Urban analytics and the role of mobile phones

markus schläpfer
massachusetts institute of technology

29 october 2013 | SMART-FM Seminar, Singapore
Ever-growing complexity of cities

A. Population growth and mass urbanization
Ever-growing complexity of cities

B. New forms of urban organization

Garreau, Anchor (1992)
Anas, Arnott & Small, J Econ Lit (1998)
Kloosterman & Musterd, Urb Stud (2001)

Levy & Gilchrist (2013)
Ever-growing complexity of cities

B. New forms of urban organization

Monocentricity

Polycentricity

Garreau, Anchor (1992)
Anas, Arnott & Small, J Econ Lit (1998)
Kloosterman & Musterd, Urb Stud (2001)
Emerging challenges

- Environmental sustainability
- Urban planning
- Transportation planning
- Infrastructure investments
- Job accessibility
- Social segregation
- ...

Urgent need for a data-driven, quantitative understanding of cities (‘urban analytics’)
The role of mobile phone data

Three examples:

1. Defining cities with people movements
2. Revealing the polycentric organization of cities
3. The impact of cities on social connectivity and spreading phenomena
1. Defining cities with people movements

Revisiting the very meaning of cities

Lewis Mumford:
Cities are „magnets that attract non-residents for intercourse, spiritual stimulus and trade.“

Step 1:
Identification of city centers as „magnets’’ that attract people

Step 2:
Delineating the full city area
City center detection algorithm (CCDA)

Assumption 1
Centers are locations that attract more people than one would expect on the basis of random chance.

Assumption 2
Centers form a hierarchy of locations with increasing regional importance: bakery, cinema, jewelry,… (aka Christaller’s Central Place Theory)
City center detection algorithm (CCDA)

classifying locations - visiting distance and frequency

Real movements (data) vs Null model

Comparing data and null model

Unsupervised clustering:
- Local center: low distances, high frequencies
- Regional center: medium distances, medium frequencies
- Metropolitan center: large distances, low frequencies
Position data from mobile phones

- Greater Boston, Singapore, Portugal
- 7 Mio. individuals
- Over 30+ days
- Partition regions into grid
- Estimate home locations
Hierarchical centers in Greater Boston

local, $r=2$ km

regional, $r=8$ km

metropolitan, $r=22$ km
Delineating the full city area

‘burning algorithm’

A

B

C

Illustration of the city delineation procedure.

Once
Figure 6. Expanding the city size (Boston) by continuously adding less connected locations to the detected centers. The colors (from blue to green) correspond to the connectivity threshold $m$ (from low to high), with $f_{\text{min}} = 1$ day/month.

3.2 Lisbon

Lisbon serves as an example for a traditional European city that started out as a medieval town (Anas, 1998). The city is characterized by a greater mixture of residences and businesses in the core and a stronger public transportation system.

3.3 Abidjan

Abidjan is the largest city of Côte d'Ivoire and is an example for a rather new and rapidly growing city of a middle-income developing country.

4 Discussion and Outlook

• The preliminary results presented here demonstrate the feasibility of using CDRs for an accurate and standardized delineation of cities.

• Further investigations are needed to find the most adequate and robust null model for the identification of city cores.

• The methodology should be particularly interesting for locations where census-based data is sparse and lagging behind reality, such as for rapidly growing cities in developing countries.

• The proposed approach can be extended to identify different subcenters within the same city (polycentric urban structures).

0.01 0.05 0.1 0.15 0.25 0.3 0.35 0.4 0.45 0.5

Cambridge*

Downtown

Salem&

Delineating the full city area

Quincy'

Delineating the full city area
2. Revealing the polycentric organization of cities

Let's test with data!
+ Develop testable framework

Burger, Meijers, Urban studies 49 (2012)
Conceptual framework

\[ P_i(r, f) = \frac{q_i(r, f)}{n_i(r)} \]

Probability that someone living at distance \( r \) from place \( i \) visits \( f \) times a month

“Attractivity” of \( i \)
Mono or Polycentricity depending on \((r, f)\)

\[
P_i(r, f) = \frac{q_i(r, f)}{n_i(r)}
\]

1. Fix \((r, f)\)
2. Plot top 100 \(P_i\)

Figure 2: Exposing the multilayered structure of urban areas. The 100 most attractive grid cells are highlighted, demonstrating the transition from polycentric to monocentric layers. A continuous transition between the layers is provided in supplementary video S1 (17).

(A, B) For a low radius highly dispersed centers are observed in Singapore \((r = 1\text{ km})\) and Greater Boston \((r = 6\text{ km})\).

(C, D) For medium distances \((r = 5\text{ km} \text{ and } r = 21 \text{ km})\) centers of regional importance emerge, for example the regional sub-centers Portsmouth, Manchester, Worchester, Providence, arranged in a hexagonal-like pattern around the city of Boston.

(E, F) A large visitation radius \((r = 10 \text{ km} \text{ and } r = 42 \text{ km})\) highlights the predominant importance of the central Downtown Core of Singapore, and the city of Boston surrounded by the “inner ring.”

(G, H) Increasing the visiting frequency reduces the size of the Singapore downtown center and exposes the industrial region of West Jurong, similar effects of dispersion appear for Greater Boston. The lighter blue tone marks all locations with available data and at least one visit. Values of \(I\) quantify the spatial clustering.
Quantify agglomerations by spatial correlation

Highly systematic & regular!

Radius $r$

and nontrivial
Importance of urban centers

Singapore Downtown

Jurong East
3. Social connectivity and spreading processes
Socio-economic indicators scale with city size

Growth, innovation, scaling, and the pace of life in cities

Luis M. A. Bettencourt*, José Lobo‡, Dirk Helbing§, Christian Kühnert†, and Geoffrey B. West*†

*Theoretical Division, MS B284, Los Alamos National Laboratory, Los Alamos, NM 87545; ‡Global Institute of Sustainability, Arizona State University, P.O. Box 873211, Tempe, AZ 85287-3211; §Institute for Transport and Economics, Dresden University of Technology, Andreas-Schubert-Strasse 23, D-01062 Dresden, Germany; and †Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

Edited by Elinor Ostrom, Indiana University, Bloomington, IN, and approved March 6, 2007 (received for review November 19, 2006)

Humanity has just crossed a major landmark in its history with the majority of people now living in cities. Cities have long been known to be society’s predominant engine of innovation and wealth creation, yet they are also its main source of crime, pollution, and other grave social problems. In the context of sustainability, cities pose three key questions: Why are they so important? How can they operate efficiently? And what is the fate of humanity as a whole in the presence of cities?

The increasing concentration of people in cities presents both opportunities and challenges (9) toward future scenarios of sustainable development. On the one hand, cities make possible

Scaling relationship

\[ Y = P^\beta, \ \beta \approx 1.15 \]

- \( Y \) Socio-economic indicator
- \( P \) Population size
- \( \beta \) Scaling exponent
Greater population – faster life and greater dividends

<table>
<thead>
<tr>
<th>Y</th>
<th>$\beta$</th>
<th>95% CI</th>
<th>Adj-$R^2$</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>&quot;Supercreative&quot; 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>

$\approx15\%$ per capita increase in wages, GDP, patents etc. for each doubling of city size
Human interactions as a unifying mechanism?

Potential scaling of personal contacts

- Upper bound:
  \[ Y = P^2 \]

- More realistic range:
  \[ \beta = 1 + \frac{\log A}{\log(P/P_0)} \]
  
  \[ A = 10-100, \quad P/P_0 = 10^7 \Rightarrow \beta = 1.14 - 1.28 \]
How to provide empirical support?

- Limited data for entire urban systems.
- Survey-based approaches:
  - small sample sizes
  - few places surveyed
  - biased towards strong social links

Inferring human interaction networks from mobile phone call records
Inferring the network of human interactions

- Anonymized call detail records (CDRs) for a period of ≈15 months
- ≈20% of the population
- Mobile phone user → node
- (Reciprocal) call between two users → link

Network of ≈2 Mio nodes and ≈10 Mio links, spanning the entire country
Assigning nodes to cities

- We assign to each user the cell tower which routed most of his calls.
- Its location is assumed to represent the caller’s home or work place.
- The location of this ‘characteristic’ cell thus determines the city to which the caller belongs to.
Scaling in excellent agreement with socio-economic indicators
Scaling relations: call volume and number of calls

The time spent on the phone and the number of calls between each pair of contacts is, on average, invariant with city size.
Non-reciprocated calls

Degree

$\beta_{Kr} = 1.24$

Volume

$\beta_{Vr} = 1.14$

Number of calls

$\beta_{Wr} = 1.13$

Rate of new ‘superficial’ contacts increases faster!
Acceleration of spreading processes

Spreading speed (SI model) vs city size

FIG. 4. Larger cities facilitate interaction-based spreading processes. (a) Spreading speed, $R$, averaged over 100 simulation trials of the SI model for each Statistical City in Portugal (circles), with nodal infection rate $\beta = 0.01$ and $I_0 = 100$ infected nodes. The solid line is the best fit to a power-law scaling relation $R \propto N^{\gamma}$, with $\gamma = 0.12 \pm 0.04$ (95% CI, Adj-$R^2 = 0.22$). (b) Corresponding simulation results for the Municipalities in Portugal. The line describes the best fit with $\gamma = 0.14 \pm 0.03$ (95% CI, Adj-$R^2 = 0.25$). Inset: association between $R$, as predicted by the power-law relation, and the number of HIV/AIDS cases per capita, $y$, for 14 Municipalities during the period of 2002 to 2010. The solid line shows the linear regression of the log-transformed data with slope $3.56 \pm 2.32$ (95% CI, Adj-$R^2 = 0.44$).
thank you for your attention

further information:

http://web.mit.edu/schlmark/www