Emergence of Urban Growth Patterns from Human Movements

Fengli Xu
Tsinghua University
Yong Li (liyong07@tsinghua.edu.cn)
Tsinghua University
Depeng Jin
Tsinghua University
Jianhua Lu
Tsinghua University
Chaoming Song
University of Miami

Article

Keywords: Urban Morphology, Scaling Laws, Correlated Percolation, Human Mobility Data, Social Couplings, Long-memory Effect, City Evolution

Posted Date: June 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-79579/v2

Version of Record: A version of this preprint was published at Nature Computational Science on December 9th, 2021. See the published version at https://doi.org/10.1038/s43588-021-00160-6.

Emergence of Urban Growth Patterns from Human Movements

Fengli Xu[a], Yong Li[a*], Depeng Jin[a], Jianhua Lu[a], & Chaoming Song[b*]

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 designs a novel computational urban growth model that offers a micro-foundation in human mobility behavior. We compare the proposed model with three empirical datasets, which evidences that strong social couplings and long-memory effect are two fundamental principles responsible for the mystical spatial correlations. The model not only 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 [1]. However, the increasingly available urban data suggests that cities grow in a bottom-up manner, calling for understandings of its

*To whom correspondence should be addressed. E-mail: c.song@miami.edu, liyong07@tsinghua.edu.cn

micro foundation [2–4]. Later, three fundamental empirical laws have been discovered [2,5,6]: First, the distribution of city size follows a scaling law with the exponent around -2, implying large cities are much rarer than small towns [5]. Second, the urban population grows super-linearly with its area due to the intense competition for spaces [7,8]. Finally, the density of occupied urban areas decreases exponentially with the radial distance to city centers [9–11]. Computational physics model, *i.e.,* diffusion-limited aggregation (DLA), has been applied to model urban growth as an aggregation of physical particles [7,12]. Further works showed that correlated percolation (CP) is a better alternative to explain the emergence of the aforementioned laws [5]. A key observation of the CP model is the requirement of strong geographical correlation to reproduce the correct scaling relations [13]. 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 computational urban growth model based on human movements, suggesting that strong social coupling and long-memory effect 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 [14,15]. Existing human movement models mainly 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 [16]. Unlike physical particles, empirical data suggests human movements are characterized by a fat-tailed jump-size distribution,

satisfying a power law,

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

where $P(\vec{r}|\vec{r}')$ is the transition probability from location $\vec{r}'$ to $\vec{r}$, with $d = 2$ for two-dimensional space [14,17]. The exponent α is observed around 0.55 ± 0.05 [18]. 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*, where the evolution of population density $\rho(\vec{r}, t)$ follows the fractional diffusion equation,

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

where D is the diffusion constant (see Methods 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 [9].

Class B models such as *Gravity model* [8] and *Radiation model* [19], 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,

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

where ρ_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 $\vec{r}$ [20]. This mechanism accounts for a mean-field background attractiveness rooted

in *social interaction*, *e.g.*, highly populated locations often offer more social opportunities [8]. 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 Methods section for details), *i.e.*, it predicts a uniformly distributed urban patterns at a large t. More recently, researches explored the correlation between human mobility and social network, finding that human movements are largely determined by the underlying social ties [21–23]. However, these models often require inputs of social network data measured from experiments, leaving the origin of these correlations unexplained.

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 scaling-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 effect*. In particular, a *preferential return mechanism* is imposed, *i.e.*, the probability returning to a previous location $\vec{r}_i$,

$$P(r_i) \propto f(r_i), \tag{4}$$

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 typical traveling distance l, *i.e.*,

the root mean square displacement, follows a logarithmic growth as

$$l \sim \log A, \tag{5}$$

where A is the total visitation area [18,24]. The logarithmic growth is one of the key ingredients for human movements, characterizing the anomalous ultra-slow diffusion and home range effect [25]. More recently, *d-EPR* model generalizes *IMM* by introducing a background field where an individual explores a new location with a probability proportional to its population density [26,27]. While *IMM* and *d-EPR* successfully captures individual movements on a daily basis, individuals move independently and do not interact with each other. Moreover, the background field in *d-EPR* is static and does not evolve with time. Therefore, these models do not capture dynamic interactions among individuals, and therefore are not capable of reproducing urban growth patterns (see SI Section S2 for more details).

Neither social interactions (class B) nor memory effects (Class C) alone reproduces the geographical correlation proposed in the CP model. It is rather curious to ask if Class D mobility models that integrate social interaction and memory effects (Fig. 1) are able to predict urban growth patterns. While Class D models are largely unexplored, there are few exceptions. A recently developed *GeoSim* model integrates the memory-aware *IMM* with a social network where each individual is more likely to explore a new location visited by his/her friends [28]. However, the social network is predetermined from the inputs of empirical data, and does not evolve with the time. Therefore, the *GeoSim* model does not track urban growth on a large time scale. To integrate memory effects with dynamic interactions, we propose *Collective Mobility Model (CMM)* that offers a minimal agent-based model connecting human movements to urban growth. We consider N

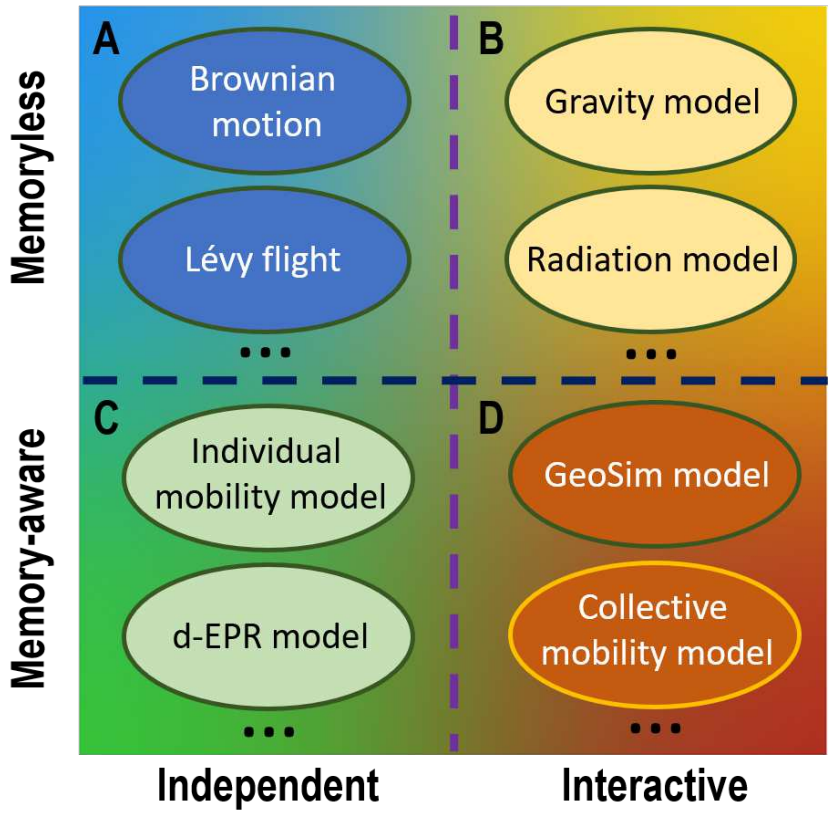

Figure 1: **The paradigm of human movement models.** Existing human movement models classified based on whether account for the memory of historic movements or the social interactions, are summarized as the paradigm with four classes: (A) *Brownian motion* and *Lévy-flight* belong to this class where 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* (IMM) and *d-EPR model* belong to this class where movements are *independent* and *memory-aware*. (D) This class is both *memory-aware* and *socially interactive*, including the *GeoSim* model and the proposed *Collective Mobility Model* (CMM). The *GeoSim* model uses a static social network measured from the experiment whereas CMM generates the social interaction dynamically from individual's locations without external inputs.

individuals moving in a $L \times L$ square lattice, where each follows a return-exploration process. Like IMM, an individuals return probability is proportional to his/her visitation frequency (4), whereas the exploration probability depends on the number of visited sites. During the exploration, the probability of choosing a new location, in contrast, relies on the populations, satisfying the *gravity law* (3). The population density $\rho(\vec{r}, t)$ in Eq. (3) is calculated instantaneously based on individuals' locations at time t. The coupling constant, ρ_0^{-1}, controls the strength of population attraction, *i.e.*, increasing ρ_0 reduces the impact of population density, and consequently, the strength of social interactions. For $\rho_0^{-1} \to 0$, *CMM* is effectively equivalent to *IMM*. Inspired by the strong geometrical correlation in *CP* models [5], we're interested in the strong coupling limit $\rho_0^{-1} \to \infty$, where *CMM* describes a strongly correlated many-body system. Therefore, the proposed *CMM* does not rely on external data and can self-organize to arbitrary time scale, which is optimized for urban growth simulation. More importantly, we find that *CMM* can simulate urban system in a self-organized manner (see Methods Section for details).

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 simulate the human migrations by four typical movement models for Classes A–D respectively, namely the *Lévy flight*, *Gravity model*, *IMM*, and the proposed *CMM* (see SI Section S2 for the details of model parameter settings. A simple simulation of large urban systems is impractical due

to the high time complexity, which is $\mathcal{O}(MA)$ for each step with M and A 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 $\mathcal{O}(M)$ (see SI Section S3 for details).

To compare the morphology of the simulated 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 [29–31].

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 simulation 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 simulation 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 simulation 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

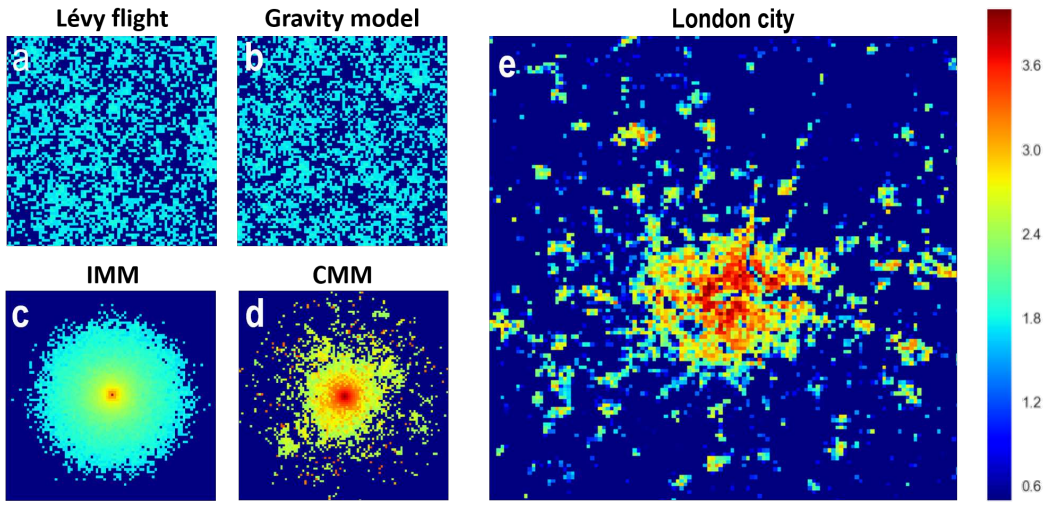

Figure 2: **The morphology of urban area generated by four different human movement models: (A) *Lévy flight*, (B) *Gravity model*, (C) *Individual mobility model (IMM)*, (D) *Collective mobility model (CMM)*, and (E) the empirical data from London city.** The population distribution of each urban system 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 simulated 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 SI Section 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 examine the model's capacity in reproducing urban growth patterns, we will focus on three fundamental empirical laws, each of which has been validated on multiple cities around the globe [2,5,6].

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

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

where the exponent τ has been reported around 2.0 [5]. Percolation theory is the prevalent narrative for this observation, with each site occupied as an urban area with a certain probability [13]. 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 [13]. The empirical data shows that city size distributions are well approximated by Eq. (6) (see Fig. 3A), with the exponent $\tau = 1.94 \pm 0.11$, 2.01 ± 0.08 and 1.91 ± 0.16 for the U.S.A. and the G.B., and the Berlin, respectively. These findings echo the theoretical predictions of site percolation theory and empirical observations in the previous research [5,13].

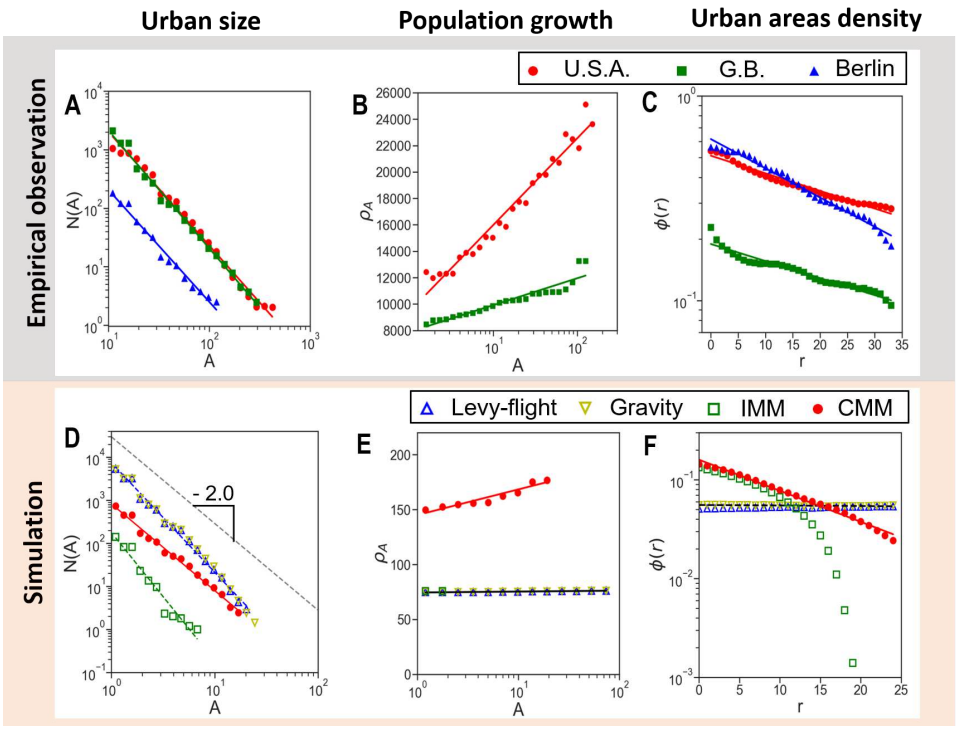

Figure 3: **The comparison of reproducing empirical urban growth patterns with urban systems driven by different human movement models.** **(A)** City size distributions for U.S.A., G.B. and Berlin. Dots and straight lines represent the empirical measurements and Eq. (6) with exponent of $\tau = 1.94 \pm 0.11$, $\tau = 2.01 \pm 0.08$ and $\tau = 1.91 \pm 0.16$, respectively. The range denotes the 95% confidence interval of estimated exponents. **(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)** City size distributions for *Lévy flight*, *Gravity model* and *IMM* models with exponents $\tau = 2.55 \pm 0.15$, $\tau = 2.58 \pm 0.17$ and $\tau = 2.98 \pm 0.51$, and for *CMM* with $\tau = 2.02 \pm 0.13$. The range denotes the 95% confidence interval of estimated exponents. **(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* reproduce exponential distribution.

Lévy flight characterizes the movement as an individual diffusion process. The urban population distributes uniformly in urban space as the urban system reaches to a stationary state. Similarly, while the *Gravity model* introduces the social interactions among individuals, the population density satisfies fractional diffusion (2). When time $t \to \infty$, the population will distribute uniformly in urban space $\rho(r) = c$, which is independent of the coupling constant ρ_0^{-1}. Therefore, the *Lévy flight* and *Gravity model* are expected to generate urban patterns as the one observed in an uncorrelated percolation. To test this hypothesis, we identify the isolated connecting components in the simulation as satellite cities and measure the area of each city [5]. Figure 3D shows that both *Lévy flight* and *Gravity model* reproduce a power-law city size distribution with exponents $\tau = 2.55 \pm 0.15$ and $\tau = 2.58 \pm 0.17$. This finding is consistent with the theoretical prediction of an uncorrelated percolation. The simulation also shows *IMM* satisfies the scaling law with an exponent $\tau = 2.98 \pm 0.51$. This 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 $\rho_0^{-1} \to 0$. In contrast, when the coupling constant $\rho_0^{-1} \to \infty$, 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. All reported exponents are ranged within a 95% confidence interval based on the Theil-Sen estimator [32] (see SI Section S7 for details). We find, however, the proposed CMM is the only model that falls within the empirical observed universal exponent -2. 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 [33,34]. Larger cities tend to have a higher urban population density, $\rho_A \equiv N(A)/A$, because they are developing into the third dimension [29,35,36]. Recent researches suggested the balance between the cost and gain of concentrating population in urban areas, would explain the observed super-linear growth [6]. This social-economic hypothesis consists of two assumptions: i) the average gain from the intense social interaction is proportional to the population density ρ_A; ii) the average living cost is proportional to the typical travel distance l to explore the city. Their balance leads to $\rho_A \sim l$. Substituting Eq. (5), we find

$$\rho_A \sim \log A. \tag{7}$$

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 ρ_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 scaling-law fitting, *i.e.*, $\rho_A \sim A^{\delta}$ with a tiny exponent $\delta \approx 0.1$ [6]. However, Within the range of magnitude of the empirical data the logarithmic function is indistinguishable with a small-exponent scaling-law. Fig. 3E compares the simulation 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 ρ_A for the other three models. This result demonstrates that both social interaction and memory are necessary for the observed scaling law (7).

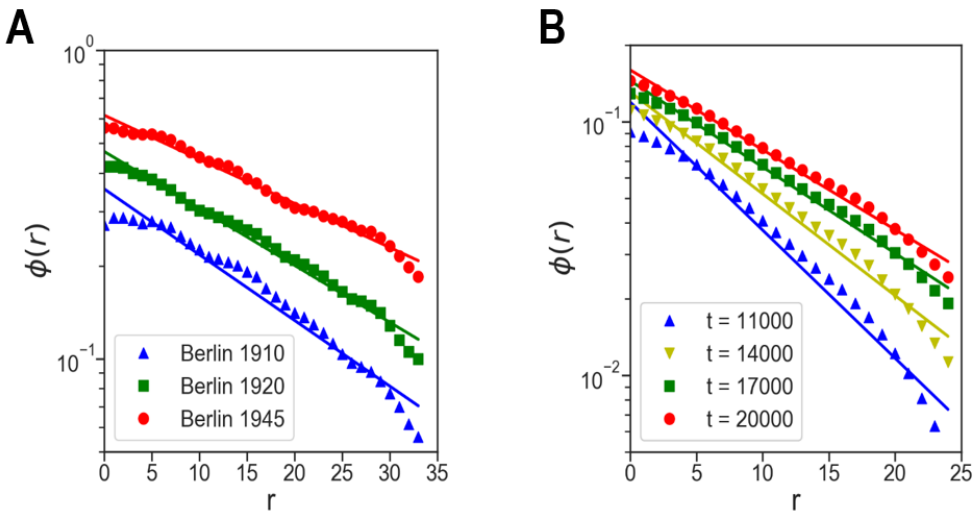

Figure 4: **The time evolution of urban occupations.** The urban occupation profile $\phi(r)$ as a function of radial distance r, for **(A)** the Berlin over three different periods, and for **(B)** *CMM* over four different periods. Straight lines represent exponential fittings for the corresponding empirical data and the simulation results, respectively.

(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 [37],

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

Figure 3C demonstrates the exponential occupation profile observed for all three empirical datasets, 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 [17,29].

This paradox can be also resolved by introducing jointly the *social interaction* and *memory* in human movement. Indeed, simulation results in Fig 3F shows the urban 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 an non-exponential decrease. Moreover, it has been suggested that the declining rate λ shall decrease as the city evolves, due to the constantly pushing forwarded frontiers of cities [5], in line with the observations in the Berlin dataset where $\phi(r)$ at three different time has been measure. Fig. 4A shows λ decreases gradually from 0.050 to 0.031. The simulation results of *CMM* precisely reproduce the evolution of the occupation profile during urban development (see Fig. 4B). We measure the occupation profile at different times for the *Lévy flight*, *Gravity model* and *IMM*, finding that they do not capture the time evolution of $phi(r)$ (see SI Section S8 for details). Indeed, the dynamic interactions in *CMM* are critical for the urban dynamics as we find that the more sophisticated *d-EPR* model with a static social interaction

[26,27] also fails to reproduce the occupation profile evolution (see SI Section S2 for details).

Discussion

The rapid urbanization process urges the demand for a more comprehensive understanding of the patterns of urban growth [38]. 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 [5]. While the connection between human mobility pattern and urban morphology has been observed [39], a self-contained framework remains unknown. In this paper, we propose a novel agent-based model that connects human movements to urban growth, 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 strongly *dynamic 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 simulation 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 statically correlated models such as *CP*, *d-EPR* and *GeoSim* models which have to take empirical

data as an input, the proposed *CMM* predicts the urban occupation profile (see Fig. 3F) and its evolution (see Fig. 4B) in a self-organized manner, without imposing additional assumptions and parameters (see SI Section S2 and S8 for more details).

Compared to existing agent-based models that often consists of a large number of microscopic details of decision making process such as the context of current locations and urban environments [40–42], the goal of this paper is to seek a minimum set of common principles shared by agent-based models that governs the empirically observed universal laws of urban growth (see SI Section S6 for more details). Our simulation results show that individual-based physical determinism models such as *IMM* and *Lévy flight* fail to capture the urban growth, which suggests human decision-making process is equally important. On the other hand, the proposed *CMM* is not solely physical deterministic, yet a highly-simplified agent-based model where individuals make decisions about their explorations based on the social attractiveness of the location. Together with the memory effect for individual return processes, *CMM* is an agent-based model with a minimal set of decision making processes that allows theoretical analysis and numeric simulations. The proposed *CMM* can be further generalized to include more detailed decision making process. However, we feel that the simplicity of the proposed model makes it a general framework that is sufficient to capture urban growth patterns, and more fine-grained decision making behavior will be investigated in future research.

It is worthy pointing out that while most human mobility models are developed for active mobility such as the daily movements, the urban growth is more closely related to residential

mobility, *i.e.*, the relocation of residence. Nevertheless, residential and active mobility are human movements on different time scales and they share common features. For instance, the gravity model is often used in both cases [43]. Moreover, while the memory effect in *IMM* was initially discovered for active mobility, it is also equally important for relocation due to numerous social-economic factors such as the effect of civic memory and change of workplaces [44,45]. Therefore, the *CMM* model provides a framework for both active and residential mobility, each corresponding to different parameter setups, *e.g.*, the relocation has much smaller transition rate whereas the daily movements have smaller exploration probability. Nevertheless, we find that the universal laws observed in the proposed *CMM* are largely independent of choosing different modeling parameters.

In all, the proposed *CMM* provides a minimal model that bridges the missing link between urban growth and human mobility. Our study may have direct implications on a wide range of applications [46,47] including city planning, resource allocation, disease controlling, etc. Existing top-down city planning strategies have been shown ineffective in governing urban growth by the previous research [2]. The bottom-up approaches as the one proposed in this paper reveals general mechanisms governing urban growth patterns, potentially leading to new city planning policies that would leverage the influence of the human movements [48].

Methods

Collective Mobility Model We start with N individuals initially placed at the center of a $L \times L$ square lattice. At each time t, individual decides to either explore a new place with probability

$P_{\text{exp}} = \delta S^{-\gamma}$ or return to a previous visited location with probability $P_{\text{ret}} = 1 - \delta S^{-\gamma}$, where S is the number of distinct locations visited previously. For the exploration process, the individual choose a new location $\vec{r}'$ with a probability $P(\vec{r}') \sim \frac{\rho(\vec{r}',t)+\rho_0}{|\vec{r}'-\vec{r}|^{(d+\alpha)}}$ (Eq. 3), where $\vec{r}$ is his/her present location and $\rho(\vec{r}', t)$ is the population density at location $\vec{r}'$ and time t. For the return process, the individual moves to one of previously visited locations $\vec{r}_i$ based on the preferential return mechanism, with a probability $P(\vec{r}_i) = \frac{f(\vec{r}_i)}{\sum_i f(\vec{r}_i)}$ (Eq. 4).

Numeric simulation is performed in a 300×300 lattice for $N = 30,000$ individuals over $T = 20,000$ total time steps. The simulation size is typical for a strongly interactive agent-based model [49], where the computational complexity is considerably high. We develop two sampling methods, alias sampling and sorted array sampling, to improve the efficiency of the simulation (see SI Section S3 for more details). These optimizations reduce computational time of each run from 690 hours to 12 hours on a workstation with 40 cores of 3.6GHz Intel i7 processor. We also perform statistical analysis to test the robustness of simulated patterns (see SI Section S7 for more details) and sensitivity of model parameters (see SI Section S5 for more details).

Diffusion Equation of Gravity Model We prove the fractional diffusion equation (2) for evolution of the density $\rho(\vec{r}, t)$ in the *Gravity model*. We start with a master equation in a lattice, and then derive the continuous equation by taking the proper limit when the lattice spacing l approaches zero. The transition rate matrix W_{ij} defines the probability rate of individuals departing from site i and arriving in site j. Equation. (3) indicates

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

where $\lambda_g(l)$ is a normalization factor that shall scale appropriately with the lattice spacing l, and $\alpha \in (0, 2]$. The corresponding master equation reads,

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

Assuming a small lattice spacing l, we approximate the summation by 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 $\widehat{(-\Delta)^{\alpha/2} f(\vec{k})} = |\vec{k}|^\alpha \hat{f}(\vec{k})$, and $c_{d,\alpha} = \frac{\pi^{d/2}|\Gamma(-\alpha/2)|}{2^\alpha \Gamma((d+\alpha)/2)}$. Taking the continuous limit $l \to 0$ requires the existence of $\lim_{l\to 0} \lambda_g(l) l^\alpha$, *i.e.*, $g(l) \sim l^{-\alpha}$ for small l. Substituting Eq. (13) to Eq. (12) leads to Eq. (2),

$$\frac{\partial \rho(\vec{x}, t)}{\partial t} = -D(-\Delta)^{\alpha/2} \rho(\vec{x}, t), \tag{12}$$

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

Data Availability The empirical urban data sets that support the findings of this study are available in GitHub, https://github.com/xfl15/Collective-Mobility-Model. The data sets for USA and GB were originally released by previous research [50] (see SI Section S1 for details).

Code Availability The source code for numeric simulation is available online: https://github.com/xfl15/Collective Mobility-Model.

Acknowledge F. Xu and Y. Li 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. Zhaocheng Wang in Tsinghua University.

Author Contributions F. Xu and Y. Li contributed to the empirical implementation and evaluation of the proposed models. C. Song performed the theoretical analysis of human mobility modeling and complex urban system. D. Jin and J. Lu offered empirical motivations and insights to this research. All authors contributed to the writing of this manuscript.

Competing Interests The authors declare no competing interests.

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., Eisele, B. & Lomax, T. Ttis 2012 urban mobility report. *Texas A&M Transportation Institute. The Texas A&M University System* **4** (2012).

16. Einstein, A. *Investigations on the Theory of the Brownian Movement* (Courier Corporation, 1956).

17. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).

18. Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nature Physics* **6**, 818–823 (2010).

19. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).

20. (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 $\rho_o \to 0$.

21. Grabowicz, P. A., Ramasco, J. J., Gonçalves, B. & Eguíluz, V. M. Entangling mobility and interactions in social media. *PloS one* **9**, e92196 (2014).

22. Deville, P. *et al.* Scaling identity connects human mobility and social interactions. *Proceedings of the National Academy of Sciences* **113**, 7047–7052 (2016).

23. Wang, D. & Song, C. Impact of human mobility on social networks. *Journal of Communications and Networks* **17**, 100–109 (2015).

24. (Note). $l \sim \log \frac{1-A^{1-\zeta}}{\zeta-1}$ in Ref [18], where ζ 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 $\zeta = 1$.

25. Burt, W. H. Territoriality and home range concepts as applied to mammals. *Journal of mammalogy* **24**, 346–352 (1943).

26. Pappalardo, L. *et al.* Returners and explorers dichotomy in human mobility. *Nature communications* **6**, 1–8 (2015).

27. Pappalardo, L., Rinzivillo, S. & Simini, F. Human mobility modelling: exploration and preferential return meet the gravity model. *Procedia Computer Science* **83**, 934–939 (2016).

28. Toole, J. L., Herrera-Yaqüe, C., Schneider, C. M. & González, M. C. Coupling human mobility and social ties. *Journal of The Royal Society Interface* **12**, 20141128 (2015).

29. Batty, M. & Longley, P. A. *Fractal cities: a geometry of form and function* (Academic press, 1994).

30. Encarnação, S., Gaudiano, M., Santos, F. C., Tenedório, J. A. & Pacheco, J. M. Fractal cartography of urban areas. *Scientific Reports* **2**, 1–5 (2012).

31. 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).

32. Sen, P. K. Estimates of the regression coefficient based on kendall's tau. *Journal of the American statistical association* **63**, 1379–1389 (1968).

33. Woldenberg, M. J. An allometric analysis of urban land use in the united states. *Ekistics* 282–290 (1973).

34. Coffey, W. J. Allometric growth in urban and regional social-economic systems. *Canadian Journal of Regional Science* **11**, 49–65 (1979).

35. Chen, Y., Wang, J. & Feng, J. Understanding the fractal dimensions of urban forms through spatial entropy. *Entropy* **19**, 600 (2017).

36. Falconer, K. *Fractal geometry: mathematical foundations and applications* (John Wiley & Sons, 2004).

37. Wilson, A. *Entropy in urban and regional modelling*, vol. 1 (Routledge, 2011).

38. Knox, P. L. & McCarthy, L. *Urbanization: an introduction to urban geography* (Pearson Prentice Hall Upper Saddle River, NJ, 2005).

39. Kang, C., Ma, X., Tong, D. & Liu, Y. Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and its Applications* **391**, 1702–1717 (2012).

40. Lu, M. Analyzing migration decisionmaking: Relationships between residential satisfaction, mobility intentions, and moving behavior. *Environment and Planning A* **30**, 1473–1495 (1998).

41. Jiang, S. *et al.* The timegeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences* **113**, E5370–E5378 (2016).

42. Manson, S. *et al.* Methodological issues of spatial agent-based models. *Journal of Artificial Societies and Social Simulation* **23** (2020).

43. Karemera, D., Oguledo, V. I. & Davis, B. A gravity model analysis of international migration to north america. *Applied economics* **32**, 1745–1755 (2000).

44. Dagger, R. Metropolis, memory, and citizenship. *American Journal of Political Science* 715–737 (1981).

45. Wissen, L. J. G. & Bonnerman, F. A dynamic model of simultaneous migration and labour market behaviour (1991).

46. 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).

47. 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).

48. Batty, M. A theory of city size. *Science* **340**, 1418–1419 (2013).

49. Millington, J. D., OSullivan, D. & Perry, G. L. Model histories: Narrative explanation in generative simulation modelling. *Geoforum* **43**, 1025–1034 (2012).

50. Makse, H. United states and great britain cities data sets (2008). URL "`https://hmakse.ccny.cuny.edu/software-and-data/`".

Supplementary Files

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

- si.pdf