Check for updates

Emergence of urban growth patterns from human mobility behavior

Fengli Xu[1], Yong Li[1]✉, Depeng Jin[1], Jianhua Lu[1] and Chaoming Song[2]✉

Cities grow in a bottom-up manner, leading to fractal-like urban morphologies characterized by scaling laws. The correlated percolation model has succeeded in modeling urban geometries by imposing strong spatial correlations; however, the origin of the underlying mechanisms behind spatially correlated urban growth remains largely unknown. Our understanding of human movements has recently been revolutionized thanks to the increasing availability of large-scale human mobility data. This paper introduces a computational urban growth model that captures spatially correlated urban growth with a micro-foundation in human mobility behavior. We compare the proposed model with three empirical datasets, discovering that strong social interactions and long-term memory effects in human movements are two fundamental principles responsible for fractal-like urban morphology, along with the three important laws of urban growth. Our model connects the empirical findings in urban growth patterns and human mobility behavior.

O ver a century ago, urban scientists envisioned that the city of tomorrow would have symmetric and regular morphologies as a result of tight top-down city planning rules that aim to optimize the accessibility of city function and space use in city development[1,2]. However, the increasingly available urban data suggest that cities grow in a bottom-up manner, calling for understandings of its micro-foundation[2–4]. Three fundamental empirical laws were later discovered[2,5,6]: first, the size distribution of the city follows a scaling law with an exponent of around –2.0, which implies that large cities are much rarer than small towns[5]. Second, 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 (r) to city centers[9–11]. A computational physics model (that is, diffusion-limited aggregation) was applied to model urban growth as an aggregation of physical particles[7,12]. Further works showed that correlated percolation is a better alternative to explain the emergence of the aforementioned laws[5]. A key finding of the correlated percolation model is the requirement for strong spatial correlation in urban growth to reproduce the scaling relations[13]. As a result, researchers proposed to model city growth as two processes: the emergence of new clusters and the growth of existing clusters[14]. A more recent work proposed a diffusion-limited gravitation model that combines diffusion-limited aggregation with the stochastic gravitation model[15] to capture spatial correlation; the model assigned the aggregating particles a stopping probability when it is closed to the developed urban sites[16].

Although these models successfully explain urban morphology, they all model city growth as the stochastic processes of physical particles, which leaves the micro-foundation that roots in human behavior largely unknown. Here we develop a computational urban growth model based on human movements; the model suggests that strong social interaction and long-term memory effects in human movements are two fundamental principles governing urban growth.

Results

Human movement models. Our understanding of human movements has been revolutionized over the past decade thanks to the availability of large-scale human mobility datasets[17,18]. Existing human movement models can be classified into a paradigm of four categories based on the mechanisms of social interaction and long-term memory (see Fig. 1a). Specifically, a human mobility model is considered socially interactive if one individual's movement affects the others; a human mobility model is considered memory-aware if an individual's future mobility behavior is affected by his/her historic movements. Category-1 models treat human movements as randomly moving particles without interactions. Brownian movement is one of such models where an individual's displacements are normally distributed[19]. Unlike physical particles, empirical data suggest human movements are characterized by a fat-tailed jump-size distribution, satisfying a power law,

$$P(\overrightarrow{v} | \overrightarrow{v}_0) \approx \frac{1}{|\overrightarrow{v} - \overrightarrow{v}_0|^{(d+\alpha)}}, \qquad (1)$$

where $P(\overrightarrow{v} | \overrightarrow{v}_0)$ is the transition probability from location $\overrightarrow{v}_0$ to $\overrightarrow{v}$, and $d = 2$ for two-dimensional space[17,20]. The exponent α is observed at around 0.55 ± 0.05 (ref. [21]). The fact that the transition probability decreases with distance characterizes the cost of travel distance of human movements, that is, most of the time people travel only over short distances, whereas occasionally people take longer trips. Neglecting social interactions and long-term memory effects, equation (1) suggests human movements follow the Lévy-flight model, in which the evolution of population density $\rho(\overrightarrow{v}, t)$ follows the fractional diffusion equation,

$$\frac{\partial \rho(\overrightarrow{v}, t)}{\partial t} = -D(-\Delta)^{\alpha/2} \rho(\overrightarrow{v}, t), \qquad (2)$$

where $(-\Delta)^{\alpha/2}$ is a fractional Laplacian and D is a diffusion constant (see the Methods for details). Nevertheless, both the Brownian movement and Lévy-flight models predict a uniform population distribution when time t approaches infinity, in contrast to empirical observations[9].

[1]Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing, China. [2]Department of Physics, University of Miami, Coral Gables, FL, USA. ✉e-mail: liyong07@tsinghua.edu.cn; c.song@miami.edu
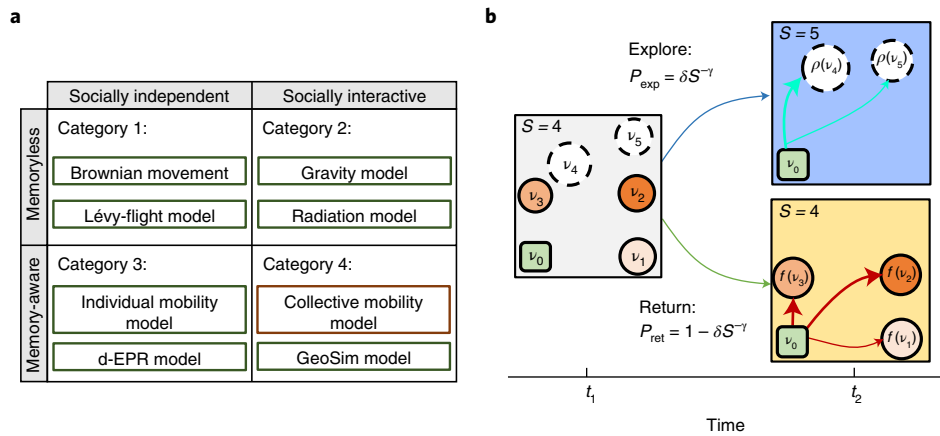
Fig. 1 | The paradigm of human movement models and illustration of the proposed collective mobility model. a, Existing human movement models can be classified into four categories based on whether they account for social interactions and the memory of historic movements: the Brownian movement and Lévy-flight models belong to the first category, where movements are socially independent and memoryless. Gravity and Radiation models are typical models of the second category, where movements are socially interactive and memoryless. IMM and d-EPR belong to the third category, where movements are socially independent and memory-aware. The fourth category is both memory-aware and social interactive, including the GeoSim model and the proposed CMM. **b**, An illustration of the proposed CMM. At time t_1, an individual is located at site v_0 (green node) and have previously visited $S = 4$ sites (left panel). Here, a larger node size represents a larger population size at that site, whereas a more intense node color represents a higher past visitation frequency. With $P_{exp} = \delta S^{-\gamma}$ probability, the individual will explore a new site (white circle) based on equation (3). Otherwise, with $P_{ret} = 1 - \delta S^{-\gamma}$ probability, the individual will return to a previously visited site (orange circle) based on equation (4).

Category-2 models such as Gravity[8] and Radiation[22] originate from the study of migrations, where the mobility flow between two locations depends on their populations. For instance, the Gravity model suggests the transition probability,

$$P(\overrightarrow{v} | \overrightarrow{v}_0) \approx \frac{\rho(\overrightarrow{v}) + \rho_0}{|\overrightarrow{v} - \overrightarrow{v}_0|^{(d+\alpha)}}, \tag{3}$$

where ρ_0 is the (inverse) coupling constant. In addition to the fat-tailed jump-size distribution equation (1), the Gravity model equation (3) also requires that the transition probability increases linearly with the population at the destination $\overrightarrow{v}$ (ref. [23]). This mechanism accounts for a mean-field background attractiveness rooted in social interaction, for example, highly populated locations often offer more social opportunities[8]. The social attractiveness provides a potential mechanism to reproduce the spatial correlation in the correlated percolation model. Unfortunately, we find that the diffusion process of the Gravity model follows the same fractional diffusion equation (2) for the Lévy-flight model (see the Methods for details), that is, it predicts a uniformly distributed urban population at a large value of t. More recently, research explored the correlation between human mobility and social network, finding that human movements are largely determined by underlying social ties[24–26]; however, these models often require inputs of social network data measured from experiments, leaving the origin of these correlations unexplained.

Unlike Category-1 and Category-2, in which individuals move freely, empirical data found notable recurrent visitation patterns in human movements[21]. To explain these observations, the individual mobility model (IMM) re-treats human movements as a two-stage return–exploration process to account for long-term memory effects[21]. In particular, a preferential return mechanism is imposed, that is, the probability of returning to a previous location $\overrightarrow{v}_i$,

$$P(\overrightarrow{v}_i) \propto f(\overrightarrow{v}_i), \tag{4}$$

which is proportional to its historic visitation frequency, $f(\overrightarrow{v}_i)$. Such a long-term memory return process slows human diffusion

drastically. In particular, IMM predicts that the typical traveling distance l_t follows a logarithmic growth as

$$l_t \approx \log A, \tag{5}$$

where A is the total number of visited locations[21,27]. Logarithmic growth is a key feature of human movements, characterizing the anomalous ultra-slow diffusion and home range effect[28]. More recently, the d-EPR model generalizes the IMM by introducing a background field where an individual explores a new location with a probability proportional to its population density[29,30]. Although IMM and d-EPR successfully capture individual movements on a daily basis, individuals move independently and do not interact with each other. Moreover, the background field in the d-EPR model is static and does not evolve with time. These models therefore cannot capture the dynamic interactions among individuals and, consequently, are not capable of reproducing urban growth patterns (see Supplementary Section 1 for more details).

Neither social interactions (Category-2) nor long-term memory effects (Category-3) alone reproduces the spatial correlation proposed in the correlated percolation model; thus, a natural question is whether the remaining Category-4 models (mobility models that integrate social interaction and long-term memory effects) can predict urban growth patterns. Although Category-4 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 their friends[31]; however, the social network is predetermined from the inputs of empirical data and does not evolve with time. It is therefore not able to track urban growth on a long timescale.

The collective mobility model. To integrate long-term memory effects with dynamic interactions, we propose the collective mobility model (CMM), which offers a minimal set of principals connecting human movements to urban growth (see Fig. 1b). We consider that individuals move according to a return–exploration process similar to IMM. Specifically, the probability that an individual decides to return to visited locations can be computed as $P_{ret} = 1 - \delta S^{-\gamma}$,
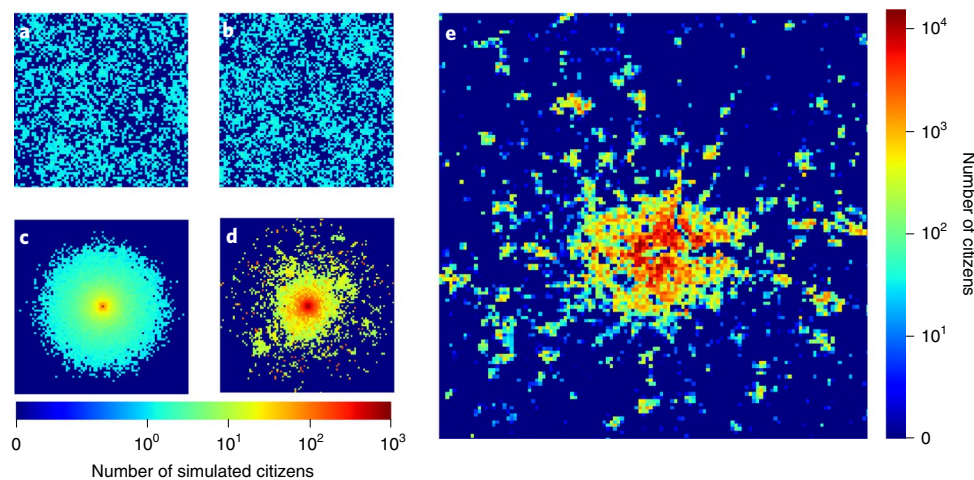
Fig. 2 | The morphologies of urban area generated by four different human movement models and in real-world city. a–e, Data generated by the Lévy-flight model (**a**), Gravity model (**b**), IMM (**c**) and CMM (**d**), as well as empirical data from the city of London (**e**). The population distribution of each urban system is visualized as a heatmap on a logarithmic scale, where blue and red colors represent underpopulated and high-population regions, respectively. The urban systems are simulated with 30,000 individuals initially situated in urban centers, which is defined as the central site of simulation space. The individuals are then simulated to move according to the mobility models until the population distribution reaches a stable state.

where S is the number of previously visited locations, δ controls the initial return probability and γ controls how the return probability changes with S. During the return, the probability to choose a specific visited location is proportional to his/her previous visitation frequency to that location, that is, equation (4). The individual also has a complementary probability $P_{exp} = \delta S^{-\gamma}$ to explore a new location. During exploration, the individual will choose previously unvisited locations on the basis of distance and population, satisfying the gravity law of equation (3). The population $\rho(\overrightarrow{v})$ in equation (3) is calculated instantaneously on the basis of all individuals's locations at time t. The coupling constant, ρ_0^{-1}, can be considered as a smoothing parameter that controls the strength of population attraction, that is, increasing ρ_0^{-1} reduces the probability of exploring unpopulated locations and, consequently, increases the strength of social interactions.

Inspired by the strong spatial correlation in correlated percolation models[5], we are interested in a CMM with the strong coupling limit $\rho_0^{-1} \to \infty$, describing a unique human mobility model with both dynamic social interaction and long-term memory effects, which has not been explored by previous research. As a result, the proposed CMM can simulate urban growth to arbitrary timescales in a self-organized manner and does not rely on any external input. It is also worth pointing out that IMM can be considered as a special case of CMM, with $\rho_0^{-1} \to 0$.

Experimental settings. We collect three public available urban development datasets, including the population and urban area of cities in (1) the United States of America (US) in 2000, (2) Great Britain (GB) in 1991 and (3) the distribution of urban area in the Berlin region in 1910, 1920 and 1945. For comparison, we simulate human mobility with four typical movement models belonging to categories 1–4, namely the Lévy-flight model, Gravity model, IMM and the proposed CMM, respectively (see Supplementary Section 1 for the details of model parameter settings). The simulations are run on a $l \times l$ two-dimensional simulation space, where l is the number of sites in each dimension. The simulated citizens are initially located at the city center, that is, the central location $(l/2, l/2)$ in simulation space; we then iteratively sample their next locations according to the mobility models at each time step until the simulation converges. A simple simulation of large urban systems is impractical due to 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 an improved sampling technique to effectively reduce the complexity to $\mathcal{O}(M)$ (see the Methods for details).

Urban morphology. To compare the morphologies of the simulated urban systems with empirical observations, we plot the obtained population distributions in Fig. 2a–d, together with the empirical distributions of the city of London in Fig. 2e. Although real-world geometry is affected by geographical features such as lakes and rivers, the city of London still exhibits prominent features of the compact city center and fractal perimeters. These observations echo previous studies on the fractal geometry of urban area[32–34].

The urban population distributions for the Lévy-flight and Gravity models follow the fractional diffusion process (equation (2)), implying that individuals will gradually diffuse away from their initial position over time. The simulation results verify the theoretical prediction that these two models cannot reproduce a compact and stable city center, as the simulated urban population is distributed uniformly in the urban space when the systems converge (see Fig. 2a,b). On the other hand, IMM predicts that 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 a similar population density (see Fig. 2c), which is consistent with the theoretical prediction and shows IMM cannot reproduce the fractal morphology of the urban area. By contrast, CMM successfully reproduces the compact city center in an urban system, where the population density is substantially higher than the peripheral urban area (see Fig. 2d). Besides, the perimeter of the city center demonstrates prominent fractal geometry and numerous subclusters are formed around it (see Supplementary Section 4 for details). These observations are in agreement with the empirical observation of the city of London, which indicates CMM can reproduce the morphology of the urban area. We also evaluate the proposed CMM on predicting the future morphology change when given the initial urban area distribution. Experiments show CMM considerably outperforms baselines in terms of the reproduced urban morphology and the prediction accuracy on whether a specific location will be developed in the future (see Supplementary Section 3 for more details).
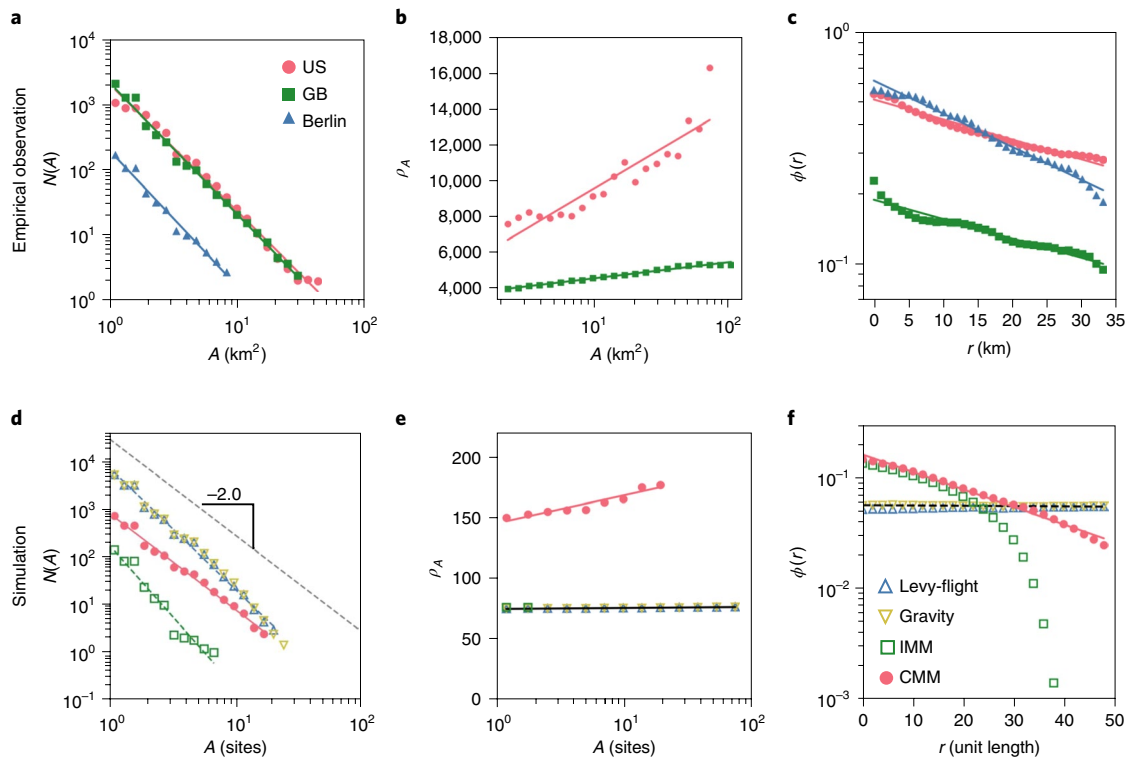
Fig. 3 | Comparisons between the empirical urban growth patterns in the US, GB and Berlin, and the simulated urban growth patterns driven by different human movement models. The dots represent the empirical and simulated data points, whereas the straight lines are the fitted urban growth patterns. **a**, The empirical city size distributions. The estimated exponents are $\tau = 1.94$ (US), $\tau = 2.01$ (GB) and $\tau = 2.08$ (Berlin). The 95% confidence intervals based on Theil–Sen estimator are [1.83, 2.05] (US), [1.93, 2.09] (GB) and [1.90, 2.26] (Berlin). **b**, The empirical relation between ρ_A and A. **c**, The empirical urban occupation profiles, where $\phi(r)$ denotes the urban occupation density at radial distance r to the city center. The estimated exponents are 0.3 (US), 0.3 (GB) and 0.4 (Berlin). **d**, The simulated city size distributions by different human movement models. The estimated exponents are $\tau = 2.55$ (Lévy-flight), $\tau = 2.58$ (Gravity), $\tau = 2.98$ (IMM) and $\tau = 2.02$ (CMM). The 95% confidence intervals based on Theil–Sen estimator are [2.40, 2.70] (Lévy-flight), [2.41, 2.75] (Gravity), [2.47, 3.49] (IMM) and [1.89, 2.15] (CMM). The dotted line shows the scaling law with an exponent of 2.0, as predicted by strongly correlated percolation theory. All four models reject the zero exponent null hypothesis via two-sided Wald Test with t-distribution: Lévy-flight (P-value < 0.05, $n = 17$), Gravity (P-value < 0.05, $n = 18$), IMM (P-value < 0.05, $n = 11$) and CMM (P-value < 0.05, $n = 16$). **e**, The simulated relations between urban population density and city size by different human movement models: Lévy-flight (P-value $= 1.0$, $n = 14$), Gravity (P-value $= 1.0$, $n = 14$) and IMM (P-value $= 1.0$, $n = 3$) cannot reject the zero exponent null hypothesis via a two-sided Wald Test with t-distribution, whereas CMM (P-value < 0.05, $n = 10$) rejects the null hypothesis with statistical significance. **f**, The simulated urban occupation profiles by different human movement models: Lévy-flight (P-value $= 1.0$, $n = 25$) and Gravity (P-value $= 1.0$, $n = 25$) cannot reject the zero exponent null hypothesis via a two-sided Wald Test with t-distribution, whereas IMM (P-value < 0.05, $n = 22$) and CMM (P-value < 0.05, $n = 25$) reject the null hypothesis with statistical significance.

Urban growth patterns. To examine the model's capability to reproduce urban growth patterns, we focus on three fundamental empirical laws, each of which has been validated on multiple cities around the globe[2,5,6].

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

$$N(A) \approx A^{-\tau}, \qquad (6)$$

where the exponent τ has been reported[5] to be around 2.0. Specifically, the scaling law of city size is equivalent to Zipf's law, with an exponent of $\tau - 1$ (ref. [35]). The scaling law with an exponent of 2.0 is therefore consistent with the previous observations on Zipf's law of city size distribution[35]. Percolation theory is the prevalent framework to characterize this observation, which can generate a probability that a certain location is occupied as an urban area[13]. It predicts the scaling law of equation (6) with exponents ranging between 2.0 and 2.5, where $\tau = 2$ corresponds to a strong correlation between different locations and $\tau = 2.5$ corresponds to a mean-field theory[13]. The empirical data shows that city size distributions are well approximated by equation (6) (see Fig. 3a), with

$\tau = 1.94 \pm 0.11$, 2.01 ± 0.08 and 2.08 ± 0.18 for the US, the GB and Berlin, respectively. These findings echo the theoretical predictions of site percolation theory and empirical observations in the previous research[5,13].

The Lévy-flight model characterizes movement as an individual diffusion process. The urban population distributes uniformly in the urban space as the urban system reaches to a stationary state. Similarly, although the Gravity model introduces social interactions among individuals, the population density satisfies fractional diffusion equation (2). When time $t \rightarrow \infty$, the population will distribute uniformly in the urban space $\rho(r) = c$, which is independent of the coupling constant ρ_0^{-1}. The Lévy-flight and Gravity models are therefore expected to generate the scaling law with an exponent of around 2.5 in an uncorrelated percolation[13]. To test this hypothesis, we identify the isolated connecting components in the simulation as cities and measure the area of each city[5]. Figure 3d shows that both the Lévy-flight and Gravity models reproduce a power-law city size distribution with $\tau = 2.55 \pm 0.15$ and $\tau = 2.58 \pm 0.17$, which is consistent with the theoretical prediction. The simulation also shows IMM satisfies the scaling law with $\tau = 2.98 \pm 0.51$. This large exponent implies the fact that individuals are localized within their own
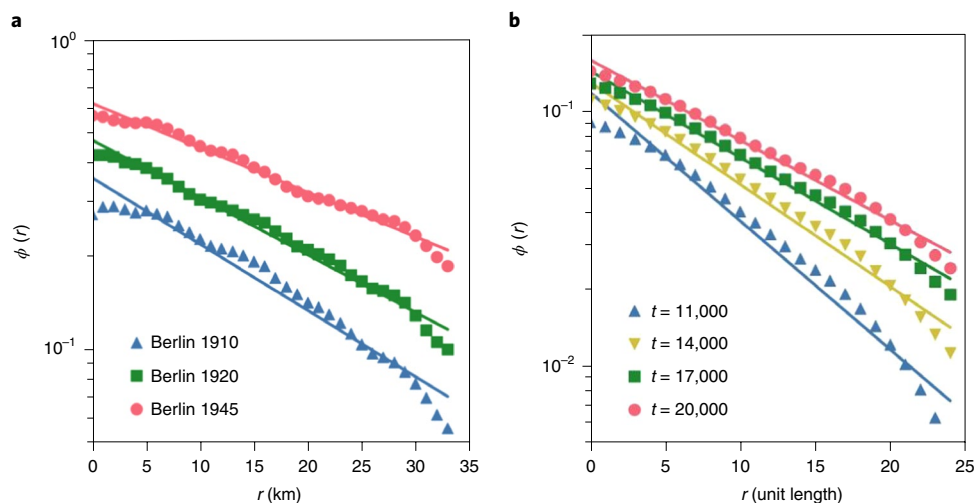
Fig. 4 | The time evolution of urban occupations. a,b, The urban occupation profile $\phi(r)$ as a function of r for Berlin (**a**) and CMM (**b**). Straight lines represent exponential fittings for the corresponding empirical data and simulation results.

home range, as the IMM is equivalent to the non-interactive limit of CMM with $\rho_0^{-1} \to 0$. By contrast, when the coupling constant $\rho_0^{-1} \to \infty$, the CMM 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 range within a 95% confidence interval based on the Theil–Sen estimator[36] (see Supplementary Section 6 for details); however, we find that the proposed CMM is the only model that contains the empirically observed universal exponent of 2.0 in its 95% confidence interval. These results suggest combining the principals of social interaction and long-term memory effects in human mobility can reproduce the empirical city size distribution, which is successfully achieved by CMM.

(ii) *Super-linear relation between population and city size:* the positive allometric population growth with city size is widely observed in cities around the globe[37,38]. Larger cities tend to have a higher urban population density defined as, $\rho_A \equiv N(A)/A$. Researchers believe that is because the competition for space is more intense in big cities and there are more high-rise buildings to handle denser populations[32,39,40]. Recent research suggested that the balance between the cost and gain of concentrating population in urban areas would explain the observed super-linear growth[6]. This socio-economic hypothesis consists of two assumptions: (1) the average gain from the intense social interaction is proportional to the population density ρ_A, (2) the average living cost is proportional to the typical travel distance l_t to explore the city. Their balance leads to $\rho_A \approx l_t$ (ref. [6]). Substituting equation (5), we have

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

The first assumption agrees with the social interaction in equation (3), whereas the second is rooted in the long-term memory effect in equation (4).

Figure 3b plots ρ_A with A across different cities in both US and GB, finding that the empirical observation agrees precisely with the predicted logarithmic law of equation (7). It is worth noting that previous studies reported most cities follow scaling laws (that is, $\rho_A \approx A^\delta$) with tiny exponents $\delta \approx 0.04 \sim 0.3$ (ref. [6]); however, within the range of magnitude of the empirical data, the logarithmic relation is indistinguishable with a small-exponent scaling law. The logarithmic relation we identified in equation (7) therefore provides an alternative explanation to the origins of population growth patterns, which is derived from human mobility behavior. Figure 3e compares the simulation results for the four models, finding that

the proposed CMM accurately reproduces the logarithmic law; however, the simulation results of the other three models are fitted by lines with a coefficient of zero, which indicates that they predict ρ_A to be independent from the city size. This result demonstrates our model can outperform baselines in reproducing the empirically observed super-linear population growth.

(iii) *Exponential occupation profile:* the urban occupation profile $\phi(r)$ is defined as the probability of finding an populated area at the distance r from the city center, where an exponential profile is observed in previous study[41],

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

where λ is the exponent of the exponential occupation profile. Empirical measurements of the exponential urban occupation profiles have been made in several past works[5,41]. Figure 3c rechecks this pattern on all three empirical datasets, demonstrating that the city center attracts most of the population, whereas the occupation probability decreases rapidly with r.

However, such a rapid decline somehow contradicts the fat-tailed nature of equation (1), which suggests humans are able to reach areas far away from the initial location[20,32]. Indeed, simulation results in Fig. 3f show that the Lévy-flight and Gravity models fail to reproduce a city center and predict $\phi(r)$ do not change with with different r. The lack of social interaction makes IMM unable to produce large subclusters on the peripheral area of city, which results in dramatic occupation density drop at the edge of city area. On the contrary, by jointly modelling the social interaction and long-term memory effects, the occupation profile reproduced by CMM agrees well with the exponential law of equation (8).

Moreover, it has been suggested that the declining rate λ shall decrease as the city evolves due to the constantly expanding frontiers of cities[5], which is in line with the observations in the Berlin dataset where $\phi(r)$ has been measured at three different time. Figure 4a shows that 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). Specifically, as more population transit from the initially located city center to the peripheral area, the peripheral area will become more attractive for incoming movement based on the social interaction mechanism. On the other hand, the long-term memory mechanism will maintain a dense area in the city center. The combination of social interaction and long-term memory mechanisms will therefore lead to the gradually flattened occupation profile; such a combination
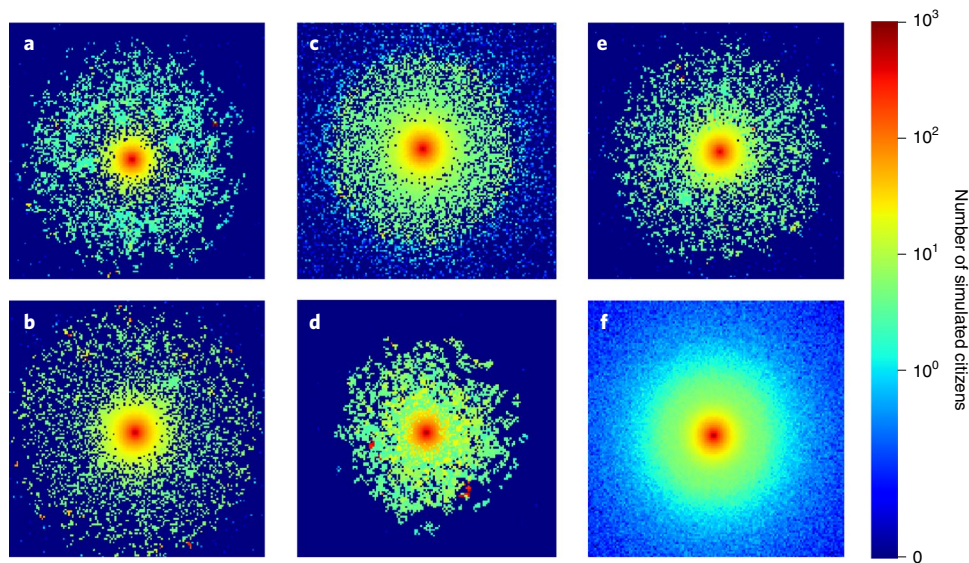
Fig. 5 | Qualitative comparison between the morphology of urban area reproduced by CMM with different parameter settings. The standard parameter setting is $\omega=1/3, \alpha=0.55, \rho_0=0$. **a-f**, We change the value of one parameter in each simulation: $\omega=1/6$ (**a**), $\omega=1/2$ (**b**), $\alpha=-0.45$ (**c**), $\alpha=1.55$ (**d**), $\rho_0=10^{-6}$ (**e**) and $\rho_0=1$ (**f**). The population distribution of the simulated urban system is visualized as heatmap on a logarithmic scale, where blue and red colors represent underpopulated and high-population regions, respectively.

demonstrated that CMM can simulate the gradually increased attractiveness of the peripheral area in a self-organizing manner. We also measure the time-varying occupation profiles of Lévy-flight, Gravity and IMM, finding that they do not capture the time evolution of $\phi(r)$ (see Supplementary Section 7 for details). Moreover, we find the more sophisticated d-EPR model with a static social interaction network[29,30] also fails to reproduce the occupation profile evolution (see Supplementary Section 1 for details). These results demonstrate that the dynamic interactions in CMM are critical for reproducing the dynamics of the urban growth.

Impact of simulation parameters. There are three important parameters that might influence the simulation of urban systems: α, ρ_0 and the average population density ω; α is the main parameter that characterizes the principal of cost of travel distance, which controls the distance distribution of each displacement. It is set as 0.55 in previous simulation based on empirical observations[21]. ρ_0 mainly controls the principal of social interaction, it determines the likelihood of individuals exploring more populated regions, and was set to $\rho_0^{-1} \to \infty$ in past simulations to reproduce the strongly interactive urban systems; ω is defined as the ratio between the number of citizens and the number of sites, that is, $\omega=M/l^2$. It represents the number of social interactions one would encounter per unit space. We now examine how these parameters may influence the characteristics of the proposed CMM.

The morphology of simulated urban areas and reproduced urban growth patterns with different parameter settings are present in Figs. 5 and 6, respectively. First, as shown in Fig. 5a,b, we observe that ω does not notably affect the fractal morphology of urban systems. Furthermore, as shown in Fig. 6a–c, the city size distributions, population growth patterns and urban occupation profiles all agree with the empirical observations under different ω settings. These results indicate the urban growth patterns are robust with different ω. Second, α has substantial impact on the compactness of simulated urban systems, which is probably because it controls the jump-size distribution of each exploration in equation (1). When α is smaller than the empirical value, the exploration will have larger jump-size; thus, as shown in Fig. 5c, the occupied areas of urban systems are more diverged and distributed further away from the city centers.

By contrast, Fig. 5d shows when α is larger than the empirical value the simulated urban system is more cohesive, and larger subclusters are produced. Moreover, Fig. 6d–f shows city size distributions, population growth patterns and urban occupation profiles all deviate from the empirical rules when α is different from the empirical value. These results show that α plays an important role in both urban morphology and urban growth patterns; thus, α is not only an important parameter in human mobility behavior, but also fundamentally governs urban growth. Third, ρ_0 largely affects the morphology of urban areas. Figure 5e shows that the simulated urban systems show prominent fractal morphology when $\rho_0^{-1} \to \infty$ as $\rho_0=10^{-6}$; however, when we increase ρ_0 to 1 in Fig. 5f, the fractal urban morphology is not reproduced and it shows homogeneous growth on the perimeter of the urban area. Furthermore, Fig. 6g shows that the exponents of city size distribution gradually increase when ρ_0 increases from 0 to 1. As shown in Fig. 6h,i, the population growth patterns and urban occupation profiles also deviate from empirical rules when ρ_0 increases to 1. These results indicate CMM can accurately reproduces the empirical rules of urban growth when $\rho_0^{-1} \to \infty$, which corresponding to a singular mobility model that has both strong social interaction and long-term memory effect. On the contrary, the CMM predictions gradually converge to IMM when $\rho_0^{-1} \to 0$, which is consistent with our theoretical analysis that IMM is a special case of CMM. These results show the parameters α and ρ_0 in CMM have important impact on the compactness of urban system and the fractal morphology, respectively. We therefore set $\alpha=0.55$ and $\rho_0^{-1} \to \infty$ in our CMM to follow the empirical observations and reproduce the strong social interactions.

Discussion

This paper seeks to establish a connection between urban growth and human mobility. It is worth pointing out that although most human mobility models are developed on the basis of the active mobility behavior that characterizes daily movements, urban growth is more closely related to the residential mobility, that is, the relocation of residence. Although residential and active mobility are human movements on different timescales, they share common features in behavior patterns. For instance, the Gravity model is often used in both scenarios[42]. Moreover, the long-term memory
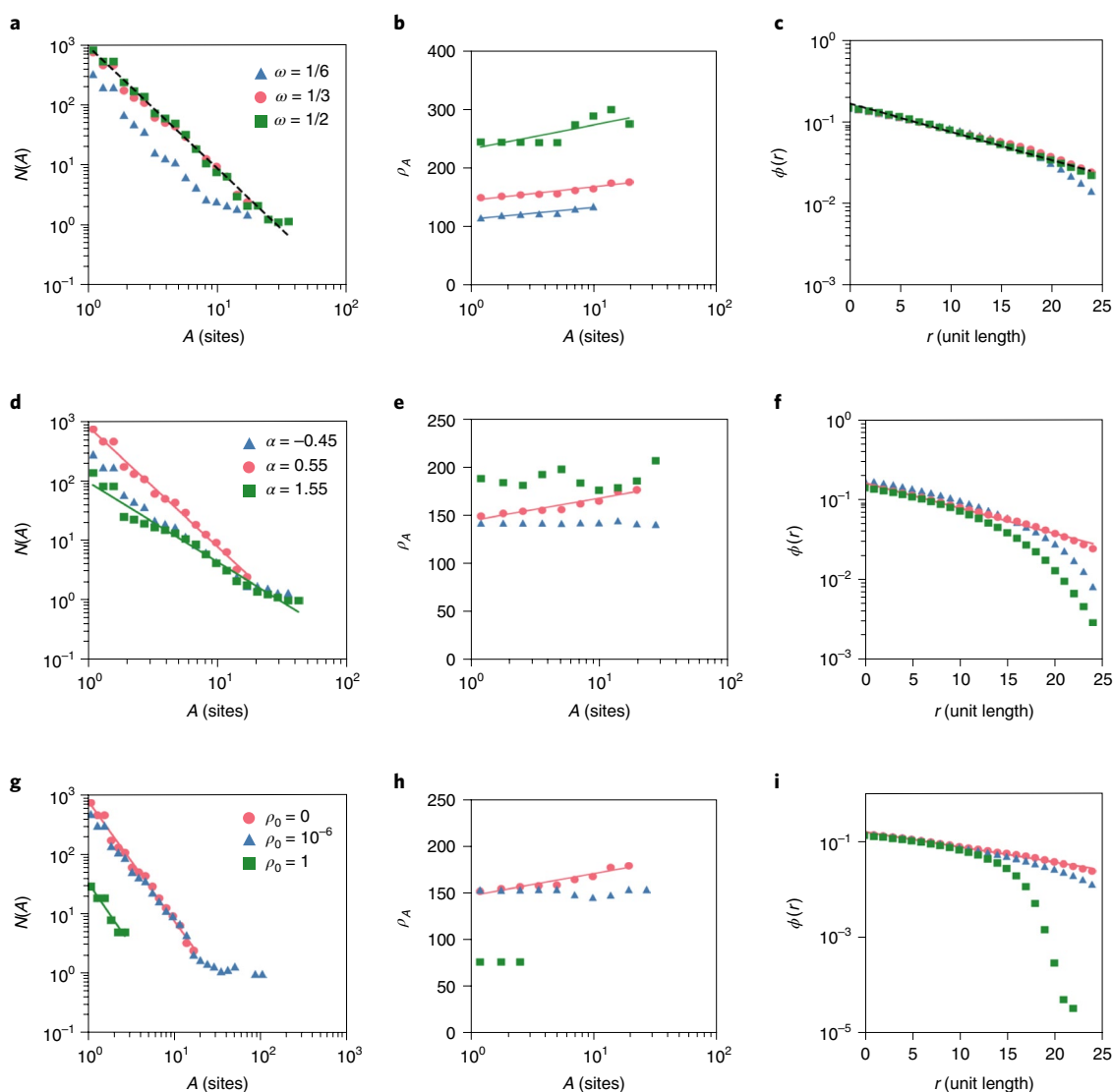
Fig. 6 | The comparison of reproducing urban growth patterns with different parameter settings. The standard parameter setting is $\omega = 1/3, \alpha = 0.55, \rho_0 = 0$. We change the value of one parameter in each panel. **a–c**, Reproduced urban growth patterns with varying ω. **d–f**, Reproduced urban growth patterns with varying α. **g–i**, Reproduced urban growth patterns with varying ρ_0.

effect is equally important for relocation due to numerous socio-economic factors such as the effect of civic memory and change of workplaces[43,44]. The CMM therefore provides a framework for both active and residential mobility, each corresponding to different experiment set-ups. For example, the relocation has a much smaller transition rate—that is, probability of movement per unit time—than the daily movements. Nevertheless, we do not set a specific timescale in the simulation and find that the universal urban laws observed in the proposed CMM are largely independent of choosing different modeling parameters (see Fig. 6). The application of simulations on different timescales will be investigated in future research. Besides, our study shows long-term memory and social interaction mechanisms are a minimum set of common principles in human mobility that governs the empirically observed universal laws of urban growth. On the other hand, the d-EPR model that leverages static social interaction network also cannot reproduce the urban growth, which shows the limitation of physical determinism models in capturing the self-organizing nature of urban system. By contrast, the proposed CMM is not solely physical deterministic, yet a highly simplified decision-making model where individuals choose their exploration locations based on the dynamic social

attractiveness of the locations (see Supplementary Section 5 for more details). Together with the long-term memory mechanism for the return processes, CMM is a human mobility model with a minimal set of decision-making processes that allows theoretical analysis and numeric simulations. We feel that the simplicity of the proposed model makes it a succinct and general framework for urban growth simulation.

Our study has direct implications on several downstream applications[45,46], including city planning, resource allocation, policy making and so on. Top-down city governing has been shown ineffective in foreseeing urban growth and optimizing urban functions by the previous research[2,47]. Predictive experiments show the proposed CMM outperform baseline models in predicting the morphology change of urban areas (see Supplementary Section 3 for details), which can help city planning and resource allocation to stay ahead of urban development and reduce inefficient investments. The proposed model provides a bottom-up perspective to understand how urban growth emerges from human mobility behavior. It allows the policy-makers to make sense of the driving force of human movements behind urban growth, which can inform the design of policies that optimize urban functions, for example, improving the

operation of transportation system and designing suitable relocation policies.

The present work can be improved from the following directions, which are left for future research. First, the proposed CMM provides a theoretical framework that connects urban growth and human movement; however, the detailed empirical mechanisms that contribute to this connection are not investigated in current research, which might require quantitative and qualitative analysis on the empirical data collected from heterogeneous sources. Second, the proposed CMM is computation intensive due its strongly interactive nature. We propose two sampling methods to improve the efficiency of simulation (see Methods for details); however, further optimization is needed to simulate large-scale urban systems with millions of citizens. Third, the proposed CMM is general and can be further optimized for specific application domain by including relevant contextual features. For example, although we demonstrate the predictive power of our model, it can be further augmented by explicitly considering the contextual features such as lakes, rivers and road networks.

Methods

Urban growth simulation. We simulate urban growth with M individual citizens moving according to the mobility models and considered the occupied sites as developed urban areas. Specifically, urban growth simulation is performed on a two-dimensional square simulation space with l sites in each dimension for T time steps. The simulated citizens are initially located at the central site $(l/2, l/2)$ of simulation space, and then iteratively sample their next sites according to the mobility models at each time step. After the system is converged, we considered the populated sites as an urban area, and examine the morphology and growth patterns of the simulated urban system. Following the settings in previous work[5], we identify the connected components of urban area as cities and group them in logarithmically spaced bins. The patterns of city size distribution and urban population growth are reported by averaging over the cities in each bin with sliding windows. As for the urban occupation profile, we use sliding windows gradually moving from city centers to frontier to capture the gradual change of urban occupation density. We perform statistical analysis to test the robustness of simulated patterns (see Supplementary Section 6 for more details).

Specifically, we set $l = 300, M = 30,000, T = 20,000$ in the main simulation experiment. The simulation size is typical for a strongly interactive agent-based model[48], where the computational complexity is considerably high. In each step, it requires to sample the next site of each individual from the entire universe of sites based on the probability distribution derived from human models. The overall time complexity of the memoryless mobility model—that is, Lévy-flight and Gravity models—is therefore $\mathcal{O}(TMA)$. As for the memory-aware mobility model, that is, IMM and CMM, we define θ_E and θ_R as the average probability of exploration and preferential return, respectively, where $\theta_E + \theta_R = 1$, and $\hat{p}$ as the average number of previously visited sites; the overall time complexity is then $\mathcal{O}(TM(\theta_E A + \theta_R \hat{p}))$. Although the simulation can be accelerated with parallel processing as the sampling at each step is independent among individuals, the computation complexity is still not scalable to simulate a large urban system. For example, it requires over 690 h to simulate a $l = 300, M = 30,000, T = 20,000$ urban system on a computer with 40 cores of 3.6 GHz Intel i7 processors. We therefore propose two improved sampling methods—alias sampling and sorted array sampling—to improve the efficiency of the simulation.

The alias sampling method aims to optimize the exploration mechanism that is important in all four mobility models, which characterizes how citizens travel to new sites. It is especially time-consuming as in each step it requires sample a site based on a certain probability distribution from the entire universe of sites for each citizen. As a result, we design an improved sampling method by exploiting the insight that the probability distributions for the citizens in the same sites are identical in each step, which allows us to reuse the probability distribution among the citizens in the same sites. To achieve this goal, we implement an alias sampling technique to draw subsequent samples from the same distribution in $\mathcal{O}(1)$ complexity[49] (see Supplementary Fig. 2). Denote the sampling space as $a \in \{1, 2, ..., A\}$, the probability of sampling a as $p[a]$. The alias sampling algorithm is described as two subprocesses: creating an alias table and generate subsequent samples. In the first subprocess, the algorithm works by reshaping the probability distribution over A sites into A pairs of sites $\{b_i | \text{ for } i \in \{1, 2, ..., A\}\}$, and the summed up probability of each pair of sites is identical, namely, A^{-1}. This subprocess is also called the Robin Hood method as it requires to break down the sites with probabilities $p[a] > A^{-1}$ and pair them with the sites with probabilities $p[a] <= A^{-1}$. In the second subprocess, the algorithm first samples a pair of sites from $\{b_i | \text{ for } i \in \{1, 2, ..., A\}\}$ based on uniform probability distribution. Then, sample a site from the site pair based on its corresponding probability. It is straightforward to prove that each sample generated by the second subprocess follows the exact probability distribution defined by $p[a]$. Note that the time

complexity of creating alias table in the first subprocess is $\mathcal{O}(A)$; however, the alias table can be reused to generate subsequent samples from the same probability distribution at $\mathcal{O}(1)$ complexity. Therefore, the amortized time complexity per draw is $\mathcal{O}(1)$ when the number of samples is large. In our simulations, a large number of citizens share the same probability distributions since they are in the same sites. Therefore, the alias sampling technique is able to effectively reduce the complexity of exploration mechanism from $\mathcal{O}(TMA)$ to approximately $\mathcal{O}(TM)$, which renders the time complexity independent with the size of site universe and substantially speeds up the simulation.

The sorted array sampling method aims to improve the efficiency of simulating preferential return mechanism in IMM and CMM. The probability of preferential return is positively correlated with the number of sites individuals previously visited S. Specifically, we can implement the preferential return mechanism by first sampling an integer from a uniform distribution on the number of total visits of an individual, $[0, \sum_i f(v_i)]$. Then, we cumulatively summed up the frequency of each site and select the site to render the sums first above the sampled integer. Therefore, the time complexity of each preferential return is proportional to S since we need to go through all visited sites, and it will be computation expensive when citizens have visited a large number of sites. An important observation is that the probability distribution is skewed among all the sites visited by individuals since the probability is proportional to the visitation frequency associated with each site. Therefore, an intuitive optimization solution is to sort the array of previously visited sites by the frequency, and first examine the site with highest frequency since they are most likely to be sampled. However, sorting an array with S elements requires a time complexity of $\mathcal{O}(S \log_2 S)$, which is more expensive than the complexity of sampling $\mathcal{O}(S)$. To address this problem, we leverage the insight that the orders of the array will not change notably between two neighboring steps. Therefore, we implement a sorted array data structure to enable the reuse of the array of sorted sites among neighboring steps. Specifically, the data structure contains the pairs of sites and the associated visitation frequency. If the citizens perform a preferential return, we update their sorted arrays by increasing the frequency of sampled sites and moving up the corresponding pairs until the arrays become sorted again. If the citizens perform an exploration, we add a new pair of sites and frequency to the bottom of the array. It is straightforward to prove that the average time complexity of maintaining a sorted array data structure is less than $\mathcal{O}(\log_2 S)$, which effectively improved the efficiency of simulating preferential return mechanism.

These optimization techniques allow us to reduce the computational time of simulating a $l = 300, M = 30,000, T = 20,000$ urban system from 690 h to 12 h on a workstation with 40 cores of 3.6 GHz Intel i7 processor, which makes it feasible to simulate large-scale urban systems.

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

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

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

$$\frac{d\rho_i(t)}{dt} = \sum_j \left[W_{ij}\rho_j(t) - W_{ji}\rho_i(t) \right] = \lambda_g \rho_0 \sum_j \frac{\rho_j(t) - \rho_i(t)}{|\vec{v}_i - \vec{v}_j|^{d+\alpha}}. \tag{10}$$

Assuming a small site spacing h, we approximate the summation by a continuous integration,

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

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 $h \to 0$ requires the existence of $\lim_{h \to 0} \lambda_g(h)h^\alpha$, that is, $g(h) \sim h^{-\alpha}$ for small h. Substituting equation (11) by equation (10) leads to equation (2), where the diffusion constant $D \equiv c_{d,\alpha} \rho_0 \lim_{h \to 0} \lambda_g(h)h^\alpha$.

For $\alpha = 2$ we recover the standard diffusion equation. Therefore, as long as the diffusion constant D is positive, the simulated urban population will diffuse to eventually follow a uniform distribution. Otherwise, negative D corresponds to a unusual negative diffusion process where urban area cannot exist stably and will sink to a single point.

Data collection and calibration. All datasets used in this paper are publicly available. Specifically, the data of the urban area and population in the US was

originally released by the US Census Bureau for the year 2000[50]. It contains the information of urban size and population in the granularity of neighborhoods or towns. Besides, the data of cities in Great Britain was originally released by the Statistical Office of the European Union for the year 1991[51]. It provides information on the city's coverage and population distribution in the resolution of discrete 200-meter squares. We use the preprocessed datasets released by previous research[52]. The dataset of the Berlin region is extracted from the telemetry images of settlement area distributions in 1910, 1920 and 1945 published by previous work (see Fig. 4.1 in ref. [53]). We follow the procedures proposed by previous researchers[5,54] to digitize the empirical data into discrete sites. The telemetry images provide information on whether a site belong to urban area, but not the urban population distribution.

Statistics and reproduciblility. We use two-sided Wald test with t-distribution[55] to examine the null hypothesis that human movement models reproduce urban growth patterns with zero exponents. The detailed statistical analysis results can be found in Supplementary Table 2.

As for the city size distribution, we find all four models reject the null hypothesis with statistical significance: Lévy-flight model (P-value $< 0.05, n = 17$), Gravity model (P-value $< 0.05, n = 18$), IMM (P-value $< 0.05, n = 11$) and CMM (P-value $< 0.05, n = 16$). Moreover, the theoretical analysis and empirical observations both indicate the exponent should be two due to the strongly correlated urban development[5]. To test this pattern, we use the Theil–Sen estimator to compute the confidence intervals of the exponent reproduced by each model[36]. As a result, the 95% confidence intervals for the Lévy-flight model, Gravity model, IMM and CMM are [2.40, 2.70], [2.41, 2.75], [2.47, 3.49] and [1.89, 2.15], respectively. Only the proposed CMM contains the empirical observed two exponents in its 95% confidence interval.

As for the super-linear urban population growth, Supplementary Table 2 shows Lévy-flight model (P-value $= 1.0, n = 14$), Gravity model (P-value $= 1.0, n = 14$), IMM (P-value $= 1.0, n = 3$) cannot reject the null hypothesis of zero exponents, whereas CMM (P-value $< 0.05, n = 10$) rejects the null hypothesis with statistical significance.

As for the urban occupation profile, we find that the Lévy-flight (P-value $= 1.0, n = 25$) and Gravity (P-value $= 1.0, n = 25$) models cannot reject the null hypothesis of 0 exponent, whereas IMM (P-value $< 0.05, n = 22$) and CMM (P-value $< 0.05, n = 25$) reject the null hypothesis with statistical significance.

Data availability
The empirical urban datasets that support the findings of this study are public available. The datasets for US and GB were released by previous research[52]. The Berlin dataset is extracted from the telemetry images in previous works[53] (see Data Collection and Calibration for details), which is available in GitHub, https://github.com/tsinghua-fib-lab/Collective-Mobility-Model[56]. Source data for Figs. 3, 4 and 6 are available with this manuscript.

Code availability
The source code for numeric simulation is available online: https://github.com/tsinghua-fib-lab/Collective-Mobility-Model (ref. [56]).

References
1. Howard, E. *To-morrow: A Peaceful Path to Real Reform* (Cambridge Univ. 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. *Phys. Rev. Lett.* **47**, 1400 (1981).
8. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
9. Clark, C. Urban population densities. *J. R. Stat Soc. A* **114**, 490–496 (1951).
10. Newling, B. E. The spatial variation of urban population densities. *Geogr. Rev.* **59**, 242–252 (1969).
11. McDonald, J. F. Econometric studies of urban population density: a survey. *J. Urban Econ.* **26**, 361 (1989).
12. Vicsek, T. *Fractal Growth Phenomena* (World scientific, 1992).
13. Essam, J. W. Percolation theory. *Rep. Prog. Phys.* **43**, 833 (1980).
14. Schweitzer, F. & Steinbrink, J. Estimation of megacity growth: simple rules versus complex phenomena. *Appl. Geogr.* **18**, 69–81 (1998).
15. Rybski, D., Ros, A. G. C. & Kropp, J. P. Distance-weighted city growth. *Phys. Rev. E* **87**, 042114 (2013).
16. Rybski, D., Li, Y., Kropp, J. P. & Born, S. Modeling urban morphology by unifying diffusion-limited aggregation and stochastic gravitation. *Urban Findings* https://doi.org/10.32866/001c.22296 (2021).
17. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
18. Schrank, D., Eisele, B. & Lomax, T. *TTI's 2012 Urban Mobility Report* Vol. 4 (Texas A&M Transportation Institute, The Texas A&M University System, 2012).
19. Einstein, A. *Investigations on the Theory of the Brownian Movement* (Courier Corporation, 1956).
20. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
21. Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nat. Phys.* **6**, 818–823 (2010).
22. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
23. 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 $\rho_o \to 0$.
24. Grabowicz, P. A., Ramasco, J. J., Gonçalves, B. & Eguíluz, V. M. Entangling mobility and interactions in social media. *PLoS ONE* **9**, e92196 (2014).
25. Deville, P. et al. Scaling identity connects human mobility and social interactions. *Proc. Natl Acad. Sci. USA* **113**, 7047–7052 (2016).
26. Wang, D. & Song, C. Impact of human mobility on social networks. *J. Commun. Networks* **17**, 100–109 (2015).
27. Note: $l \approx \log\frac{1-A^{1-\zeta}}{\zeta-1}$ in ref. [21], 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 $\zeta < 1$, $l \approx \log(A)$ captures unbounded movements, that is, a migration process. We ignore the marginal double-logarithmic case when $\zeta = 1$.
28. Burt, W. H. Territoriality and home range concepts as applied to mammals. *J. Mammal.* **24**, 346–352 (1943).
29. Pappalardo, L. et al. Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**, 1–8 (2015).
30. Pappalardo, L., Rinzivillo, S. & Simini, F. Human mobility modelling: exploration and preferential return meet the gravity model. *Proc. Comput. Sci.* **83**, 934–939 (2016).
31. Toole, J. L., Herrera-Yaqüe, C., Schneider, C. M. & González, M. C. Coupling human mobility and social ties. *J. R. Soc. Interface* **12**, 20141128 (2015).
32. Batty, M. & Longley, P. A. *Fractal Cities: A Geometry of Form and Function* (Academic, 1994).
33. Encarnação, S., Gaudiano, M., Santos, F. C., Tenedório, J. A. & Pacheco, J. M. Fractal cartography of urban areas. *Sci. Rep.* **2**, 1–5 (2012).
34. Benguigui, L., Czamanski, D., Marinov, M. & Portugali, Y. When and where is a city fractal? *Environ. Plann. B* **27**, 507–519 (2000).
35. Rozenfeld, H. D., Rybski, D., Gabaix, X. & Makse, H. A. The area and population of cities: new insights from a different perspective on cities. *Am. Econ. Rev.* **101**, 2205–25 (2011).
36. Sen, P. K. Estimates of the regression coefficient based on kendall's tau. *J. Am. Stat. Assoc.* **63**, 1379–1389 (1968).
37. Woldenberg, M. J. An allometric analysis of urban land use in the united states. *Ekistics* **36**, 282–290 (1973).
38. Coffey, W. J. Allometric growth in urban and regional social-economic systems. *Canadian J. Region. Sci.* **11**, 49–65 (1979).
39. Chen, Y., Wang, J. & Feng, J. Understanding the fractal dimensions of urban forms through spatial entropy. *Entropy* **19**, 600 (2017).
40. Falconer, K. *Fractal Geometry: Mathematical Foundations and Applications* (John Wiley & Sons, 2004).
41. Wilson, A. *Entropy in Urban and Regional Modelling* Vol. 1 (Routledge, 2011).
42. Karemera, D., Oguledo, V. I. & Davis, B. A gravity model analysis of international migration to North America. *Applied Econ.* **32**, 1745–1755 (2000).
43. Dagger, R. Metropolis, memory, and citizenship. *Am. J. Political Sci.* **25**, 715–737 (1981).
44. Wissen, L. J. G. & Bonnerman, F. *A Dynamic Model of Simultaneous Migration and Labour Market Behaviour* (Faculty of Economics and Business Administration, Free University, 1991).
45. Camagni, R., Gibelli, M. C. & Rigamonti, P. Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion. *Ecol. Econ.* **40**, 199–216 (2002).
46. Krajzewicz, D., Erdmann, J., Behrisch, M. & Bieker, L. Recent development and applications of sumo-simulation of urban mobility. *Int. J. Adv. Syst. Measurements* **5**, 128–138 (2012).
47. Jacobs, J. *The Death and Life of Great American Cities* (Vintage, 2016).
48. Millington, J. D., O'Sullivan, D. & Perry, G. L. Model histories: Narrative explanation in generative simulation modelling. *Geoforum* **43**, 1025–1034 (2012).
49. Marsaglia, G. et al. Fast generation of discrete random variables. *J. Stat. Softw.* **11**, 1–11 (2004).

50. *US Census Data in 2000* (accessed 3 March 2018); https://www.census.gov/
51. *Great Britain Census Data in 1991* (accessed 5 March 2018); http://ec.europa.eu/eurostat
52. Rozenfeld, H. D. et al. Laws of population growth. *Proc. Natl Acad. Sci. USA* **105**, 18702–18707 (2008).
53. Frankhauser, P. *La Fractalité des Structures Urbaines* (Economica, 1994).
54. Makse, H. A. et al. Modeling urban growth patterns with correlated percolation. *Phys. Rev. E* **58**, 7054 (1998).
55. Engle, R. F. Wald, likelihood ratio, and lagrange multiplier tests in econometrics. *Handbook Economet.* **2**, 775–826 (1984).
56. Xu, F., Li, Y., Jin, D., Lu, J., & Song, C. *tsinghua-fib-lab/Collective-Mobility-Model: First Release* (Zenodo, 2021); https://doi.org/10.5281/zenodo.5722743

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.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-021-00160-6.

Correspondence and requests for materials should be addressed to Yong Li or Chaoming Song.

Peer review information *Nature Computational Science* thanks Pu Wang, Marta Gonzalez and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Editor recognition statement: Handling editor: Fernando Chirigati, in collaboration with the *Nature Computational Science* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.