Lawrence Berkeley National Laboratory
LBL Publications

Title
Deconstructing laws of accessibility and facility distribution in cities.

Permalink
https://escholarship.org/uc/item/7782w7d0

Journal
Science Advances, 6(37)

Authors
Xu, Yanyan
Olmos, Luis
Abbar, Sofiane
et al.

Publication Date
2020-09-01

DOI
10.1126/sciadv.abb4112

Peer reviewed

NETWORK SCIENCE

Deconstructing laws of accessibility and facility distribution in cities

Yanyan Xu[1,2,3], Luis E. Olmos[1,2,3], Sofiane Abbar[4], Marta C. González[1,2,3,5]*

The era of the automobile has seriously degraded the quality of urban life through costly travel and visible environmental effects. A new urban planning paradigm must be at the heart of our road map for the years to come, the one where, within minutes, inhabitants can access their basic living needs by bike or by foot. In this work, we present novel insights of the interplay between the distributions of facilities and population that maximize accessibility over the existing road networks. Results in six cities reveal that travel costs could be reduced in half through redistributing facilities. In the optimal scenario, the average travel distance can be modeled as a functional form of the number of facilities and the population density. As an application of this finding, it is possible to estimate the number of facilities needed for reaching a desired average travel distance given the population distribution in a city.

INTRODUCTION

At a time of the very visible effects of the climate impact on our urban lives, some cities have become unbreathable, and greenhouse gas emissions are produced by buildings heating and cooling networks and all-round petrol transport. At a time when transport has become the first emitter of CO_2, we need to imagine, propose, other ways of occupying urban space. This calls for a better understanding of the spatial distributions of facilities and population (1–7). The information age and the online mapping revolution allow us to globally study the interactions of humans with their built and natural environment (8–13). Pioneering work in multicity studies has uncovered scaling laws relating population to distribution of facilities and socioeconomic activities at macroscopic scale (3, 6, 14–16). It has been asserted, for example, that more populated cities are more efficient in their per capita consumption (3, 4), and their occupation diversity can be modeled as social networks embedded in space (10). However, a systematic understanding of the interplay of the urban form, their facilities distribution, and their accessibility at multiple scales remains an elusive task.

At the country scale, when maximizing for the accessibility of population to a fixed number of facilities, Gastner and Newman (17) demonstrated a simple two-thirds power law between the optimal density of facilities d and their population density ρ. The power law was fitted by allocating 5000 facilities in the continental United States using population data within more than 8 million census blocks. In this case, each facility covers an area about the size of a county ($\sim$1000 km^2). In a follow up study, Um et al. (18) proposed distinct optimization goals to differentiate public services, such as fire stations and public schools, from commercial facilities, such as banks and restaurants. Public service facilities aim to minimize the overall distance between people and the facilities and follow $d \propto \rho^{2/3}$. However, in the case of profit-driven facilities, which have the goal of maximizing the number of potential customers, the power law has an exponent close to 1, that is, $d \propto \rho$. The authors found align-

ment in the analytical optimization and empirical distributions in the United States and South Korea, confirming the 2/3 exponent for public services and the 1 exponent for profit-driven facilities. The simple power law at city scale reveals the equilibrium of empirical allocation of resources across cities with different population. However, distributing facilities at fine scale within cities, where the coverage area per facility is of few blocks ($\sim$10 km^2), results in more heterogeneous settlements of population with different socioeconomic characteristics. Studies of accessibility within cities merit attention for science-informed land use planning and the redistribution of public services after disasters and evacuations (19–23). Forward-looking approaches for planning facilities in cities would also consider individuals' preferences to facilities via mining mobility patterns. Zhou et al. (24) introduced a location-based social network dataset to derive the demand for different types of cultural resources and identified the urban regions with lack of venues. While efforts have been devoted to address the optimal allocation problem in specific cities (24–27), systematic understanding of the optimal distribution of facilities is still lacking from the urban science perspective.

To contribute in this direction, we propose a multicity study that measures the accessibility of city blocks to different types of facilities through their road networks and investigate the role of population distributions. While at large scale, travel cost can be substituted by the Euclidean distance from residents to the facilities, road networks and geographic constraints play important roles for human mobility within cities (28–31). It has been well established that road network properties affect the daily journeys of residents (32–34), their urban form (35, 36), and their accessibility (29, 37). As a complement to most studies devoted to travel costs of commuters, we analyze in this work the road network distance of individuals to the nearest amenity of various types, dividing the space in high-resolution blocks of constant area of 1 km^2. For each city and facility type, we optimally redistribute the existing facilities and compare the result with their empirical distribution. We observe that, in the redistribution, some blocks increase their accessibility and others decrease it. This implies that, to make the best use of the existing facilities for a more equitable accessibility, some blocks would benefit, whereas others would have facilities removed. At the city level, the gap between the empirical facility distribution and the optimal planning offers the opportunity to assess the planning quality of facilities in

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of City and Regional Planning, University of California, Berkeley, CA 94720, USA. [3]Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [4]Qatar Computing Research Institute, HBKU, Doha 5825, Qatar. [5]Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, USA.
*Corresponding author. Email: martag@berkeley.edu

diverse cities. We also revisit the power law between facility and population densities and observe that the two-thirds power law is not followed by the empirical cases, and it is observed in the optimal scenario only when the number of facilities is small compared to the total number of blocks in the city.

We further investigate optimal distributions of facilities by modeling its average travel distance in different cities as a function of the number of facilities to assign. A model of this quantity is derived on both synthetic and real-world cities and fits different cities well with only two free parameters. Furthermore, this gives us a universal function between the average travel distance and the number of facilities controlled by the urban form derived from the population distribution. As an application case, we estimate the number of facilities required to achieve a given accessibility via the proposed function in 12 real-world cities.

RESULTS
Empirical distribution of facilities
We select three cities [Boston, Los Angeles (LA), and New York City (NYC)] in the United States and three cities (Doha, Dubai, and Riyadh) in the Gulf Cooperation Countries (GCC) to study the empirical distribution of facilities. For each city, we collect the population in blocks with a spatial resolution of 30 arc sec (1 km^2 near the equator) from LandScan (38), road networks with the OpenStreetMap (39), and facilities from the Foursquare (40) service application. These novel, rich, and publicly available datasets have proven value in transportation planning (34, 41, 42), land use studies (43, 44), and human activity modeling (45–47). The boundary of each city is drawn along with the Metroplex, encompassing both urban and rural regions. Figure 1 depicts the road network, population density, and

10 selected types of facilities (e.g., hospitals and schools) in NYC and Doha. The statistical information of six cities is summarized in Table 1. For clarity, all variables and notations introduced in this work are summarized in note S1. The distribution of all available facilities in Foursquare data for the six cities is presented in fig. S1. Details of the data sets are described in Materials and Methods and note S2. Figure S2 presents the distribution of population and different facility categories as a function of the distance from the central business district, indicating the diversity of the selected cities. Specifically, it can be observed that Doha and Dubai have more facilities that are located in the highly populated areas, whereas Boston has the majority of facilities located near the city center where fewer people reside. Discrepancy in the distributions between population and facilities can also be observed in LA, NYC, and Riyadh. In these three cities, the population density peaks near the city center, but the facilities are distributed more uniformly across the city.

It is noteworthy that, for calculating facility density and total number of facilities, we first merge the same type of facilities (e.g., hospitals) located in the same block as one facility. Thus, the number of facilities thereafter refers to the number of blocks accommodating a given type of facility, denoted by N. We define N_{max} as the total number of blocks in one city. Furthermore, as nearly unpopulated blocks do not weigh in the calculations of accessibility, we define N_{occ} as the number of occupied blocks given by the blocks with population over a threshold. We set the threshold as 500 in real-world cities, which is commonly used to distinguish between urban and rural regions. The ratio between the number of blocks occupied by facilities N and populated blocks N_{occ} is denoted by D_{occ}. Table 1 reports D_{occ} of the 10 selected types of facilities in the six cities of study. As an example, D_{occ} of hospital in Boston equals to 0.11, indicating that about 11% of populated blocks are occupied by hospitals.
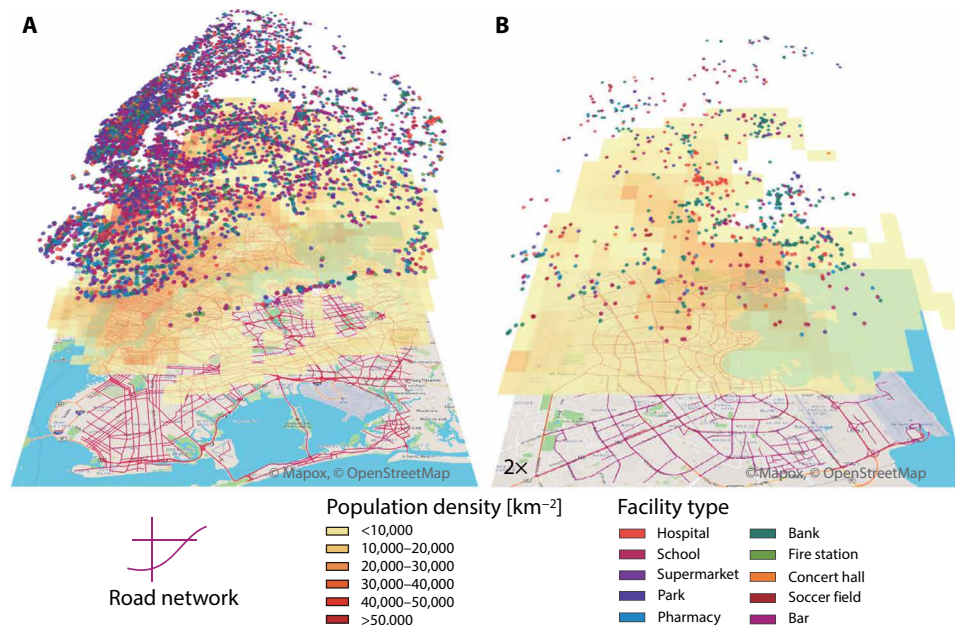
Fig. 1. Illustration of the three datasets. (**A**) Datasets in NYC. The lower layer depicts the map and road networks of the city. The middle and upper layers illustrate the population density and the locations of the 10 selected types of facilities in the same region, respectively. (**B**) Datasets in Doha. Compared with very dense NYC, Doha has a simpler road network, less population, and sparser facilities. Illustration of the three datasets in NYC (A) and Doha (B). The lower layer depicts the map and road networks of the city. The middle and upper layers illustrate the population density and the locations of the 10 selected types of facilities in the same region, respectively. Compared with very dense NYC, Doha has a simpler road network, less population, and sparser facilities.

Table 1. The statistical information of the six cities.

		Boston	LA	NYC	Doha	Dubai	Riyadh
No. of facilities [×1000]		66.3	319.8	207.5	12.9	38.7	64.0
Population [million]		2.36	9.59	8.12	0.99	2.91	5.17
Area [km²]		1649	3869	1132	309	1402	1091
N_{occ}		1125	3696	947	217	657	847
N_{max}		1947	5505	1357	380	1635	1305
UCI		0.21	0.17	0.19	0.24	0.35	0.26
D_{occ}	Hospital	0.11	0.13	0.27	0.27	0.20	0.33
	School	0.32	0.33	0.74	0.37	0.18	0.29
	Supermarket	0.06	0.08	0.38	0.13	0.12	0.14
	Park	0.42	0.45	0.74	0.22	0.17	0.26
	Pharmacy	0.23	0.30	0.67	0.22	0.24	0.37
	Bank	0.37	0.35	0.67	0.42	0.33	0.47
	Fire station	0.13	0.09	0.32	0.03	0.02	0.01
	Concert hall	0.05	0.07	0.14	0.05	0.03	0.07
	Soccer field	0.08	0.04	0.11	0.14	0.07	0.10
	Bar	0.30	0.26	0.66	0.18	0.15	0.19

To quantify the accessibility of the population to facilities, previous work used the Voronoi cell around each facility as a proxy of the tendency of individuals to select the closest facility in a Euclidean distance (*17*, *18*). However, within cities, the distance that people travel in the road networks is constrained by the infrastructure and the landscape. In this context, the routing distance is a better proxy of the accessibility from the place of residence to each amenity. Figure S2C compares the distributions of routing distance of the actual and optimal locations of facilities versus the Euclidean distances, respectively. Our findings confirm that the optimal strategy based on the Euclidean distance achieves similar costs to the actual distribution of facilities, which is much less effective than the strategy that optimizes for routing distance.

Optimal distribution of facilities to maximize overall accessibility

Accessibility indicates the level of service of facilities to the residents. In network science, accessibility is defined as the ease of reaching points of interest within a given cost budget (*48–50*). How to allocate the facilities to maximize the overall accessibility in cities is one of the most essential concerns of facility planning. From this point of view, we redistribute the facilities by minimizing the total routing distance of population to their nearest facilities. In the following, we refer to this redistribution as the optimal scenario. Likewise, the empirical distribution of facilities is referred to as the actual scenario. Specifically, among the N_{max} blocks of one city, we denote as facility-tagged the N blocks that are occupied by a given type of facility in the actual scenario and redistribute the same number of facilities in the optimal scenario. The shortest distance between any pair of two blocks is calculated using the Dijkstra's algorithm in the road network. The idea is to find a new set of N blocks and label them as facility-tagged such as it minimizes the total population-weighted travel distance from all N_{max} blocks to the newly selected N blocks. This optimal allocation problem in

networks is known as the *p*-median problem and, here, it is solved with an efficient algorithm proposed by Resende and Werneck (*51*) (Materials and Methods).

The difference of the travel distance between the actual and the optimal scenarios assesses the quality of the distribution and therefore, of the accessibility in different cities. In each scenario, each residential block is associated with the facility that can be reached in the shortest routing distance. The block is linked to itself if it is occupied by a facility. It is important to note that we do not consider in the present study the capacity of facilities as a constraint, i.e., the number of people using the same facility is not limited. We group the set of blocks served by the same facility and define them as a service community. Considering hospitals as an example, we present in Fig. 2 (A and B) the service communities in Boston in the actual and optimal scenarios, respectively. The color of each cluster depicts the total population p_j^S in the service community of the *j*th facility. The communities in optimal scenario are more uniform in both size and population compared to those in the actual scenario. Particularly in the actual scenario, the communities have small area in downtown Boston but large in the rural area, revealing the uneven distribution of hospitals.

To quantify the disparities between blocks in the level of service for a given type of facility, we compare the actual and optimal travel distances to facilities. We define a gain index of the *i*th block as

$$r_i = \frac{l_i}{\hat{l}_i} \qquad (1)$$

where $\hat{l}_i$ and l_i are the shortest travel distances from the *i*th block to its nearest facility in the actual and optimal scenarios, respectively. An $r_i > 1$ identifies that the block is better served by the facility in the actual than the optimal scenario. Residents living in these blocks benefit more from the distribution of facilities than they would in the scenario of social optimum. In Fig. 2C, we illustrate in Boston the r_i of each block to hospitals in a logarithmic scale. The blocks in
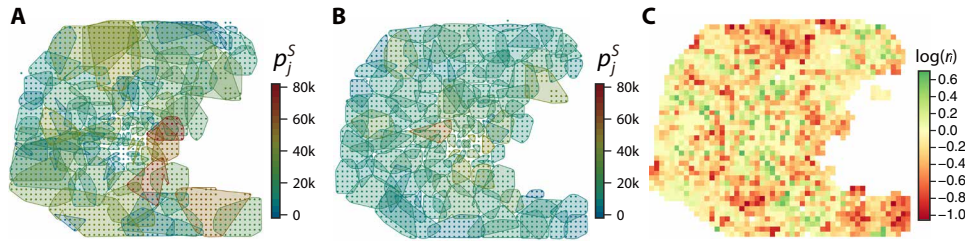
Fig. 2. Service communities of hospitals in the actual and optimal scenarios in Boston and gain of travel distance per block. (**A**) Service communities of the hospitals, as measured from empirical data in Boston. The dots refer to the blocks in the city. The color indicates the population in each community. (**B**) Service communities of the optimally distributed hospitals in Boston. In this optimal scenario, both of the area and the population of communities are generally more equable than those in the actual scenario. (**C**) The block gain index r_i in logarithmic scale in Boston. The blocks in green indicate that the residents are better served in reality; r_i does not have units. The blocks in red indicate that their actual travel distance to the hospital is larger than the optimal distance and they are underserved, such as in the northern and southeastern areas.

green, near to hospitals, are located in the central, southern, and northeastern areas, while the blocks in red have lower accessibility to hospitals when compared with the optimal scenario and are located in the northern, southwestern, and southeastern areas. This has some resemblance with the spatial distribution of wealth in Boston metropolitan area (*52*). The actual travel distance $\hat{l}_i$ and the gain index r_i in the ith block to hospitals for six cities are presented in fig. S3.

Although the inequality of the distribution of facilities can be visually observed from Fig. 2C, for comparing the inequality across facility types and between cities, we compute the Gini coefficient of r_i of all blocks per facility type per city, as illustrated in fig. S4A. We observe that the Gini coefficients of all selected facility types in Boston are similar and around 0.5. NYC has the most discrepancies in the Gini coefficients over the 10 facility types, where the distributions of schools, parks, pharmacies, banks, and bars are more equitable than others due to their high densities (see Table 1). In the GCC cities, fire stations are the most equitably distributed facilities, while bars, hospitals, parks, and pharmacies are distributed less equitably than others. The Lorenz curves and the values of the Gini coefficients per facility type are presented in fig. S4B. The three cities in the United States are generally planned more equally than the GCC cities.

Thereafter, we compare the difference in accessibility across cities to various facility types. Figure 3A presents the average travel distances in the actual scenario ($\hat{L}$) and optimal scenario (L) to the 10 selected types of amenities. The first row displays the facilities with higher densities in the United States cities: banks, pharmacies, schools, parks, and bars. Next come hospitals and supermarkets, followed by concert halls, soccer fields, and fire stations, which have the lowest densities. As expected, the lower the density, the longer the travel distance there is to them. Note that the accessibilities to parks, fire stations, and bars have the largest differences between the United States and GCC cities mainly due to lower availability in the latter. To compare the travel distance in different cities in the same order, we exhibit the scatterplots of $\hat{L}$ and L versus D_{occ}, the ratio between N and N_{occ}, in Fig. 3 (B and C). The discrepancy of actual travel distance $\hat{L}$ among the six cities is mainly caused by the difference in facility planning strategy and urban form. As expected, the optimal travel distance L displays a more uniform tendency than $\hat{L}$, revealing the potential of modeling L with the number of facilities N.

An interesting measure is the improvement of overall accessibility if the locations of facilities are optimally redistributed at a city scale. To that end, we define the optimality index R for a given type of facility at city level as the ratio between the average travel distance to the nearest facilities in the optimal and actual scenarios

$$R = \frac{L}{\hat{L}} = \frac{\sum_{i=1}^{N_{max}} p_i l_i}{\sum_{i=1}^{N_{max}} p_i \hat{l}_i} \qquad (2)$$

where p_i is the population in the ith block. R ranges from 0 to 1, with 1 indicating that the facilities are optimally distributed in reality. In note S3 and fig. S4C, we discuss the change of R with N/N_{max} by introducing two extreme planning strategies, random and population-weighted assignments, described in note S3. We observe that R score of actual planning is mostly between the two extreme strategies, except Riyadh, in which R is even lower than random assignment. This suggests the imbalance between facility locations and service delivery in Riyadh (*53*). Besides, we observe that the R score of actual planning is the highest when N/N_{max} is the smallest for cities, except LA and NYC. For the two extreme strategies, we observe R is u-shaped as a function of N, except for LA. This suggests a higher R for both small and large N values. This is because, for a small N, simply allocating the facilities in the most crowded blocks would shorten the total travel cost to a great extent, while for a large N, most blocks are occupied by facilities. The R score of LA keeps flat compared to other cities mainly due to the polycentric distribution of population, indicating that a small number of facilities cannot efficiently serve most of the population.

Figure 3D depicts the box plot of R of the 10 types of facilities in the six cities. R is generally lower with larger facility density, suggesting that the gaps between actual and optimal distribution are larger. For example, hospitals and fire stations have much lower density than bars, but their R scores are larger. Public services need to be uniformly distributed, while commercial ones do not. The more available facilities are banks, pharmacies, schools, parks, and bars, with an R between 0.4 and 0.5 on average, revealing that the average travel distance could be reduced by 50% if all facilities are planned in the optimal locations.

Revisiting scaling law between facility and population densities

Previous work has related the facility density to population density as a power function both in the actual and optimal scenarios at the national scale (*18*). Here, through introducing the road networks, we dissect these power laws in the two scenarios in diverse cities. We calculate both facility and population densities in the service communities, as shown in Fig. 2 (A and B). Specifically, $d_j^S = 1/a_j^S$ and $\rho_j^S = p_j^S/a_j^S$, where a_j^S is approximated by the product of the number of blocks n_j^S and the average block area in the city, that is, $a_j^S = n_j^S \bar{a}$. Taking hospitals as an example, their densities versus the population
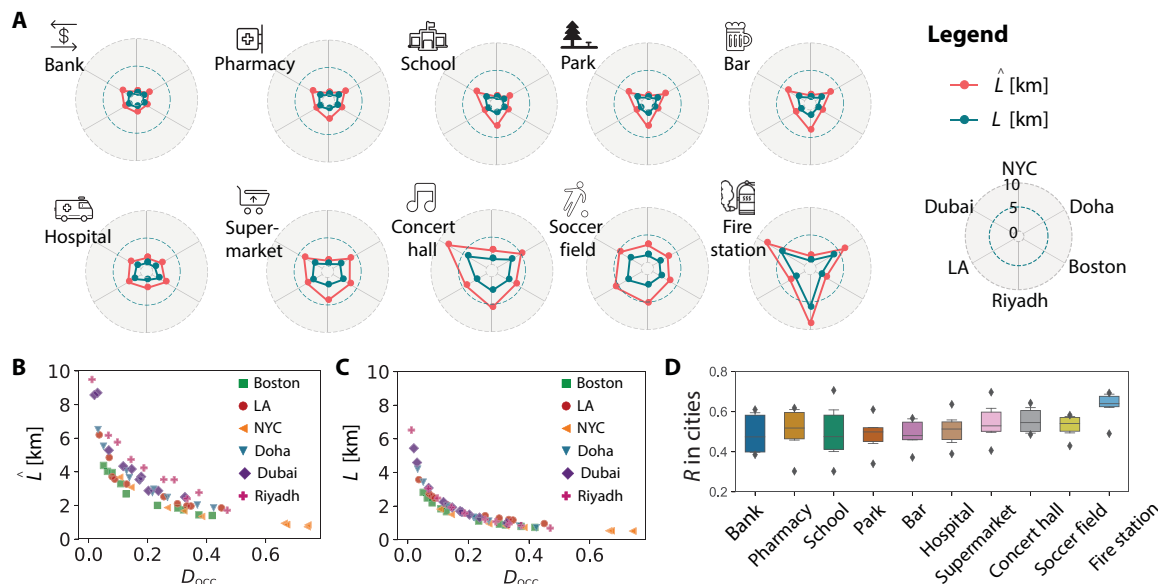
Fig. 3. Optimality of planning by facility type and city. (**A**) The average travel distance in actual scenario, $\hat{L}$, and optimal scenario, L, for the 10 types of facilities in the six cities. (**B**) $\hat{L}$ as a function of D_{occ}, the ratio between N and N_{occ}, in the six cities. The dot refers to one of the 10 types of facilities in a given city. Cities display different descending rates. (**C**) L as a function of D_{occ} in the six cities. All cities show similar descending rates. (**D**) Box plot of the optimality index, R, by facility type. Facility types are ranked by their average densities in the six cities in the descending order. Among the facility types, fire station is the most optimally distributed and bank and school are the worst. In general, facility type with lower density is better located than dense facility type from the perspective of collective benefit maximization.

densities of the service communities in the actual scenario over the six cities are illustrated in Fig. 4A. The full lines represent the fitted power law functions with least squares method and with communities with more than 500 residents. Cities have different exponents and the r^2 scores of the fitting are less than 0.5 in most cases. These results show that, although the two-thirds power law was found for public facilities at county resolution (*18*), we do not find a uniform law between facility and population densities at finer resolutions, i.e., intra-city community level.

Once facilities are optimally redistributed in the city, the service communities are reorganized accordingly. The fitted power laws between the distribution of hospitals and population in optimal scenario of the six cities are shown in Fig. 4B. The fitted exponents are closer to 2/3 and have a larger r^2, and the 95% confidence intervals are narrower than those in Fig. 4A, depicting the actual scenario. The exponents for the 10 selected types of facilities in the actual and optimal scenarios are reported in table S1. As expected, cities have different exponents for both actual and optimal scenarios. In all cases, we observe that the optimal exponents deviate from the analytical 2/3 previously reported when the facilities are optimally distributed by a Euclidean distance at national case (*17*). Sources of difference are both the constraints introduced by the road networks and the higher density of facilities to be distributed.

For a comprehensive understanding of the existence of the power laws, we optimally allocate varying number of facilities N in our six cities of study and in synthetic cities. In Fig. 4 (C and D), we relate the β to D_{occ}, the ratio of N to N_{occ}, and observe 2/3 when $D_{occ} <$ 0.2(0.1) for the real-world (synthetic) cities. We simulate controlled scenarios via four synthetic or toy cities of size 100×100, with population distributions depicted in Fig. 5A. Note that the population threshold is set as 50 in toy cities to count N_{occ}, and the total population is fixed as half million, which is about $^1/_{10}$ of the studied cities.

We find that the curves of diverse cities collapse into a single one, indicating that the difference in the change of β across cities is mainly caused by different N_{occ}. In the toy cities, we notice that the change of β is not monotonous. It stays around 2/3 when D_{occ} is below 0.1. Subsequently, β decreases with D_{occ} as more facilities are assigned to the low-density regions and then increases as facilities start to refill the high-density regions. After all high-density blocks are assigned with facilities, β starts to drop to zero, implying that all blocks are filled with facilities. The same fluctuation of β is not observed in real-world cities because the large and low-density regions are not segregated like in the synthetic cities. In summary, in the optimal scenario, the two-thirds power law can be found for a limited number of facilities but tends to disappear for larger values of N.

Modeling accessibility to optimally distributed facilities
In Fig. 3B, we see that D_{occ} is the most determinant factor to decrease the average distance $\hat{L}$ to a facility independent of its type and city. In Fig. 3C, we observe that these decreasing functions $L(D_{occ})$ collapse for the optimal distributions in each city. Following up on this observation, we explore further the relation between travel distance in optimal scenario L and the number of facilities N for diverse cities with various geographic constraints and population distributions. To this end, we designed 17 toy cities with different levels of urban centrality; 4 of these are illustrated in Fig. 5A. Population distributions of the toy cities are generated by a two-dimensional Gaussian function (e.g., cities *a* and *b*) or a mixture of several two-dimensional Gaussian functions (e.g., cities *c* and *d*). The toy cities have the same population of half million and are equal sized, consisting of 100×100 blocks. The size of each block is set to 1 km^2, and the travel cost between two blocks is calculated with the Euclidean distance between their centroids. We measure the centrality of a city by computing the urban centrality index (UCI), proposed by Pereira *et al.* (*54*), of
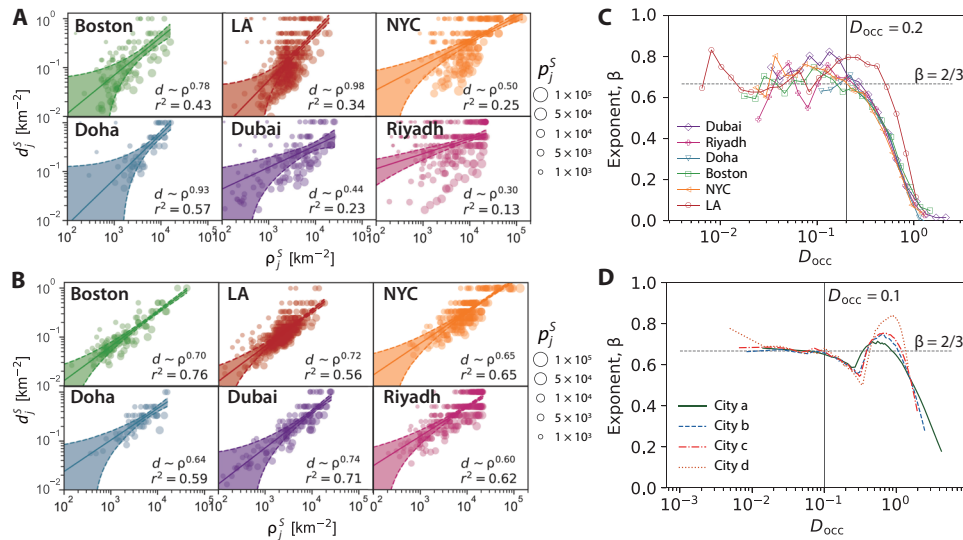
Fig. 4. Fitted power law for the distribution of hospitals per city. (**A**) Actual hospital density (inverse of the area of community) versus population density at service community level. Each colored dot refers to a service community and the size represents its population, as shown in Fig. 2A. The full line represents the best fitted power law. The colored shadow represents the 95% confidence interval. The low r^2 indicates that a clear power law can not be found and the fitted exponent differs from the empirically observed 2/3 at larger scales (*18*). (**B**) Hospital density versus population density in the optimal scenario. The full lines show clear power laws, with exponents close to 2/3 in all cities. The r^2 values are higher and confidence intervals are narrower than the actual scenario. The fitted exponents for other facility types are numerically provided in table S1. (**C**) Change of fitted exponent β with D_{occ} in the optimal scenario in the six cities. Each β is calculated by optimally distributing a given number of facilities in the city. The gray dashed and full lines indicate that β = 2/3 and $D_{occ} = 0.2$, respectively. (**D**) Change of fitted exponent β with D_{occ} in the optimal scenario in four toy cities.

the population distribution (Materials and Methods). UCI ranges from 0 to 1, with 0 indicating the totally polycentric—with the population of the city uniformly distributed—and 1 indicating totally monocentric—with all the population residing in one block. In addition, we include 12 real-world cities for further exploration, the six aforementioned to which we add: Paris, Barcelona, London, Dublin, Mexico City, and Melbourne. Population distributions of four selected cities are illustrated in Fig. 5B. Paris is the most monocentric, with a UCI of 0.50 and most residents residing in the urban region, while Melbourne is the most polycentric, with a UCI of 0.08 and residents dispersed over the city.

For an estimate of the optimal travel distance L in each city, we first assume that the in-block travel distance is constant $l_{min} = 0.5$ km, and the average travel distance within a service community approximates to $g_j^S \sqrt{a_{j,occ}^S}$, where g_j^S denotes the geometric factor in the community; $a_{j,occ}^S$ denotes the area of the occupied blocks (*17*). Then, L is expressed as the sum of two terms, the first for the population in the N blocks with facilities and the second for the population in the N_{max}-N blocks without facilities

$$L = \frac{1}{P} \cdot \left(l_{min} \cdot \sum_{j=1}^{N} p_j + \sum_{j=1}^{N} g_j^S \widetilde{p}_j^S \left(a_{j,occ}^S \right)^{0.5} \right) \quad (3)$$

where P is the total population in the city, and $\widetilde{p}_j^S$ denotes the population in the service community of the jth facility after removing the block where the jth facility is located, that is, $\widetilde{p}_j^S = p_j^S - p_j$. We find that $a_{j,occ}^S$ follows power law relation to the total area in community a_j^S in most cities, that is, $a_{j,occ}^S \propto \left(a_j^S \right)^\gamma$ (fig. S5A). We assume that g_j^S is constant in each city, written as g_{city}, and $a_j^S \approx \overline{a^S} = \overline{a} \cdot N_{max}/N$, with $\overline{a}$ denoting the average block area in the city. Then, we can rewrite Eq. 3 as

$$L(N) = l_{min} \cdot p(N) + A \cdot N^{-\lambda} \cdot (1 - p(N)) \quad (4)$$

where $p(N)$ denotes the share of population in blocks with facilities; A and λ are both constant. More details of this derivation can be found in note S4.1.

We further study how the share of population in blocks without facilities is related to the number of facilities N, and find that $1 - p(N) \approx e^{-\alpha N}$ when $N \ll N_{occ}$ (see details in fig. S5B and notes S4.2 and S4.3). Thereby, we could model L as

$$L(N) = l_{min} \cdot (1 - e^{-\alpha N}) + A \cdot N^{-\lambda} \cdot e^{-\alpha N} \quad (5)$$

where the number of facilities N is the main variable that determines L. While α controls the relation between $p(N)$ and N, A and λ are two free parameters to calibrate. The model of $L(N)$ summarizes the fact that to model L, the only two essential ingredients are the number of facilities to allocate N and the distribution of population in space.

Next, we numerically assign the optimal distribution of facilities given varying number of facilities for both toy and real-world cities. We present the average travel distance L versus the number of facilities N in the toy and real-world cities in log-log plots in Fig. 5 (C and D). In Fig. 5C, we see that for the same N, the global travel costs in polycentric cities are larger than in the monocentric ones. To validate the proposed function $L(N)$, we first calibrate α by fitting $1 - p(N) = e^{-\alpha N}$ per city (fig. S5B), and then calibrate the two free parameters A and λ in Eq. 5 with the simulated L. All parameters are presented in table S2. The fitted $L(N)$ values are shown with lines in Fig. 5 (C and D). The simulated and modeled L values are presented separately for each city in fig. S6, showing good results under various empirical conditions.
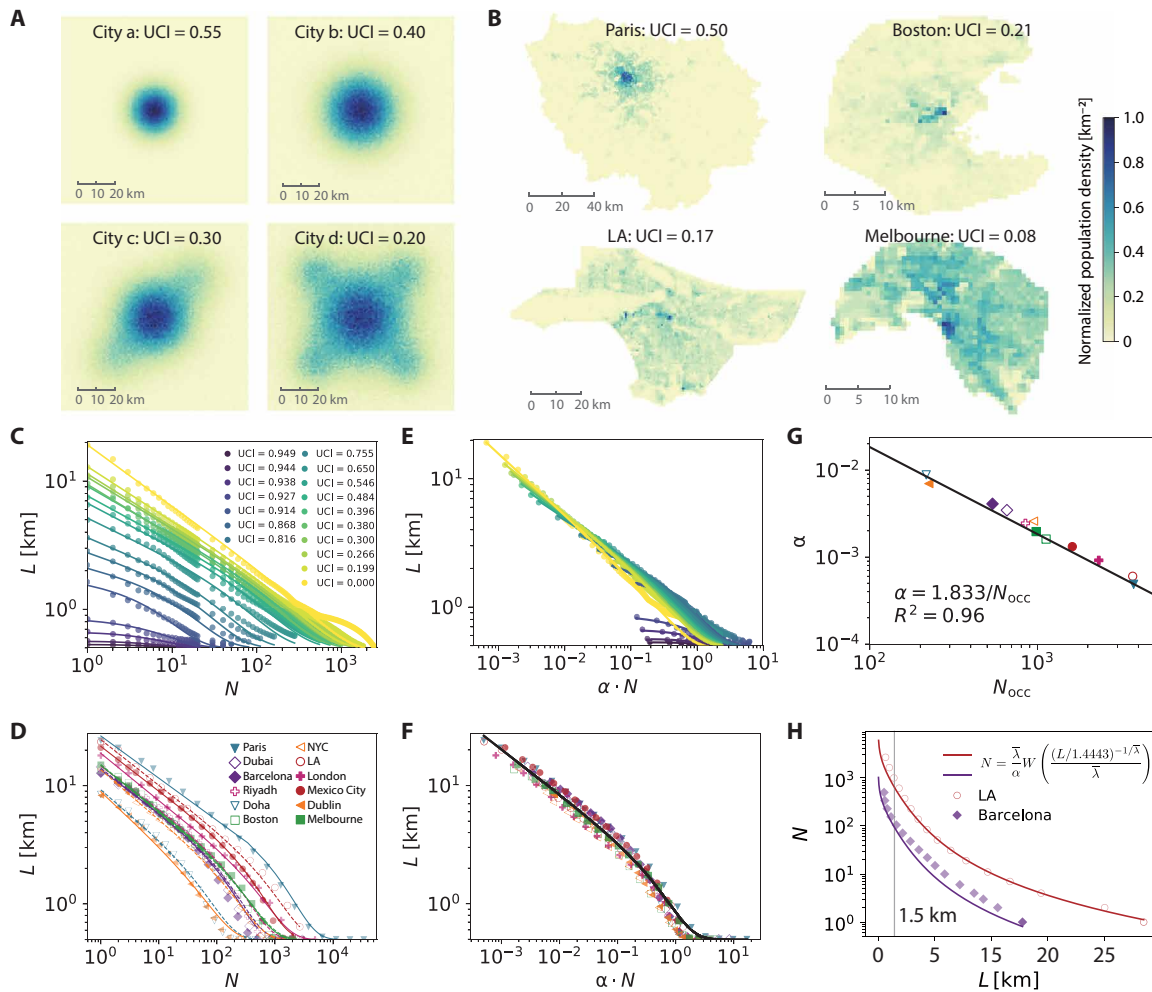
Fig. 5. Modeling of optimal travel distance to facilities in toy and real-world cities. (**A**) Population distributions of four selected toy cities. (**B**) Population distributions of four selected real-world cities, among which Paris is the most centralized and Melbourne is the most polycentric. (**C** and **D**) Simulated and modeled optimal travel distance, L, versus the number of facilities, N, in toy cities and real-world cities, respectively. The dots represent the simulated L with varying N. The lines represent the fitted model $L(N)$ in Eq. 5. Cities are ranked by urban centrality (UCI) in the descending order in the legend. (**E**) Simulated and modeled L versus αN in toy cities. By scaling N with α, we collapse the curves of L with UCI lower than 0.9 into a single one. (**F**) Simulated and modeled L versus αN in real-world cities. The black line represents the function in Eq. 6, approached by the simulated L in all cities. (**G**) Relation between α and N_{occ}. α can be well fitted with N_{occ}, suggesting that the decay rate of shared population in blocks without facilities is in inverse proportion to the urban area in one city for a small N. (**H**) Validation of the universal function in LA and Barcelona.

For seeking a universal function to approach the simulated L in diverse cities, we use λ in Eq. 5 as a constant, fixing its average empirical value $\bar{\lambda} = 0.382$ in the 12 real-world cities. Combining the observation that N_{max} is inversely proportional to α and $A \approx g_{city} \bar{a}^{\bar{\lambda}} N_{max}^{\bar{\lambda}}$ (note S4.1), we can expect that $A \propto \alpha^{-\bar{\lambda}}$. Figure S7C confirms this, showing that $A = 1.4443 \alpha^{-\bar{\lambda}}$. We can rewrite Eq. 5 as follows

$$L(N) = l_{min} \cdot (1 - e^{-\alpha N}) + 1.4443 \cdot (\alpha N)^{-\bar{\lambda}} \cdot e^{-\alpha N} \qquad (6)$$

This function with only one free parameter α suggests that we are able to rescale N with α to collapse the curves of L in all cities into one, as shown in Fig. 5F that depicts Eq. 6 as solid line. The same rescaling of N in toy cities is presented in Fig. 5E, where the collapse is not as good as in the real cities due to the divergent values of λ of toy cities in table S2. Next, we go beyond the average distance

L and plot the distribution of travel distances when keeping αN fixed (fig. S7, F and G). In all cases, the travel distance follows a gamma distribution. This universality suggests that (i) given a certain αN, all real-world cities can reach comparable accessibility; and (ii) the overall accessibility in the optimal scenario depends not only on the availability of the resources but also on the settlement of population, independently from the road network and total area of the city.

Empirically, the decay of population share in blocks without facilities α depends on the population distribution in space. Taking into account that unpopulated blocks are not ideal when optimizing accessibility, N_{occ} is a better variable to express α. A good agreement $\alpha = 1.833/N_{occ}$ ($R^2 = 0.96$) over the 12 real-world cities is shown in Fig. 5G, suggesting that α can be estimated by N_{occ}. Given that $\alpha = 1.833/N_{occ}$ and the universal relation of $L(\alpha N)$, we can explain the collapses found in Figs. 3C and 4 (C and D).

As a concrete application of this universal model for optimal distance of facilities, in Eq. 6, we can plan for facilities by, for example, extracting how many facilities are needed for varying levels of accessibility to a given type of service. In this context, the number of facilities N can be estimated with the inverse function of Eq. 6. As the second term in Eq. 6 dominates the L for a limited N, we simply invert the second term to estimate N, given by $N(L;\alpha) = \frac{\bar{\lambda}}{\alpha} \cdot W\left(\frac{(L/1.4443)^{-1/\bar{\lambda}}}{\bar{\lambda}}\right)$, where $W(\cdot)$ is the ProductLog or the Lambert W function (note S4.5). Figure 5H presents the estimated and simulated N versus L for two limiting cases, LA, in which the approximation agrees well with the simulation, and Barcelona, in which the approximation underestimates N. The results of other real-world cities are depicted in fig. S8, showing in general a good agreement between the analytical approximation via the Lambert W function and the numerical simulations.

DISCUSSION

As cities differ in their form, economy, and population distribution, the interplay between population and facility distributions is challenging to plan. The accessibility of facilities is constrained by their availability, the road network, and means of transportation. While efforts are devoted to managing daily commuting and transit-oriented developments, the planning of the distribution of different urban facilities deserves attention to a paradigm shift toward walkable cities. We present a framework that uses publicly available data to compare the optimal and the actual accessibility of various facility types at the resolution of urban blocks. This allows us to efficiently pinpoint blocks that are underserved, i.e., those where people have to travel longer distances to reach the facilities they need compared to the social optimum. By relocating the facilities to optimize the global travel distance, we find that the relation between facility and population densities follows the scaling law, $d \propto \rho^\beta$ only in the limit of few or limited number of facilities, regardless of the differences in road network structures. The observed exponent β is generally around 2/3 if the number of facilities is diluted or less than 10% of the occupied blocks, and it starts to decay for larger number of facilities. This confirms the continuous limit for diluted number of facilities presented at national scale (18). We observe that the empirical conditions within cities do not follow the continuous approximation for the power law with population density because facilities are not equally planned, and the number of facilities is large in comparison with the number of populated blocks.

To gain further insights when the number of facilities is large, we analytically model the average travel distance L in the optimal scenario versus the number of facilities N and three parameters. Parameter α represents the rate of the population share in blocks without facilities, and the other two parameters can be approximated as constant among cities. A universal expression $L(\alpha N)$ is verified with 17 synthetic cities and 12 real-world cities depicting diverse urban forms. Furthermore, the travel distance to optimally distributed facilities follows a gamma distribution for all cities once αN is fixed. This function can be applied to estimate the number of facilities needed to offer services to people within a given accessibility in average. The results estimated with the derived function find a good match to the numerical simulations that require solving the optimal distribution of facilities. When relating α to the urban form, we uncover that centralized cities require less facilities than polycentric cities to

achieve the same levels of accessibility. Applications of this framework could be to optimally reallocate resources that provide emergency services, such as the placement of shelters, ambulances, or mobile petrol stations in the event of natural disasters.

The optimal planning of facilities in this work supposes that all residents equally need the resources, and the accessibility is measured from their places of residence. In reality, the socioeconomic segregation in cities results in heterogeneous needs for resources. Cities in different social systems and economic development levels also exhibit different needs for various types of facilities that would need to be taken into account for economic considerations. On the other hand, people's needs are naturally dynamic and change in time and space owing to their time-varying mobility behavior. All these factors result in complex interactions between the allocation of facilities and settlements of residents and can be considerable avenues for future research. Another important avenue is to consider the limited capacity of facilities in the optimal planning. This became ever more evident when distributing the health care system resources during the outbreak of a pandemic, such as the COVID-19 in 2020.

MATERIALS AND METHODS
Datasets description
Population density
The population with a spatial resolution of 30 arc sec (approximately 1 km² near the equator) of each city was obtained from the LandScan (38) in 2015. The average population density varies from 1431 per km² in Boston to 7175 per km² in NYC.
Facility data
Facilities were crawled from Foursquare using their public application programming interfaces (APIs) in 2017 (40). In the dataset, each facility is associated with a name, geographical location, a facility type (e.g., hospital, supermarket, and bank), and a category (e.g., arts and entertainment, nightlife spot, shop, and service). Facilities are then assigned to the blocks defined by the LandScan population data with their locations. The total number of facilities in the six cities is given in Table 1. The distributions of facility categories for each city are presented in fig. S1. We select 10 types of facilities to inspect their actual and optimal planning, and they are given in the legend of Fig. 1. After merging facilities of the same type located in the same block, the occupancy of each type of facility, N/N_{occ}, is presented in Table 1. The United States cities generally have more dense facilities than the GCC cities.
Road networks
We extract the road networks from OpenStreetMap (39). The road network is represented as a directed graph, in which edges indicate road segments and nodes indicate intersections. Each edge is associated with a weight representing its length. The travel distance between two blocks is computed by finding the shortest path between two randomly selected nodes in these blocks using the Dijkstra's algorithm (55).

Optimal distribution of facilities in space
Finding the optimal locations of facilities to minimize the total travel cost is essentially an optimal placement problem in network theory, which is non-deterministic polynomial (NP) hard and known as p-median problem. The problem in this work is formalized as follows: "Given a set of blocks N_{max} in a city, a set of residential blocks $X \in N_{max}$ is with population, and each block in N_{max} can only accommodate one facility. The goal is to open N facilities in N_{max} so

as to minimize the sum of population-weighted travel distances from each residential block to its nearest open facility." (*56*).

For simplicity, the *p*-median problem is written as a linear programming problem

$$
\begin{aligned}
\text{minimize} \quad & \sum_i \sum_j c_{i,j} x_{i,j} \\
\text{subject to} \quad & \sum_j x_{i,j} = 1 && \forall i \\
& \sum_j y_j = N \\
& x_{i,j} \leq y_j && \forall i,j \\
& x_{i,j} \in \{0,1\} && \forall i,j \\
& y_j \in \{0,1\} && \forall j
\end{aligned}
\tag{7}
$$

where i and j are indices of the blocks; $x_{i,j} = 1$ means that people living in block i are assigned to their nearest facility in block j, and $i = j$ signifies that there is a facility located in residential block i; $y_j = 1$ if there is a facility in block j, else $y_j = 0$; N is the number of facilities to assign, and we assume that one block can only accommodate one facility of the same type; $c_{i,j}$ is the travel cost from block i to block j, which equals to the total routing distance of all population residing in block i. In this work, we solve the *p*-median problem with a fast algorithm based on swap-based local search procedure implemented by Resende and Werneck (*51*).

Urban centrality index

We adopt the UCI proposed by Pereira *et al.* (*54*) to measure the centrality of the population distribution in cities. UCI is the product of two components, the location coefficient (LC) and the proximity index (PI). The former is introduced to measure the inhomogeneity of population distribution in space. The latter is introduced to measure the difference between the current distribution and the most decentralized scenario. The calculation of LC and PI are as follows

$$
\text{LC} = \frac{1}{2} \sum_{1}^{N_{\max}} \left(s_i - \frac{1}{N_{\max}} \right)
\tag{8}
$$

$$
\text{PI} = 1 - \frac{V}{V_{\max}}
\tag{9}
$$

where $V = \mathcal{S}' \times \mathcal{D} \times \mathcal{S}$. $\mathcal{S}$ is a vector of population fraction in block i, $s_i = p_i/P$, signifying the share of population in block i (p_i) of the total population of the city (P); $\mathcal{D}$ is the distance matrix between blocks. $V_{\max}$ is calculated by assuming that the total population are uniformly settling on the boundary of the city, which indicates an extreme sprawl. UCI ranges from 0 to 1. Large UCI values indicate more centralized population distributions.

SUPPLEMENTARY MATERIALS

REFERENCES AND NOTES

1. H. A. Makse, S. Havlin, H. E. Stanley, Modelling urban growth patterns. *Nature* **377**, 608–612 (1995).
2. W. Pan, G. Ghoshal, C. Krumme, M. Cebrian, A. Pentland, Urban characteristics attributable to density-driven tie formation. *Nat. Commun.* **4**, 1961 (2013).
3. L. M. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G. B. West, Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7301–7306 (2007).
4. L. M. Bettencourt, G. West, A unified theory of urban living. *Nature* **467**, 912–913 (2010).
5. E. Glaeser, Cities, productivity, and quality of life. *Science* **333**, 592–594 (2011).
6. L. M. Bettencourt, The origins of scaling in cities. *Science* **340**, 1438–1441 (2013).
7. D. Weiss, A. Nelson, H. S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, E. Poyart, S. Belchior, N. Fullman, B. Mappin, U. Dalrymple, J. Rozier, T. C. D. Lucas, R. E. Howes, L. S. Tusting, S. Y. Kang, E. Cameron, D. Bisanzio, K. E. Battle, S. Bhatt, P. W. Gething, A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333–336 (2018).
8. H. D. Rozenfeld, D. Rybski, X. Gabaix, H. A. Makse, The area and population of cities: New insights from a different perspective on cities. *Am. Econ. Rev.* **101**, 2205–2225 (2011).
9. T. Tsekeris, N. Geroliminis, City size, network structure and traffic congestion. *J. Urban Econ.* **76**, 1–14 (2013).
10. L. M. Bettencourt, H. Samaniego, H. Youn, Professional diversity and the productivity of cities. *Sci. Rep.* **4**, 5393 (2015).
11. M. Schläpfer, L. M. A. Bettencourt, S. Grauwin, M. Raschke, R. Claxton, Z. Smoreda, G. B. West, C. Ratti, The scaling of human interactions with city size. *J. R. Soc. Interface* **11**, 20130789 (2014).
12. C. Brelsford, J. Lobo, J. Hand, L. M. A. Bettencourt, Heterogeneity and scale of sustainable development in cities. *Proc. Natl. Acad. Sci. U.S.A.* 8963–8968 (2017).
13. R. Li, L. Dong, J. Zhang, X. Wang, W.-X. Wang, Z. Di, H. E. Stanley, Simple spatial scaling rules behind complex cities. *Nat. Commun.* **8**, 1841 (2017).
14. M. E. Newman, Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351 (2005).
15. J. Depersin, M. Barthelemy, From global scaling to the dynamics of individual cities. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2317–2322 (2018).
16. M. Barthelemy, The statistical physics of cities. *Nat. Rev. Phys.* **1**, 406–415 (2019).
17. M. T. Gastner, M. Newman, Optimal design of spatial distribution networks. *Phys. Rev. E* **74**, 016117 (2006).
18. J. Um, S.-W. Son, S.-I. Lee, H. Jeong, B. J. Kim, Scaling laws between population and facility densities. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14236–14240 (2009).
19. K.-W. Tsou, Y.-T. Hung, Y.-L. Chang, An accessibility-based integrated measure of relative spatial equity in urban public facilities. *Cities* **22**, 424–435 (2005).
20. P. Apparicio, A.-M. Séguin, Measuring the accessibility of services and facilities for residents of public housing in Montreal. *Urban Stud.* **43**, 187–211 (2006).
21. S. Macintyre, L. Macdonald, A. Ellaway, Do poorer people have poorer access to local resources and facilities? The distribution of local resources by area deprivation in Glasgow, Scotland. *Soc. Sci. Med.* **67**, 900–914 (2008).
22. H. Dadashpoor, F. Rostami, B. Alizadeh, Is inequality in the distribution of urban facilities inequitable? Exploring a method for identifying spatial inequity in an Iranian city. *Cities* **52**, 159–172 (2016).
23. C. Brelsford, T. Martin, J. Hand, L. M. Bettencourt, Toward cities without slums: Topology and the spatial evolution of neighborhoods. *Sci. Adv.* **4**, eaar4644 (2018).
24. X. Zhou, A. Noulas, C. Mascolo, Z. Zhao, Discovering latent patterns of urban cultural interactions in Wechat for modern city planning, in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (ACM, New York, U.S.A., 19 to 23 August 2018), pp. 1069–1078.
25. A. N. White, Accessibility and public facility location. *Econ. Geogr* **55**, 18–35 (1979).
26. Z. Tao, Y. Cheng, T. Dai, M. W. Rosenberg, Spatial optimization of residential care facility locations in Beijing, China: Maximum equity in accessibility. *Int. J. Health Geogr.* **13**, 33 (2014).
27. L. E. Olmos, M. S. Tadeo, D. Vlachogiannis, F. Alhasoun, X. Espinet Alegre, C. Ochoa, F. Targa, M. C. González, A data science framework for planning the growth of bicycle infrastructures. *Transp. Res. Part C Emerg. Tech.* **115**, 102640 (2020).
28. H. Samaniego, M. E. Moses, Cities as organisms: Allometric scaling of urban road networks. *J. Transp. Land Use* **1**, 10.5198/jtlu.v1i1.29 (2008).
29. R. Louf, M. Barthelemy, How congestion shapes cities: From mobility patterns to scaling. *Sci. Rep.* **4**, 5561 (2014).
30. T. Louail, M. Lenormand, O. G. Cantu Ros, M. Picornell, R. Herranz, E. Frias-Martinez, J. J. Ramasco, M. Barthelemy, From mobile phone data to the spatial structure of cities. *Sci. Rep.* **4**, 5276 (2014).
31. M. A. Florez, S. Jiang, R. Li, C. H. Mojica, R. A. Rios, M. C. Gonzáles, Measuring the impact of economic well being on commuting networks–A case study of Bogota, Colombia, in *Transportation Research Board 96th Annual Meeting, no. 17-03745* (Washington DC, U.S.A., 8 to 12 January 2017).
32. R. Louf, P. Jensen, M. Barthelemy, Emergence of hierarchy in cost-driven growth of spatial networks. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8824–8829 (2013).
33. J. Wang, D. Wei, K. He, H. Gong, P. Wang, Encapsulating urban traffic rhythms into road networks. *Sci. Rep.* **4**, 4141 (2014).
34. S. Çolak, A. Lima, M. C. González, Understanding congested travel in urban areas. *Nat. Commun.* **7**, 10793 (2016).
35. E. Strano, V. Nicosia, V. Latora, S. Porta, M. Barthélemy, Elementary processes governing the evolution of road networks. *Sci. Rep.* **2**, 296 (2012).

36. R. Louf, M. Barthélemy, A typology of street patterns. *J. R. Soc.SS Interface* **11**, 20140924 (2014).

37. R. Louf, M. Barthelemy, Modeling the polycentric transition of cities. *Phys. Rev. Lett.* **111**, 198702 (2013).

38. LandScan Global Population Database (2017); http://web.ornl.gov/sci/landscan/ [accessed 1 April 2017].

39. OpenStreetMap (2017); https://www.openstreetmap.org [accessed 1 April 2017].

40. Foursquare (2017); https://foursquare.com/ [accessed 15 April 2017].

41. J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, M. C. González, The path most traveled: Travel demand estimation using big data resources. *Transp. Res.Part C Emerg. Technol.* **58**, 162–177 (2015).

42. Y. Xu, M. C. González, Collective benefits in traffic during mega events via the use of information technologies. *J. R. Soc. Interface* **14**, 20161041 (2017).

43. D. Quercia, D. Saez, Mining urban deprivation from foursquare: Implicit crowdsourcing of city land use. *IEEE Pervasive Comput.* **13**, 30–36 (2014).

44. S. Spyratos, D. Stathakis, M. Lutz, C. Tsinaraki, Using foursquare place data for estimating building block use. *Environ. Plan. B Urban Anal. City Sci.* **44**, 693–717 (2017).

45. J. Frith, Communicating through location: The understood meaning of the foursquare check-in. *J. Comput. Mediated Commun.* **19**, 890–905 (2014).

46. R. Krueger, D. Thom, T. Ertl, Semantic enrichment of movement behavior with foursquare–A visual analytics approach. *IEEE Trans. Vis. Comput. Graph.* **21**, 903–915 (2015).

47. Y. Xu, S. Jiang, R. Li, J. Zhang, J. Zhao, S. Abbar, M. C. González, Unraveling environmental justice in ambient $PM_{2.5}$ exposure in Beijing: A big data approach. *Comput. Environ. Urban Syst.* **75**, 12–21 (2019).

48. A. Antunes, Á. Seco, N. Pinto, An accessibility–maximization approach to road network planning. *Comput. Aided Civ. Inf. Eng.* **18**, 224–240 (2003).

49. M. Barthélemy, Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).

50. D. Levinson, Network structure and city size. *PLOS ONE* **7**, e29721 (2012).

51. M. G. Resende, R. F. Werneck, A fast swap-based local search procedure for location problems. *Ann. Oper. Res.* **150**, 205–230 (2007).

52. United States Census Bureau, Census data, (2016); https://www.census.gov/data.html. [accessed 15 Oct 2016].

53. A. Alhomaidhi, Geographic distribution of public health hospitals in Riyadh, Saudi Arabia. *Geogr. Bull.* **60**, 45–48 (2019).

54. R. H. M. Pereira, V. Nadalin, L. Monasterio, P. H. Albuquerque, Urban centrality: A simple index. *Geogr. Anal.* **45**, 77–89 (2013).

55. E. W. Dijkstra, A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271 (1959).

56. D. Dohan, S. Karp, B. Matejek, *K-Median Algorithms: Theory in Practice* (Working Paper, Princeton, Computer Science, 2015).

57. M. Batty, P. Longley, *Fractal Cities: A Geometry of Form and Function* (Academica Press, 1994).

58. W. Alonso, *Location and Land Use* (Harvard Univ. Press, 1994).

59. E. S. Mills, B.-N. Song, *Urbanization and Urban Problems* (Harvard, 1979).

60. R. F. Muth, *Cities and Housing* (University of Chicago Press, 1969).

61. C. Clark, Urban population densities. *J. R. Stat. Soc.* **114**, 490–496 (1951).

Citation: Y. Xu, L. E. Olmos, S. Abbar, M. C. González, Deconstructing laws of accessibility and facility distribution in cities. *Sci. Adv.* **6**, eabb4112 (2020).