The Origins of Scaling in Cities

Luís M. A. Bettencourt

Despite the increasing importance of cities in human societies, our ability to understand them scientifically and manage them in practice has remained limited. The greatest difficulties to any scientific approach to cities have resulted from their many interdependent facets, as social, economic, infrastructural, and spatial complex systems that exist in similar but changing forms over a huge range of scales. Here, I show how all cities may evolve according to a small set of basic principles that operate locally. A theoretical framework was developed to predict the average social, spatial, and infrastructural properties of cities as a set of scaling relations that apply to all urban systems. Confirmation of these predictions was observed for thousands of cities worldwide, from many urban systems at different levels of development. Measures of urban efficiency, capturing the balance between socioeconomic outputs and infrastructural costs, were shown to be independent of city size and might be a useful means to evaluate urban planning strategies.

Cities exist, in recognizable but changing forms, over an enormous range of scales (*1*), from small towns with just a few people to the gigantic metropolis of Tokyo, with more than 35 million inhabitants. Many parallels have been suggested between cities and other complex systems, from river networks (*2*) and biological organisms (*3*–*6*) to insect colonies (*1*, *7*) and ecosystems (*8*). The central flaw of all these arguments is their emphasis on analogies of form rather than function, which limit their ability to help us understand and plan cities.

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA.
E-mail: bettencourt@santafe.edu

Recently, our increasing ability to collect and share data on many aspects of urban life has begun to supply us with better clues to the properties of cities, in terms of general statistical patterns of land use, urban infrastructure, and rates of socioeconomic activity (*6*, *9*–*13*). These empirical observations have been summarized across several disciplines, from geography to economics, in terms of how different urban quantities (such as the area of roads or wages paid) depend on city size, usually measured by its population, N.

The evidence from many empirical studies over the past 40 years points to there being no special size to cities, so that most urban properties, Y, vary continuously with population size and are well described mathematically on average by power-law scaling relations of the form $Y = Y_0 N^{\beta}$, where Y_0 and β are constants in N. The surprise, perhaps, is that cities of different sizes do have very different properties. Specifically, one generally observes that rates of social quantities (such as wages or new inventions) increase per capita with city size (*11*, *12*) (superlinear scaling, $\beta = 1 + \delta > 1$, with $\delta \simeq 0.15$), whereas the volume occupied by urban infrastructure per capita (roads, cables, etc.) decreases (sublinear scaling, $\beta = 1 - \delta < 1$) (Fig. 1). Thus, these data summarize familiar expectations that larger cities are not only more expensive and congested, but also more exciting and creative when compared to small towns.

These empirical results also suggest that, despite their apparent complexity, cities may actually be quite simple: Their average global properties may be set by just a few key parameters (*12*, *13*). However, the origin of these observed scaling relations and an explanation for the interdependences between spatial, infrastructural, and social facets of the city have remained a mystery.

Here, I develop a unified and quantitative framework to understand, at a theoretical level, how cities operate and how these interdependencies arise. Consider first the simplest model of a city with circumscribing land area A and

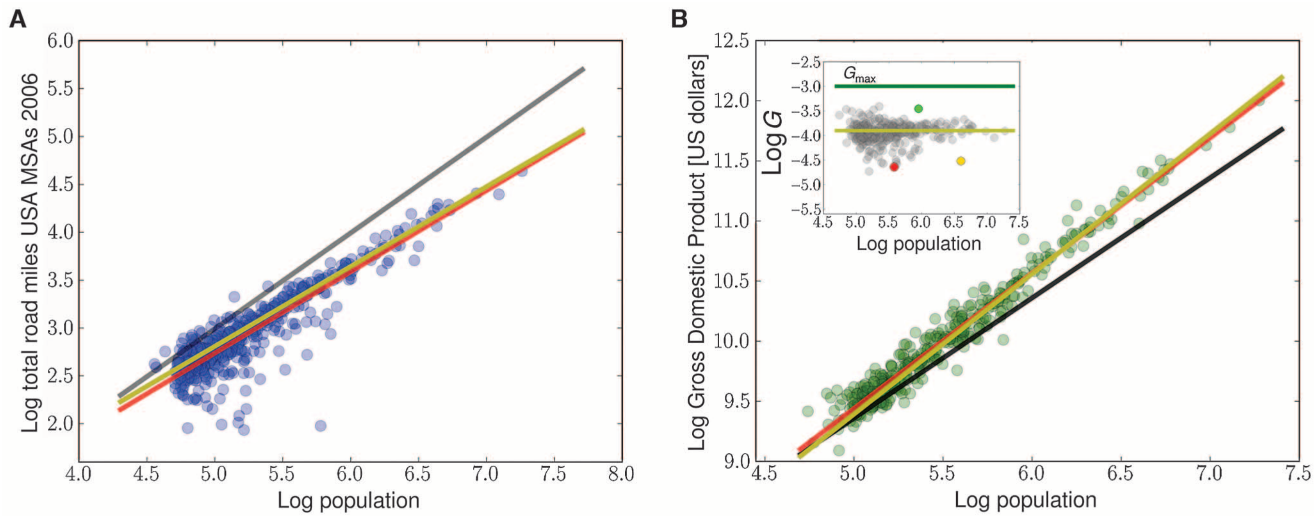

Fig. 1. Scaling of urban infrastructure and socioeconomic output. (A) Total lane miles (volume) of roads in U.S. metropolitan areas (MSAs) in 2006 (blue dots). Data for 415 urban areas were obtained from the Office of Highway Policy Information from the Federal Highway Administration (*14*). Lines show the best fit to a scaling relation $Y(N) = Y_0 N^{\beta}$(red), with $\beta = 0.849 \pm 0.038$ [95% confidence interval (CI), $R^2 = 0.65$]; the theoretical prediction, $\beta = 5/6$ (yellow); and linear scaling $\beta = 1$ (black). (**B**) Gross metropolitan product of MSAs in 2006 (green dots). Data obtained for 363 MSAs from U.S. Bureau of Economic Analysis (*14*). Lines describe best fit (red) to data, $\beta = 1.126 \pm 0.023$ (95% CI, $R^2 = 0.96$); the theoretical prediction, $\beta = 7/6$ (yellow); and proportional scaling, $\beta = 1$ (black). The two best-fit parameters in each scaling relation were estimated by means of ordinary least-squares minimization to the linear relation between logarithmically transformed variables (*14*). The inset shows the estimate of G for 313 U.S. MSAs and the conservation law $\frac{d \ln G}{d \ln N} = 0$ ($R^2 = 0.003$). G is measured as the product of gross domestic product and road volume, both per capita. As predicted by the theory, observed values of G for different cities cluster around its most likely value (mode, yellow line), which gives an estimate of the optimum G^*, and are bounded by the maximum $G_{max} \simeq 8G^*$(green line); see also Fig. 2B. Several metropolitan areas, such as Bridgeport, Connecticut (green circle); Riverside, California (yellow circle); or Brownsville, Texas (red circle), are outliers, suggesting that they are suboptimal in terms of their transportation efficiency or amount of social mixing.

population N. I write the interactions between people i, j in terms of a social network, F_k^{ij}, and assume that social interactions [e.g., friendship, employment, acquaintance, etc.] are local, take place over an interaction area a_0 (a cross section in the language of physics), and have strength g_k, where k describes social link types (*14*). The parameters, g_k, can be either positive (attractive, expressing a social benefit, e.g., mutually beneficial economic relations) or negative (repulsive, expressing a social cost, e.g., crime). All these processes share the same average underlying dynamics of social encounters in space and time, against the background of the city and its infrastructure networks.

The average number of local interactions per person is given by the product of the volume spanned by their movement, $a_0\ell$, times the population density $n = N/A$, where ℓ is the typical length traveled by people, goods, and information (*14*). The total average social output of a city can be obtained by multiplying the total number of interactions by the average outcome per interaction, $\bar{g}$, leading to $Y = G\frac{N^2}{A}$, with the parameter $G \equiv \bar{g}a_0\ell$ measuring the product of average social output times area, both per capita (Fig. 1). Each urban socioeconomic output, Y, has physical units set by g_k, but it is useful to think of all quantities ultimately expressed in terms of energy per unit time (power).

Another crucial property of cities is that they are mixing populations. That is, even if people in the city explore different locations at different times, anyone can in principle be reached by anyone else. This concept, familiar from population biology (*15*), is the basis of definitions of functional cities as metropolitan statistical areas (MSAs), e.g., by the U.S. census bureau. In practice, this means that the cost per person of a mixing population is proportional to the transverse dimension (diameter), L, of the city $L \sim A^{1/2}$. Thus, the total power spent in transport processes to keep the city mixed is $T = \varepsilon LN = \varepsilon A^{1/2}N$, where ε is a force per unit time. This cost must be covered by each individual's budget, $y = Y/N$, requiring $y \simeq T/N$, which implies $A(N) = aN^{\alpha}$ with $\alpha = 2/3$ and $a = (G/\varepsilon)^{\alpha}$. The baseline area, a, increases with more productive interactions, e.g., due to economic growth, and decreasing transportation costs, as is observed in worldwide patterns of urban sprawl over time (*16*). Thus, I obtain $Y = Y_0N^{\beta}$, where $\beta = 2 - \alpha = 1 + 1/3 > 1$ and $Y_0 = G^{1-\alpha}\varepsilon^{\alpha}$. This simple model leads to area, A, varying sublinearly with N ($\alpha = 2/3 < 1$), and socioeconomic outputs, Y, varying superlinearly ($\beta = 4/3 > 1$). However, this overestimates β because as cities grow, space becomes occupied and transportation of people, goods, and information is channeled into networks. The space created by these networks gives the correct measure of the social interactions that can occur in cities.

I propose a more realistic model by generalizing these ideas in terms of four simple assumptions:

1) Mixing population. The city develops so that citizens can explore it fully given the resources at their disposal. I formalize this principle as an entry condition (*17*), by requiring that the minimum resources accessible to each urbanite, $Y_{min}/N \sim GN/A$, match the cost of reaching anywhere in the city. Because travel paths need not be linear, I generalize their geometry via a fractal dimension, H, so that distance travelled $\propto A^{H/D}$ (*14*). Matching interaction density to costs, I obtain a generalized area scaling relation, $A(N) = aN^{\alpha}$, with a as before and $\alpha = \frac{2}{2+H}$ [$\alpha = \frac{D}{D+H}$ in D dimensions]. $H = 1$ allows individuals to fully explore the city within the smallest distance traveled, implying that N scales like a physical volume (*14*, *18*).

2) Incremental network growth. This assumption requires that infrastructure networks develop gradually to connect people as they join, leading to decentralized networks (*6*, *19*). Specifically, the scaling of Fig. 1A is obtained when the average distance between individuals $d = n^{-1/2} = (A/N)^{1/2}$ equals the average length of infrastructure network per capita so that the total network area, $A_n(N) \sim Nd = A^{1/2}N^{1/2}$. Together with the first assumption, this implies that $A_n \sim a^{1/2}N^{1-\delta}$ with $\delta = 1/6$ [$A_n \sim A^{1/D}N^{(D-1)/D} = a^{1/D}N^{1-\delta}$, with $\delta = \frac{H}{D(D+H)}$ in D dimensions]. This has been observed in U.S. and German road networks (*6*, *12*, *19*) and tracks the average built area of more than 3600 large cities worldwide (*16*), measured through remote sensing.

3) Human effort is bounded, which requires that G is, on average, independent of N, i.e., $dG/dN = 0$ (Fig. 1B, inset). The increasing mental and physical effort that growing cities can demand from their inhabitants has been a pervasive concern to social scientists (*20*). Thus, this assumption is necessary to lift an important objection to any conceptualization of cities as scale-invariant systems. Bounded effort is also observed in urban cell phone communication networks (*21*) and is in general a function of human constraints and urban services and structure.

4) Socioeconomic outputs are proportional to local social interactions, so that $Y = GN^2/A_n \sim N^{1+\delta}$. From this perspective, cities are concentrations not just of people, but rather of social interactions. This point was emphasized by Jacobs (*22*, *23*), but has been difficult to quantify. The prediction that social interactions scale with $\beta = 1 + \delta \simeq 7/6$ was observed recently in urban telecommunication networks (*21*). Together these assumptions predict scaling exponents for a wide variety of urban indicators, from patterns of human behavior and properties of infrastructure to the price of land (*6*, *9*–*12*, *16*, *21*, *24*, *25*), summarized in Table 1 (*14*).

Thus far, I obtained estimates for scaling exponents without the need for a detailed model of infrastructure. Next, I show how network models of infrastructure can help to illuminate urban planning issues. Consider the infrastructure in a city described by a network with h hierarchical levels (Fig. 2A). The network branching, b, measures the average ratio of the number of units of infrastructure at successive levels, $N_i = b^i$, e.g., number of paths to small roads, or larger roads to highways. I assume that the number of infrastructure units at the lowest level, $i = h$, equals the number of people, so that $N_h = N$ and $h = \ln N/\ln b$. These networks are not hierarchical trees (*26*) (Fig. 2A). The length of a network segment (such as a road) at level i is l_i, crossing a land area a_i, and its transverse dimension is s_i, an area in 3D networks and a length in 2D. To obtain the above scaling relations, I assume that the transverse dimension of the smallest network units, s_*, is independent of N. This leads to the scaling of network width, $s_i = s_* b^{(1-\delta)(h-i)}$, which says that highways or water mains are much wider than building corridors or household

Table 1. Urban indicators and their scaling relations. Columns show measured exponent ranges (see table S3 for details). Also shown are predicted values for $D = 2$, $H = 1$ (the simplest theoretical expectation) and for general D, H. Agglomeration effects vanish as $H \to 0$ (*14*). The larger range for the observed land-area exponent is likely the result of different definitions of the city in space and distinct measurement types. See table S3 and supplementary text for specific values of observed exponents, discussion, and additional data sources.

Urban scaling relations	Observed exponent range	Model ($D = 2$, $H = 1$)	Model D, H
Land area $A = aN^{\alpha}$	[0.56,1.04]	$\alpha = \frac{2}{3}$	$\alpha = \frac{D}{D+H}$
Network volume $A_n = A_0N^{\nu}$	[0.74,0.92]	$\nu = \frac{5}{6}$	$\nu = 1 - \delta$
Network length $L_n = L_0N^{\lambda}$	[0.55,0.78]	$\lambda = \frac{2}{3}$	$\lambda = \alpha$
Interactions per capita $\bar{I}_i = I_0N^{\delta}$	[0.00,0.25]	$\delta = \frac{1}{6}$	$\delta = \frac{H}{D(D+H)}$
Socioeconomic rates $Y = Y_0N^{\beta}$	[1.01,1.33]	$\beta = \frac{7}{6}$	$\beta = 1 + \delta$
Network power dissipation $W = W_0N^{\omega}$	[1.05,1.17]	$\omega = \frac{7}{6}$	$\omega = 1 + \delta$
Average land rents $P_L = P_0N^{\delta_L}$	[0.46,0.52]	$\delta_L = \frac{1}{2}$	$\delta_L = 1 - \alpha + \delta$

pipes, $s_0 = s_* b^{(1-\delta)h} \gg s_h = s_*$. Additionally, because infrastructure must reach everyone in the city (*6*, *18*), total network length is area filling, $l_i = a_i/l$, with $a_i = ab^{(\alpha-1)i}$. This means that the land area per person, $a_h = aN^{\alpha-1}$, and shortest network distance, $l_h = (a/l)N^{\alpha-1}$, which defines l, decrease with N. The total network length L_n and network area A_n follow from the sum of the geometric series over levels

$$L_n = \sum_{i=0}^{h} l_i N_i = \frac{a}{l}\sum_{i=0}^{h} b^{\alpha i} = \frac{a}{l}\frac{b^{\alpha(h+1)}-1}{b^{\alpha}-1} \simeq L_0 N^{\alpha}, L_0 = a/l \quad (1)$$

$$A_n = \sum_{i=0}^{h} s_i l_i N_i = s_* \frac{a}{l} b^{(1-\delta)h}\sum_{i=0}^{h} b^{(\alpha+\delta-1)i} \simeq A_0 N^{1-\delta}, \quad A_0 = \frac{s_* a}{l(1-b^{\alpha+\delta-1})} \quad (2)$$

where I took $\alpha + \delta < 1$, which holds for $D > 1$.

I can now compute the cost of maintaining the city connected as the energy necessary for moving people, goods, and information across its infrastructure networks. These movements form a set of currents, transporting various quantities across the city and can be quantified by means of the language of circuits. The scaling of s_i together with total current, J, conservation across levels $J_i = s_i \rho_i v_i N_i = s_{i-1}\rho_{i-1}v_{i-1}N_{i-1} = J_{i-1}$ for all i, sets the scaling for $\rho_i v_i$, the current density at level i, where ρ_i is the density of carriers in the network and v_i their average velocity. This quantity is interesting because it controls the dissipation mechanisms in any network. I obtain $\rho_i v_i = b^{-\delta}\rho_{i-1}v_{i-1}$, which implies that the current density decreases with increasing i, so that highways are faster and/or more densely packed than smaller roads (*27*, *28*). Making the additional assumption that individual needs, $\rho_h v_h = \rho_* v_*$, are independent of N (*12*) leads to $\rho_i v_i = b^{\delta(h-i)}\rho_* v_*$. Then, the total current $J_i = J = J_0 N$, with $J_0 = s_* \rho_* v_*$, which is a function only of individuals' characteristics.

There are many forms of energy dissipation in networks, including those that occur at large velocity or density. Here, I make the standard assumption that the resistance per unit length per transverse network area, r, is constant (*2*, *5*), leading to the resistance per network segment, $r_i = r\frac{l_i}{s_i}$. For N_i parallel resistors this gives the total resistance per level, $R_i = \frac{r_i}{N_i} = \frac{ar}{ls_*}b^{-(1-\alpha+\delta)i-(1-\delta)h}$. The total power dissipated, W, follows from summing $W_i = R_i J_i^2$ over levels,

$$W = J^2\sum_{i=1}^{h} R_i = J^2\frac{ar}{ls_*}b^{-(1-\delta)h}\frac{1-b^{-(1-\alpha+\delta)(h+1)}}{1-b^{-1+\alpha-\delta}} \simeq W_0 N^{1+\delta}, \quad W_0 = \frac{arJ_0^2}{ls_*(1-b^{-1+\alpha-\delta})} \quad (3)$$

which scales superlinearly, with exponent 1+ δ = 1+1/6 in $D = 2$, $H = 1$. Thus, energy dissipation scales with population like social interactions, as observed in German urban power grids (*12*), so that the ratio Y/W, a measure of urban efficiency, is independent of city size.

Finally, I show that these results can be derived by maximizing net urban output, $\mathcal{L}$, as the difference between social interaction outcomes, Y, and infrastructure energy dissipation, W, under settlement and network constraints,

$$\mathcal{L} = Y - W + \lambda_1(\varepsilon A^{H/D} - GN/A) + \lambda_2(A_n - cNd) \underset{d\mathcal{L}/dG=0}{\rightarrow} \frac{2\alpha-1}{\alpha}G^*\frac{N^2}{A_n(N)} \quad (4)$$

where $c = A_0 a^{-1/D}$ and λ_1, λ_2 are Lagrange multipliers. Equation 4 gives the basis for the derivation of the properties of every segment in the network, through Eqs. 1 and 2, in analogy with (*2*, *4*, *5*). The novelty in Eq. 4 is the prediction of an optimal $G = G^*$, through $d\mathcal{L}/dG = 0$, and the expectation that values of G for different cities fluctuate around this value, as observed in Fig. 1B (inset).

To see this, consider that, keeping ε fixed and $a = (G/\varepsilon)^{\alpha}$, both Y and W grow with G, because $Y_0 \sim G^{1-\alpha}$ and $W_0 \sim G^{\alpha}$. This tension between social interactivity, transportation costs, and spatial settlement patterns is at the root of most urban planning and policy. The limiting values of G follow from the solutions to $\mathcal{L} = 0 : G = 0$ and $G = G_{\max} = \left[\frac{(\varepsilon l)^{2\alpha}}{r' J_0^2} l^{2(1-\alpha)}\right]^{\frac{1}{2\alpha-1}}$, where $r' \approx r$ (*14*). It follows that $G^* = \left(\frac{1-\alpha}{\alpha}\right)^{1/(2\alpha-1)} G_{\max} \simeq G_{\max}/8$, with $\alpha \simeq 2/3$ (Fig. 1B, inset). Thus, cities will form if the balance of social interactions is positive, $\bar{g} > 0$. However, there is an upper value of $G = G_{\max}$ (Fig. 1B, inset) beyond which dissipation costs overcome social benefits and a city may split up into regions. For $G < G^*$, the social interaction potential of a city is underdeveloped. Such places tend to be poorer and have less advanced infrastructure. Thus, I would expect that cities such as Riverside, California, or Brownsville, Texas (Fig. 1B), where estimates of G are less than average, would typically benefit from measures

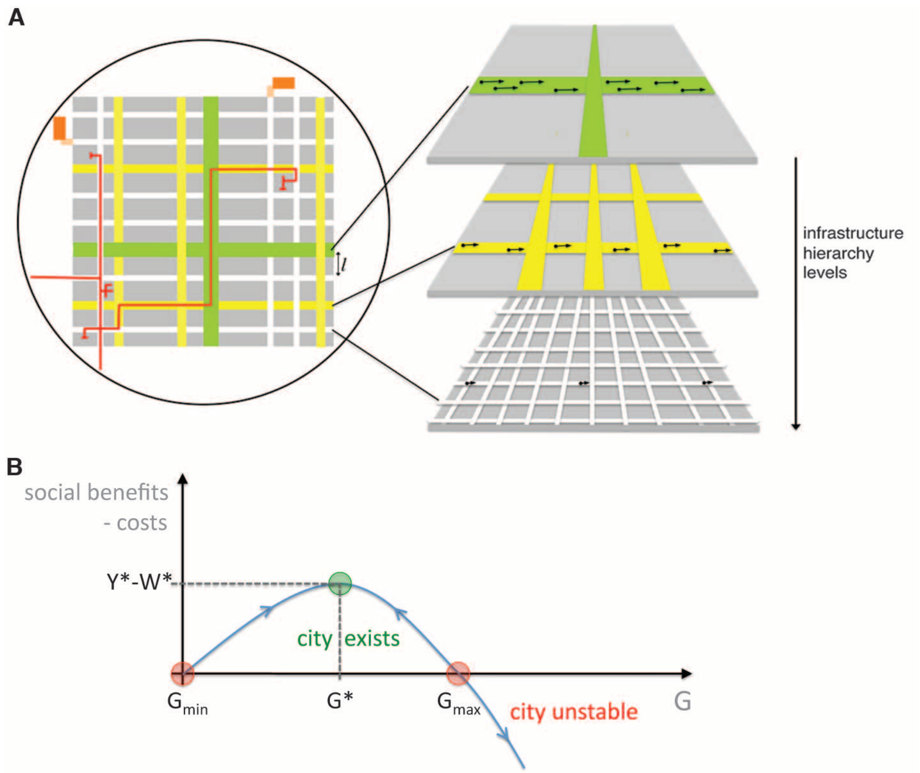

Fig. 2. The spatial city and its social and dissipative processes. (**A**) Gray blocks denote settled areas, and spaces in between (white, yellow, green) represent infrastructure networks, treated in terms of a size hierarchy. Total network length $L_n = 2(n_b + 1)L \simeq A/l$ is area filling (circle), where n_b is the number of blocks across the city (*14*). Red lines denote the volume of public space spanned by an individual, which determines his or her average number of social interactions. As the city grows and new land is settled (orange blocks), the infrastructure network grows incrementally (orange segments). The flux $\rho_i v_i$ in larger network segments is higher (black dots plus arrows), controlling the energy dissipation in the city. (**B**) There is an optimal value of G at which cities are most productive. Cities can exist when social interactions are positive $G > G_{\min} = 0$, and less than an upper value $G < G_{\max}$ (red circles), at which point dissipation costs overcome benefits. The optimal $G = G^*$ (green circle) corresponds to the most efficient city.

that promote greater mobility or density, in order to achieve more intense and beneficial city-wide social contact. Conversely, cities with $G > G^*$ become victims of their socioeconomic success by incurring escalating mobility costs. Bridgeport, Connecticut's, MSA (Fig. 1B) may be developed in terms of its economic functions and infrastructure, but might generally benefit from more compact urban living or from increases in transportation energy efficiency. That is, cities may be suboptimal either because they do not realize their full social potential or because they do so in a manner that renders transportation costs too high. In either case, this approach shows how urban planning must take into account the delicate net balance between density, mobility, and social connectivity and thus provides a general framework for the iterative development and assessment of urban policies.

That many cities are becoming more global in their economic relations and political and cultural influence (*29*) does not alter the basic premises of the theory. The internal dynamics and organization of cities (as social networks of people and institutions) produces new socioeconomic functions that allow cities to exchange goods, services, people, and information within and across national borders (*22*, *23*, *30*). Thus, even if some singular places such as Hong Kong, Singapore, or Dubai are primarily part of international economies, the majority of the world's most global cities, such as Tokyo, New York, Los Angeles, Beijing, Shanghai, Berlin, or Frankfurt, show clear scaling effects in line with their own national urban systems (Fig. 1 and figs. S1 to S3).

All cities have spatial and social pockets of greater and lower mobility, social integration, better or worse services, and so forth (*1*, *17*). It should be emphasized that the theory does not predict density profiles or socioeconomic differences inside the city, but the scaling for the properties of the city as a whole. None of these pockets exist in absolute isolation; they are just more or less "connected," so they must be understood with reference to the rest of the city (*17*).

The interactions between people also provide the basis for institutional relationships via the appropriate groupings of individuals in social or economic organizations and by the consideration of the resulting links between such entities. Institutions and industries that benefit from strong mutual interactions may aggregate in space and time within the city in order to maximize their $Y - W$, a point first made by Marshall (*23*) in the context of industrial districts. Other organizations may benefit primarily from the general effects that result from being in the wider city and collecting a diversity of interactions, an argument often attributed to Jacobs (*22*). These results establish necessary conditions for urban areas to express certain levels of socioeconomic productivity, but it remains a statistical question (*21*, *25*) how well they are realized in specific places.

Most urban systems for which reliable data exist confirm almost exactly the simplest predictions of the theory developed here. Examples are the scaling of area for about 1800 cities in Sweden (*14*, *18*), or for roads in several hundred American (Fig. 1A) and Japanese metropolitan areas (fig. S3). One of the most spectacular agreements is for the scaling of total area of paved surfaces for all cities worldwide above 100,000 people (over 3600 cities) (*14*, *16*). These examples illustrate the result derived above that urban infrastructure volume scales faster with population than land area (and both are sublinear). This effect is visually apparent in large, developed cities, where roads, cables, and pipes become ubiquitous and eventually migrate into the third dimension, above or below ground.

Measurements of electrical cable length and dissipative losses in German urban power grids (*12*) further confirm these expectations and support another key result obtained above: The energy loss in transport processes scales like socioeconomic rates (and both are superlinear). This shows how cities are fundamentally different from other complex systems, such as biological organisms (*4*, *5*) or river networks (*2*), which are thought to have evolved to minimize energy dissipation. Thus, the framework developed here also brings into focus efforts for sustainable urban development, by showing what kind of energy budget must be expended in order to keep cities of varying sizes socially connected.

The predictions of the theory are further supported by data on the size of urban economies from hundreds of cities in several continents, such as those in the United States (Fig. 1B), Japan (fig. S3), China (fig. S2A), or Germany (fig. S2B). In particular, the specific result that scaling exponents remain invariant over time, and are independent of population size and level of development, is confirmed by data for wages in U.S. metropolitan areas spanning 40 years (fig. S3). Direct empirical tests on the predictions made here for individual properties remain more difficult, but are confirmed, for example, by measurements for the scaling of social interactions with city size in the cell phone networks of two European nations (*21*), and for certain other patterns of individual behavior (*12*, *20*, *31*). Nevertheless, for most nations, we cannot yet access all predicted urban quantities simultaneously, especially in developing countries. This provides many future tests and applications for the theory, especially where understanding urbanization is most critical.

The spatial concentration and temporal acceleration of social interactions in cities has some striking qualitative parallels in other systems that are also driven by attractive forces and become denser with scale (*20*, *30*). The most familiar are stars, which burn faster and brighter (superlinearly) with increasing mass. Thus, although the form of cities may resemble the vasculature of river networks or biological organisms, their primary function is as open-ended social reactors. This view of cities as multiple interconnected networks that become denser with increasing scale (*32*) may also help to elucidate the function of other systems with similar properties, from ecosystems to technological information networks, despite their different relationships to physical space.

References and Notes

1. L. Mumford, *The City in History: Its Origins, Its Transformations, and Its Prospects* (Harcourt, Brace & World, Inc., New York, 1961).
2. I. Rodríguez-Iturbe, A. Rinaldo, *Fractal River Basins: Chance and Self-Organization* (Cambridge Univ. Press, New York, 1997).
3. E. N. Bacon, *The Design of Cities* (Penguin Group USA, New York, 1976).
4. G. B. West, J. H. Brown, B. J. Enquist, *Science* **276**, 122 (1997).
5. G. B. West, J. H. Brown, B. J. Enquist, *Science* **284**, 1677 (1999).
6. H. Samaniego, M. E. Moses, *J. Transport. Land Use* **1**, 21 (summer 2008).
7. J. S. Waters, C. T. Holbrook, J. H. Fewell, J. F. Harrison, *Am. Nat.* **176**, 501 (2010).
8. D. S. Dendrinos, H. Mullally, *Urban Evolution: Studies in the Mathematical Ecology of Cities* (Oxford Univ. Press, Oxford, 1985).
9. L. Sveikauskas, *Q. J. Econ.* **89**, 393 (1975).
10. E. L. Glaeser, B. Sacerdote, *J. Polit. Econ.* **107**, S225 (1999).
11. L. M. A. Bettencourt, J. Lobo, D. Strumsky, *Res. Policy* **36**, 107 (2007).
12. L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G. B. West, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7301 (2007).
13. M. Batty, *Science* **319**, 769 (2008).
14. See supplementary materials for details.
15. R. Anderson, R. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, Oxford, 1991).
16. S. Angel, J. Parent, D. L. Civco, A. Blei, D. Potere, *Prog. Plann.* **75**, 53 (2011).
17. D. Saunders, *Arrival City* (Pantheon Books, New York, 2010).
18. S. Nordbeck, *Geogr. Ann.* **53**, 54 (1971).
19. M. Barthélemy, *Phys. Rep.* **499**, 1 (2011).
20. S. Milgram, *Science* **167**, 1461 (1970).
21. M. Schläpfer *et al.*, *The Scaling of Human Interactions with City Size*; http://arxiv.org/abs/1210.5215.
22. J. Jacobs, *The Economy of Cities* (Vintage Books, New York, 1969).
23. A. Marshall, *Principles of Economics* (MacMillan, London, 1920). See book IV, chapter X.
24. E. L. Glaeser, J. D. Gottlieb, *J. Econ. Lit.* **47**, 983 (2009).
25. A. Gomez-Lievano, H. Youn, L. M. A. Bettencourt, *PLoS ONE* **7**, e40393 (2012).
26. C. Alexander, *Archit. Forum* **122**, 58 (1965).
27. A. Downs, *Traffic Quarterly* **16**, 393 409 (1962).
28. G. Duranton, M. A. Turner, *Am. Econ. Rev.* **101**, 2616 (2011).
29. S. Sassen, *The Global City: New York, London, Tokyo* (Princeton Univ. Press, Princeton NJ, 1991).
30. M. Fujita, P. Krugman, A. J. Venables, *The Spatial Economy: Cities, Regions, and International Trade* (MIT Press, Cambridge MA, 2001).
31. F. Lederbogen *et al.*, *Nature* **474**, 498 (2011).
32. J. Leskovec, J. Kleinberg, C. Faloutsos, "Graphs over time: Densification laws, shrinking diameters and possible explanations." In *KDD '05 Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pp. 177–187 (2005); http://dl.acm.org/citation.cfm?id=1081893.

Acknowledgments: I thank J. Lobo, G. West, and H. Youn for discussions. This research was supported by grants from the Rockefeller Foundation, James S. McDonnell Foundation (grant 220020195), NSF (grant 103522), Bill and Melinda Gates Foundation (grant OPP1076282), John Templeton Foundation (grant 15705), and Bryan J. and June B. Zwan Foundation.

Supplementary Materials

www.sciencemag.org/cgi/content/full/340/6139/1438/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
Tables S1 to S3
References (*33*–*63*)

29 January 2013; accepted 29 April 2013
10.1126/science.1235823