Bettencourt *et al*. 10.1073/pnas.0610172104.

Supporting Information

Files in this Data Supplement:

SI Figure 5

Fig. 5. Population of New York City's metropolitan service area over time (1780-2004; data are from the U.S. Census Bureau). Population evolution is punctuated by periods of superexponential growth, separated by brief periods of deceleration.

SI Figure 6

Fig. 6. Relative growth rate of New York City's population experiences successive superexponential growth periods, punctuated by brief decelerations. During these periods, the relative rate grows with the population as predicted under super linear scaling. Note that exponential growth would result in a horizontal line, independent of population, which is not a feature of the observed dynamics.

SI Text

Data Sources for Specific Indicators and Additional Remarks

· New patents refers to number of new patents granted by the U.S. Patent and Trademark Office over the period of 1 year to authors residing in a given metropolitan statistical area. Inventors refers to number of patent authors over a given year, inferred from the same source. Data are courtesy of Deborah Strumsky (Harvard Business School, Cambridge, MA) (see also ref. 1). We studied the scaling relations for patents and inventors over the period 1980-2001 and have found systematic scaling with metropolitan size and exponents that are commensurate within 95% confidence intervals over this time period.

· Private R&D employment (U.S.) and private R&D establishments (U.S.) data are taken from the U.S. Economic Census, which provides information on the employment levels and number of establishments engaged in conducting original investigation undertaken on a systematic basis to gain new knowledge (research) and/or the application of research findings or other scientific knowledge for the creation of new or significantly improved products or processes (experimental development). Data are available from the Economic Census, U.S. Census Bureau (www.census.gov/econ/census02).

· R&D employment (China) is analogous to private R&D employment U.S. but also includes public establishments, which, in China comprise a substantial fraction of employment in the sector. Data are from the National Bureau of Statistics, China (www.stats.gov.cn), courtesy of Shannon Larsen (Santa Fe Institute, Santa Fe, New Mexico).

· Total wages (U.S.) refers to total income from wages over the period of a year. Data are from the Bureau of Economic Analysis, Regional Economic Accounts (www.bea.gov/bea/regional/data.htm).

· GDP (China) refers to gross domestic product of metropolitan areas in China (Urban Administrative Units). Data are from the National Bureau of Statistics, China (www.stats.gov.cn), courtesy of Shannon Larsen.

· GDP (Europe and EU countries) is analogous to metropolitan GDP (China). Data are from Urban Audit, Eurostat (http://epp.eurostat.cec.eu.int/). German metropolitan road surface was obtained from the same source.

· Total housing (U.S.) refers to total dwellings per metropolitan statistical area. Data are from the County and City Data Book, U.S. Census Bureau (www.census.gov/statab/www/ccdb.html).

· Total employment (U.S.) refers to full-time and part-time wage positions. Data are from the Bureau of Economic Analysis, Regional Economic Accounts (www.bea.gov/bea/regional/data.htm).

· Total household electricity consumption (China) refers to total energy consumed in households in urban administrative units in 2002. Data are from the National Bureau of Statistics, China (www.stats.gov.cn), courtesy of Shannon Larsen.

· Total household water consumption (China) is the total volume of water consumed by metropolitan households in urban administrative units in 2002. Data are from the National Bureau of Statistics, China (www.stats.gov.cn), courtesy of Shannon Larsen.

· Fuel sales by gasoline station (U.S.) refers to the amount of fuel sales in dollars per year sold by gasoline stations located in metropolitan areas. Data are available from the Economic Census, U.S. Census Bureau (www.census.gov/econ/census02).

· New AIDS cases in selected American cities were obtained from the Center for Disease Control (U.S.), Divisions of AIDS/HIV Prevention (www.cdc.gov/hiv/surveillance.htm).

· According to the definition put forward by Florida (2), "supercreative" professions are computer and mathematical; architecture and engineering; life, physical, and social science occupations; education; training and library; arts, design, entertainment, sports, and media occupations. The occupational classifications were derived from the Standard Occupational Classification System (SOC) introduced by U.S. Bureau of Labor Statistics in 1998. The SOC classification data are constructed by using the North American Industrial Classification System (NAICS). Data are courtesy of Richard Florida and Kevin Stolarick (George Mason University, Manassas, VA).

· German electricity data refers to consumption, generation, and distribution for German cities. Data were compiled by the Verband der Elektrizitätswirtschaft (VDEW) and published by VWEW (Verlags- und Wirtschaftsgesellschaft der Energiewirtschaft) Energieverlag (www.vwew.de). The data contain variables collected from German electricity producers, most of which are local electricity suppliers to a single city. The electricity market was opened in 1998, which means that that local power plants can be owned by their city authority and that a city can now buy its electricity from other nonlocal producers. In practice, however, according to information from the VDEW, suppliers still deliver today nearly 100% of their generated power to their local cities and meet their needs almost completely.

Estimation of the Magnitude of Superlinear Scaling Exponents

The generality of superlinear exponents associated with social indicators leads to the question of the magnitude of their values. Here we present an argument that leads to a semiquantitative estimate of the observed numerical ranges. We emphasize that a detailed predictive theory for the scaling exponents b, integrating interactions between people and institutions, is the ultimate long-term goal of any picture of urban scaling. The argument below simply integrates some of its necessary ingredients to produce a rough estimate of the exponent's plausible ranges.

The exponents b are commensurate for many social quantities, but there is no strong indication that they must be identical for different urban systems. Can we determine b from the formulation of a maximization or minimization principle, as was done in biology for the properties of networks of resource distribution? There is little doubt that human interactions in a city may be represented in terms of networks, and it is difficult to foresee their general structural properties. What we do know is that if a city provides an enlarged space of opportunities for effective interactions between people but also that the number and intensity of such interactions is constrained by time and effort and by limits on individual cognition. This is what Milgram, writing on the experience of living in cities (3), referred to as information saturation. This observation can be used to produce an estimate of the values of b.

First, consider the total number of effective contacts *C* between individuals in a population of size *N*. The maximal value that *C* can take is $C = N(N - 1)/2$, implying a bound on $\beta \leq 2$. This upper bound corresponds to every individual in a city knowing everyone else, which is clearly not realistic as cities grow large. Instead, consider that the quantities Y of Table 1 are proportional to the number of effective contacts so that $C(N) = C_0 N^{\beta}$. Let's now define P as the ratio of productive contacts per capita between the largest city with population N_{max} and the smallest city with population N_{min}, so that

$$P = \left(\frac{N_{max}}{N_{min}}\right)^{\beta-1} \rightarrow \beta = 1 + \frac{\log(P)}{\log\left(\frac{N_{max}}{N_{min}}\right)} \quad \textbf{[1]}$$

P expresses by how much an individual's time, effort, and cognitive ability can be expanded in response to the greater demands of the largest city, relative to those of the smallest town. If we assume P = 10-100, and $N_{max}/N_{min} = 10^7$, we obtain b = 1.14-1.28, which is in qualitative agreement with the observations.

Additional Figures for Superexponential Growth Periods of New York City Population

Note also that periods of superexponential growth have been identified for the total human population of the world in (4-6).

Generalized Growth Equation

Here, we dispense with the assumption that the costs of maintenance are linear in N. As such we write the growth equation as

$$\frac{dN(t)}{dt} = \left(\frac{Y_0}{E}\right) N^{\beta}(t) - \left(\frac{R}{E}\right) N^{\alpha}(t). \quad \textbf{[2]}$$

Qualitatively, the solution has the same properties discussed in the main text. If both b and a <1 the solution is always sigmoidal. If at least one a or b >1 there are always two distinct possibilities (*i*) b >1, b > a, and (*ii*) a >1, a > b.

Regime (*i*), which includes the interesting possibility that b >a > 1, has two possible solutions:

· $N(0) > (R/Y_0)^{1/(b-a)}$: The rate of growth is dominated by available resources and the solution is a growing superexponential, which reaches infinity in finite time.

· $N(0) < (R/Y_0)^{1/(b-a)}$: Costs dominate the rate and the population collapses.

Consequently, in this regime, the fixed point $N = (R/Y_0)^{1/(b-a)}$ is unstable, and small perturbations around it grow either toward a finite time singularity or collapse.

Regime (*ii*), analogously, has two possible solutions

· $N(0) > (R/Y_0)^{1/(b-a)}$: costs dominate the growth rate, and the solution decreases toward the carrying capacity N¥ $= (R/Y_0)^{1/(b-a)}$.

· $N(0) < (R/Y_0)^{1/(b-a)}$: Available resources dominate the rate, and the solution grows toward *N*¥.

Thus, in Regime (*ii*) the carrying capacity *N* = *N*¥ is a stable fixed point to which population trajectories in both regimes converge to at long times.

For completeness, we give an analytical method to solve the generalized growth equation.

Define $A = Y_0/E$ and $B = Y_0/E$. The growth equation can be simplified by a change of variables $y = N^{1-b}$, which results in

$$\frac{dy}{dt} = (1-\beta)A - (1-\beta)By^{\lambda}$$, **[3]**

where l = (a - b)/(1 - b). This equation can be solved easily given the function *f* such that

$$\frac{df}{dy} = \frac{1}{A - By^{\lambda}}$$. **[4]**

The solution then obeys $f(y) = f(N^{1-b}) = (1-b)t$.

In particular we see that if l = 1 (i.e., a = 1) then $f(y) = -B^{-1} \log (A - By)/K$, where *K* is an integration constant, and a little algebra reveals the solution of the main text.

In general, we obtain

$$f(y) = \frac{y}{A}\,{}_2F_1\left(\lambda^{-1}, 1, 1+\lambda^{-1}, -\tfrac{B}{A}y^{\lambda}\right) - K$$ ', **[5]**

where ${}_2F_1$ is Gauss' hypergeometric function and *K*¢ is an integration constant. The function ${}_2F_1$ does not have a simple closed form in terms of elementary functions.

However, it does have an interesting transformation property due to Euler

$${}_2F_1(a,b,c,z) = (1-z)^{-b}\,{}_2F_1\left(c-a, b, c, \frac{z}{z-1}\right)$$. **[6]**

With this transformation we can write the solution as

$$\frac{N^{1-\beta}}{A + BN^{\alpha-\beta}}\,{}_2F_1\left(1, 1, 1+\lambda^{-1}, \frac{\frac{B}{A}N^{\alpha-\beta}}{1+\frac{B}{A}N^{\alpha-\beta}}\right) = K + (1-\beta)t$$, **[7]**

where *K* equals the left side of the equation at time *t* = 0. Although nontrivial in general, the argument of the hypergeometric function goes to a pure number in the limit of large *N*. We are then left with a transcendental equation for *N*, which depends both on the sign of 1 - b and of a - b.

1. Bettencourt LMA, Lobo J, Strumsky D (2006) Santa Fe Institute Working Paper 04-12-038, *Res Policy*, 86:107-120.

2. Florida R (2002) *The Rise of the Creative Class* (Basic Books, Cambridge MA), pp 327-329.

3. Milgram S (1970) *Science* 167:1461-1468.

4. Cohen JE (1995) (1995) *Science* 269:341-346.

5. Kremer M (1993) *Q J Econ* 108:681-716.

6. Kapitza SP (1996) *Phys-Usp* 39:57.