Scaling laws of cities

Markus Schläpfer
Future Cities Laboratory, ETH-Singapore Centre
School of Computer Science and Engineering, NTU

NTU Winter School on Complexity Science
March 14, 2017
Ever-increasing complexity of cities

A. Population growth and mass urbanization

UN Population Division, 2015
Ever-increasing complexity of cities

B. New forms of urban organization

Monocentricity

Univ. Munster

Polycentricity

MIT/Senseable City Lab, Kael Greco
Increasing uncertainties in urban planning and design

- Urban mobility
- Infrastructure design
- Social sustainability (social segregation, job accessibility)
- …

Urgent need for a quantitative understanding of cities
Growing availability of human activity data

- Mobile phone data
- Smart card data from public transportation
- GPS traces from vehicular devices
- Location-based social networks (Foursquare, Twitter, Flickr, Running Apps, etc.)
- User-generated mapping projects (OpenStreetMap)
- Open data provided by city governments
- ....
Content

1. Urban scaling laws
2. Urban structure: building heights and shapes
3. Urban dynamics: the movement of people in cities
4. Application: infrastructure design
1. Urban Scaling Laws
The scaling of socio-economic quantities with city size

\[ Y \propto N^\beta \]

Exponent

City population size

Socio-economic quantity (wages, patents, crime, AIDS cases etc.)

\[ \beta \approx 1.15 > 1 \]

L.M.A. Bettencourt et al.
Greater population — „faster life and greater dividends“

<table>
<thead>
<tr>
<th>Y</th>
<th>β</th>
<th>95% CI</th>
<th>Adj-R²</th>
<th>Observations</th>
<th>Country–year</th>
</tr>
</thead>
<tbody>
<tr>
<td>New patents</td>
<td>1.27</td>
<td>[1.25,1.29]</td>
<td>0.72</td>
<td>331</td>
<td>U.S. 2001</td>
</tr>
<tr>
<td>Inventors</td>
<td>1.25</td>
<td>[1.22,1.27]</td>
<td>0.76</td>
<td>331</td>
<td>U.S. 2001</td>
</tr>
<tr>
<td>Private R&amp;D employment</td>
<td>1.34</td>
<td>[1.29,1.39]</td>
<td>0.92</td>
<td>266</td>
<td>U.S. 2002</td>
</tr>
<tr>
<td>“Supercreative” employment</td>
<td>1.15</td>
<td>[1.11,1.18]</td>
<td>0.89</td>
<td>287</td>
<td>U.S. 2003</td>
</tr>
<tr>
<td>R&amp;D establishments</td>
<td>1.19</td>
<td>[1.14,1.22]</td>
<td>0.77</td>
<td>287</td>
<td>U.S. 1997</td>
</tr>
<tr>
<td>R&amp;D employment</td>
<td>1.26</td>
<td>[1.18,1.43]</td>
<td>0.93</td>
<td>295</td>
<td>China 2002</td>
</tr>
<tr>
<td>Total wages</td>
<td>1.12</td>
<td>[1.09,1.13]</td>
<td>0.96</td>
<td>361</td>
<td>U.S. 2002</td>
</tr>
<tr>
<td>Total bank deposits</td>
<td>1.08</td>
<td>[1.03,1.11]</td>
<td>0.91</td>
<td>267</td>
<td>U.S. 1996</td>
</tr>
<tr>
<td>GDP</td>
<td>1.15</td>
<td>[1.06,1.23]</td>
<td>0.96</td>
<td>295</td>
<td>China 2002</td>
</tr>
<tr>
<td>GDP</td>
<td>1.26</td>
<td>[1.09,1.46]</td>
<td>0.64</td>
<td>196</td>
<td>EU 1999–2003</td>
</tr>
<tr>
<td>GDP</td>
<td>1.13</td>
<td>[1.03,1.23]</td>
<td>0.94</td>
<td>37</td>
<td>Germany 2003</td>
</tr>
<tr>
<td>Total electrical consumption</td>
<td>1.07</td>
<td>[1.03,1.11]</td>
<td>0.88</td>
<td>392</td>
<td>Germany 2002</td>
</tr>
<tr>
<td>New AIDS cases</td>
<td>1.23</td>
<td>[1.18,1.29]</td>
<td>0.76</td>
<td>93</td>
<td>U.S. 2002–2003</td>
</tr>
<tr>
<td>Serious crimes</td>
<td>1.16</td>
<td>[1.11,1.18]</td>
<td>0.89</td>
<td>287</td>
<td>U.S. 2003</td>
</tr>
</tbody>
</table>

≈15% per capita increase in wages, GDP, patents etc. for each doubling of city size

L.M.A. Bettencourt et al.
Network of human interactions as a unifying mechanism?
Correspondence and requests for materials should be addressed to A.P. (email: pentland@mit.edu).

Motivated by empirical evidence on the interplay between geography, population density and urban characteristics attributable to

ARTICLE

Received 29 September 2008; published 30 January 2009

Superlinear scaling for innovation in cities
Samuel Arbesman
Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, Massachusetts 02115, USA
Jon M. Kleinberg
Computer Science, Cornell University, Ithaca, New York 14853, USA
Steven H. Strogatz
Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853, USA

Superlinear scaling in cities, which appears in sociological quantities such as economic productivity and creative output relative to urban population size, has been observed, but not been given a satisfactory theoretical explanation. Here we provide a network model for the superlinear relationship between population size and innovation found in cities, with a reasonable range for the exponent.

Urban characteristics attributable to density-driven tie formation
Wei Pan1, Gourab Ghoshal1,2, Coco Krumme1, Manuel Cebrian1,2,3 & Alex Pentland4

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
Growing availability of human activity data

- Mobile phone data
Mobile phone data - exemplary data sources

• Open data
  • Italy - Telecom Italia Open BigData Initiative
    http://theodi.fbk.eu/openbigdata

• Big data research competitions
  • Ivory Coast - Orange D4D Challenge 2013
    http://www.d4d.orange.com/en/Accueil
  
  • Senegal - Orange D4D Challenge 2015
    http://www.d4d.orange.com/en/Accueil

  • Italy - Telecom Italia BigData Challenge 2015

• (Telco providers and data analytics companies)
Several millions of anonymized call detail records (CDRs) from Portugal for a period of ≈15 months
Call detail records (CDRs)

- Anonymized ID (surrogate number) of the caller
- Anonymized ID of the callee
- Start time of the call
- Duration of the call
- The locations of the antennas routing the call
Inferring the interaction network

Mobile phone user → Node

Reciprocal call → Link
between two users

Portugal data:
1.6 Mio nodes
6.8 Mio links
Human interactions
<table>
<thead>
<tr>
<th>city definition</th>
<th>number</th>
<th>network type</th>
<th>ΔT (days)</th>
<th>Y</th>
<th>β</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portugal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>statistical city</td>
<td>140</td>
<td>reciprocal</td>
<td>409</td>
<td>degree ($K_d$)</td>
<td>1.12</td>
<td>[1.11, 1.14]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>call volume ($V_d$)</td>
<td>1.11</td>
<td>[1.09, 1.12]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>number of calls ($W_d$)</td>
<td>1.10</td>
<td>[1.09, 1.11]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>92</td>
<td>degree ($K_d$)</td>
<td>1.10</td>
<td>[1.09, 1.11]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>call volume ($V_d$)</td>
<td>1.10</td>
<td>[1.08, 1.11]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>number of calls ($W_d$)</td>
<td>1.08</td>
<td>[1.07, 1.10]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>non-reciprocal</td>
<td>409</td>
<td>degree ($K_d$)</td>
<td>1.24</td>
<td>[1.22, 1.25]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>call volume ($V_d$)</td>
<td>1.14</td>
<td>[1.12, 1.15]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>number of calls ($W_d$)</td>
<td>1.13</td>
<td>[1.12, 1.14]</td>
</tr>
<tr>
<td>larger urban zone</td>
<td>9(8)</td>
<td>reciprocal</td>
<td>409</td>
<td>degree ($K_d$)</td>
<td>1.05</td>
<td>[1.00, 1.11]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>call volume ($V_d$)</td>
<td>1.11</td>
<td>[1.02, 1.20]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>number of calls ($W_d$)</td>
<td>1.10</td>
<td>[1.05, 1.15]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>non-reciprocal</td>
<td>409</td>
<td>degree ($K_d$)</td>
<td>1.13</td>
<td>[1.08, 1.18]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>call volume ($V_d$)</td>
<td>1.14</td>
<td>[1.05, 1.23]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>number of calls ($W_d$)</td>
<td>1.13</td>
<td>[1.08, 1.18]</td>
</tr>
<tr>
<td>municipality</td>
<td>293</td>
<td>reciprocal</td>
<td>409</td>
<td>degree ($K_d$)</td>
<td>1.13</td>
<td>[1.11, 1.14]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>call volume ($V_d$)</td>
<td>1.15</td>
<td>[1.13, 1.17]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>number of calls ($W_d$)</td>
<td>1.13</td>
<td>[1.11, 1.14]</td>
</tr>
<tr>
<td>UK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>urban audit city</td>
<td>24</td>
<td>reciprocal</td>
<td>31</td>
<td>degree ($K$)</td>
<td>1.08</td>
<td>[1.05, 1.12]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>degree, land-mobile ($K_{lm}$)</td>
<td>1.14</td>
<td>[1.11, 1.17]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>call volume ($V$)</td>
<td>1.10</td>
<td>[1.07, 1.17]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>number of calls ($W$)</td>
<td>1.08</td>
<td>[1.05, 1.11]</td>
</tr>
</tbody>
</table>
Nodal clustering

Clustering coefficient:
Probability that one’s contacts are also connected with each other.

\[ C_i \equiv \frac{2z_i}{k_i(k_i - 1)} \]

- \( z_i \): Number of links between the \( k_i \) neighbours
- \( k_i \): Degree of node \( i \)

As larger cities provide a larger pool of people, the clustering coefficient should decrease if contacts were established at random.
- Average clustering is an invariant of city size.
- Even in large cities we live in groups that are as tightly knit as those in small towns or ‘villages’.

Acceleration of spreading processes

Susceptible-infected (SI) model

\[ P_{ij} \propto v_{ij} \]

- \( P_{ij} \): Transmission probability
- \( v_{ij} \): Call volume between user \( i \) and user \( j \)

Graphs showing the number of infected nodes over time for Lisbon and Meda.
Potential "hidden" biases

Test on different data sets

• UK, mobile phones and landlines (Schläpfer et al. 2014)
• Ivory Coast, mobile phones (Andris and Bettencourt, 2014)
• „Unnamed“ European Country, mobile phones (Llorente, 2015)
• US and Europe, Twitter data (Tizzoni, 2015)
• Switzerland, mobile phones (Büchel and von Ehrlich, 2016)
2. Urban Structure: Building Heights and Shapes
Building functional cities

J. Vernon Henderson, Anthony J. Venables, Tanner Regan, Ilia Samsonov

The literature views many African cities as dysfunctional with a hodgepodge of land uses and poor “connectivity.” One driver of inefficient land uses is construction decisions for highly durable buildings made under weak institutions. In a novel approach, we model the dynamics of urban land use with both formal and slum dwellings and ongoing urban redevelopment to higher building heights in the formal sector as a city grows. We analyze the evolution of Nairobi using a unique high–spatial resolution data set. The analysis suggests insufficient building volume through most of the city and large slum areas with low housing volumes near the center, where corrupted institutions deter conversion to formal sector usage.

Fig. 1. City of Nairobi building height and distribution. Nairobi shows average built height in 2015 as 150-m by 150-m cells split across the formal and slum sectors. The compass (top left) points north. The location of the Kibera slum and the CBD are marked. The boundary of the city spans about 22 km east to west and 11 km north to south; the map tilt may distort the appearance of distances. Modified from HRV. [Background imagery Airbus Defense and Space 2016, taken from the SPOT5 satellite 20 September 2004].
Generating simple 3D city models

DSM

DSM - DEM

Building polygons

DSM: Digital surface model
DEM: Digital elevation model

Building heights

New York NY, N = 19.6M
Los Angeles CA, N = 12.8M
Toronto ON, N = 5.6M
Boston MA, N = 4.6M
Portland OR, N = 2.2M
Santa Fe NM, N = 0.1M

Building heights

A

Log average height (m)

Log population

Moncton NB
Cape Coral FL
North Port FL
Chicago IL
Los Angeles CA

Building heights

Log average height (m) vs. Log population

- Moncton NB
- Cape Coral FL
- North Port FL
- Los Angeles CA
- Chicago IL

Height prediction from urban scaling theory

For cities to be **functional:**

\[ h \propto N^{1/6} \]

Building height

Population size

Scaling exponent

Building heights distribution

![Graph showing the distribution of building heights in different cities](image)

- **New York**
- **Los Angeles**
- **Boston**
- **Portland**
- **Santa Fe**

The x-axis represents the log10 of building height in meters, while the y-axis represents the density.
Building shapes

Building volume

![Graph A](graph_a.png)

- Detroit MI
- Kitchener ON

Building shape

![Graph B](graph_b.png)

- Boston MA
- Moncton NB
- Cape Coral FL
- North Port FL

Allometric scaling

![Graph C](graph_c.png)

- Cape Coral FL
- North Port FL

3. Urban Dynamics: 
Movement of People in Cities
'Collective' movements in cities

Individual trajectories from mobile phone data

<table>
<thead>
<tr>
<th>User ID, Timestamp, Cell tower ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2013-01-24 11:30:00, 599</td>
</tr>
<tr>
<td>1, 2013-01-24 12:30:00, 608</td>
</tr>
<tr>
<td>1, 2013-01-24 13:00:00, 608</td>
</tr>
<tr>
<td>1, 2013-01-24 19:00:00, 446</td>
</tr>
<tr>
<td>1, 2013-01-24 20:10:00, 323</td>
</tr>
<tr>
<td>1, 2013-01-24 20:30:00, 323</td>
</tr>
<tr>
<td>1, 2013-01-24 22:00:00, 323</td>
</tr>
<tr>
<td>1, 2013-01-24 22:20:00, 323</td>
</tr>
</tbody>
</table>
1. *How many* people visit a given location?
2. From *how far* do they come?
3. *How often* do they visit?
Greater Boston area
≈ 2 Mio. mobile phone users over 4 months
≈ 10^9 location based records per month (triangulation)
46,210 locations (500m x 500m grid cells)

Lets look into the data!
Quantifying the attractiveness of locations

1 visit per month

2 visits per month

3 visits per month
$r$ visiting distance (km)  
$f$ visiting frequency (visits per month)  

Brightness of pixel: number of visitors, $q(r,f)$
Increasing visiting distance
Increasing visiting frequency
Increasing visiting frequency

Increasing visiting distance
Dimensional analysis

\[ q = q[r, f, u, Y(N)] \]

Travel speed

Socioeconomic features

\[ \Rightarrow q(r, f) = G\left(\frac{rf}{u}\right) = F(rf) \]
Increasing visiting frequency

Increasing visiting distance
Example (v = 20km/month):

number of visitors coming from 5 km and 4 times a month

= number of visitors coming from 10 km and 2 times a month

= number of visitors coming from 20 km and once a month
What is the **functional relation** between:

- number of visitors,
- their travel distance from home,
- their visiting frequency?

**Fluid dynamic model**

**Dimensional argument**

\[ q(r, f) = G\left(\frac{rf}{u}\right) = F(rf) \]

**Theoretical expectation**

**Number of visitors**

\[ \propto [\text{travel distance x visiting frequency}]^{-2} \]
Number of visitors

\[ \propto \left[ \text{travel distance} \times \text{visiting frequency} \right]^{-2} \]
1. **How many** people visit a given location?
2. From **how far** do they come?
3. **How often** do they visit?
1. How many people visit a given location?

From how far do they come.

How often do they visit.
Greater Boston
Schläpfer, Szell, Ratti, West (in preparation)
Locations with 'anomalous' behavior
4. Application: Infrastructure design
Electrification planning in developing countries
A correlation between the mobile phone data characteristics of the corresponding area density (i.e., average distance between households) is expected to highlight the electricity needs of the mobile phone users who are currently available for electrification planning practices.

This study is expected to provide an accurate estimation of potential future migration trends towards electrified areas within the country, particularly when combined with other data sources used in electrification planning practices. The aggregated electricity profiles can be assessed with reference to the differentiated mobile phone data based on the level of electrification in Senegal, their location within the area, and the size and location of villages and their communities.

The differentiated mobile phone data is prepared for the World Bank and is categorized into different levels of electrification (0%, 0% - 25%, 25% - 50%, 50% - 75%, 75% - 100%, and 100%). This differentiation of the mobile phone data is critical to quantify the mobile phone activity and migration and to classify based on the state-of-the-art extent on the differentiation of the mobile phone data based on the size and location of villages.

The differentiation of the mobile phone data is considered for average villages of different population sizes in Senegal, their location corresponding to each Voronoi area comprises all points that are closer to that tower than to any other tower. Therefore, the differentiation of the mobile phone data based on the size and location of villages is an example of the assessment of electrification options and development of dwelling patterns within the area.

Analysis of the abovementioned information can facilitate better electrification planning practices for villages of different sizes. The differentiated mobile phone data is obtained from the SoDa solar energy services database, and the aggregated electricity demand profiles, solar radiation, and temperature are represented in the map.

Table I: Example of the maximum amount of institutional settings for the World Bank in Senegal

<table>
<thead>
<tr>
<th>Setting</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schools</td>
<td>1</td>
</tr>
<tr>
<td>Markets</td>
<td>1</td>
</tr>
<tr>
<td>Hospitals</td>
<td>3</td>
</tr>
<tr>
<td>Public Lighting points</td>
<td>3</td>
</tr>
</tbody>
</table>

This map of Senegal shows the level of electrification rates in various settlements, including cities, towns, and villages. The map uses different colors to represent the level of electrification, with 0% being the least electrified and 100% being fully electrified. The map also indicates the presence of different institutional settings such as schools, markets, hospitals, and public lighting points.
Using information from mobile phone infrastructure to facilitate electrification
Mobile phone data as a proxy for electricity demand

(a) Number of calls, $A_i$ [-]
(b) Number of calls, $A_i$ [-]
(c) Call duration, $A_i$ [h]
(d) Call duration, $A_i$ [h]
(e) Number of users, $A_i$ [-]
(f) Number of users, $A_i$ [-]
But not only…
Electrification technology optioneering: techno-economic analysis

- Medium voltage electricity grid extension
- Diesel-based microgrid installation
- Traditional and minimalistic solar photovoltaic system

The model and equations are in the paper ;}
Electrification recommendations

- City
- Town
- Village

Recommended option:
- Yellow: MV grid extension
- Orange: Microgrid
- Light green: PV or Microgrid

Martinez-Cesena, Mancarella, Ndiaye, Schläpfer
D4D Challenge 2015, First Prize (best overall) and Energy Prize
Take home: urban 'big' data

„Don‘t“s

give new insights per se

„Do“s

are free of 'hidden' biases
Take home: urban 'big' data

„Don‘ts“
- give new insights per se
- are free of ‚hidden‘ biases

„Do‘s“
- allow testing „old“ ideas
- cover large parts of the population
- objective measurements
Thank you!

Markus Schläpfer — schlaepfer@arch.ethz.ch