

# Emergence of Urban Growth Patterns from Human Movements

Fengli Xu

Tsinghua University

Yong Li (✉ [liyong07@tsinghua.edu.cn](mailto:liyong07@tsinghua.edu.cn))

Tsinghua University

Chaoming Song

University of Miami

---

## Article

**Keywords:** Applied Physics, Statistical Physics, Social Couplings, Long-memory Effects

**Posted Date:** September 23rd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-79579/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Emergence of Urban Growth Patterns from Human Movements

Fengli Xu<sup>a</sup>, Yong Li<sup>a\*</sup> & Chaoming Song<sup>b\*</sup>

*a) Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing, China*

*b) Department of Physics, University of Miami, Coral Gables, FL, USA*

**Cities grow in a bottom-up manner, leading to fractal-like urban morphology characterized by scaling laws. Correlated percolation has succeeded in modeling urban geometries by imposing strong spatial correlations. However, the origin of such correlations remains largely unknown. Very recently, our understanding of human movements has been revolutionized thanks to the increasing availability of large-scale human mobility data. This paper proposes a novel human movement model that offers a micro-foundation for the dynamics of urban growth. We compare the proposed model with three empirical datasets, which evidences that strong social couplings and long-memory effects are two fundamental principles responsible for the mystical spatial correlations. The model accounts for the empirically observed scaling laws, but also allows us to understand the city evolution dynamically.**

Over a century ago, urban scientists envisioned the future cities being perfectly symmetric as a result of well designed top-down city planning strategies <sup>1</sup>. However, the increasingly available

---

<sup>\*</sup>Fengli Xu and Yong Li offer the empirical motivations and conduct the empirical analysis. Chaoming Song contributes to the theoretical analysis in this study. To whom correspondence should be addressed. E-mail: c.song@miami.edu, liyong07@tsinghua.edu.cn

urban data suggests that cities grow in a bottom-up manner, calling for understandings of its micro foundation <sup>2-4</sup>. Later, three fundamental empirical laws have been discovered <sup>2,5,6</sup>: First, the distribution of city size follows a power law with a scaling exponent around two, implying large cities are much rarer than small towns <sup>5</sup>. Second, the urban area grows super-linearly with the population, due to intense competition for spaces in previous studies <sup>7,8</sup>. Finally, the density of occupied urban areas decreases exponentially with the radial distance to city centers <sup>9-11</sup>. Physicists have applied diffusion-limited aggregation (DLA) to model urban growth as an aggregation of physical particles <sup>7,12</sup>. Further works showed that correlated percolation (CP) is a better alternative to explain the emergence of the aforementioned laws <sup>5</sup>. A key observation of the CP model is the requirement of strong geographical correlation to reproduce the correct scaling relations <sup>13</sup>. While the CP model successfully explains the urban morphology, it has little connections with human activities at the micro-level. The micro-foundation of such geographical correlation remains a mystery. Here, we develop a novel urban growth model based on human movements, suggesting that strong social coupling and long memory are two fundamental principles governing urban growth.

Thanks to the availability of large-scale movement datasets, our understanding of human movements has been revolutionized over the past decade <sup>14-16</sup>. Existing human movement models fall into three classes, as depicted in Fig. 1: *Class A* models treat human movements as randomly moving particles without interactions. *Brownian movement* is one of such prototype models where an individual's displacements are normal-distributed <sup>17</sup>. Unlike physical particles, empirical data suggests human movements are characterized by large jumps between two consecutive steps, sat-

isfying a power law,

$$P(\vec{r}|\vec{r}') \sim \frac{1}{|\vec{r} - \vec{r}'|^{(d+\alpha)}}, \quad (1)$$

where  $P(\vec{r}|\vec{r}')$  is the transition probability from location  $\vec{r}'$  to  $\vec{r}$ , with  $d = 2$  for two-dimensional space<sup>14,18</sup>. The exponent  $\alpha$  is observed around  $0.55 \pm 0.05$ <sup>19</sup>. The fact that the transition probability decreases with distance characterizes the *cost of travel distance* of human movements, *i.e.*, most of the time people travel only over short distances, whereas occasionally people take longer trips. Neglecting social interactions and memory effects, Eq. (1) suggests human movements follow a *Lévy-flight*, with the population density  $\rho(\vec{r})$  satisfying the fractional diffusion equation,

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} = -D(-\Delta)^{\alpha/2} \rho(\vec{r}, t), \quad (2)$$

where  $D$  is the diffusion constant (see Method section for details). Nevertheless, both *Brownian motion* and *Lévy flight* predict a uniform population distribution when time  $t$  approaches infinity, in contrast to empirical observations<sup>9</sup>.

*Class B* models such as *Gravity model*<sup>8</sup> and *Radiation model*<sup>20</sup>, originate from the study of migrations, where the traffic flow between two locations depends on their populations. For instance, the *Gravity model* suggests the transition probability,

$$P(\vec{r}|\vec{r}') \sim \frac{\rho(\vec{r}) + \rho_0}{|\vec{r} - \vec{r}'|^{(d+\alpha)}}, \quad (3)$$

where  $\rho_0$  is the (inverse) coupling constant. In addition to the fat-tailed jump size distribution (1), the *gravity model* (2) also requires the transition probability increases linearly with the population at the destination  $\vec{r}$ <sup>21</sup>. This mechanism accounts for a mean-field background attractiveness rooted

in *social interaction*, e.g., highly populated locations often offer more social opportunities<sup>8</sup>. One would hope that this social attractiveness being responsible for the mystical geographical correlation in the CP model. Unfortunately, we find that the diffusion process of the *gravity model* follows the same fractional diffusion Eq. (2) for *Lévy flights* (see Method section for details), i.e., it predicts a uniformly distributed urban patterns at a large  $t$ .

*Class C* models have been developed during the recent study of human mobility. Unlike *Class A* where individuals move freely, empirical data found notable recurrent-visitation patterns in human movements. Consequently, individuals show an ultra-slow diffusion, in contrast to a regular power-law diffusion in the *Brownian motion* and *Lévy flight*. To explain these new findings, *Individual mobility model (IMM)* retreats human movements as a two-stage return-exploration process to account for *long memory effects*. In particular, a *preferential return mechanism* is imposed, i.e., the probability returning to a previous location  $\vec{r}_i$ ,

$$P(r_i) \propto f(r_i), \quad (4)$$

proportional to its historic visitation frequency  $f(r_i)$ . Such long-memory return process slows the human diffusion drastically. In particular, *IMM* predicts that the traveling distance  $l$  follows,

$$l \sim \log A, \quad (5)$$

where  $A$  is the total visitation area<sup>22</sup>. The logarithmic growth is one of the key ingredients for human movements, characterizing the anomalous ultra-slow diffusion and home range effect<sup>23</sup>. Unfortunately, since *IMM* only models individual movements without involving social interactions, it fails to capture urban growth patterns (see Result Section).

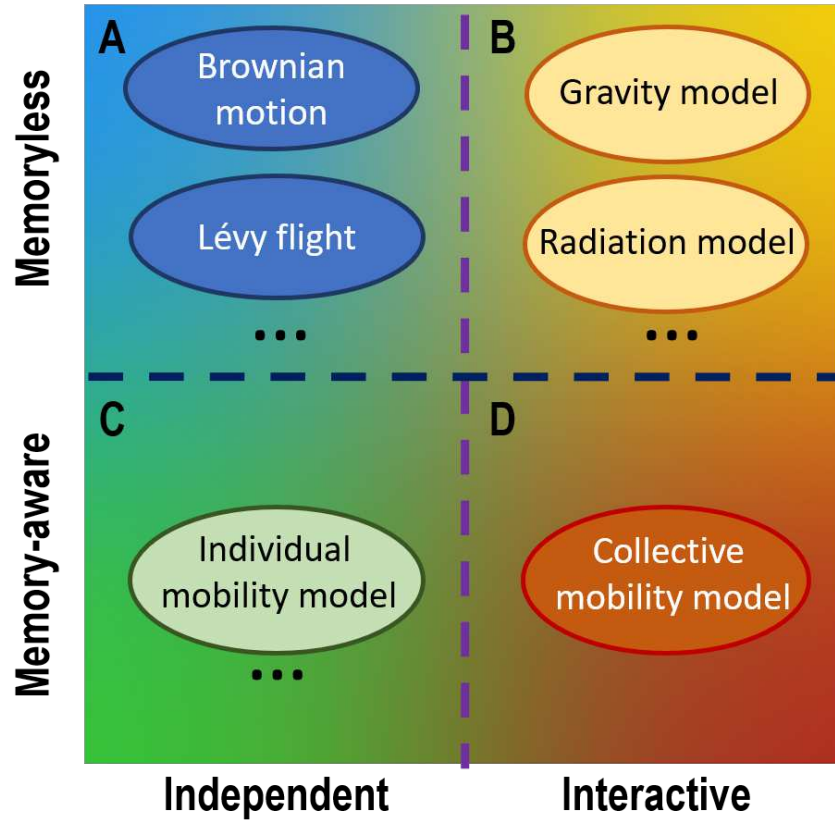


Figure 1: **The paradigm of human movement models.** Existing human movement models classified based on whether account for the memory of historic movements or the social interactions, are summarized as the paradigm with four classes: (A) *Brownian motion* and *Lévy-flight* belong to this class where movements are *independent* and *memoryless*; (B) *Gravity model* and *Radiation model* are the typical models where movements are *socially correlated* and *memoryless*; (C) *Individual mobility model* belongs to this class where movements are *independent* and *memory-aware*. (D) Missing corner of the current paradigm. The proposed *Collective Mobility Model* fills in this class where movements are both *socially correlated* and *memory-aware*.

In summary, neither social interactions (class B) nor memory effects (Class C) alone reproduce the geographical correlation required in the CP model. However, there is a missing corner (Class D) in the current paradigm of human movement models (Fig. 1) that integrates social interactions and long-memory simultaneously. It is rather curious to ask if these new class models can predict urban growth patterns. To fill the gap, we propose a *Collective Mobility Model (CMM)* that accounts in a minimal fashion both for memory effects and for interactions. Consider a  $N$ -individual migration process in a  $L \times L$  square lattice. Each individual follows a return-exploration process. Like *IMM*, an individual's return probability is proportional to his/her visitation frequency (4). The probability to explore a new site, in contrast, relies on the populations, satisfying the *gravity law* (3) (see supplementary material S2 for details). The coupling constant,  $\rho_0^{-1}$ , controls the strength of population attraction, *i.e.*, increasing  $\rho_0$  reduces the impact of population density, and consequently, the strength of social interactions. For  $\rho_0^{-1} \rightarrow 0$ , *CMM* is effectively equivalent to *IMM*. Inspired by the strong geometrical correlation in CP models <sup>5</sup>, we're interested in the strong coupling limit  $\rho_0^{-1} \rightarrow \infty$ , where *CMM* describes a strongly correlated many-body system.

## Results

We collect three public available urban development datasets, including the population and urban area of cities in i) United States of America (U.S.A.) at 2000, ii) Great Britain (G.B.) at 1991, and iii) the distribution of urban area in Berlin region at 1910, 1920 and 1945. For comparison, we emulate the human migrations by four typical movement models for Classes A–D respectively, namely the *Lévy flight*, *Gravity model*, *IMM*, and the proposed *CMM*. A simple emulation of large

urban systems is impractical due to the high time complexity, which is  $\mathcal{O}(Ml^2)$  for each epoch with  $M$  and  $l$  denoting the number of citizens and the size of the urban area, respectively. We address this problem by designing improved sampling techniques to effectively reduce the complexity to  $\mathcal{O}(M)$

To compare the morphology of the emulated urban systems to the empirical observations, we plot population distributions in Fig. 2A–D for all models, together with the empirical distribution of London city in where Fig. 2E. While the real-world geometry is affected by geographical features, *e.g.*, lakes and rivers, London city still exhibits prominent features of the compact city center and fractal perimeters. These observations echo previous studies on the fractal geometry of urban area<sup>24–26</sup>.

The urban population distribution for *Lévy flight* and *Gravity model* follows the fractional diffusion process (2), implying that individuals will gradually diffuse away from their initial position over time. The emulation verifies this prediction with the urban population distributed uniformly in urban space when the systems converge(see Fig. 2A–B). It indicates these two models fail to reproduce compact and stable city centers. On the other hand, *IMM* predicts urban systems grow homogeneously in the perimeter. The emulation result shows the perimeter of the urban area is a standard circle, and the urban areas that have a similar radial distance to the city center have similar population density (see Fig. 2C), which is in consistence with the theoretical prediction. Therefore, the emulation suggests *IMM* cannot reproduce the fractal morphology of the urban area. On the contrary, *CMM* successfully reproduces the compact city center in the urban system, where the



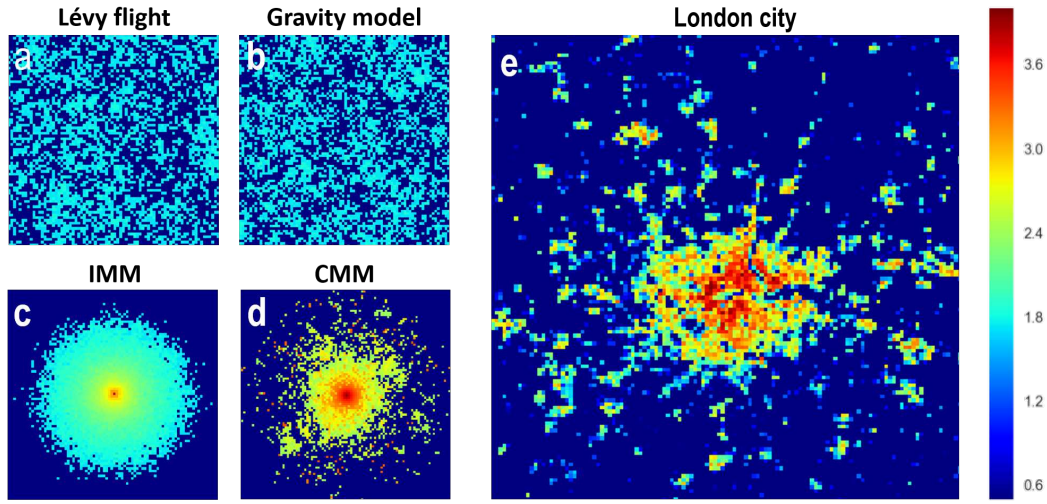


Figure 2: **The morphology of urban area generated by four different human movement models: (A) *Lévy flight*, (B) *Gravity model*, (C) *Individual mobility model (IMM)*, (D) *Collective mobility model (CMM)*, and (E) the empirical data from London city.** The population distribution of each urban system is visualized as a heatmap in log scale, where blue color represents underpopulated regions and red colors corresponding to regions of high population. The urban systems are emulated with 30,000 individuals initially situated in urban centers and then move according to mobility models until reaching a stable state. The *Lévy flight* and *Gravity model* fail to reproduce a compact urban center, while *IMM* predicts the urban system grows in a homogeneous manner, where fractal perimeter and sub-clusters are absent. *CMM* accurately reproduces the compact urban center, fractal perimeter of urban area, and sub-clusters. The morphology predicted by *CMM* is consistent with the empirical observation of London city.

population density is significantly higher than the peripheral urban area (see Fig. 2D). Besides, the perimeter of the city center demonstrates prominent fractal geometry, and numerous sub-clusters are formed around it (see Supplementary Material S4 for details). These observations are in agreement with the empirical observation on London city, which indicates *CMM* can reproduce the morphology of the urban area.

To compare models to real-world urban growth quantitatively, we will focus on three fundamental empirical laws, each of which has been validated on multiple cities around the globe<sup>2,5,6</sup>:

(A) *City size distribution*: The number of cities  $N(A)$  decreases with their areas  $A$ , following a power law,

$$N(A) \sim A^{-\tau}, \quad (6)$$

where the exponent  $\tau$  has been reported around 2.0<sup>5</sup>. Percolation theory is the prevalent narrative for this observation, with each site occupied as an urban area with a certain probability<sup>13</sup>. It predicts the scaling law (6) with the exponent ranging between 2 and 2.5, where  $\tau = 2$  corresponds to a strong correlation between different sites and  $\tau = 2.5$  corresponds to a mean-field theory<sup>13</sup>. The empirical urban datasets show that city size distributions are well approximated by Eq. (6) (see Fig. 3A), with  $\tau = 2.09 \pm 0.09$  for U.S.A.,  $\tau = 2.01 \pm 0.08$  for G.B. and  $\tau = 1.91 \pm 0.16$  for Berlin. These findings echo the theoretical predictions of site percolation theory and empirical observations in the previous research<sup>5,13</sup>.

*Lévy flight* characterizes the movement as an individual diffusion process. The urban popula-

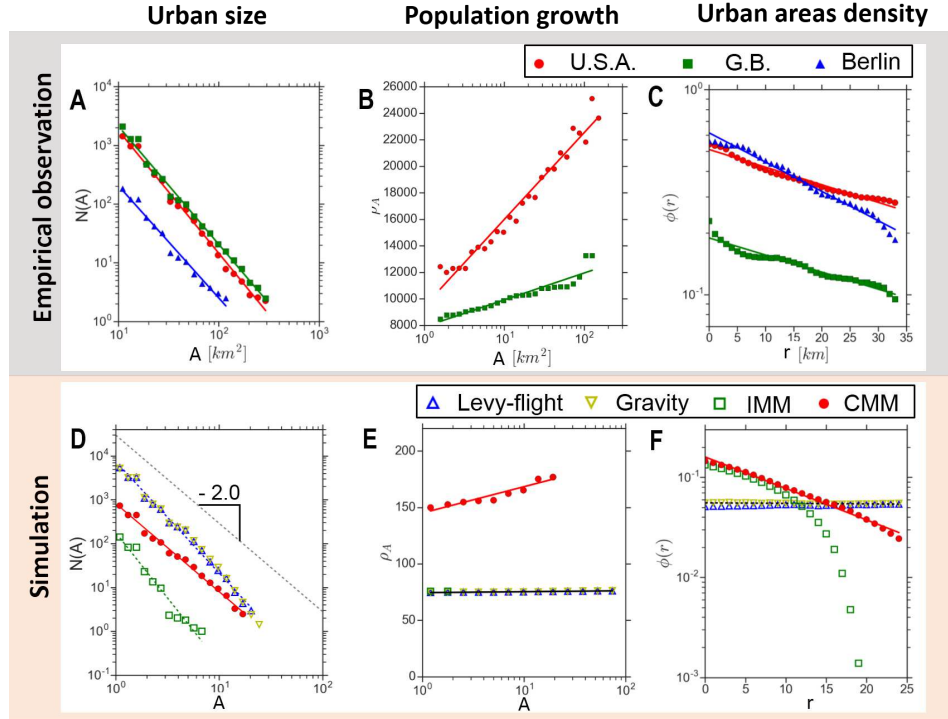


Figure 3: **The comparison of reproducing empirical urban growth patterns with urban systems driven by different human movement models.** (A) Empirical city size distributions follows scaling laws with exponent of  $-2.09 \pm 0.09$ ,  $-2.01 \pm 0.08$  and  $-1.91 \pm 0.16$  for U.S.A., G.B. and Berlin region, respectively. (B) Empirical urban population density increases logarithmically with city sizes. (C) Populated area density decreases exponentially with the distance to city center with exponents of  $-0.3$ ,  $-0.3$  and  $-0.4$  for U.S.A., G.B. and Berlin region, respectively. (D) *Lévy flight*, *Gravity model* and *IMM* predict the scaling law with exponent of  $-2.55 \pm 0.15$ ,  $-2.58 \pm 0.17$  and  $-2.98 \pm 0.51$ , while *CMM* predicts the exponent to be  $-2.02 \pm 0.13$ . (E) *Lévy flight*, *Gravity model* and *IMM* predict urban population density to be invariant with city size, while *CMM* reproduces the logarithmic correlations. (F) *Lévy flight* and *Gravity model* predict the populated area density is invariant with distance to urban center, while *IMM* predicts populated area density decreases slower than exponential function. *CMM* reproduces exponential distribution.

tion will distribute uniformly in urban space as the urban system reaches a stable state. On the other hand, although the *Gravity model* introduces the correlation among individuals through Eq. (3), our analysis shows it also predicts the population distribution follows a fractional diffusion as Eq. (2). When time  $t \rightarrow \infty$ , the population will distribute uniformly in urban space  $\rho(r) = c$ , which is independent of the coupling constant  $\rho_0^{-1}$ . Therefore, both the *Lévy flight* and *Gravity model* are equivalent to the uncorrelated percolation. The emulated urban systems of *Lévy flight* and *Gravity model* reproduce the scaling law distribution with  $-2.55 \pm 0.15$  and  $-2.58 \pm 0.17$  (see Fig. 3D), which is consistent with the theoretical prediction of completely uncorrelated percolation. While the analytical prediction of the memory-aware and social independent *IMM* model is unclear, the emulation shows it also satisfies the scaling law, yet with an exponent  $\tau = 2.98 \pm 0.51$ . The large exponent implies the fact that individuals are localized within their own home-range since the *IMM* is equivalent to the non-interactive limit of *CMM* with  $\rho_0^{-1} \rightarrow 0$ . In contrast, when the coupling constant  $\rho_0^{-1} \rightarrow \infty$ , the *CMM* model becomes strongly-correlated. As a result, it reproduces the scaling law with  $\tau = 2.02 \pm 0.13$ , which agrees with the theoretical predictions and empirical patterns observed in real-world data. These results suggest both the principals of *social interaction* and *memory* are essential components of reproducing the empirical city size distribution, while *CMM* successfully integrates them into a unified movement model.

(B) *Super-linear relation between population and city size*: The positive allometric population growth with the urban area is widely observed in cities around the globe<sup>27,28</sup>. Larger cities tend to have a higher urban population density,  $\rho_A \equiv N(A)/A$ , because they are developing into the third dimension<sup>24,29,30</sup>. Recent researches suggested the balance between the cost and gain of

concentrating population in urban areas, would explain the observed super-linear growth <sup>6</sup>. This social-economic hypothesis consists of two assumptions: i) the average gain from the intense social interaction is proportional to the population density  $\rho_A$ ; ii) the average living cost is proportional to the typical travel distance  $l \sim \log A$  (see Eq. (5)) to explore the city. Their balance leads to,

$$\rho_A \sim \log A. \quad (7)$$

The assumption i) agrees with the *social interaction* in Eq. (3), whereas the assumption ii) is rooted in the *memory effect* in Eq. (4).

Fig. 3B plots the population density  $\rho_A$  with city area  $A$  across different cities for both U.S and G.B, finding that the empirical observation agrees precisely with the predicted logarithmic law (7). It is worth noting that previous studies reported a power-law fitting, *i.e.*,  $\rho_A \sim A^\delta$  with a tiny exponent  $\delta \approx 0.1$  <sup>6</sup>. However, Within the range of magnitude of the empirical data the logarithmic function is indistinguishable with a small-exponent power-law. Fig. 3E compares the emulation results for the four prototype models, finding that the proposed *CMM* reproduces the logarithmic law, whereas there is no area-dependence of the population density  $\rho_A$  for the other three models. This result demonstrates that both social interaction and memory are necessary for the observed scaling law (7).

(C) *Exponential occupation profile*: The urban occupation profile  $\phi(r)$  is defined as the probability of finding an inhabited area at the distance  $r$  from the city center. Empirical studies suggested an exponential profile <sup>31</sup>,

$$\phi(r) \sim \phi_c e^{-\lambda r}. \quad (8)$$

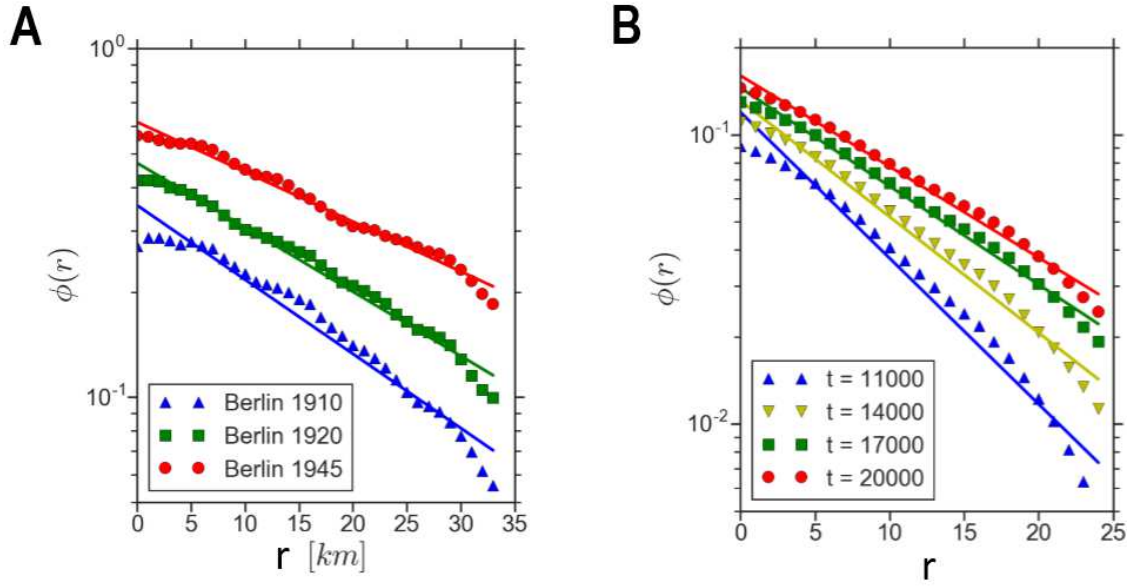


Figure 4: **The qualitative examination on reproducing the populated area density profile in different periods.** (A) The exponent of exponential function between populated area density, and distance to the city center gradually decreases as the cities evolve in the Berlin region. (B) *CMM* predicts the exponent of exponential function between populated area density and distance to the city center gradually decreases as the emulation proceeds.

We also observe the same exponential law for all three empirical datasets, shown in Fig. 3C, indicating that the city center attracts most of the population, whereas the occupation probability decreases rapidly with the radial distance. However, such a rapid decline somehow contradicts with the fat-tailed nature of the human movement (1) that suggests the human travels being able to reach areas far away from the initial location<sup>18,24</sup>.

This paradox can be also resolved by introducing jointly the *social interaction* and *memory* in human movement. Indeed, emulation results in Fig 3F shows the occupation profile in *CMM* agrees very well with the exponential law (8). In contrast,  $\phi(r)$  is independent with  $r$  for *Lévy flight* and *Gravity model*, whereas *IMM* shows a non-exponential decrease. Moreover, it has been suggested that the declining rate  $\lambda$  shall decrease as the city evolves, due to the constantly pushing forwarded frontiers of cities<sup>5</sup>, in line with the observations in the Berlin dataset where  $\phi(r)$  at three different time has been measure. Fig. 4A shows  $\lambda$  decreases gradually from 0.050 to 0.031. The emulation results of *CMM* precisely reproduce the evolution of the occupation profile during urban development (see Fig. 4B).

**Discussion** The rapid urbanization process urges the demand for a more comprehensive understanding of the patterns of urban growth<sup>32,33</sup>. Correlated percolation model (CP) has reproduced successfully urban morphology by introducing a strong geographical correlation to percolation theory, leaving the origin of such correlations a mystery<sup>5</sup>. In this paper, we propose a novel urban growth model that roots in human movements, providing a solid micro-foundation for the mystical geographical correlation in the CP model. It offers a bottom-up approach towards understanding

the observed urban morphology and scaling laws. Two principles, namely, the strong *social interaction* and *memory* of historical movements, are shown to be the key ingredients governing human migration, and consequently the urban development. Unlike existing human movement models where individual movements are either uncorrelated or memoryless and fail to capture the urban growth, the proposed *Collective mobility model (CMM)* demonstrates both principals play essential roles. Theoretical analysis and emulation results show the *memory* principal is essential to reproduce compact and stable city centers in urban systems. In particular, *CMM* reproduces three major empirical laws: city size distribution, super-linear population-area relation, and the exponential occupation profile, consistent with the established CP model and social-economic model at the macroscopic level. Unlike the CP model that is purely static, and has to take the exponential occupation profile as input, the proposed *CMM* predicts the occupation profile (see Fig. 3F) and its evolution (see Fig. 4B) in a self-contained manner, without imposing additional assumptions (see supplementary material S5 for more details about model parameters). In all, the *CMM* not only fills the missing gap in the paradigm of existing movement models (Fig. 1), but at the same time complements previous urban growth models, shedding light on the underlying mechanisms urban growth in the macroscopic level.

In addition to the scientific findings, our research also may have direct implications on wide range downstream applications<sup>34–37</sup>, *i.e.*, city planning, resource allocation, disease controlling, etc. First, conventional top-down city planning strategies have been shown ineffective in governing urban growth by the previous research<sup>2</sup>. We propose an alternative bottom-up urban growth model that predicts the urban growth process from the perspective of human movement. These findings



may inspire novel city planning policies that leverage the principals of human movement <sup>38</sup>. For example, promoting the development of urban regions by reducing the cost of traveling to them. Second, the proposed urban model facilitates us to emulate the patterns of urban dynamics with high accuracy, which may shed light on the preemptive allocation of critical resources, such as transportation infrastructures <sup>39</sup> and medical supplies <sup>40</sup>. Third, the proposed model reveals the underlying correlations between urban structures and human movement. Therefore, the findings improve our understandings of urban movement, which plays a crucial role in controlling contagious diseases <sup>41,42</sup>.

## Methods

**Diffusion Equation of Gravity Model** In this section, we develop the diffusion equation (2) for the *Gravity model*. We start with a lattice model to work out the master equation and then find the equation at the continuous limit by taking the lattice spacing approaching zero.

Equation. (3) indicates that the transition matrix  $W_{ij}$  from the lattice site  $j$  to  $i$  follows,

$$W_{ij} = \lambda_g(l) \frac{\rho_i(t) + \rho_0}{|\vec{r}_i - \vec{r}_j|^{d+\alpha}}, \quad (9)$$

where  $\lambda_g(l)$  is the transition rate with an appropriate scaling with the lattice spacing  $l$ , and  $\alpha \in (0, 2]$ .

The corresponding master equation reads,

$$\begin{aligned}\frac{d\rho_i(t)}{dt} &= \sum_j [W_{ij}\rho_j(t) - W_{ji}\rho_i(t)] \\ &= \lambda_g \rho_0 \sum_j \frac{\rho_j(t) - \rho_i(t)}{|\vec{r}_i - \vec{r}_j|^{d+\alpha}}\end{aligned}\tag{10}$$

Assuming a small lattice spacing  $l$ , we approximate the summation with a continuous integration,

$$\begin{aligned}\sum_j \frac{\rho_j(t) - \rho_i(t)}{|\vec{r}_i - \vec{r}_j|^{d+\alpha}} &\approx l^\alpha \int \frac{\rho(\vec{y}, t) - \rho(\vec{x}, t)}{|\vec{x} - \vec{y}|^{d+\alpha}} dy \\ &= -c_{d,\alpha} l^\alpha (-\Delta)^{\alpha/2} \rho(\vec{x}, t),\end{aligned}\tag{11}$$

where  $(-\Delta)^{\alpha/2}$  is the fractional Laplacian satisfying,

$$(-\Delta)^{\alpha/2} f(\vec{k}) = |\vec{k}|^\alpha \hat{f}(\vec{k}),\tag{12}$$

and

$$c_{d,\alpha} = \frac{\pi^{d/2} |\Gamma(-\alpha/2)|}{2^\alpha \Gamma((d+\alpha)/2)}.\tag{13}$$

The continuous equation satisfies the fractional diffusion equation,

$$\frac{\partial \rho(\vec{x}, t)}{\partial t} = -D (-\Delta)^{\alpha/2} \rho(\vec{x}, t),\tag{14}$$

where the diffusion constant  $D \equiv c_{d,\alpha} \rho_0 \lim_{l \rightarrow 0} \lambda_g(l) l^\alpha$ , in line with Eq. (2). For  $\alpha = 2$  we recover the standard diffusion equation.

**Model Emulation** We evaluate the effectiveness of the proposed movement models in predicting urban growth patterns through emulating urban systems with citizens governed by them. Specifically, the urban system is emulated in a two-dimensional square space with  $l$  sites in width and

length and each site represents a unit area. The urban system consists of  $N$  citizens with their movement trajectories generated based on the corresponding urban movement models. The emulated citizens initially located at the centers of urban systems, and then iteratively sample their next visit sites at each epoch until the system reaches a stable state. After the system is converged, we considered the occupied sites as populated urban areas, and examine the morphology and growth patterns of the emulated urban system. Emulating agent-based urban systems according to urban movement models is time expensive <sup>43,44</sup>. To address these challenges, we propose two improved sampling methods, alias sampling and sorted array sampling, to accelerate the emulation (see supplementary material S3 for details).

**Data Availability** All the empirical urban data we examine are publicly available in government websites and previous research (see supplementary material S1 for details). Specifically, the data of U.S.A cities can be accessed through the website of U.S.A. Census Bureau <sup>45</sup>. The data of cities in G.B. is available on the website of the Statistical Office of the European Union <sup>46</sup>. Both of these two datasets provide information on urban area coverage and population distribution. On the other hand, the dataset of the Berlin region is collected from the telemetry images of urban area distribution in 1910, 1920, and 1945, which is released in the previous research <sup>47</sup>. Note that the telemetry images only provide information on the coverage area of cities but not the urban population distribution.

**Acknowledge** F.X and Y.L. were supported in part by The National Key Research and Development Program of China under grant SQ2020AAA010130, the National Nature Science Foundation

of China under U1936217, 61971267, 61972223, 61941117, 61861136003, Beijing National Research Center for Information Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology. We also thanks for the insightful discussion with Prof. Depeng Jin, Prof. Jianhua Lu and Prof. Zhaocheng Wang in Tsinghua University.

1. Howard, E. *To-morrow: A peaceful path to real reform* (Cambridge University Press, 2010).
2. Batty, M. The size, scale, and shape of cities. *science* **319**, 769–771 (2008).
3. Batty, M. *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals* (The MIT press, 2007).
4. Batty, M. *The new science of cities* (MIT press, 2013).
5. Makse, H. A., Havlin, S. & Stanley, H. E. Modelling urban growth patterns. *Nature* **377**, 608–612 (1995).
6. Bettencourt, L. M. The origins of scaling in cities. *Science* **340**, 1438–1441 (2013).
7. Witten Jr, T. & Sander, L. M. Diffusion-limited aggregation, a kinetic critical phenomenon. *Physical review letters* **47**, 1400 (1981).
8. Barthélemy, M. Spatial networks. *Physics Reports* **499**, 1–101 (2011).
9. Clark, C. Urban population densities. *Journal of the Royal Statistical Society. Series A (General)* **114**, 490–496 (1951).
10. Newling, B. E. The spatial variation of urban population densities. *Geographical Review* 242–252 (1969).
11. McDonald, J. F. Econometric studies of urban population density: a survey. *Journal of urban economics* **26**, 361 (1989).
12. Vicsek, T. *Fractal growth phenomena* (World scientific, 1992).

13. Essam, J. W. Percolation theory. *Reports on progress in physics* **43**, 833 (1980).
14. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *nature* **453**, 779–782 (2008).
15. Schrank, D., Lomax, T. *et al.* 2009 urban mobility report. (2009).
16. Schrank, D., Eisele, B. & Lomax, T. Ttis 2012 urban mobility report. *Texas A&M Transportation Institute. The Texas A&M University System* **4** (2012).
17. Einstein, A. *Investigations on the Theory of the Brownian Movement* (Courier Corporation, 1956).
18. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
19. Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nature Physics* **6**, 818–823 (2010).
20. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
21. (Note). The traffic flow from location  $\vec{r}'$  to  $\vec{r}$ ,  $T(\vec{r}, \vec{r}') \equiv P(\vec{r}|\vec{r}')\rho(\vec{r}') = (\rho(\vec{r}) + \rho_0)\rho(\vec{r}')/|\vec{r} - \vec{r}'|^{d+\alpha}$ , in line with the convectional form at the strong coupling limit  $\rho_o \rightarrow 0$ .
22. (Note).  $l \sim \log \frac{1-A^{1-\zeta}}{\zeta-1}$  in Ref <sup>19</sup>, where  $\zeta$  is the zipf's exponents. For  $\zeta > 1$ ,  $l$  saturates at large  $A$ , which accounts for the home range effect for daily movements. For  $\zeta > 1$ ,  $l \sim \log(A)$

captures unbounded movements, *i.e.*, a migration process. We ignore the marginal double-logarithmic case when  $\zeta = 1$ .

23. Burt, W. H. Territoriality and home range concepts as applied to mammals. *Journal of mammalogy* **24**, 346–352 (1943).
24. Batty, M. & Longley, P. A. *Fractal cities: a geometry of form and function* (Academic press, 1994).
25. Encarnação, S., Gaudiano, M., Santos, F. C., Tenedório, J. A. & Pacheco, J. M. Fractal cartography of urban areas. *Scientific Reports* **2**, 1–5 (2012).
26. Benguigui, L., Czamanski, D., Marinov, M. & Portugali, Y. When and where is a city fractal? *Environment and planning B: Planning and design* **27**, 507–519 (2000).
27. Woldenberg, M. J. An allometric analysis of urban land use in the united states. *Ekistics* 282–290 (1973).
28. Coffey, W. J. Allometric growth in urban and regional social-economic systems. *Canadian Journal of Regional Science* **11**, 49–65 (1979).
29. Chen, Y., Wang, J. & Feng, J. Understanding the fractal dimensions of urban forms through spatial entropy. *Entropy* **19**, 600 (2017).
30. Falconer, K. *Fractal geometry: mathematical foundations and applications* (John Wiley & Sons, 2004).
31. Wilson, A. *Entropy in urban and regional modelling*, vol. 1 (Routledge, 2011).

32. Harvey, D. The urbanization of capital: Studies in the history and theory of capitalist urbanization (1987).
33. Knox, P. L. & McCarthy, L. *Urbanization: an introduction to urban geography* (Pearson Prentice Hall Upper Saddle River, NJ, 2005).
34. Camagni, R., Gibelli, M. C. & Rigamonti, P. Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion. *Ecological economics* **40**, 199–216 (2002).
35. Krajzewicz, D., Erdmann, J., Behrisch, M. & Bieker, L. Recent development and applications of sumo-simulation of urban mobility. *International journal on advances in systems and measurements* **5** (2012).
36. Lima, A., De Domenico, M., Pejovic, V. & Musolesi, M. Disease containment strategies based on mobility and information dissemination. *Scientific reports* **5**, 10650 (2015).
37. Chaintreau, A. *et al.* Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing* **6**, 606–620 (2007).
38. Batty, M. A theory of city size. *Science* **340**, 1418–1419 (2013).
39. Vazifteh, M. M., Santi, P., Resta, G., Strogatz, S. H. & Ratti, C. Addressing the minimum fleet problem in on-demand urban mobility. *Nature* **557**, 534–538 (2018).



40. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. Predictability and epidemic pathways in global outbreaks of infectious diseases: the sars case study. *BMC medicine* **5**, 34 (2007).
41. Eubank, S. *et al.* Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004).
42. Toroczkai, Z. & Guclu, H. Proximity networks and epidemics. *Physica A: Statistical Mechanics and its Applications* **378**, 68–75 (2007).
43. Behrisch, M., Bieker, L., Erdmann, J. & Krajzewicz, D. Sumo—simulation of urban mobility: an overview. In *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation* (ThinkMind, 2011).
44. Krajzewicz, D. Traffic simulation with sumo—simulation of urban mobility. In *Fundamentals of traffic simulation*, 269–293 (Springer, 2010).
45. *US Census Data in 2000* (2000 (accessed March 3, 2018)). <https://www.census.gov/>.
46. *Great Britain Census Data in 1991* (1991 (accessed March 5, 2018)). <http://ec.europa.eu/eurostat>.
47. Frankhauser, P. Fractal geometry of urban patterns and their morphogenesis. *Discrete Dynamics in Nature and Society* **2** (1998).

## Figures

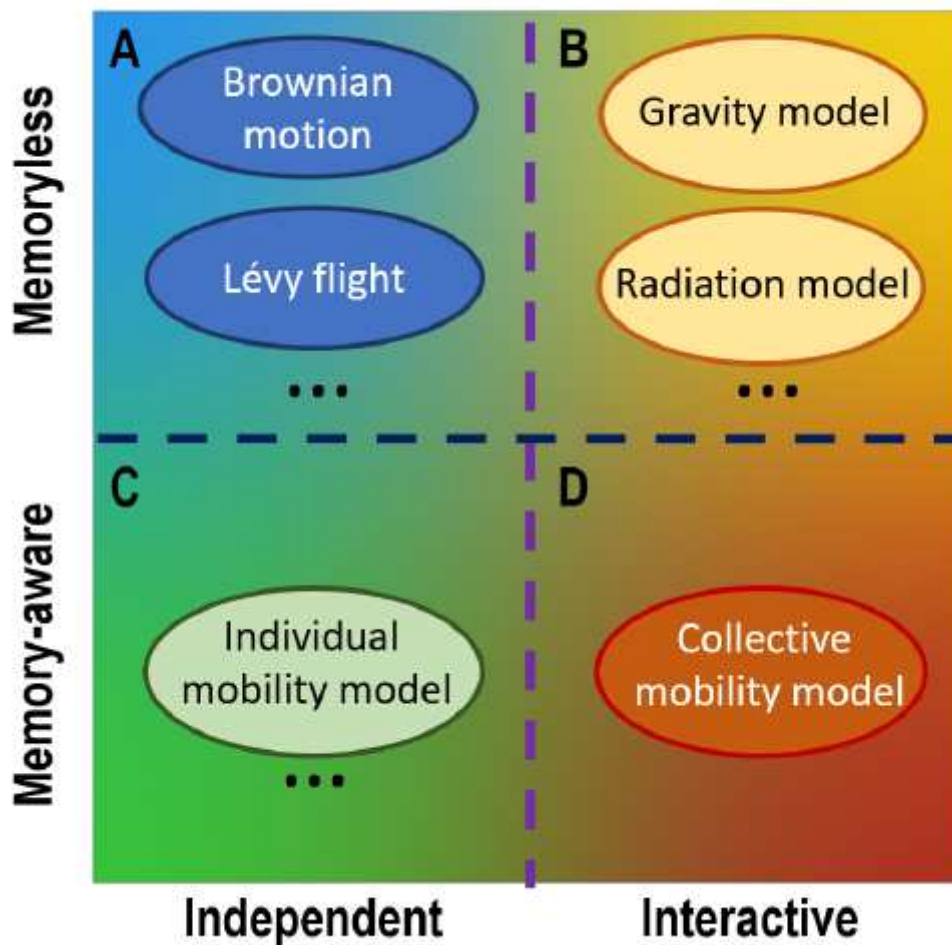
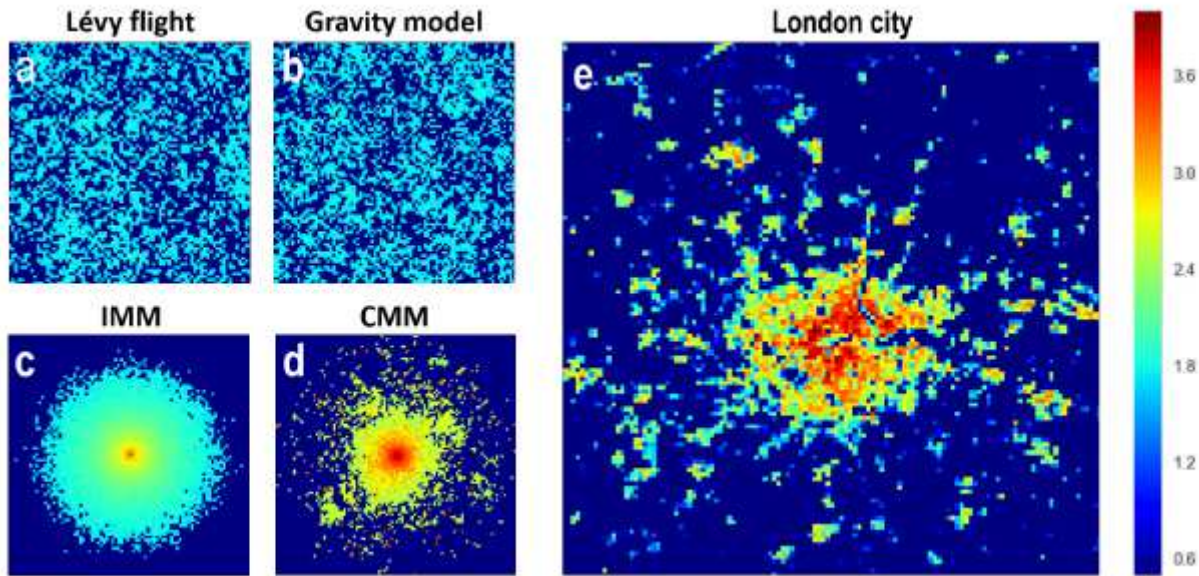


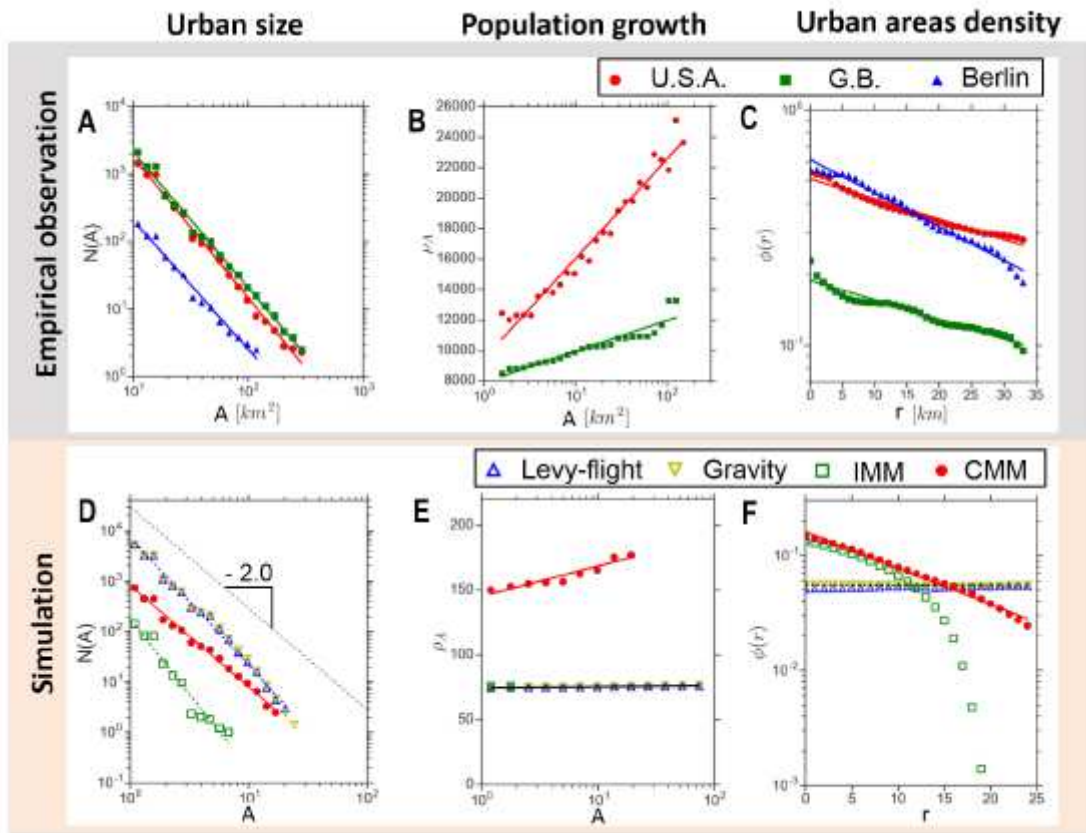
Figure 1

The paradigm of human movement models. Existing human movement models classified based on whether account for the memory of historic movements or the social interactions, are summarized as the paradigm with four classes: (A) Brownian motion and Lévy-flight belong to this class where movements are independent and memoryless; (B) Gravity model and Radiation model are the typical models where movements are socially correlated and memoryless; (C) Individual mobility model belongs to this class where movements are independent and memory-aware. (D) Missing corner of the current paradigm. The proposed Collective Mobility Model fills in this class where movements are both socially correlated and memory-aware.



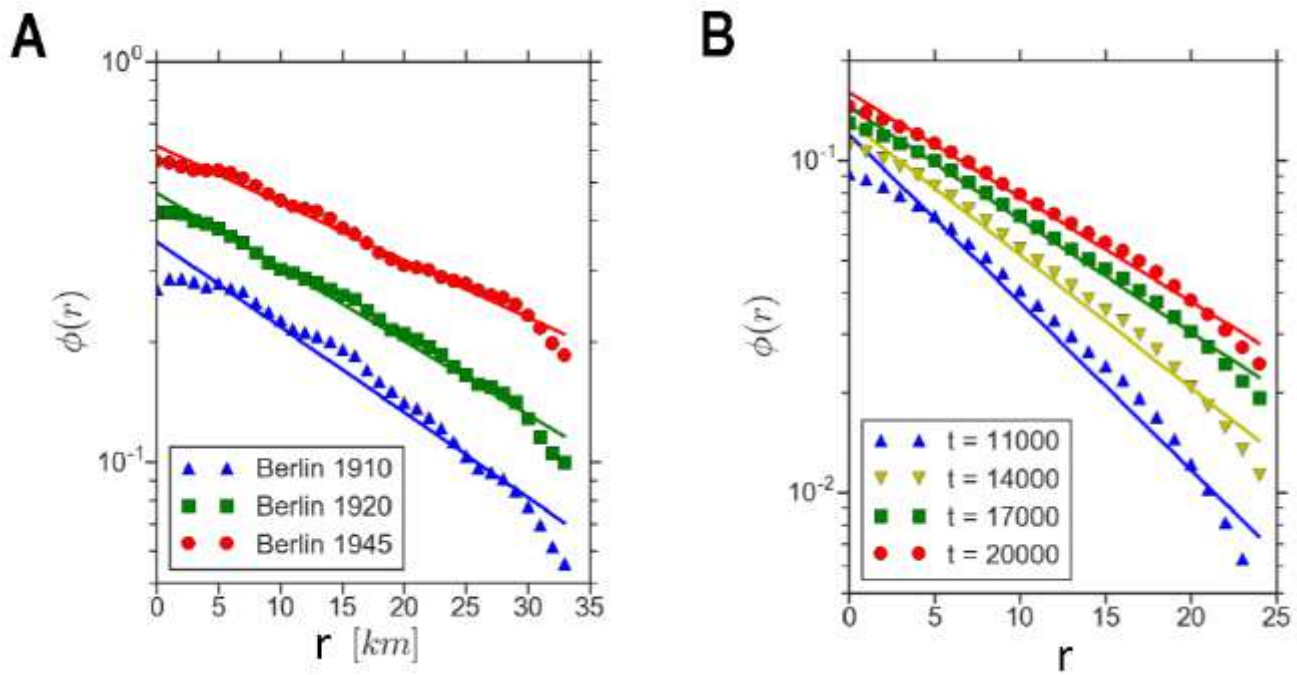
**Figure 2**

The morphology of urban area generated by four different human movement models: (A) Lévy flight, (B) Gravity model, (C) Individual mobility model (IMM), (D) Collective mobility model (CMM), and (E) the empirical data from London city. The population distribution of each urban system is visualized as a heatmap in log scale, where blue color represents underpopulated regions and red colors corresponding to regions of high population. The urban systems are emulated with 30,000 individuals initially situated in urban centers and then move according to mobility models until reaching a stable state. The Lévy flight and Gravity model fail to reproduce a compact urban center, while IMM predicts the urban system grows in a homogeneous manner, where fractal perimeter and sub-clusters are absent. CMM accurately reproduces the compact urban center, fractal perimeter of urban area, and sub-clusters. The morphology predicted by CMM is consistent with the empirical observation of London city.



**Figure 3**

The comparison of reproducing empirical urban growth patterns with urban systems driven by different human movement models. (A) Empirical city size distributions follows scaling laws with exponent of  $-2.09 \pm 0.09$ ,  $-2.01 \pm 0.08$  and  $-1.91 \pm 0.16$  for U.S.A., G.B. and Berlin region, respectively. (B) Empirical urban population density increases logarithmically with city sizes. (C) Populated area density decreases exponentially with the distance to city center with exponents of  $-0.3$ ,  $-0.3$  and  $-0.4$  for U.S.A., G.B. and Berlin region, respectively. (D) L'evy flight, Gravity model and IMM predict the scaling law with exponent of  $-2.55 \pm 0.15$ ,  $-2.58 \pm 0.17$  and  $-2.98 \pm 0.51$ , while CMM predicts the exponent to be  $-2.02 \pm 0.13$ . (E) L'evy flight, Gravity model and IMM predict urban population density to be invariant with city size, while CMM reproduces the logarithmic correlations. (F) L'evy flight and Gravity model predict the populated area density is invariant with distance to urban center, while IMM predicts populated area density decreases slower than exponential function. CMM reproduce10 exponential distribution.



**Figure 4**

The qualitative examination on reproducing the populated area density profile in different periods. (A) The exponent of exponential function between populated area density, and distance to the city center gradually decreases as the cities evolve in the Berlin region. (B) CMM predicts the exponent of exponential function between populated area density and distance to the city center gradually decreases as the emulation proceeds.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [si.pdf](#)