.com. The resumes contain job hunter work experience, from which we can obtain a job hunter's former workplaces. All the workplaces within a city are regarded as the same with an identical zone label.

(4) Internal migrations in the US (USM). The USM data set is extracted from the Statistics of Income Division of the Internal Revenue Service (IRS) in the US from 2011 to 2012 and was downloaded from https://www.irs.gov/statistics/soi-tax-stats-migration-data. The IRS contains records of all individual income tax forms filed in each year, from which we can determine who has or has not, moved residence/workplace locations in the intervening fiscal year[31]. All the residence/workplace locations within a state are regarded as the same with an identical zone label.

(5) Intercity travels. The intercity travels data sets include intercity travels in China (CNT), intercity travels in the US (UST) and intercity travels in Belgium (BLT). The CNT data set is extracted from check-in records of the Sina Weibo website for users in mainland China[40]. The UST data set is extracted from check-in records of the Foursquare website for users in the continental US[63]. The BLT data set is extracted from check-in records of the website Gowalla for users in Belgium[64]. These data sets contain each user's spatial and temporal information, from which we can obtain the user's location. All the check-in locations within a city are regarded as the same with an identical zone label.

(6) Intracity trips. The intracity trips data sets include intracity trips in Suzhou (SZT), intracity trips in Beijing (BJT), intracity trips in Shenzhen (SHT), intracity trips in London (LOT) and intracity trips in Berlin (BET). The SZT data set is extracted from the mobile phone call detail records in Suzhou, a city of China. The data contains the time and positions of users making phone calls or sending text messages. The BJT data set[65] and the SHT data set[65] are extracted from the travel records of taxi passengers in Beijing and Shenzhen, respectively. When a passenger gets on or gets off a taxi, the coordinates and time are recorded automatically by a GPS-based device installed in the taxi. The LOT data set[64] and the BET data[64] set are extracted from checkin records at Gowalla in London and Berlin. Because of the absence of natural partitions in cities (in contrast to states or counties), the city is divided into zones, each of which is 1 km $\times$ 1 km (for SZT is 0.01 longitude $\times$ 0.01 latitude). All the locations within a zone are regarded as the same with an identical zone label[38].

Normalized entropy. We use normalized entropy to reflect the heterogeneity of individual destination selection

$$E_i = \frac{-\sum_{j=1}^{N} p_{ij} \log\left(p_{ij}\right)}{\log\left(p_{ij}\right)}, \tag{9}$$

where E_i is the normalized entropy of location i, p_{ij} is the probability that the individual at location i chooses location j as his/her destination, and N is the number of locations.

Sørensen similarity index. The Sørensen similarity index[66] is a similarity measure between two samples. Here, we apply a modified version[38] of the index to measure whether real fluxes are correctly reproduced (on average) by theoretical models, defined as

$$\mathrm{SSI} = \frac{1}{N(N-1)} \sum_{i}^{N} \sum_{j \neq i}^{N} \frac{2 \min\left(T_{ij}, T'_{ij}\right)}{T_{ij} + T'_{ij}} \tag{10}$$

where N is the number of locations, T_{ij} is the predicted flux from location i to j and T'_{ij} is the empirical flux. Obviously, if each T_{ij} is equal to T'_{ij} the index is 1, and if all T_{ij} are far from the real values, the index is close to 0.

Data availability

Data available on request from the authors.

Received: 6 December 2019; Accepted: 1 March 2020;
Published online: 13 March 2020

References

1. Barthelemy, M. The statistical physics of cities. *Nat. Rev. Phys* **1**, 406–415 (2019).
2. Barthelemy, M. Spatial Networks. *Phys. Rep.* **499**, 1–101 (2011).
3. Ortúzar, J. D. & Willumsen, L. G. *Modelling transport*. (John Wiley & Sons, New York, 2011).
4. Huang, Z. *et al*. Modeling real-time human mobility based on mobile phone and transportation data fusion. *Transport. Res. Part C.* **96**, 251–269 (2018).
5. Huberman, B. A., Romero, D. M. & Wu, F. Social networks that matter: Twitter under the microscope. arXiv:0812.1045 (2009).
6. Roy, J. R. & Thill, J. C. Spatial interaction modelling. *Pap. Reg. Sci.* **83**, 339–361 (2003).
7. Lenormand, M. *et al*. Multiscale socio-ecological networks in the age of information. *PLoS One* **13**, e0206672 (2018).
8. Karemera, D., Oguledo, V. I. & Davis, B. A gravity model analysis of international migration to North America. *Appl. Econ.* **32**, 1745–1755 (2000).
9. Patuelli, R., Reggiani, A., Gorman, S. P., Nijkamp, P. & Bade, F. J. Network analysis of commuting flows: A comparative static approach to German data. *Netw. Spat. Econ.* **7**, 315–331 (2007).
10. Batty, M. The size, scale, and shape of cities. *Science* **319**, 769–771 (2008).
11. Li, R. *et al*. Simple spatial scaling rules behind complex cities. *Nat. Commun.* **8**, 1841 (2017).
12. Tobler, W. Migration: Ravenstein, thornthwaite, and beyond. *Urban Geogr.* **16**, 327–343 (1995).
13. Kaluza, P., Kölzsch, A., Gastner, M. T. & Blasius, B. The complex network of global cargo ship movements. *J. R. Soc. Interface* **7**, 1093–1103 (2010).
14. Helbing, D. Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.* **73**, 1067 (2001).
15. Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci. USA* **101**, 15124–15129 (2004).
16. Eubank, S. *et al*. Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180 (2004).
17. Balcan, D. *et al*. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA* **106**, 21484–21489 (2009).
18. Wang, P., González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding the spreading patterns of mobile phone viruses. *Science* **324**, 1071–1076 (2009).
19. Bagrow, J. P., Wang, D. & Barabasi, A. L. Collective response of human populations to large-scale emergencies. *PLoS One* **6**, e17680 (2018).
20. Lu, X., Bengtsson, L. & Holme, P. Predictability of population displacement after the 2010 Haiti earthquake. *Proc. Natl. Acad. Sci. USA* **109**, 11576–11581 (2012).
21. Rutherford, A. *et al*. Limits of social mobilization. Proceedings of the National Academy of Sciences. *Proc. Natl. Acad. Sci. USA* **110**, 6281–6286 (2013).
22. Zipf, G. K. The P_1P_2/D hypothesis: On the intercity movement of persons. *Am. Sociol. Rev.* **11**, 677–686 (1946).
23. Goh, S., Lee, K., Park, J. S. & Choi, M. Y. Modification of the gravity model and application to the metropolitan Seoul subway system. *Phys. Rev. E* **86**, 026102 (2012).

24. Jung, W. S., Wang, F. & Stanley, H. E. Gravity model in the Korean highway. *EPL* **81**, 48005 (2008).
25. Grosche, T., Rothlauf, F. & Heinzl, A. Gravity models for airline passenger volume estimation. *J. Air Transp. Manag.* **13**, 175–183 (2007).
26. Viboud, C. *et al*. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
27. Niedercorn, J. H. & Bechdolt, B. V. Jr. An economic derivation of the "gravity law" of spatial interaction. *J. Regional Sci.* **9**, 273–282 (1969).
28. Domencich, T. A. & Mcfadden, D. *Urban travel demand: A behavioral analysis*. (North-Holland, Amsterdam, 1975).
29. Yan, X. Y. & Zhou, T. Destination choice game: A spatial interaction theory on human mobility. *Sci. Rep.* **9**, 1–9 (2019).
30. Stouffer, S. A. Intervening opportunities: A theory relating mobility and distance. *Am. Sociol. Rev.* **5**, 845–867 (1940).
31. Barbosa-Filho, H. *et al*. Human mobility: Models and applications. *Phys. Rep.* **734**, 1–74 (2018).
32. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
33. Simini, F., Maritan, A. & Néda, Z. Human mobility in a continuum approach. *PLoS One* **8**, e60069 (2013).
34. Kang, C., Liu, Y., Guo, D. & Qin, K. A generalized radiation model for human mobility: Spatial scale, searching direction and trip constraint. *PLoS One* **10**, e0143500 (2015).
35. Varga, L., Tóth, G. & Néda, Z. Commuting patterns: The flow and jump model and supporting data. *EPJ Data Sci.* **7**, 37 (2018).
36. Varga, L., Tóth, G. & Néda, Z. An improved radiation model and its applicability for understanding commuting patterns in Hungary. *Regional Statistics* **6**, 27–38 (2016).
37. Ren, Y., Ercsey-Ravasz, M., Wang, P., Gonzáles, M. C. & Toroczkai, Z. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nat. Commun.* **5**, 5347 (2014).
38. Yan, X.-Y., Zhao, C., Fan, Y., Di, Z.-R. & Wang, W.-X. Universal predictability of mobility patterns in cities. *J. R. Soc. Interface* **11**, 20140834 (2014).
39. Montroll, E. W. & Weiss, G. H. Random walks on lattices. *J. Math. Phys.* **6**, 167–181 (1965).
40. Yan, X.-Y., Wang, W.-X., Gao, Z.-Y. & Lai, Y.-C. Universal model of individual and population mobility on diverse spatial scales. *Nat. Commun.* **8**, 1639 (2017).
41. Sim, A., Yaliraki, S. N., Barahona, M. & Stumpf, M. P. Great cities look small. *J. R. Soc. Interface* **12**, 20150315 (2015).
42. Liu, E. & Yan, X. New parameter-free mobility model: Opportunity priority selection model. *Physica A* **526**, 121023 (2019).
43. Pan, W., Ghoshal, G., Krumme, C., Cebrian, M. & Pentland, A. Urban characteristics attributable to density-driven tie formation. *Nat. Commun.* **4**, 1961 (2013).
44. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
45. González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
46. Yan, X.-Y., Han, X.-P., Wang, B.-H. & Zhou, T. Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Sci. Rep.* **3**, 2678 (2013).
47. Roth, C., Kang, S. M., Batty, M. & Barthélemy, M. Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PLoS One* **6**, e15923 (2011).
48. Eagle, N., Macy, M. & Claxton, R. Network diversity and economic development. *Science* **328**, 1029–1031 (2010).
49. Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl. Acad. Sci. USA* **109**, 11157–11162 (2005).
50. Kramer, A. M. *et al*. Spatial spread of the West Africa Ebola epidemic. *R. Soc. Open Sci.* **3**, 160294 (2016).
51. Kitsak, M. *et al*. Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888 (2010).
52. Viboud, C. *et al*. Synchrony, waves and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
53. Illenberger, J., Nagel, K. & Flötteröd, G. The role of spatial interaction in social networks. *Netw. Spat. Econ.* **13**, 255–282 (2013).
54. Pan, R. K., Kaski, K. & Fortunato, S. World citation and collaboration networks: Uncovering the role of geography in science. *Sci. Rep.* **2**, 902 (2012).
55. Song, C., Qu, Z., Blumm, N. & Barabási, A. L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).
56. Bagrow, J. P. & Lin, Y. R. Mesoscopic structure and social aspects of human mobility. *PLoS One* **7**, e37676 (2012).
57. Pappalardo, L. *et al*. Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**, 8166 (2015).
58. Gallotti, R., Bazzani, A., Rambaldi, S. & Barthelemy, M. A stochastic model of randomly accelerated walkers for human mobility. *Nat. Commun.* **7**, 12600 (2016).
59. Zhao, Y. M., Zeng, A., Yan, X. Y., Wang, W. X. & Lai, Y. C. Unified underpinning of human mobility in the real world and cyberspace. *New J. Phys.* **18**, 053025 (2016).
60. Louail, T. *et al*. Uncovering the spatial structure of mobility networks. *Nat. Commun.* **6**, 6007 (2015).
61. Lian, D. & Xie, X. Learning location naming from user check-in histories. *In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* 112–121 (2011).
62. Lian, D. et al. GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation. *In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* 831–840 (2014).
63. Levandoski, J. J., Sarwat, M., Eldawy, A. & Mokbel, M. F. Lars: A location-aware recommender system. *In 2012 IEEE 28th international conference on data engineering* 450–461 (2012).
64. Cho, E., Myers, S. A. & Leskovec, J. Friendship and mobility: User movement in location-based social networks. *In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* 1082–1090 (2011).
65. Liang, X., Zheng, X., Lv, W., Zhu, T. & Xu, K. The scaling of human mobility by taxis is exponential. *Physica A* **391**, 2135–2144 (2012).
66. Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.* **5**, 1–34 (1948).

Acknowledgements

X.-Y.Y. was supported by NSFC under grant nos. 71822102, 71621001 and 71671015.

Author contributions

X.-Y.Y. designed the research; X.-Y.Y. and E.-J.L. performed the research; E.-J.L. analysed the empirical data; and E.-J.L. and X.-Y.Y. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-61613-y.

Correspondence and requests for materials should be addressed to X.-Y.Y.