Urban Scaling, Economic Complexity and Cultural Evolution

Andres Gomez-Lievano
Postdoc at the Center for International Development, Harvard University

Contact info:
Email: andres_gomez@hks.harvard.edu
Twitter: @GomezLievano
An urban species

https://sites.google.com/a/mmrsschool.org/mmr-cities/home
% Urbanization, 2010

Log of P.C. GDP 2010 PPP

Fitted values

Graph by Prof. Edward Glaeser
Stark differences in levels of urbanization, wealth, innovation, health, crime, etc...

Why?
A presumptuous goal?

Big Data Needs a Big Theory to Go with It

Just as the industrial age produced the laws of thermodynamics, we need universal laws of complexity to solve our seemingly intractable problems

By Geoffrey West on May 1, 2013

Collaborators:
Luis M.A. Bettencourt
Jose Lobo
HyeJin Youn
Rachata Muneepeerakul
Shade Shutters
Deborah Strumsky
Kevin Stolarick
Oscar Patterson-Lomba
Ricardo Hausmann
Three takeaways

1. There is a statistical regularity present in urban systems called "Urban Scaling".

2. Every social phenomenon has a "complexity" that summarizes many of its statistical properties.

3. Ideas from Cultural Evolution are needed in order to account for the differences in development across cities.
Let’s internalize the types of questions we’re asking

<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Larceny-theft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>3,141</td>
</tr>
</tbody>
</table>
### Why 3,141?

<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Larceny-theft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>3,141</td>
</tr>
</tbody>
</table>
Why 768?

<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Larceny-theft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>768</td>
</tr>
</tbody>
</table>

Source: https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/
## Why 20?

<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Robbery</th>
<th>Larceny-theft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>20</td>
<td>768</td>
</tr>
</tbody>
</table>

### Metropolitan Statistical Area (MSA) Year Population Robbery Larceny-theft

<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Robbery</th>
<th>Larceny-theft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>20</td>
<td>768</td>
</tr>
<tr>
<td>Bridgeport-Stamford-Norwalk, CT M.S.A.</td>
<td>2010</td>
<td>895,941</td>
<td>991</td>
<td>10,986</td>
</tr>
</tbody>
</table>

Source: https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/
Numbers change... but maybe *proportions* do not....

<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Robbery</th>
<th>Larceny-theft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>20</td>
<td>768</td>
</tr>
<tr>
<td>Bridgeport-Stamford-Norwalk, CT M.S.A.</td>
<td>2010</td>
<td>895,941</td>
<td>991</td>
<td>10,986</td>
</tr>
</tbody>
</table>

Source: https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/
Numbers change... but maybe *proportions* do not....

<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Robbery Rate (cases per 100,000 inhabitants)</th>
<th>Larceny-theft (cases per 100,000 inhabitants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>36.3</td>
<td>1,393.3</td>
</tr>
<tr>
<td>Bridgeport-Stamford-Norwalk, CT M.S.A.</td>
<td>2010</td>
<td>895,941</td>
<td>110.6</td>
<td>1,226.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Robbery Rate (cases per 100,000 inhabitants)</th>
<th>Larceny-theft (cases per 100,000 inhabitants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>36.3</td>
<td>1,393.3</td>
</tr>
<tr>
<td>Bridgeport-Stamford-Norwalk, CT M.S.A.</td>
<td>2010</td>
<td>895,941</td>
<td>110.6</td>
<td>1,226.2</td>
</tr>
</tbody>
</table>

Source: https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/
<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Robbery Rate (cases per 100,000 inhabitants)</th>
<th>Larceny-theft (cases per 100,000 inhabitants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>36.3</td>
<td>1,393.3</td>
</tr>
<tr>
<td>Bridgeport-Stamford-Norwalk, CT M.S.A.</td>
<td>2010</td>
<td>895,941</td>
<td>110.6</td>
<td>1,226.2</td>
</tr>
</tbody>
</table>

Source: https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/
<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Robbery Rate (cases per 100,000 inhabitants)</th>
<th>Larceny-theft (cases per 100,000 inhabitants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>36.3</td>
<td>1,393.3</td>
</tr>
<tr>
<td>Bridgeport-Stamford-Norwalk, CT M.S.A.</td>
<td>2010</td>
<td>895,941</td>
<td>110.6</td>
<td>1,226.2</td>
</tr>
</tbody>
</table>

Source: https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/
<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Robbery Rate (cases per 100,000 inhabitants)</th>
<th>Larceny-theft (cases per 100,000 inhabitants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>36.3</td>
<td>1,393.3</td>
</tr>
<tr>
<td>Bridgeport-Stamford-Norwalk, CT M.S.A.</td>
<td>2010</td>
<td>895,941</td>
<td>110.6</td>
<td>1,226.2</td>
</tr>
</tbody>
</table>

Source: https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/
One approach
One approach

Theory: $Y = f_1(X)$
One approach

Theory: \( Y = f_1(X) \)  \quad \rightarrow \quad \text{Empirics:} \quad y_i = \beta_0 + \beta_1 x_{i,1} + \epsilon_i
One approach

Theory: $Y = f_1(X)$  \quad \rightarrow \quad \text{Empirics: } y_i = \beta_0 + \beta_1 x_{i,1} + \epsilon_i

\downarrow

\text{Theory: } Y = f_2(X, W)$
One approach

Theory: $Y = f_1(X)$ \quad \Rightarrow \quad Empirics: $y_i = \beta_0 + \beta_1 x_{i,1} + \epsilon_i$

Empirics: $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i$ \quad \Rightarrow \quad Theory: $Y = f_2(X, W)$
One approach

Theory: $Y = f_1(X)$

Empirics: $y_i = \beta_0 + \beta_1 x_{i,1} + \epsilon_i$

Empirics: $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i$

Theory: $Y = f_2(X, W)$

$Y = f_3(X, W, Z)$
One approach

Theory: \( Y = f_1(X) \) \( \rightarrow \) Empirics: \( y_i = \beta_0 + \beta_1 x_{i,1} + \epsilon_i \)

Empirics: \( y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i \) \( \rightarrow \) Theory: \( Y = f_2(X, W) \)

\[ Y = f_3(X, W, Z) \] \( \rightarrow \) \( \ldots \)
One approach

Theory: $Y = f_1(X)$  \quad \Rightarrow \quad \text{Empirics: } y_i = \beta_0 + \beta_1 x_{i,1} + \epsilon_i$

Empirics: $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i$  \quad \Leftarrow \quad \text{Theory: } Y = f_2(X, W)$

$Y = f_3(X, W, Z) \quad \Rightarrow \quad \ldots \quad \Rightarrow \quad y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \epsilon_i$
One approach

Theory: \( Y = f_1(X) \) \quad \Rightarrow \quad \text{Empirics: } \quad y_i = \beta_0 + \beta_1 x_{i,1} + \epsilon_i

Empirics: \quad y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i \quad \Leftarrow \quad \text{Theory: } \quad Y = f_2(X, W)

\[ Y = f_3(X, W, Z) \quad \Rightarrow \quad \ldots \quad \Rightarrow \quad y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \epsilon_i \]
• Collective systems are full of interdependencies, interactions, feedback loops, etc...
A methodological decision

- The conventional approach:

Multivariate regression analysis
( → Identifying effects/r.h.s.)
A methodological decision

• The conventional approach:

  Multivariate regression analysis
  (→ Identifying effects/r.h.s.)

• An alternative approach:

  Scaling analysis
  (→ Identifying mechanisms/l.h.s.)
“Scaling Analysis”

- Studying a phenomenon as it changes with scale (e.g., size):
  - Regularities
  - Important underlying mechanisms
“Scaling Analysis”

- **Studying a phenomenon as it changes with scale (e.g., size):**
  - Regularities
  - Important underlying mechanisms
“Scaling Analysis”

• Studying a phenomenon as it changes with scale (e.g., size):
  – Regularities
  – Important underlying mechanisms
“Scaling Analysis”

- Studying a phenomenon as it changes with scale (e.g., size):
  - Regularities
  - Important underlying mechanisms
“Scaling Analysis”

• **Studying a phenomenon as it changes with scale (e.g., size):**
  – Regularities
  – Important underlying mechanisms

\[ T \sim r^{3/2} \]
“Scaling Analysis”

• **Studying a phenomenon as it changes with scale (e.g., size):**
  – Regularities
  – Important underlying mechanisms

1. **Slope** of the line: Gravitational interaction decays with the *square* of the distance.

   $$T \sim r^{3/2}$$

2. **Intercept** of the line: Gravitational force is proportional to *mass*. 
If interested...

- “In the beat of a heart”, John Whitfield (BIOLOGY)

- “Critical Phenomena in Natural Sciences”, Didier Sornette (PHYSICS & COMPLEX SYSTEMS)

- “Fractals, Chaos, Power Laws”, Manfred Schroeder (GENERAL, ENGINEERING & PHYSICS)

### Table 1: Scaling relationships and corresponding theories.

<table>
<thead>
<tr>
<th>Quantities</th>
<th>Scaling Law</th>
<th>Name</th>
<th>Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orbital period $T$ and distance to the Sun $r$</td>
<td>$T = T_0 \ r^{3/2}$</td>
<td>Kepler’s third law</td>
<td>Theory of planetary motion</td>
</tr>
<tr>
<td>Average radius of diffusion $r$ and time $t$</td>
<td>$r = r_0 \ t^{1/2}$</td>
<td>Law of diffusion</td>
<td>Theory of Brownian motion</td>
</tr>
<tr>
<td>Metabolic rate $B$ and body mass $M$</td>
<td>$B = B_0 \ M^{3/4}$</td>
<td>Kleiber’s law</td>
<td>Metabolic Theory of Ecology</td>
</tr>
<tr>
<td>Socioeconomic rates $Y$ and population size $N$</td>
<td>$Y = Y_0 \ N^\beta$</td>
<td>Urban Scaling Laws</td>
<td>???.</td>
</tr>
</tbody>
</table>
Table 1: Scaling relationships and corresponding theories.

<table>
<thead>
<tr>
<th>Quantities</th>
<th>Scaling Law</th>
<th>Name</th>
<th>Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orbital period $T$ and distance to the Sun $r$</td>
<td>$T = T_0 \ r^{3/2}$</td>
<td>Kepler’s third law</td>
<td>Theory of planetary motion</td>
</tr>
<tr>
<td>Average radius of diffusion $r$ and time $t$</td>
<td>$r = r_0 \ t^{1/2}$</td>
<td>Law of diffusion</td>
<td>Theory of Brownian motion</td>
</tr>
<tr>
<td>Metabolic rate $B$ and body mass $M$</td>
<td>$B = B_0 \ M^{3/4}$</td>
<td>Kleiber’s law</td>
<td>Metabolic Theory of Ecology</td>
</tr>
<tr>
<td>Socioeconomic rates $Y$ and population size $N$</td>
<td>$Y = Y_0 \ N^\beta$</td>
<td>Urban Scaling Laws</td>
<td>???</td>
</tr>
</tbody>
</table>
Allometric laws in urban systems
Allometric laws in urban systems

The Statistics of Urban Scaling and Their Connection to Zipf’s Law

Andres Gomez-Lievano¹*, HyeJin Youn², Luis M. A. Bettencourt²

¹ School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, United States of America, ² Santa Fe Institute, Santa Fe, New Mexico, United States of America

![Graphs showing urban scaling in Colombia, Mexico, and Brazil](image)

Figure 1. Annual number of homicides in cities of Colombia, Mexico and Brazil versus population size (2007). Large cities are defined in terms of metropolitan areas which are aggregations of municipalities (red circles) while non-metropolitan municipalities are shown separately (green squares). The solid blue line fits only the scaling of homicides for metropolitan areas. Large variations, especially among the smaller population units, and the fact that many municipalities have $Y = 0$ (not shown) prevent a direct scaling analysis. However, it is possible to analyze the data consistently through the estimation of conditional probabilities.

doi:10.1371/journal.pone.0040393.g001
Urban scaling in Europe

Luís M. A. Bettencourt, José Lobo

Published 16 March 2016. DOI: 10.1098/rsif.2016.0005

Colombia, Mexico and Brazil versus population size (2007). Large cities are defined as municipalities (red circles) while non-metropolitan municipalities are shown separately. Large variations, especially among the smaller population sizes, prevent a direct scaling analysis. However, it is possible to analyze the data.
Urban scaling in Europe

Luís M. A. Bettencourt, José Lobo

Published 16 March 2016, DOI: 10.1098/rsif.2016.0005

The Pre-History of Urban Scaling

Scott G. Ortman, Andrew H. F. Cabaniss, Jennie O. Sturm, Luís M. A. Bettencourt

Published: February 12, 2014 • http://dx.doi.org/10.1371/journal.pone.0087902

The relationships between population size and various urban characteristics are shown in the graphs. The data suggest that the relationship between population size and urban area follows a power law, with a scaling exponent of approximately 0.736. This exponent is consistent across different regions and time periods, indicating a fundamental scaling law for urban growth.

The graphs also show that the relationship between population size and settled area is linear, with a slope close to 1. This suggests that the growth of urbanized areas is proportional to the increase in population size, supporting the notion of urban scaling.

Furthermore, the relationship between population size and residential mound area is also linear, with a slope close to 1, indicating that the growth of residential areas is also proportional to the increase in population size.

These findings have implications for urban planning and policy, as they suggest that urban growth is not random but follows predictable patterns. Understanding these scaling laws can help in better planning and resource allocation in urban environments.
Population/Area Relationship for Medieval European Cities

Rudolf Cesaretti¹, José Lobo², Luís M. A. Bettencourt³, Scott G. Ortman³,⁴, Michael E. Smith¹

¹ School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, 85281, United States of America, ² School of Sustainability, Arizona State University, Tempe, Arizona, 85281, United States of America, ³ Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, New Mexico, 87501, United States of America, ⁴ Department of Anthropology, University of Colorado Boulder, Boulder, Colorado, 80309, United States of America

Fig 1. Estimation of Area—Population Scaling Relation for All Settlements. The Area—Population scaling relation for the entire data set of all medieval cities (n = 172). The black line represents proportionate (linear) scaling; the yellow line the theoretical prediction where α = 5/2, and the red line the best-fit line from OLS regression of the log-transformed data.
The Non-linear Health Consequences of Living in Larger Cities

Luis E C Rocha, Anna E. Thorson, and Renaud Lambiotte

FIG. 1  a Relative change in the outcome $Q_2 (Q_2 - Q_1)/Q_1$ (y-axis), for different values of the exponent $\alpha$ (x-axis). For $\alpha = 1$, we recover the usual linear relation between population size and the outcome variable. The relation between the population size (x-axis) and b the number of deaths by heart attack or c number of abortions (y-axis). Both axes are in log-scale.

Log Area (ha) vs Log Population

Fig 4. Estimation of Area—Population Scaling Relation for All Settlements. The Area—Population scaling relation for the entire data set of all medieval cities ($n = 172$). The black line represents proportionate (linear) scaling; the yellow line the theoretical prediction where $e = 58$; and the red line the best-fit line from DLS regression of the log-transformed data.
The Non-linear Health Consequences of Living in Larger Cities

Luis E C Rocha, Anna E. Thorson, and Renaud Lambiotte

Per capita incidence of sexually transmitted infections increases systematically with urban population size: a cross-sectional study

Oscar Patterson-Lomba1, Edward Goldstein2, Andrés Gómez-Liévano3, Carlos Castillo-Chavez4, Sherry Towers4

Figure 2: Scaling of STD incidence with MSA population with Negative Binomial regression for chlamydia (left), gonorrhoea (centre) and syphilis (right) using model (4), as reported in Table 1. Comparing the blue lines (with slopes ) with the dotted lines (with slope 1) shows the departures from the linear pattern in each case.
Urban Scaling: $Y = Y_0 N^{(1.1667)}$

<table>
<thead>
<tr>
<th>Scaling Relation</th>
<th>Exponent</th>
<th>Error</th>
<th>Observations</th>
<th>Region/Nation</th>
<th>Urban Unit</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socioeconomic rates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP</td>
<td>$\beta = 1.13$</td>
<td>$[1.11,1.15]$</td>
<td>363</td>
<td>USA</td>
<td>MSA</td>
<td>2006</td>
</tr>
<tr>
<td>GDP</td>
<td>$\beta = 1.22$</td>
<td>$[1.11,1.33]$</td>
<td>273</td>
<td>China</td>
<td>Prefectural Cities</td>
<td>2005</td>
</tr>
<tr>
<td>GDP</td>
<td>$\beta = 1.10$</td>
<td>$[1.01,1.18]$</td>
<td>35</td>
<td>Germany</td>
<td>LUZ</td>
<td>2004</td>
</tr>
<tr>
<td>income</td>
<td>$\beta = 1.12$</td>
<td>$[1.07,1.17]$</td>
<td>12</td>
<td>Japan</td>
<td>MA</td>
<td>2005</td>
</tr>
<tr>
<td>wages</td>
<td>$\beta = 1.12$</td>
<td>$[1.07,1.17]$</td>
<td>363</td>
<td>USA</td>
<td>MSA</td>
<td>1969-2009</td>
</tr>
<tr>
<td>violent crime</td>
<td>$\beta = 1.16$</td>
<td>$[1.11,1.19]$</td>
<td>287</td>
<td>USA</td>
<td>MSA</td>
<td>2003</td>
</tr>
<tr>
<td>violent crime</td>
<td>$\beta = 1.20$</td>
<td>$[1.07,1.33]$</td>
<td>12</td>
<td>Japan</td>
<td>MA</td>
<td>2008</td>
</tr>
<tr>
<td>violent crime</td>
<td>$\beta = 1.20$</td>
<td>$[1.15,1.25]$</td>
<td>27; 5,570</td>
<td>Brazil</td>
<td>MA; Municipios</td>
<td>2003-07</td>
</tr>
<tr>
<td>new AIDS cases</td>
<td>$\beta = 1.23$</td>
<td>$[1.17,1.29]$</td>
<td>93</td>
<td>USA</td>
<td>MSA</td>
<td>2002-3</td>
</tr>
<tr>
<td>new patents</td>
<td>$\beta = 1.27$</td>
<td>$[1.22,1.32]$</td>
<td>331</td>
<td>USA</td>
<td>MSA</td>
<td>1980-2001</td>
</tr>
<tr>
<td>supercreative jobs</td>
<td>$\beta = 1.15$</td>
<td>$[1.13,1.17]$</td>
<td>331</td>
<td>USA</td>
<td>MSA</td>
<td>1999-2001</td>
</tr>
<tr>
<td>R&amp;D employment</td>
<td>$\beta = 1.19$</td>
<td>$[1.12,1.26]$</td>
<td>227-278</td>
<td>USA</td>
<td>MSA</td>
<td>1987-2002</td>
</tr>
<tr>
<td><strong>Average socioeconomic rates</strong></td>
<td>$\beta = 1.17$</td>
<td>$[1.01,1.33]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Urban Scaling: \( Y = Y_0 N^{(1.1667)} \)

<table>
<thead>
<tr>
<th>Scaling Relation</th>
<th>Exponent</th>
<th>Error</th>
<th>Observations</th>
<th>Region/Nation</th>
<th>Urban Unit</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socioeconomic rates</td>
<td>( \beta = 1.13 )</td>
<td>[1.11,1.15]</td>
<td>363</td>
<td>USA</td>
<td>MSA</td>
<td>2006</td>
</tr>
<tr>
<td>GDP</td>
<td>( \beta = 1.22 )</td>
<td>[1.11,1.33]</td>
<td>273</td>
<td>China</td>
<td>Prefectural Cities</td>
<td>2005</td>
</tr>
<tr>
<td>GDP</td>
<td>( \beta = 1.10 )</td>
<td>[1.01,1.18]</td>
<td>35</td>
<td>Germany</td>
<td>LUZ</td>
<td>2004</td>
</tr>
<tr>
<td>income</td>
<td>( \beta = 1.12 )</td>
<td>[1.07,1.17]</td>
<td>12</td>
<td>Japan</td>
<td>MA</td>
<td>2005</td>
</tr>
<tr>
<td>wages</td>
<td>( \beta = 1.12 )</td>
<td>[1.07,1.17]</td>
<td>363</td>
<td>USA</td>
<td>MSA</td>
<td>1969-2009</td>
</tr>
<tr>
<td>violent crime</td>
<td>( \beta = 1.16 )</td>
<td>[1.11,1.19]</td>
<td>287</td>
<td>USA</td>
<td>MSA</td>
<td>2003</td>
</tr>
<tr>
<td>violent crime</td>
<td>( \beta = 1.20 )</td>
<td>[1.07,1.33]</td>
<td>12</td>
<td>Japan</td>
<td>MA</td>
<td>2008</td>
</tr>
<tr>
<td>violent crime</td>
<td>( \beta = 1.20 )</td>
<td>[1.15,1.25]</td>
<td>27; 5,570</td>
<td>Brazil</td>
<td>MA; Municipios</td>
<td>2003-07</td>
</tr>
<tr>
<td>new AIDS cases</td>
<td>( \beta = 1.23 )</td>
<td>[1.17,1.29]</td>
<td>93</td>
<td>USA</td>
<td>MSA</td>
<td>2002-3</td>
</tr>
<tr>
<td>new patents</td>
<td>( \beta = 1.27 )</td>
<td>[1.22,1.32]</td>
<td>331</td>
<td>USA</td>
<td>MSA</td>
<td>1980-2001</td>
</tr>
<tr>
<td>supercreative jobs</td>
<td>( \beta = 1.15 )</td>
<td>[1.13,1.17]</td>
<td>331</td>
<td>USA</td>
<td>MSA</td>
<td>1999-2001</td>
</tr>
<tr>
<td>R&amp;D employment</td>
<td>( \beta = 1.16 )</td>
<td>[1.12,1.26]</td>
<td>227-278</td>
<td>USA</td>
<td>MSA</td>
<td>1987-2002</td>
</tr>
</tbody>
</table>

Average socioeconomic rates: \( \beta = 1.17 \) [1.01,1.33]

The question of Individual vs. Systemic effects

- Analytic:
  - Are individuals in larger cities more educated, innovative, productive, entrepreneurial (… but also more violent, unhealthy), etc...?
The question of Individual vs. Systemic effects

- Analytic:
  - Are individuals in larger cities are more educated, innovative, productive, entrepreneurial (... but also more violent, unhealthy), etc...?

- Productive individuals self-select into larger urban areas.
The question of Individual vs. Systemic effects

• Analytic:
  – Are individuals in larger cities are more educated, innovative, productive, entrepreneurial (... but also more violent, unhealthy), etc...?

• Productive individuals self-select into larger urban areas.

• However, additional assumptions would be needed to explain the scaling law.
The question of Individual vs. Systemic effects

• **Analytic:**
  – Are individuals in larger cities are more educated, innovative, productive, entrepreneurial (... but also more violent, unhealthy), etc...?

• Productive individuals self-select into larger urban areas.

• However, additional assumptions would be needed to explain the scaling law.

• It wouldn’t explain the disproportionate concentration of crime and disease in larger urban areas.
The question of Individual vs. Systemic effects

• Holistic:
  – Is output (aggregate GDP, patents, crime, disease, etc...) a function of the number of social connections?
The question of Individual vs. Systemic effects

- Holistic:
  - Is output (*aggregate* GDP, patents, crime, disease, etc...) a function of the number of social connections?

\[ E = N \times \left( \frac{p \times (N - 1)}{2} \right) \]

*Interactions \( \propto \) Individuals^2*

\[ Y \propto E \propto N^2 \]
The question of Individual vs. Systemic effects

• Holistic:
  – Is output (*aggregate* GDP, patents, crime, disease, etc...) a function of the number of social connections?

\[ E = N \times \left( \frac{p \times (N - 1)}{2} \right) \]

Interactions \( \propto \) Individuals\(^2\)

\[ Y \propto E \propto N^2 \]

---

Size and the Density of Interaction in Human Aggregates\(^1\)

Bruce H. Mayhew
University of South Carolina

Roger L. Levinger
Temple University

Size and the Density of Interaction in Human Aggregates
Author(s): Bruce H. Mayhew and Roger L. Levinger
The question of Individual vs. Systemic effects

• Holistic:
  – Is output (*aggregate* GDP, patents, crime, disease, etc...) a function of the number of social connections?

\[ E = N \times \left( \frac{p \times (N - 1)}{2} \right) \]

\[ Interactions \propto Individuals^2 \]

\[ Y \propto E \propto N^2 \]

• More individuals engage in productive *as well* as harmful, disadvantageous activities.

---

Size and the Density of Interaction in Human Aggregates

Bruce H. Mayhew
*University of South Carolina*

Roger L. Levinger
*Temple University*

Size and the Density of Interaction in Human Aggregates
Author(s): Bruce H. Mayhew and Roger L. Levinger
Urban Scaling: \( Y = Y_0 N^{(1.1667)} \)

Why do different urban activities scale differently?!

Three takeaways

1. There is a statistical regularity present in urban systems called “Urban Scaling”. \[ Y \sim N^\beta \]

2. Every social phenomenon has a “complexity” that summarizes many of its statistical properties.

3. Ideas from Cultural Evolution are needed in order to account for the differences in development across cities.
Three takeaways

1. There is a statistical regularity present in urban systems called “Urban Scaling”. \[ Y \sim N^\beta \]

2. Every social phenomenon has a “complexity” that summarizes many of its statistical properties.

3. Ideas from Cultural Evolution are needed in order to account for the differences in development across cities.
Economics

\[ y = f(\text{LAND, LABOR, CAPITAL, HUMAN CAPITAL}) \]
International trade

SITC-4 Rev 2: 772 Products, 129 Countries (Year 2000)

... it is the type of pattern you’d expect if countries were playing a game...


The product space predicts which products will be produced next

- Entry into new product is easier if you already produce related products
- Products appear in cascades (Klimek/Hausmann/Thurner 2012)

Source:
Huge divergence in income  

Investment in education
Thailand vs. Ghana in the Product Space
1965

Total Value: $615,728,000

Total Value: $294,604,000

Thailand

Ghana
1970

Total Value: $721,421,000

Thailand

Total Value: $432,140,000

Ghana
1985

Thailand

Ghana

Total Value: $7,354,613,137

Total Value: $462,524,204
1995

Total Value: $50,644,730,628

Thailand

Total Value: $1,294,057,269

Ghana
2000

Total Value: $67,126,271,442

Thailand

Total Value: $1,206,161,694

Ghana
2005

Total Value: $111,099,204,052

Total Value: $1,871,625,725

Thailand

Ghana
The Theory of Economic Complexity

• Economic processes arise from a multiplicity of factors.

• More complex processes require more factors.
  – # of factors required = complexity = $q$

• Richer countries have more factors.
  – Endowment of factors = diversity = $r$
= \( r \)

= small \( q \)

= large \( q \)
\[ \mathbb{E} \left[ \frac{Y}{N} \right] = e^{-M(1-r)q} = e^{-Mq} e^{Mrq} \]
\[ \mathbb{E} \left[ \frac{Y}{N} \right] = e^{-Mq} e^{Mrq} \]

<table>
<thead>
<tr>
<th>Metropolitan Statistical Area (MSA)</th>
<th>Year</th>
<th>Population</th>
<th>Robbery Rate (cases per 100,000 inhabitants)</th>
<th>Larceny-theft (cases per 100,000 inhabitants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson City, NV M.S.A.</td>
<td>2010</td>
<td>55,119</td>
<td>36.3</td>
<td>1,393.3</td>
</tr>
<tr>
<td>Michigan City-La Porte, IN M.S.A.</td>
<td>2010</td>
<td>111,553</td>
<td>81.6</td>
<td>2,656.1</td>
</tr>
<tr>
<td>Chico, CA M.S.A.</td>
<td>2010</td>
<td>222,130</td>
<td>70.7</td>
<td>1,582.4</td>
</tr>
<tr>
<td>Lansing-East Lansing, MI M.S.A.</td>
<td>2010</td>
<td>450,078</td>
<td>79.1</td>
<td>1,763.5</td>
</tr>
<tr>
<td>Bridgeport-Stamford-Norwalk, CT M.S.A.</td>
<td>2010</td>
<td>895,941</td>
<td>110.6</td>
<td>1,226.2</td>
</tr>
<tr>
<td>Las Vegas-Paradise, NV M.S.A.</td>
<td>2010</td>
<td>1,951,609</td>
<td>240.6</td>
<td>1,580.0</td>
</tr>
<tr>
<td>Phoenix-Mesa-Glendale, AZ M.S.A.</td>
<td>2010</td>
<td>4,229,275</td>
<td>124.0</td>
<td>2,400.4</td>
</tr>
<tr>
<td>Los Angeles-Long Beach-Santa Ana, CA M.S.A.</td>
<td>2010</td>
<td>12,912,749</td>
<td>189.5</td>
<td>1,428.3</td>
</tr>
<tr>
<td>New York-Northern New Jersey-Long Island, NY-NJ-PA M.S.A.</td>
<td>2010</td>
<td>19,042,526</td>
<td>164.2</td>
<td>1,300.3</td>
</tr>
</tbody>
</table>

Source: https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/
• Robbery rates do not seem to increase exponentially with population size.
Three takeaways

1. There is a statistical regularity present in urban systems called “Urban Scaling”. \( Y \sim N^\beta \)

2. Every social phenomenon has a “complexity” that summarizes many of its statistical properties.

3. Ideas from Cultural Evolution are needed in order to account for the differences in development across cities.
Three takeaways

1. There is a statistical regularity present in urban systems called “Urban Scaling”. \[ Y \sim N^{\beta} \]

2. Every social phenomenon has a “complexity” that summarizes many of its statistical properties.

3. Ideas from Cultural Evolution are needed in order to account for the differences in development across cities.
• Robbery rates do not seem to increase exponentially with population size.
Cumulative culture

- Robbery rates do not seem to increase exponentially with population size.

→ The diversity of factors (i.e., cultural traits) does not change in proportion to population size.

Cumulative culture

• Robbery rates do not seem to increase exponentially with population size.

The diversity of factors (i.e., cultural traits) does not change in proportion to population size.

See also:
Models of economic complexity

Models of cultural evolution

&
Models of economic complexity

![Images of economic complexity models]

\[ E[Y] = e^{-Mq} Ne^{Mq} r(N) \]

Models of cultural evolution

![Diagram of cultural evolution]

\[ r(N) = a + b \ln(N) \]
Models of economic complexity

\[ \mathbb{E} [Y] = e^{-Mq} Ne^{Mq} r(N) \]

Urban Scaling: \[ \mathbb{E} [Y] = e^{-Mq(1-a)} N^{1+Mqb} = Y_0 N^\beta \]

Models of cultural evolution

\[ r(N) = a + b \ln(N) \]
Urban Scaling:  

\[
E[Y] = e^{-Mq} Ne^{Mq} r(N)
\]

Models of economic complexity

Models of cultural evolution

\[
r(N) = a + b \ln(N)
\]
What the Theory Predicts

With the prevalence of a phenomenon in a single city, the theory predicts what the prevalence in the rest of cities is likely to be.

**Procedure:**

Given coefficients \( s_1 \) and \( s_2 \) and the populations of all cities \( n_1, n_2, \ldots \).

For a phenomenon of interest:

Pick a random city \( c \) with known population size and prevalence:

\[ (n_c, Y_c) \]

Apply the equations:

\[
\beta_{\text{pred.}} = \frac{1 - s_1 \ln(Y_c)}{1 - s_2 \ln(n_c)}
\]

\[
\ln(Y_c)_{\text{pred.}} = 1 - \frac{\beta_{\text{pred.}}}{s_1}
\]

\[
\sigma_{\text{pred.}} = \frac{\beta_{\text{pred.}} - 1}{s_2}
\]

Use the populations \( n_1, n_2, \ldots \), to predict the prevalence of the phenomenon in the rest of cities within some **prediction bands**:

\[
Y_i = \exp \left\{ \ln(Y_{i\text{pred.}}) + \beta_{\text{pred.}} \ln(n_i) \pm z_0 \sigma_{\text{pred.}} \right\}, \quad \text{for all } i = 1, 2, \ldots
\]

To test the predictions, we simulated the **Procedure** 50 times for each phenomenon, for a total of 2150 simulations.

**Results of Simulations**

For all phenomena:

1. Apply the **Procedure**
2. Compute the fraction \( f \) of cities within the predicted bands

\[
f = \ldots \ldots - - - - - - - - - - - \ldots
\]

![Image of results showing the fraction of cities within predicted 90% bands.](image)
THANK YOU

Contact info:
Email: andres_gomez@hks.harvard.edu
Twitter: @GomezLievano
Learning at Different Levels

• Individuals learn by going to school, by doing, by interacting.
Learning at Different Levels

• Individuals learn by going to school, by doing, by interacting.

• Cities “learn” by attracting new individuals with more diverse skills (Florida, 1995, Futures).
Learning at Different Levels

- Individuals *are limited* by the amount of knowledge that they can learn;
Learning at Different Levels

• Individuals *are limited* by the amount of knowledge that they can learn;

• **Societies, however, are not.**
### Example

<table>
<thead>
<tr>
<th>Getting a patent requires:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Having a technological problem</td>
</tr>
<tr>
<td>2) Having a solution</td>
</tr>
<tr>
<td>3) Presenting the idea clearly</td>
</tr>
<tr>
<td>4) Applying for a patent</td>
</tr>
<tr>
<td>5) Including corrections from examiners</td>
</tr>
<tr>
<td>6) Satisfying all legal requirements</td>
</tr>
</tbody>
</table>
### Example

<table>
<thead>
<tr>
<th>Getting a patent requires:</th>
<th>City c</th>
<th>Person 1</th>
<th>Person 2</th>
<th>Person 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Having a technological problem</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2) Having a solution</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Presenting the idea clearly</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>4) Applying for a patent</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>5) Including corrections from examiners</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>6) Satisfying all legal requirements</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gets the patent?</th>
<th>Person 1</th>
<th>Person 2</th>
<th>Person 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>
### Getting a patent requires:

<table>
<thead>
<tr>
<th></th>
<th>City c</th>
<th>Person 1</th>
<th>Person 2</th>
<th>Person 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Having a technological problem</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2) Having a solution</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>3) Presenting the idea clearly</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Applying for a patent</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>5) Including corrections from examiners</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) Satisfying all legal requirements</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Gets the patent?**

- Person 1: **YES**
- Person 2: **NO**
- Person 3: **YES**
Example

<table>
<thead>
<tr>
<th>Getting a patent requires:</th>
<th>City c</th>
<th>Person 1</th>
<th>Person 2</th>
<th>Person 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Having a technological problem</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2) Having a solution</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3) Presenting the idea clearly</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4) Applying for a patent</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>5) Including corrections from examiners</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>6) Satisfying all legal requirements</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gets the patent?</th>
<th>Person 1</th>
<th>Person 2</th>
<th>Person 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gets the patent?</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>

City c has **2 inventors** (from a population of 3).
The Model

• The activity in question requires $N_a$ substeps.
• The substeps provided by the city as a $N_a \times 1$ vector, $\vec{C}$, 1 for the substeps provided, 0 otherwise.
• We represent the substeps that person $j$ needs using a $N_a \times 1$ vector, $\vec{p}_j$, with 1 for the missing substeps, 0 otherwise.

<table>
<thead>
<tr>
<th>City $c$</th>
<th>Person 1</th>
<th>Person 2</th>
<th>Person 3</th>
<th>$\vec{C}$</th>
<th>$\vec{p}_1$</th>
<th>$\vec{p}_2$</th>
<th>$\vec{p}_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
The Model

- The activity in question requires $N_a$ substeps.
- The substeps provided by the city as a $N_a \times 1$ vector, $\vec{C}$, 1 for the substeps provided, 0 otherwise.
- We represent the substeps that person $j$ needs using a $N_a \times 1$ vector, $\vec{p}_j$, with 1 for the missing substeps, 0 otherwise.

<table>
<thead>
<tr>
<th>City $c$</th>
<th>Person 1</th>
<th>Person 2</th>
<th>Person 3</th>
<th>$\vec{C}$</th>
<th>$\vec{p}_1$</th>
<th>$\vec{p}_2$</th>
<th>$\vec{p}_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>1</td>
<td>0</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>1</td>
<td>0</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>1</td>
<td>0</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
The Model

Table 1: Parameters of the model. The parameters $M$, $q$ and $r$ are in principle phenomenon-dependent.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N &gt; 0$</td>
<td>City population size susceptible of participating of a given phenomenon.</td>
</tr>
<tr>
<td>$M &gt; 0$</td>
<td>Number of possible factors required for the given phenomenon.</td>
</tr>
<tr>
<td>$q \in (0, 1)$</td>
<td>Probability that an individual needs any given factor from the environment.</td>
</tr>
<tr>
<td>$r \in (0, 1)$</td>
<td>Probability that the city facilitates any one of the factors to the individual.</td>
</tr>
</tbody>
</table>

• The main parameters, $r$ and $q$:
  — $r$: the average “diversity” of the city.
  — $q$: the average “complexity” of the activity.

• Prevalence of the activity in the city: $Y = \sum_{j=1}^{N} X_j$

• $E[Y] = f(N, r, q) =$ ?