

Scaling laws between population and facility densities

Jaegon Um^{a,1}, Seung-Woo Son^{b,1}, Sung-Ik Lee^c, Hawoong Jeong^{b,2}, and Beom Jun Kim^{d,2}

^aDepartment of Physics, Pohang University of Science and Technology, Pohang 790-784, Korea; ^bDepartment of Physics, Institute for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea; ^cNational Creative Research Initiative Center for Superconductivity, Department of Physics, Sogang University, Seoul, 121-742, Korea; and ^dBK21 Physics Research Division and Department of Energy Science, Sungkyunkwan University, Suwon 440-746, Korea

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved June 24, 2009 (received for review February 20, 2009)

When a new facility like a grocery store, a school, or a fire station is planned, its location should ideally be determined by the necessities of people who live nearby. Empirically, it has been found that there exists a positive correlation between facility and population densities. In the present work, we investigate the ideal relation between the population and the facility densities within the framework of an economic mechanism governing microdynamics. In previous studies based on the global optimization of facility positions in minimizing the overall travel distance between people and facilities, it was shown that the density of facility D and that of population ρ should follow a simple power law $D \sim \rho^{2/3}$. In our empirical analysis, on the other hand, the power-law exponent α in $D \sim \rho^\alpha$ is not a fixed value but spreads in a broad range depending on facility types. To explain this discrepancy in α , we propose a model based on economic mechanisms that mimic the competitive balance between the profit of the facilities and the social opportunity cost for populations. Through our simple, microscopically driven model, we show that commercial facilities driven by the profit of the facilities have $\alpha = 1$, whereas public facilities driven by the social opportunity cost have $\alpha = 2/3$. We simulate this model to find the optimal positions of facilities on a real U.S. map and show that the results are consistent with the empirical data.

optimal positioning | social opportunity cost | microdynamics model

Starbucks Coffee Company, at the height of its success, opened 4.5 new stores every day, and the convenience store chain Seven Eleven opened 6. Interesting questions for these companies to consider when opening a new store might include where should we locate it, and what factors do we need to consider as we select a location? Floating populations, flow-by traffic, and nearby commercial areas are often important factors in determining optimal positions with respect to economics and business growth. Clearly, locating stores is a complicated but very important issue because the provision of better services to more people with fewer facilities is a challenge not only for various kinds of commercial service industries but also for government social welfare agencies. Although various fields such as business economics, systems engineering, computer science, geography, and even biology have approached this issue from an optimal positioning standpoint (1–8), in most cases ad-hoc local solutions without a theoretical basis have been applied in practice.

An unevenly distributed population makes this positioning problem even more difficult. Because the world's populations are aggregated in urban districts centered around old settlement points, the distribution patterns of populations are very uneven. The population density is nearly log-normal distribution, not only on a global scale but also on subnational ones (1, 9). Economic geography has been investigating the problem of optimal positioning of facilities given uneven population densities (2–4), and has found that the relation between the population density ρ and the facility density D follows the simple scaling $D \sim \rho^\alpha$ in a broad range of ρ , and that the exponent α depends on types of facilities (3–5). Intuitively, within a given country, the more populated an area is the more facilities exist, whereas there are fewer facilities in places where the population density is lower. In that sense, the exponent $\alpha = 1$, which represents a linearly proportional relation

between population and the number of facilities, looks natural, but the empirical data do not support this argument. For example, in the case of public service facilities, the exponents have been observed to be $2/3$. Although the exponent $\alpha = 2/3$ is derived analytically by minimizing the total travel distance between people and facilities from the global optimization scheme (5, 10, 11), the theory has failed to explain $\alpha = 1$ in the same approach, and only a phenomenological reformulation of the theory has been tried (5).

Generally, commercial facilities like small stores have to attract large numbers of people to make a profit. If they cannot do this, such stores will need to close and move to a place with larger population. On the other hand, public facilities such as fire stations and public schools need to be built in positions where they are as close as possible to their clients, because the consumed travel time or spatial distance yields the social opportunity cost by depriving visitors of time for producing other products or services. For example, public schools that exist in big cities require students who live in nearby small towns or villages to commute for a long period of time. To reduce the social opportunity cost caused by such a long-distance commute, new schools should be built in places where the population is sparse but not negligible. Therefore, one can expect that the public facilities should be more evenly distributed spatially than commercial ones. This inference is consistent with the global optimization of facility positioning to minimize the total travel time in the previous studies (5, 10, 11). However, the global optimization is not applicable to the real economic system because it costs a great deal to regulate the locations of all individual facilities. Therefore, a simple but more realistic model is needed to explain the positions of facilities in general. In this work, we propose a simple model that incorporates relevant economic factors suitable both for public and commercial facilities to explain the two different exponents, $2/3$ and 1 , in a unified argument.

Empirical Data

To explore the scaling relation between population and the facility densities with up-to-date data, we have gathered and analyzed extensive empirical data on the positions of various facilities in 2 socially and culturally different countries: the United States (US) and South Korea (SK). For the U.S. data, we used the databases of the U.S. Census Bureau (www.census.gov/) and Fedstats (www.fedstats.gov/), and, especially for educational data such as that site of the National Center for Education Statistics (<http://nces.ed.gov/>). We gathered the Korean data from the Korean Statistical Information System (<http://kosis.nso.go.kr/>) and the National Geographic Information Institute (<http://www.ngii.go.kr/>). Ideally, if we knew the exact position of each facility, we could introduce the set of points closer to a specific facility than to any other facilities, which is the so-called

Author contributions: S.-I.L., H.J., and B.J.K. designed research; J.U. and S.-W.S. performed research; and J.U., S.-W.S., H.J., and B.J.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹J.U. and S.-W.S. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: hjeong@kaist.edu or beomjun@skku.edu.

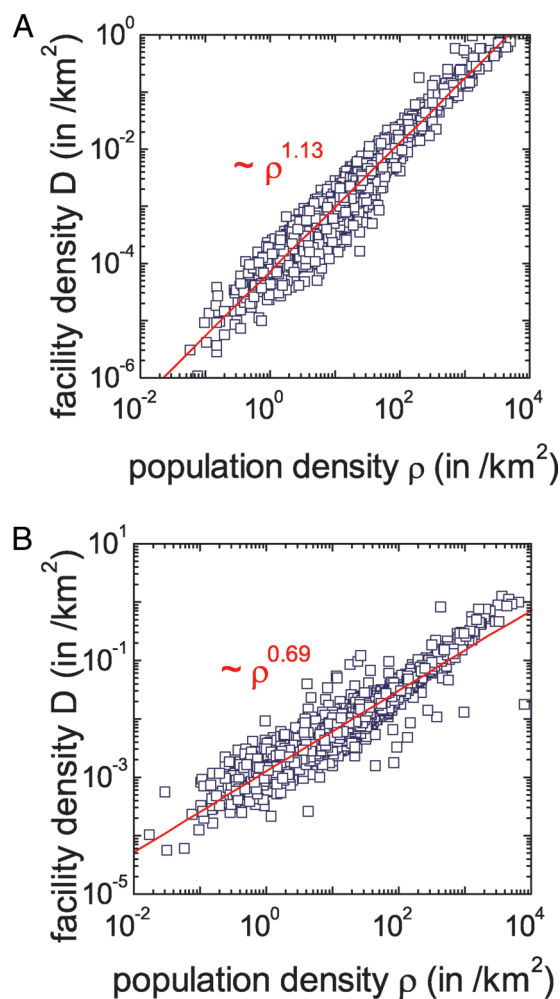


Fig. 1. Scatter plots of population and facility density obtained from empirical data. (A) Facility density D versus population density ρ for ambulatory hospitals in the US. (B) D versus ρ for the public schools in the US. For ambulatory hospital the exponent is 1.13 close to 1, which shows clearly a different distribution from the exponent 0.69 for public schools. The public school shows roughly 2 regimes of different exponents around $\rho \sim 100/\text{km}^2$. The region above $100/\text{km}^2$ of population density shows the exponent as ≈ 1 , and the region below $100/\text{km}^2$ shows $2/3$. The population density $100/\text{km}^2$ corresponds to the cross-over point of the facility density $0.03/\text{km}^2$, which means that 1 school covers about 33 km^2 . If we assume the geometry is nearly a circle, the radius of the attending distance is $\approx 3 \text{ km}$. For attending distances $< 3 \text{ km}$, public school distribution also shows similar behaviors to those of the private schools.

Voronoi cell. However, due to the resolution limitation of the facility positions, we measure the scaling relation by using the number of people and facilities of each county instead as a coarse-grained scheme (see the *Materials and Methods* for details). From an analysis of the data, it was found in both the US and SK that although the commercial facilities tend to have an exponent $\alpha \approx 1$, the public facilities commonly have an exponent $\alpha \approx 2/3$. For commercial facilities such as ambulatory hospitals, the exponent α was found to be 1.13 (top), for public facilities such as schools, the exponent was 0.69 (bottom), as shown in Fig. 1. The other results for various facilities in US and SK are summarized in Table 1.

Interestingly, as a border of exponent 0.8 in Table 1, each facility is clearly classified into 2 categories, commercial and public. In the US data, ‘ambulatory hospital,’ ‘beauty care,’ ‘laundry,’ ‘automotive repair,’ ‘private school,’ ‘restaurant,’ ‘accommodation,’ ‘bank,’ and ‘gas station’ are categorized as commercial facilities more or less, whereas ‘death care,’ ‘fire station,’ ‘police station,’ and ‘public school’ are categorized as public facilities, which provide

social welfare services. Even though the 2 countries have very different physical, economic, sociocultural, and political environments, (such as different topography, urban formations, population densities, standards of living, and cultures), similar facilities, like the ambulatory hospital in the US and the primary clinic in SK, have similar exponents.

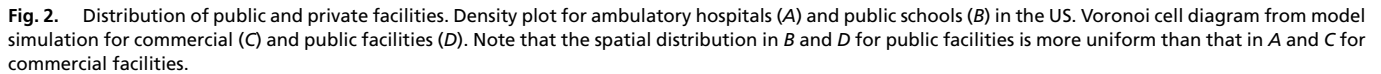
It should be noted, however, that due to the differences in the educational systems in the US and SK, the result-related school system is somewhat different. Because the US has a K-12 educational system both in the public and private sectors, the table shows the exponent 0.95 for private schools and 0.69 for public schools. In SK, even though there is a distinction between public and private schools, this distinction is deceptive, as most students are allocated to public and private schools by the ministry of education without any preference between private and public school. Therefore, the exponent of primary and secondary schools in SK is the same, 0.77, close to the public facilities and proving no distinction between private and public schools in SK. However, ‘university/college,’ which is a more profit-driven type of school in SK shows $\alpha \sim 0.93$, close to the commercial facility.

As a result, the social facilities, such as police and fire stations and government offices, have respectively low exponents close to $2/3$. One of major differences between public and commercial facilities is probably the distance between these facilities and their clients. This is especially the case with respect to social facilities. For example, fire stations need to be located close to people’s residences, so that fire fighters can serve quickly if needed, and the same applies to police stations. For a school student, the commute distance is also important. As an extreme case, public health

Table 1. Summary of the exponents

US facility	α (SE)	R^2
Ambulatory hospital	1.13(1)	0.93
Beauty care	1.08(1)	0.86
Laundry	1.05(1)	0.90
Automotive repair	0.99(1)	0.92
Private school	0.95(1)	0.82
Restaurant	0.93(1)	0.89
Accommodation	0.89(1)	0.70
Bank	0.88(1)	0.89
Gas station	0.86(1)	0.94
Death care	0.79(1)	0.80
* Fire station	0.78(3)	0.93
* Police station	0.71(6)	0.75
Public school	0.69(1)	0.87
SK facility	α (SE)	R^2
Bank	1.18(2)	0.96
Parking place	1.13(2)	0.91
* Primary clinic	1.09(2)	1.00
* Hospital	0.96(5)	0.97
* University/college	0.93(9)	0.89
Market place	0.87(2)	0.90
* Secondary school	0.77(3)	0.98
* Primary school	0.77(3)	0.97
Social welfare org.	0.75(2)	0.84
* Police station	0.71(5)	0.94
Government office	0.70(1)	0.93
* Fire station	0.60(4)	0.93
* Public health center	0.09(5)	0.19

Summary of the values of α in $D \sim \rho^\alpha$ for various facilities in the US and SK. The coefficient of determination R^2 is obtained through the least-squares analysis. The value for each facility type is obtained from information in the county level, except for the asterisk(*)-marked values, which are from the state level (US) and the province level (SK). The numbers in parentheses are the standard errors in the last digits.



n_i , the number of visitor n_i becomes almost the same with each other, and facilities no longer have to change their location. By using the Voronoi cell and its area s_i , we obtain the expressions of the population density at certain position \mathbf{r} , $\rho(\mathbf{r}) = n(\mathbf{r})/s(\mathbf{r})$ in continuous form and the facility density $D(\mathbf{r}) = 1/s(\mathbf{r})$, where $n(\mathbf{r}) = n_i$ and $s(\mathbf{r}) = s_i$ if $\mathbf{r} \in V_i$. At the steady state, $n(\mathbf{r}) \approx N_p/N_f$ leads to

$$D(\mathbf{r}) \approx \frac{N_f}{N_p} \rho(\mathbf{r}). \quad [1]$$

From Eq. 1, we reach the conclusion that the exponent $\alpha = 1$ for the commercial facilities is consistent with the empirical data.

$$c_i = \sum_{k \in P(V_i)} |\mathbf{r}_{ik}| = n_i \langle r_i \rangle, \quad [2]$$

If optimization of facility positions is achieved by consecutive relocations of facilities to positions near the facility of the highest cost, the cost of every facility becomes the same in the steady state, i.e., $c_i = c$. With a plausible assumption that $\langle r_i \rangle \approx g\sqrt{s_i}$ with a geometrical constant g ($\sim \mathcal{O}(1)$), the summation of travel distance for V_i can be written as $c_i = n_i g\sqrt{s_i}$. By using $\rho(\mathbf{r}) = n(\mathbf{r})/s(\mathbf{r})$ and $D(\mathbf{r}) = 1/s(\mathbf{r})$, we get the following expression from $c = n_i g\sqrt{s_i}$,

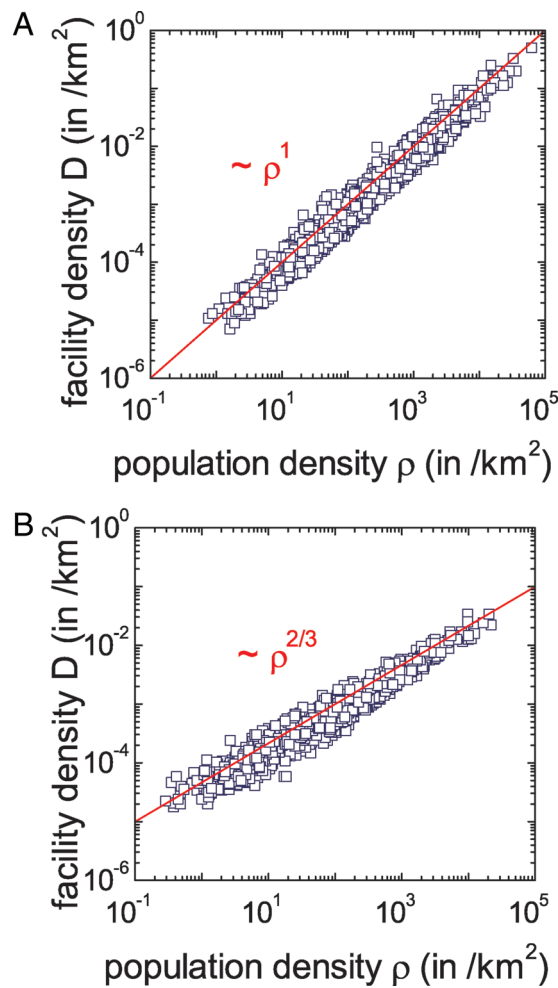


Fig. 3. Simulation results. (A) Facility density D versus population density ρ for the commercial facility. (B) D versus ρ for the public facility.

$$D(\mathbf{r}) = 1/s(\mathbf{r}) = (g/c)^{2/3} [\rho(\mathbf{r})]^{2/3}. \quad [3]$$

To obtain c as the form of ρ , we rewrite c as

$$c = g/N_f \int_A \rho(\mathbf{r}) \sqrt{s(\mathbf{r})} d^2r, \quad [4]$$

where we have used $\int_{V_i} 1/s(\mathbf{r}) d^2r = 1$ and $\int_A 1/s(\mathbf{r}) d^2r = N_f$ with V_i and A denoting the region of integration. From Eqs. 3 and 4, we obtain

$$D(\mathbf{r}) = N_f \frac{[\rho(\mathbf{r})]^{2/3}}{\int_A [\rho(\mathbf{r})]^{2/3} d^2r}, \quad [5]$$

which gives $\alpha = 2/3$. Therefore, our model precisely predicts $\alpha = 1$ for the commercial facility, and $\alpha = 2/3$ for the public facility. Especially for the case of the public facility, our result is equivalent to the scaling behavior found in the global optimization (5, 10, 11).

Not every facility has the exact exponents $\alpha = 1$ or $2/3$; they have a broad range of exponents as shown in Table 1. This range exists because in reality, each facility can have both commercial and public characteristics. In this respect, in order to explain various values of the exponent α , we suggest the generalized objective function v_i for the i th facility, which can be written as $v_i = n_i \langle r_i \rangle^\beta$ to cover above mentioned various types of facilities. From the definition of the Voronoi cell, we write v_i as

$$v_i = g^\beta \frac{\rho(\mathbf{r})}{[D(\mathbf{r})]^{(\beta+2)/2}}. \quad [6]$$

Then, finally we obtain the following relation from the optimization condition

$$D(\mathbf{r}) = N_f \frac{[\rho(\mathbf{r})]^{2/(\beta+2)}}{\int_A [\rho(\mathbf{r})]^{2/(\beta+2)} d^2r}, \quad [7]$$

which gives $\alpha = \frac{2}{\beta+2}$. Consequently, when $\beta = 0$, corresponding to the objective function $v_i = n_i$ for the commercial facility, one gets $\alpha = 1$, whereas when $\beta = 1$, corresponding to the public facility, $\alpha = 2/3$ is obtained. By tuning the exponent β , it is possible to get the various exponent. For example, a public health center in SK shows $\alpha \approx 0.1$ in Table 1. Because the most important factor is distance in this case, the exponent β goes to infinite and α converges to zero.

Simulation Result

We perform extensive computer simulations of the above model with a real US population distribution (see *Materials and Methods*). Consistent with both the analytic prediction and our empirical analysis for the US and SK, commercial facilities are found to have $\alpha \approx 1$, whereas public ones have $\alpha \approx 2/3$. In Fig. 2A and C, the density of ambulatory hospitals in the US and the Voronoi cell diagram from a computer simulation for commercial facilities are displayed, and in Fig. 2B and D, the density of public schools in the US and the simulation result for public facilities are shown. The difference of the spatial unevenness between the commercial and the public facility distributions is clearly observed, which is consistent with different values of α in Fig. 3 for facility densities versus population densities. In Fig. 2A and C, the density of the facilities is more uneven than that of Fig. 2B and D, and the size of the Voronoi cell in Fig. 2D is more regular and uniform than that of Fig. 2C.

Social Opportunity Cost

From the simulation result, we find that the facility distribution is changed dramatically in accordance with the facility type, and the facility densities are proportional to the power of the population densities with an exponent of not only 1 but also 2/3. If the public facilities behave like commercial ones for their profit, how much more do we have to spend in terms of the social opportunity cost? From our comparison between public and commercial facilities with the same N_f , the visitors' total travel distance to commercial facilities is 1.5 times longer than to public facilities, which means that students have to spend 1.5 times more time attending school and that firefighters and policemen have to go 1.5 times farther, on average, to fight a fire or crime. This interesting result can be interpreted as a location version of the price of anarchy for public facility (13).

The expression of total distance between visitors and commercial facilities can be obtained from Eq. 4:

$$\bar{c} N_f = g \frac{N_p^{1/2}}{N_f^{1/2}} \int_A \sqrt{\rho(\mathbf{r})} d^2r, \quad [8]$$

where \bar{c} represents the average cost. Therefore, we obtain the ratio r of the total travel distances for commercial and public facilities,

$$r \approx \frac{[\int_A \rho(\mathbf{r}) d^2r]^{1/2} \cdot \int_A [\rho(\mathbf{r})]^{1/2} d^2r}{[\int_A [\rho(\mathbf{r})]^{2/3} d^2r]^{3/2}}. \quad [9]$$

From Eq. 9, one can find easily that $r > 1$ comes from the uneven population distribution: If the population is distributed uniformly, $r = 1$ because the densities of public and commercial facility will have the same uniform distribution, and there is no social opportunity cost caused from the self-interest of a profit-oriented commercial facility.

Summary and Discussion

We have found from the empirical analysis of distributions for a variety of different facilities in the US and SK that the facility density D and the population density ρ are positively correlated to each other, described by $D \sim \rho^\alpha$. Commercial and public facilities have been observed to have different distributions: $\alpha \approx 1$ for the former and $\alpha \approx 2/3$ for the latter. We have proposed a simple model with a focus on the microscopic activity of 2 different types of facilities, driven by profit and social opportunity cost, and successfully reproduced as $\alpha = 1$ and $\alpha = 2/3$ for the commercial and the public facilities, respectively.

If the process of removing a facility is prohibited in numerical simulation, the distribution of facilities is found to differ from Fig. 3. In the graph of facility density versus population density, the data scatter in a broader area above the scaling line $D \sim \rho^\alpha$. The data scatters because the addition process alone makes the points in the lower part of the curve $D \sim \rho^\alpha$ move to the upper part and not vice versa. On the other hand, if removals of facilities are allowed in addition, the upper points move to the lower part, making the scaling law $D \sim \rho^\alpha$ more visible. The difference in results comes from the presence of regulating processes such as removal or relocation, which are also shown in other studies of optimal networks (11, 14, 15).

Materials and Methods

For empirical study, we use compilations of data (see *Empirical Data* for details) for the US and SK, which contain information of the number of various facilities and population for 3, 147 counties (mean area 2,916 km²) for the US including Alaska and Hawaii and 234 municipal counties (mean area 426 km²) for SK. Most of the data are analyzed at county level, and when the data of county level, are not accessible, we use the coarse-grained data of subnational level, 50 states and the District of Columbia for the US and 9 provinces and 7 metropolises for SK. In our simulation model, we first put

N_p people according to the US population density distribution and randomly distribute N_0 facilities, where N_0 is less than the final number of facilities N_f . With the plausible assumption that people will visit their nearest facility, we compute n_i and c_i for the case of commercial and public facilities, respectively. At each step, we either eliminate a commercial (public) facility of the lowest $n_i(c_i)$ or create one near the facility of the highest $n_i(c_i)$, and then all $n_i(c_i)$ are updated. We also checked that when we create a facility, even in a randomly chosen place, during the process, it does not change the main result but only changes the time to the steady state. The number of facilities is increased and decreased periodically during simulations with the average number of facilities fixed at N_f , which is analogous to the expansion and contraction of a market in reality. For Fig. 2, we set $N_p = 100,000$ and $N_f = 2,000$ and simulation usually starts from $N_0 = N_f/10$ or N_f , which does not change the result. While we repeatedly create and eliminate 1 facility following the above procedure, the averages \bar{n} and \bar{c} and standard deviations are calculated. As our optimization procedure proceeds, the system is observed to approach the steady state with $\sigma_n/\bar{n} \sim 0.1$ or $\sigma_c/\bar{c} \sim 0.1$, where $\sigma_n(\sigma_c)$ is the standard deviation of $n(c)$, and optimal positions of N_f facilities within our framework are obtained to give us Voronoi cell plots such as in Fig. 2 C and D. Each Voronoi cell, by definition, contains 1 facility, and thus the density D_i of the facility in V_i is nothing but the inverse of the area s_i of the Voronoi cell containing the i th facility. By using the gridded 1 km population data of US, we compute the population density within the i th Voronoi cell. The scatter plots for D and ρ like Fig. 3 are made, and the exponent α in $D \sim \rho^\alpha$ is computed. A Java applet version of our simulation is available at <http://statphys.skku.ac.kr/bjkim/Applet/opof.html>.

ACKNOWLEDGMENTS. We thank S. Combe, M. T. Gastner and P.-J. Kim for useful discussions. This work was supported by Korea Science and Engineering Foundation Grant R01-2007-000-20084-0 (to B. J. K.), the Korean Systems Biology Program from the Ministry of Education, Science, and Technology through Korea Science and Engineering Foundation Grant M10309020000-03B5002-00000 (to H. J.). S. I. L. was supported by the Creative Research Initiative Program. We also acknowledge support from Korea Research Council of Fundamental Science and Technology and Samsung Economic Research Institute.

1. Yook SH, Jeong H, Barabási AL (2002) Modeling the Internet's large-scale topology. *Proc Natl Acad Sci USA* 99:13382–13386.
2. Lloyd PE, Dicken P (1977) *Location in space: A theoretical approach to economic geography* (Joanna Cotler, London).
3. Stephan G (1977) Territorial division: The least-time constraint behind the formation of subnational boundaries. *Science* 196:523–524.
4. Vining DR, Yang CH, Yeh ST (1979) Political subdivision and population density. *Science* 205:219.
5. Stephan G (1988) The distribution of service establishments. *J Region Sci* 28:29–40.
6. Hamilton MJ, Milne BT, Walker RS, Brown JH (2007) Nonlinear scaling of space use in human hunter-gatherers. *Proc Natl Acad Sci USA* 104:4765–4769.
7. Jetz W, Carbone C, Fulford J, Brown JH (2004) The scaling of animal space use. *Science* 306:266–268.
8. Haskell JP, Ritchie ME, Olff H (2002) Fractal geometry predicts varying body size scaling relationships for mammal and bird home ranges. *Nature* 418:527–530.
9. Engen S, Lande R (1996) Population dynamics models generation the lognormal species abundance distribution. *Math Biosci* 132:169–183.
10. Gusein-Zade SM (1982) Bunge's problem in central place theory and its generalizations. *Geogr Anal* 14:246–252.
11. Gastner MT, Newman MEJ (2006) Optimal design of spatial distribution networks. *Phys Rev E* 74:016117.
12. Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Econ Geogr* 46:234–240.
13. Youn H, Gastner MT, Jeong H (2008) Price of anarchy in transportation networks: Efficient and optimality control. *Phys Rev Lett* 101:128701.
14. Barthélemy M, Flammini A (2006) Optimal traffic networks. *J Stat Mech Theory E*, 10.1088/1742-5468/2006/07/L07002.
15. Xie YB, et al. (2007) Geographical networks evolving with an optimal policy. *Phys Rev E* 75:036106.