Big Data and Collaborative Research via a Visual Lab

The strength of weak ties

Jiangping Zhou
Assistant Professor
Iowa State University
Associate Director
Beijing City Lab
Outline

• An introduction to BCL (Possibly a ICL?)
• Big data: What, Why and How
• Some examples:
  --Big data and collaborative research
  --Big data, me and Urban China
• Summary
• Q&A
The Beijing City Lab (BCL) is a virtual research community, dedicated to studying, but not limited to, China’s capital Beijing. The Lab focuses on employing interdisciplinary methods to quantify urban dynamics, generating new insights for urban planning and governance, and ultimately producing the science of cities required for sustainable urban development. The lab’s current mix of planners, architects, geographers, economists, and policy analysts lends unique research strength.
Beijing City Lab, BCL

• **Organization structure**
  – Lead researchers ($\times 7$)
  – Honorary Directors ($\times 11$)
  – Core researchers ($\times 24$)
  – Student members ($\times 38$)
  – Followers（6000+）

• **Missions**
  – A network for quantitative urban studies
  – A platform for sharing (40 working papers+24 datasets)
  – An attempt to scientifically understand cities
  – Visuals involving public participation
  – Concentration on Beijing but care for China and beyond
BCL visitors

BCL URL: WWW.BEIJINGCITYLAB.ORG OR LONGY.JIMDO.COM

In the sequence of 2014 BCL conference presentations
Sample projects by BCLers
(Traditional, big, open and big open data)

1 BUDEM
2 Urban Growth Boundaries
3 Bus Landscapes
4 Population China
5 Planning Support Systems
5 Urban Form
6 Population Synthesis
7 Social Network Mining
8 Big Model
9 Beijing Parking
10 Urban Network Analysis
Sample Data by BCLers

Physical-demographic:

– Chinese cities’ administrative boundaries, road network, existing parcels, urbanized areas, planning permission (not all cities), land use maps, DEM, water, urban land by RS and natural features

– Population

– Street-level density, parcel-level population and associated attributes

Quality of life evaluation

Urban environmental info (including PM2.5) POI, public facilities, housing prices, bus routes and stations and restaurants
BCL’s Open Data and Big Data

• **Human activities and movements**
  – Hotspots, check-ins, location-exposed Weibos, traffic flows between cities, smartcard data for transit (Beijing), household travel surveys (Beijing) and taxi travel data
  – Forecasts
  – Master plans (200+ cities)
  – Scenarios for urban expansion
Connections among Cities
(based on trains between/across cities)
Road intersection density
Attempt to scientifically understand cities

- Informationized planning? ✗
- New planning technologies ✗
- A Science of Cities? ✓
Big data: What, Why and How
What is big data?

- Speadsheet that MS Excel cannot handle?
- Data we generated since we have the Internet? (The data we generate daily are the same as those our ancestors did for hundred of years)
- User-generated data (Some call them people sensing data)?
What is big data?

• Regardless small or big, the end goal for us to collect and analyze big data is to generate knowledge and wisdom.

• Big data are not the only way to generate knowledge and wisdom (e.g., we have numerous great scientists and philosophers before big data emerged).
Why we need big data?

• Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...
Why we need big data?

- We want to more information, knowledge, wisdom and efficacy from data.
How we can best use big data?

• Theoretical developments about data
• Mechanisms of knowledge discovery
• Big data standard, transferability, scalability, measurement, analysis and methodological questions
• Institutional issues, e.g., organizations, networks and infomediaries
Big data and urban research
Tackle the Big challenges in Big cities using Big data!

Visualization of Taxi Pick-ups (Orange) and Drop-offs (Blue) in New York City
(NYU Center for Urban Science and Progress)
Humanity's migration and cultural history

https://www.youtube.com/watch?v=4gIhRkCcD4U#t=95
Emergence of big data
Big data and Urban China
Spatial patterns of human settlements/movements
- Both Employment and Residential Subcenter

Employment Subcenter (Employment #)
- 197 - 500
- 501 - 1000
- 1001 - 1500
- 1501 - 2000

Residential Subcenter (Residence #)
- 197 - 500
- 501 - 1000
- 1001 - 1500
- 1501 - 2000
- Tian'an Men

Arterials
Hot spots where there are problems
Optimize commuting and traffic
Identify losers and winners in optimization
Losers’ Distribution When Bus Commuting Optimized for the Whole City

Average Loss for Home-based Trips (in km)

-5.5
-5.4 - 2.4
-2.3 - 1.1
-1.0 - 0.3
-0.2 - 0.02

Average Loss for Journeys to Work (in km)

-5.3 - 3.6
-3.5 - 2.4
-2.3 - 1.4
-1.3 - 0.7
-0.6 - 0.01

★ Tian’anmen (Epicenter of Beijing)
Construct policy scenarios to understand impacts of different policies
Policy Scenario 1: Doing nothing

Policy Scenario 2: Beijing adopts comprehensive travel demand management measures and sees 0-20% decrease in traffic and travel cost between TAZs.

Policy Scenario 3: In light of large volumes of bus riders to several employment centers (TAZs 97, 216, 284, 651 and 694) where there are more than 2,000 bus commuters per day, Beijing now operates bus rapid transit (BRT) from these centers and consolidates services of certain existing bus routes.
Bus Journeys Taken by Commuters in Beijing with BRT
Commuting economy under different scenarios

- Base
- TDM
- BRT
<table>
<thead>
<tr>
<th>Trip Distance (Km)</th>
<th>Number of Resident Workers</th>
<th>Before the Optimization</th>
<th>After the Optimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;4.05</td>
<td>1</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>4.05-8.1</td>
<td>93</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8.1-12.2</td>
<td>8</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>12.2-16.3</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>&gt;16.3</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Even with big data, we cannot forget traditional and open data
<table>
<thead>
<tr>
<th>Prototype</th>
<th>Characteristics</th>
<th>Landmarks and Directions</th>
<th>Spatial Index in Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>University campuses, hotels and old Danwei compounds left with mostly apartment buildings</td>
<td>Minzu and Jiaotong Universities and areas in between; Beijing Technology and Business University and Capital Normal University (east campus) and areas adjacent to them; Beijing University of Aeronautics and Astronautics and Beijing University of Science and Technology and adjacent areas</td>
<td>A (Areas around and areas to its northeast)</td>
</tr>
<tr>
<td>2</td>
<td>Parks with luxury hotels, high-end apartments, specialized research institutes, hospitals and some mixed-use residential areas</td>
<td>Areas north to Yu Yuan Tan Park; Area adjacent to Tian Tan Park in the east</td>
<td>A (South to A, the darkest area); The U-shaped area south to the star</td>
</tr>
<tr>
<td>3</td>
<td>Traditional Hu’tong with old, cheap, small, shared and underserviced rental housing units</td>
<td>Areas in between Qian Men Da Jie and Zhu Shi Kou Da Jie</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>Residential areas with mixed-age housing units adjacent to freeway interchanges or arterials, railways within the fifth ring road</td>
<td>Areas near Yong Ding Men and Nan Sha Wo Bridges</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>Low density, developing areas with relatively cheap housing units in the suburb</td>
<td>Areas adjacent to the sixth ring road and Jingshi Freeway interchange; Areas adjacent to Yan Chun Railway Station</td>
<td>D</td>
</tr>
</tbody>
</table>
Verify and even extend theories
Reliable data from traditional sources

Smart card data and derived info based on them

Sampling, weighting, \( t \) test, etc.

Reliable combined data and info, if applicable

If applicable

Existing theories of urban formation

If quality is good

Yes

Studies of urban formation

New insights and findings
Figure 1. Flow distribution. Loglog plot of the histogram of the number of trips between two stations of the tube system. The line is a power law fit with exponent $\approx 1.3$.

doi:10.1371/journal.pone.0015923.g001
\[ P = 0.139 \times N^{-0.473}, \text{ R-Square}=0.926 \]
Summary

• Big data have the potential to be a much more dynamic source of data for planning and policy studies than traditional data.

• When enhanced by traditional data, big data can be used to generate new knowledge and insights.

• Geo-visualization can help publicize the above knowledge and insights.
Summary

• Urban China provides unlimited opportunities for those interested in big data and associated studies

• Visual labs/communities like BCL would enable us to take full advantage of those opportunities (“The strength of weak ties”!)