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**

BCL

Beijing City Lab

- Home
- Projects
- Members
- Working papers
- Slides
- Data released
- Ranking
- Blogs
- Links&Partners
- About

Maintained by Dr Ying Long
BeijingCityLab(at)gmail(dot)com

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 buildings visualized by GeoCanvas

www.beijingcitylab.org OR longy.jimdo.com

Beijing City Lab, BCL

- **Organization structure**
  - Lead researchers (×7)
  - Honorary Directors(×11)
  - Core researchers(×24)
  - Student members(×38)
  - Followers（6000+）
- **Missions**
  - A network for quantitative urban studies
  - A platform for sharing (40 working papers+24datasets)
  - 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








Sample projects by BCLers

(Traditional, big, open and big open data)

Projects

1 BUDEM

2 Urban Growth Boundaries

3 Bus Landscapes

4 Population China

5 Planning Support Systems

5 Urban Form

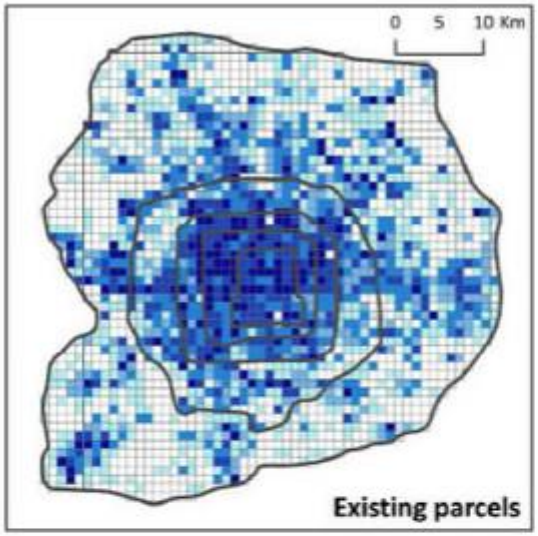
6 Population Synthesis

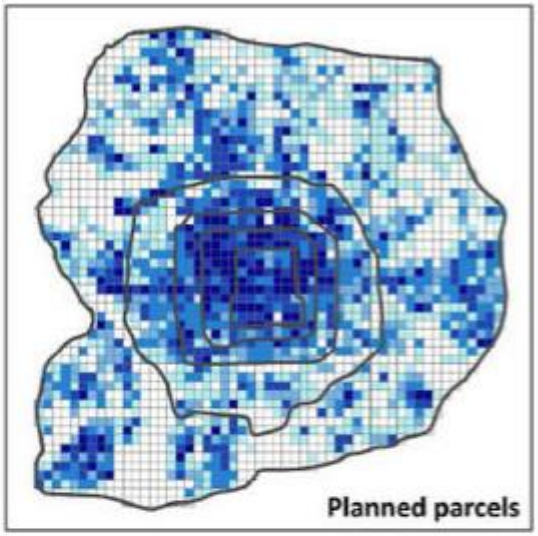
7 Social Network Mining

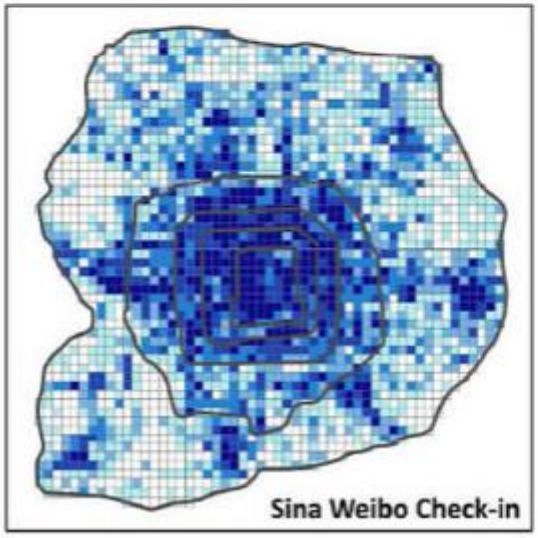
8 Big Model

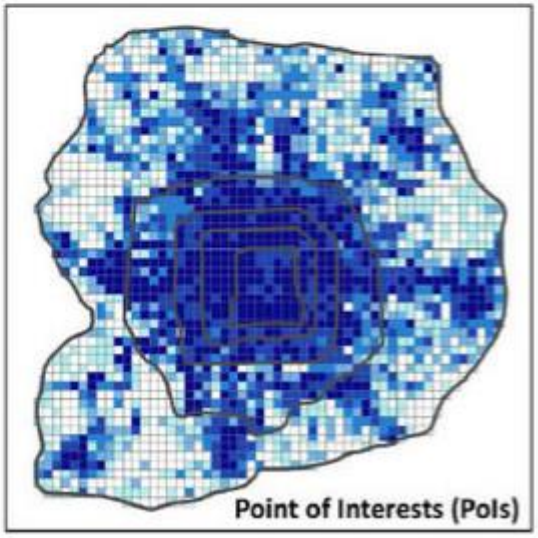
9 Beijing Parking

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

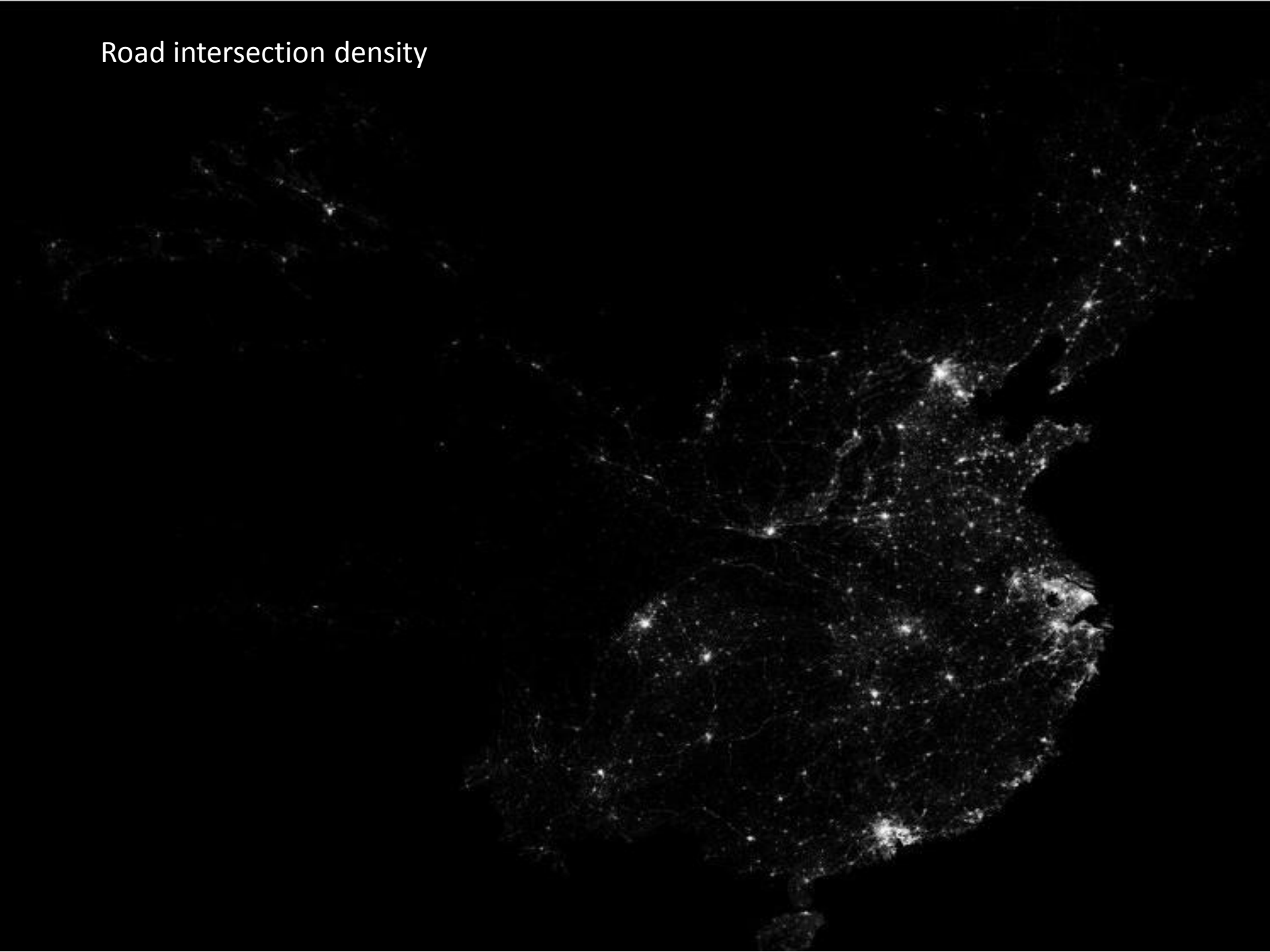
Urban environmental info (including PM2.5) POI, public facilities, housing prices, bus routes and stations and restaurants

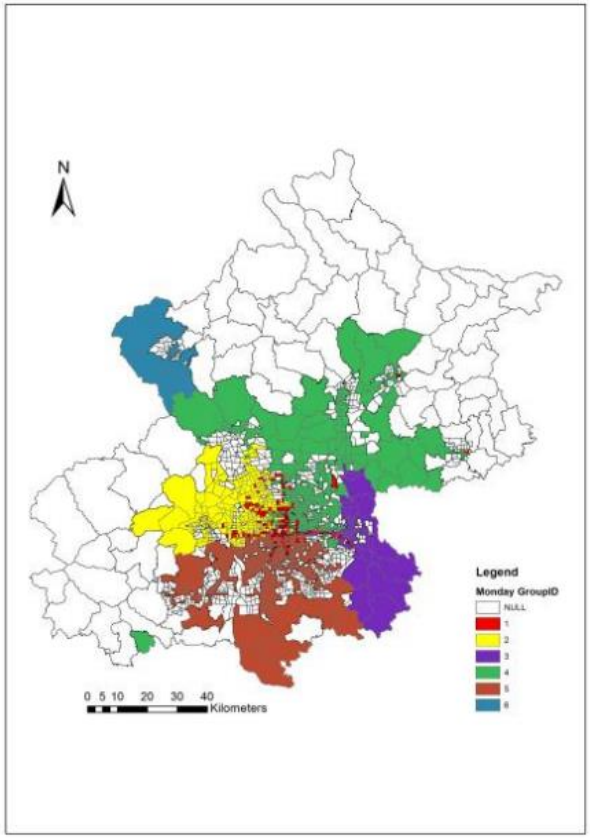
BCL's Open Data and Big Data

- **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

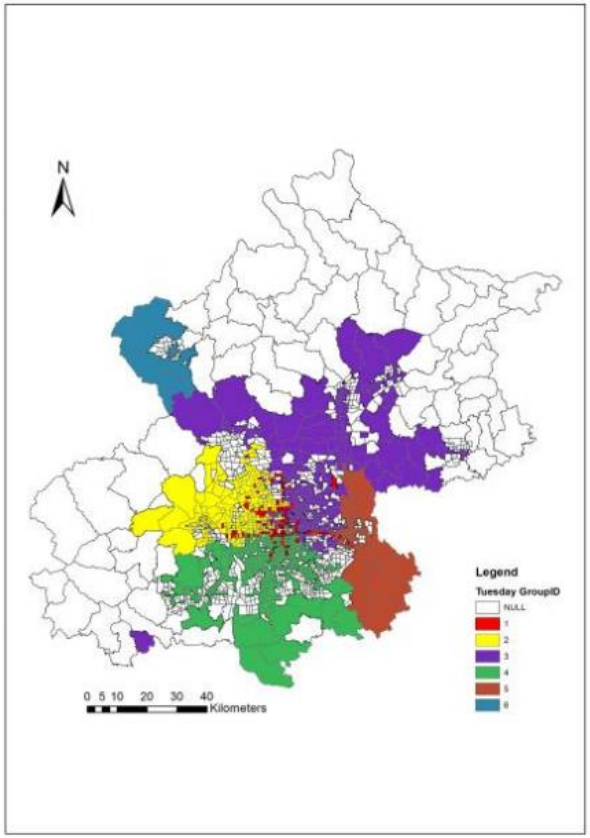
Connections among Cities
(based on trains between/across cities)

Road intersection density

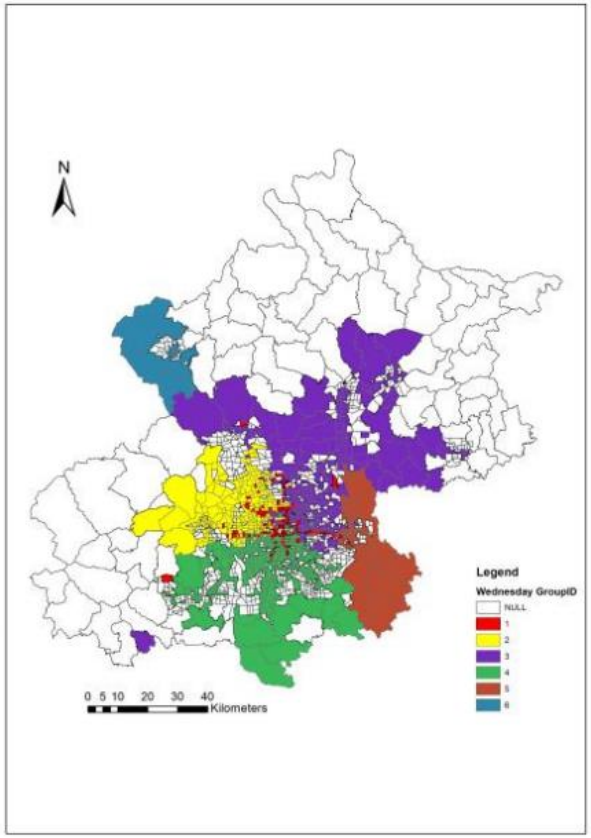




Monday



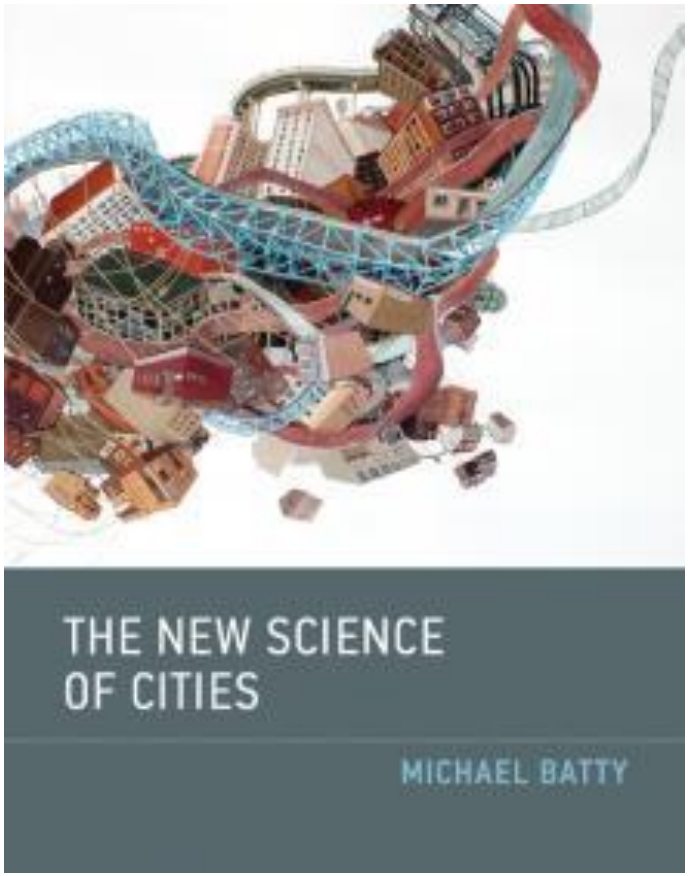
Tuesday



Wednesday

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?

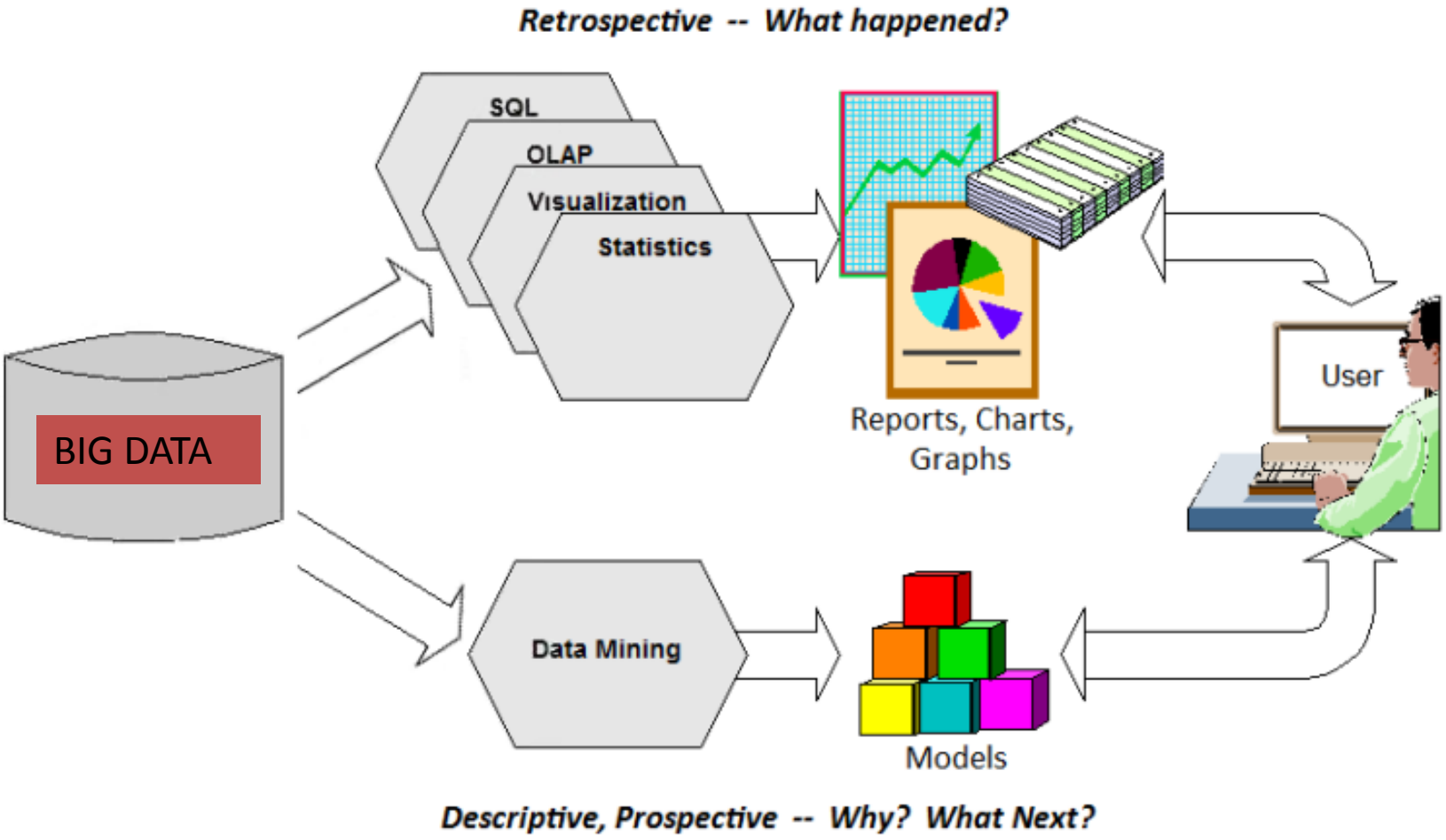
- **Regardless small or big, the end goal for us to collect and analyze big data is to generate knowledge and wisdom**
- **Big data are not the only way to generate knowledge and wisdom (e.g., we have numerous great scientists and philosophers before big data emerged)**

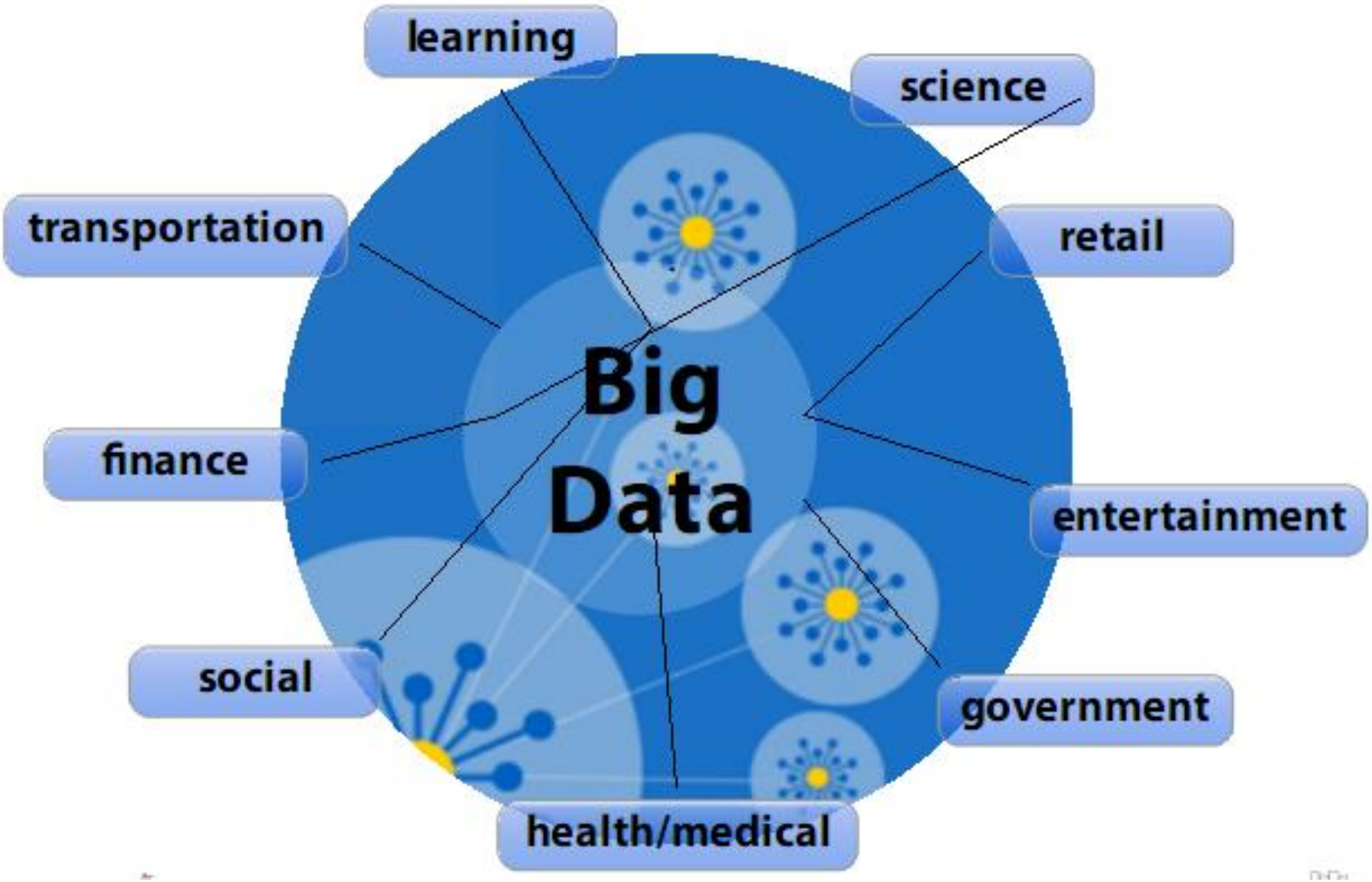
Why we need big data?

- **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**

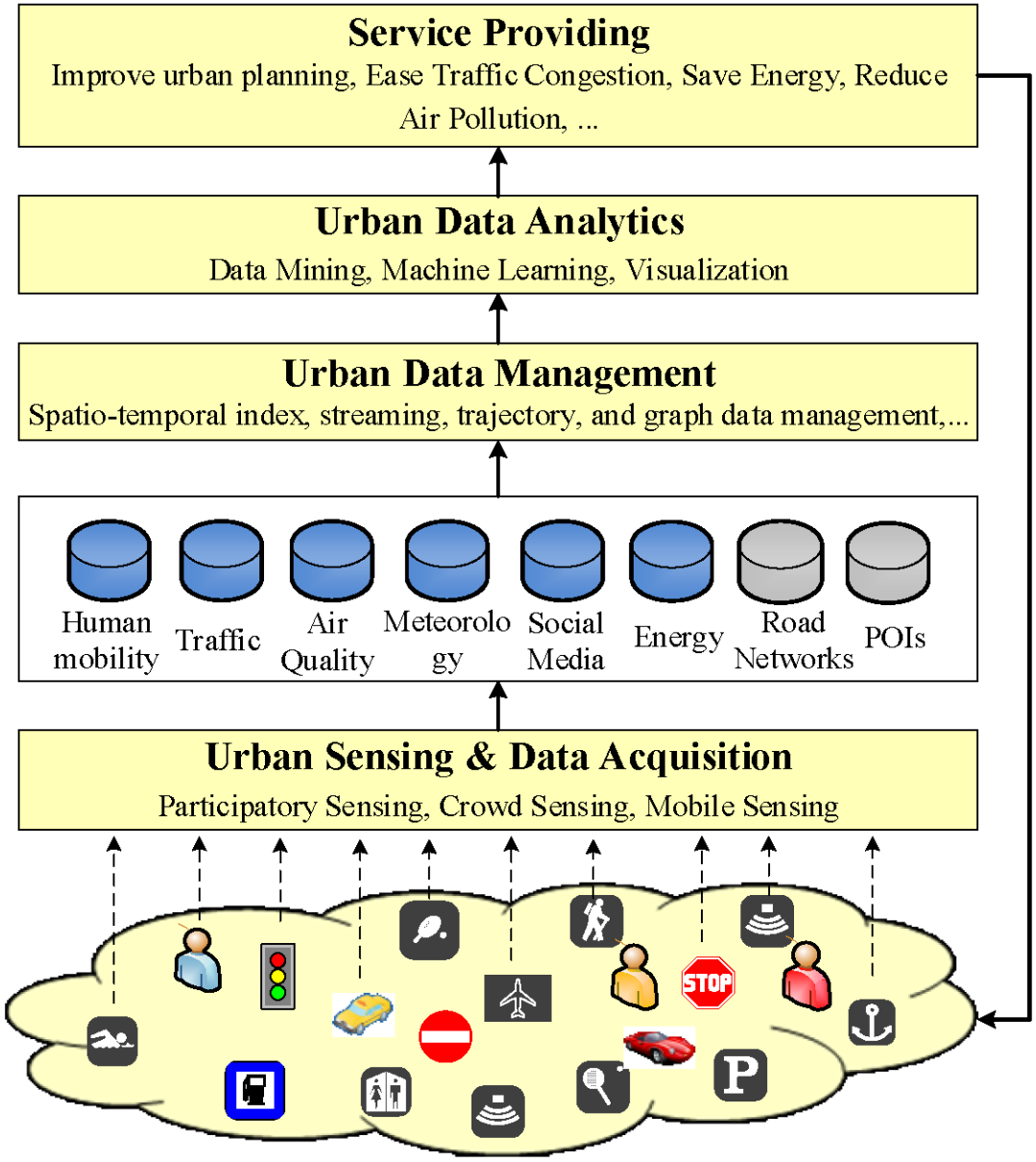


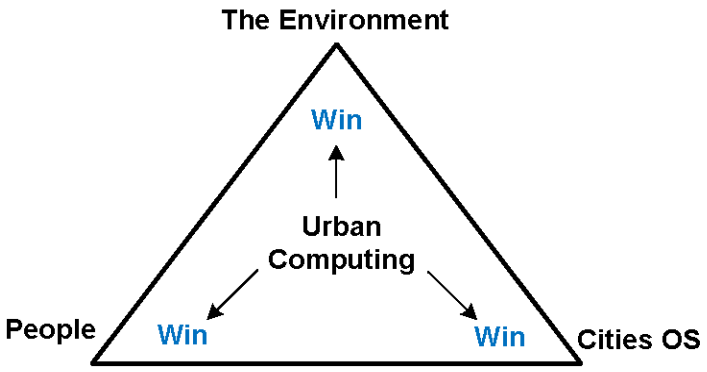

learning
science
transportation
retail
Big
Data
finance
entertainment
social
government
health/medical

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**

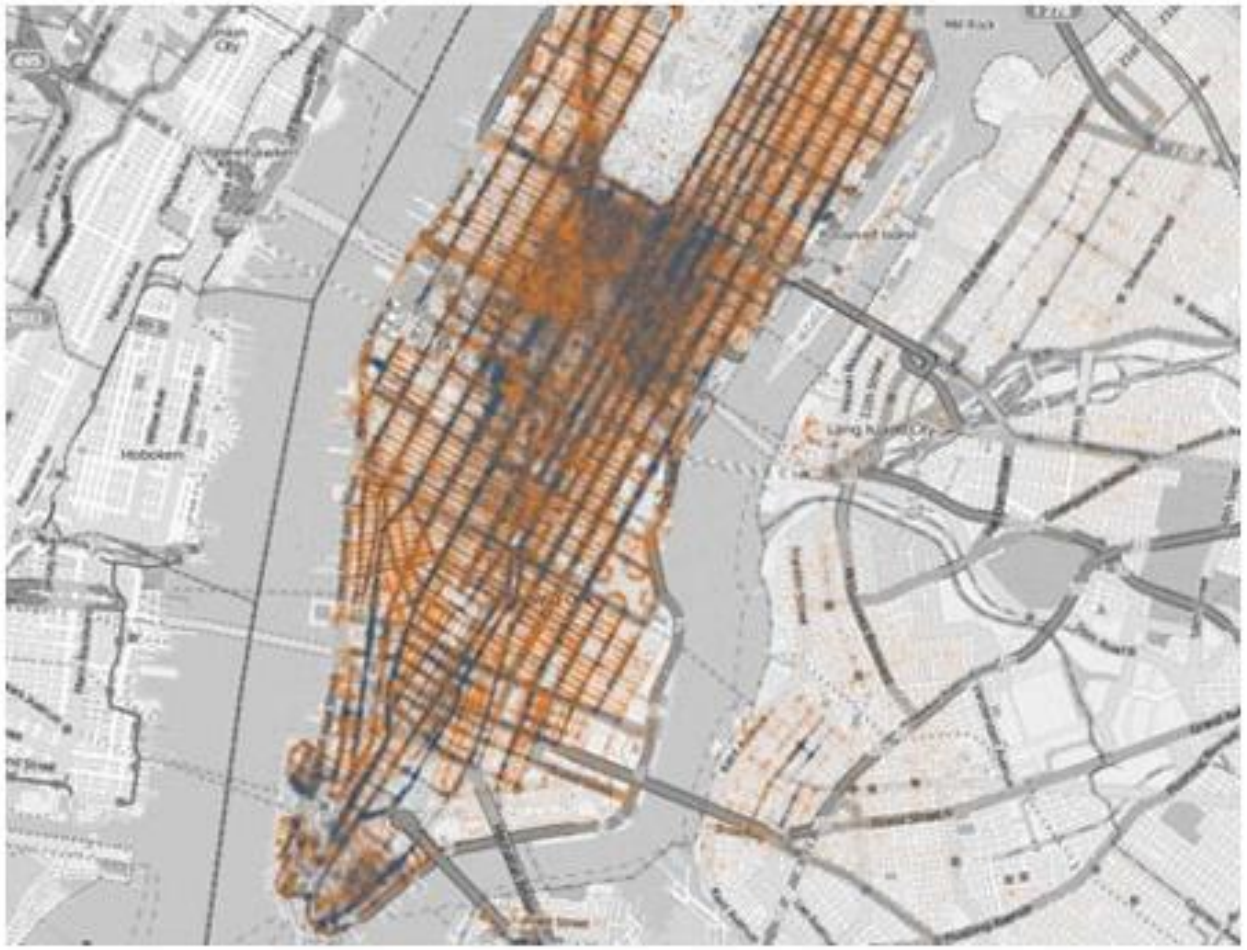
Big data and urban research





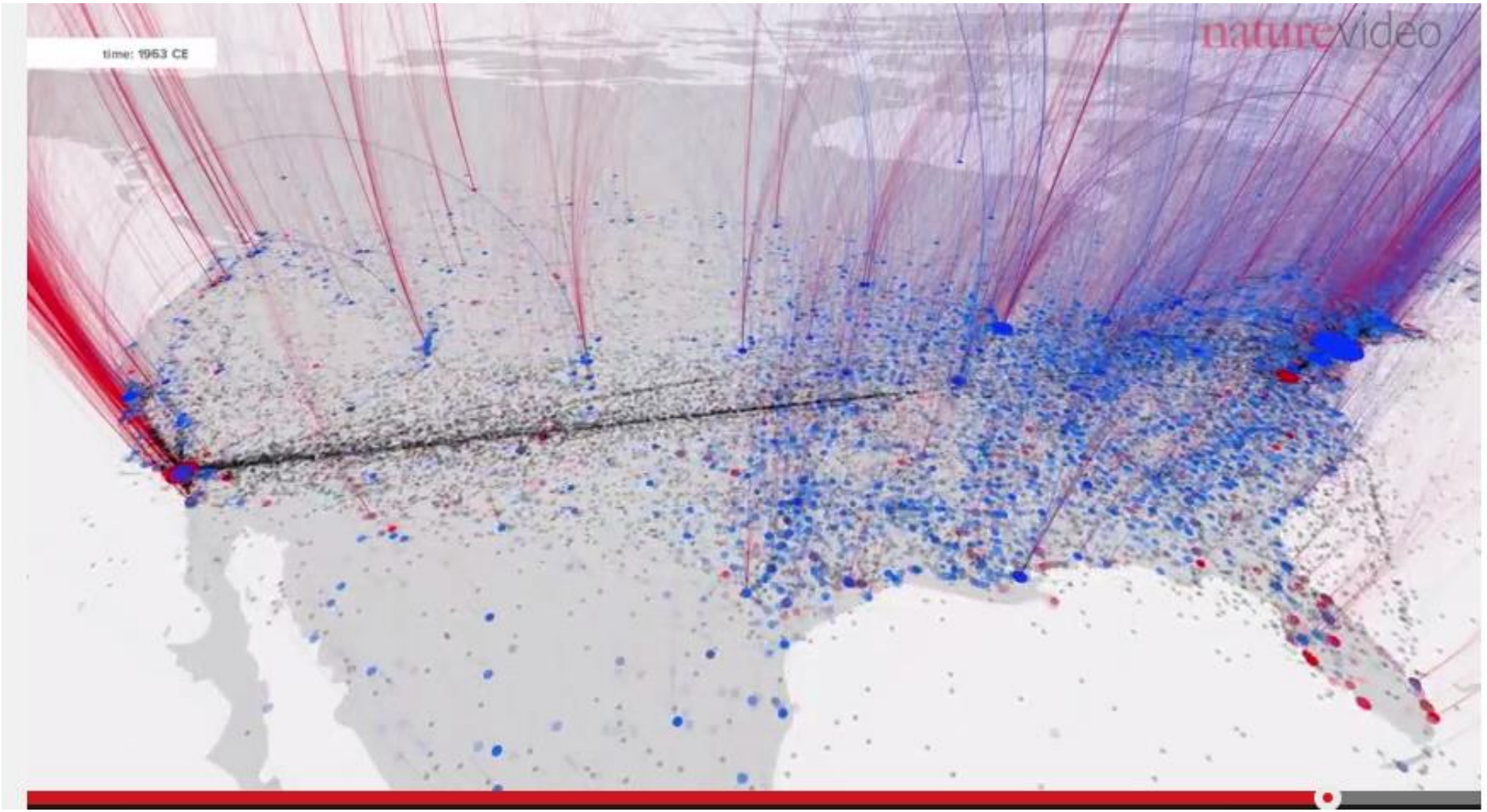
Tackle the Big challenges in Big cities using Big data!

Zheng, Y., et al. Urban Computing: concepts, methodologies, and applications. *ACM transactions on Intelligent Systems and Technology*.

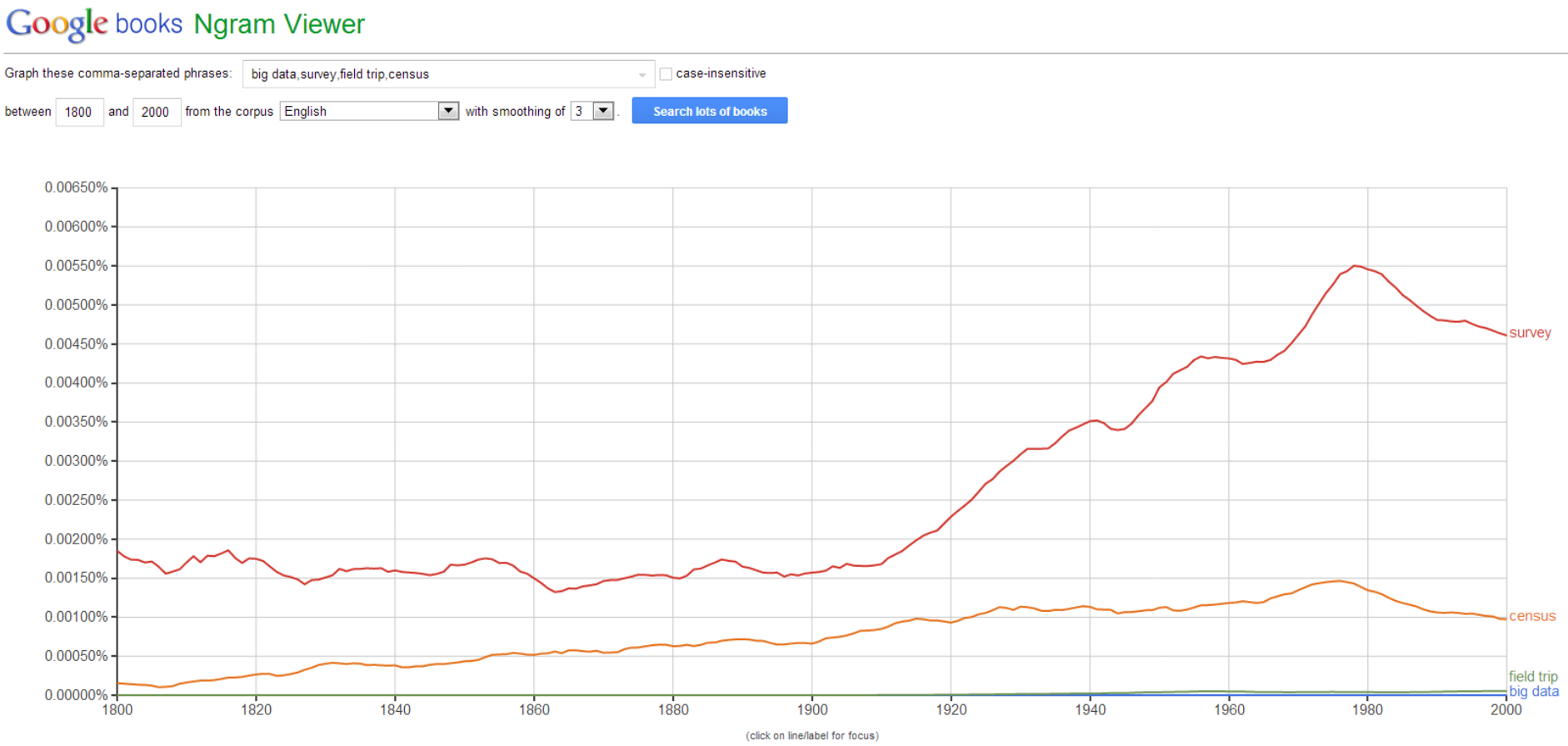


Visualization of Taxi Pick-ups (Orange) and Drop-offs (Blue) in New York City (NYU Center for Urban Science and Progress)

Humanity's migration and cultural history



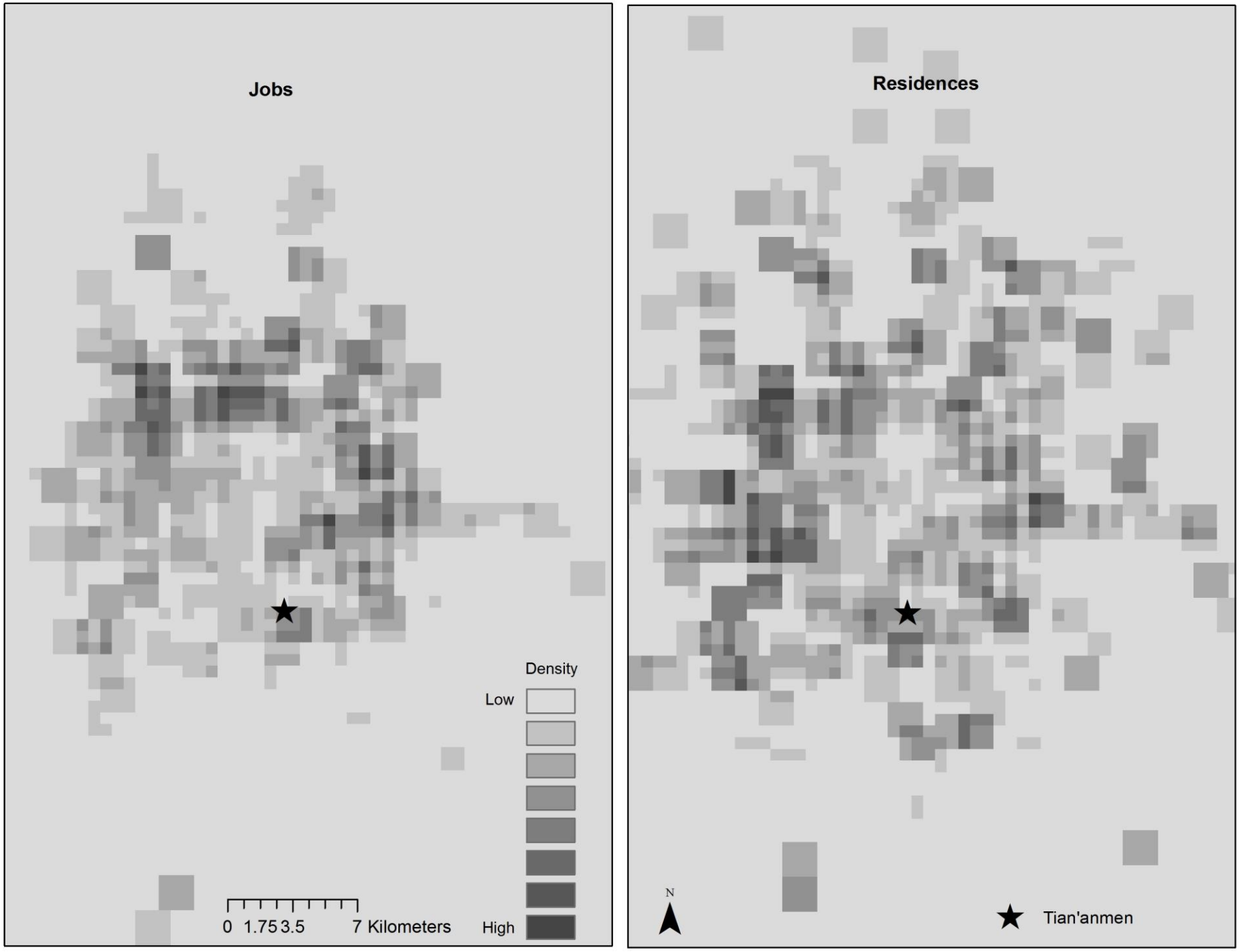
https://www.youtube.com/watch?v=4gIhRkCcD4U#t=95

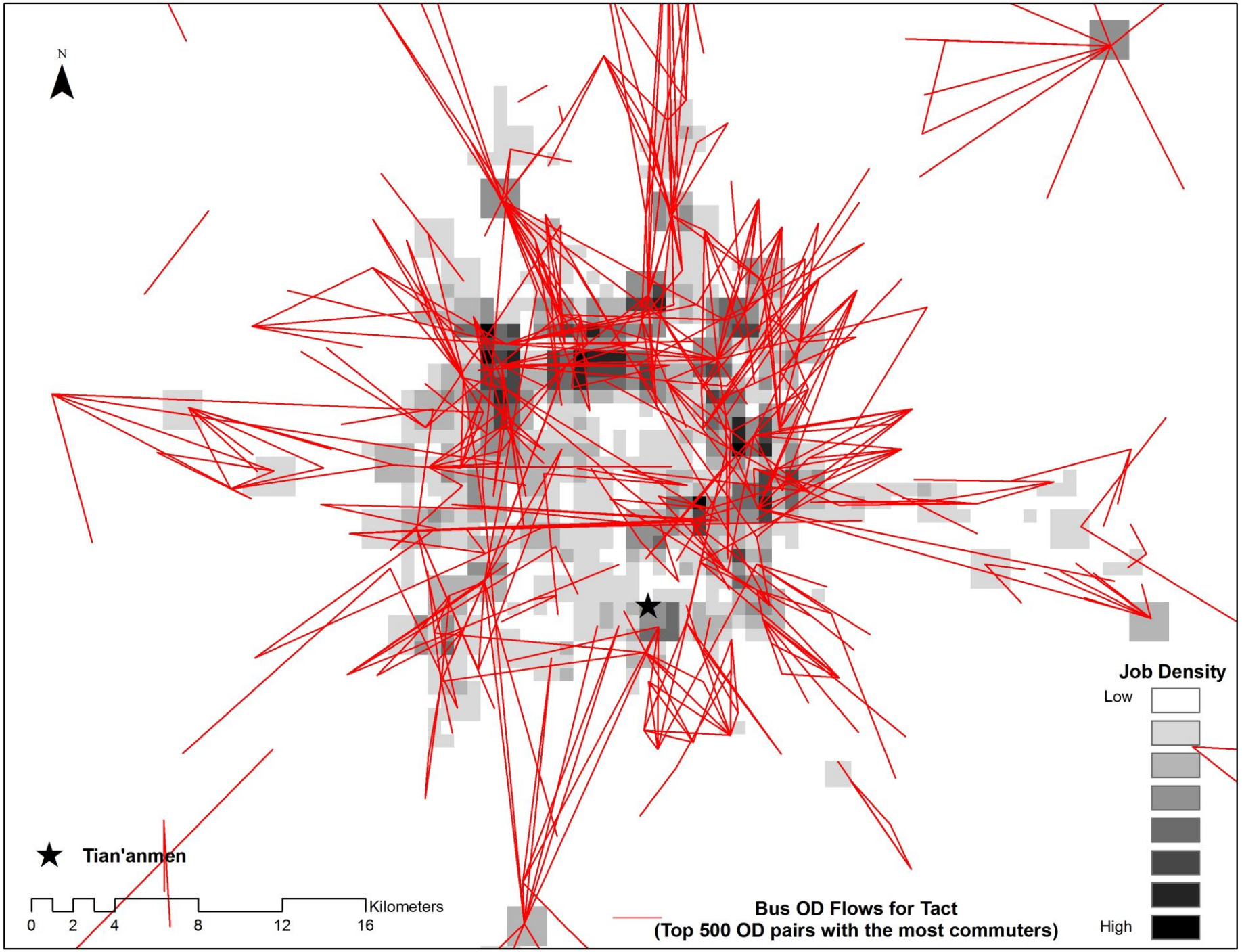


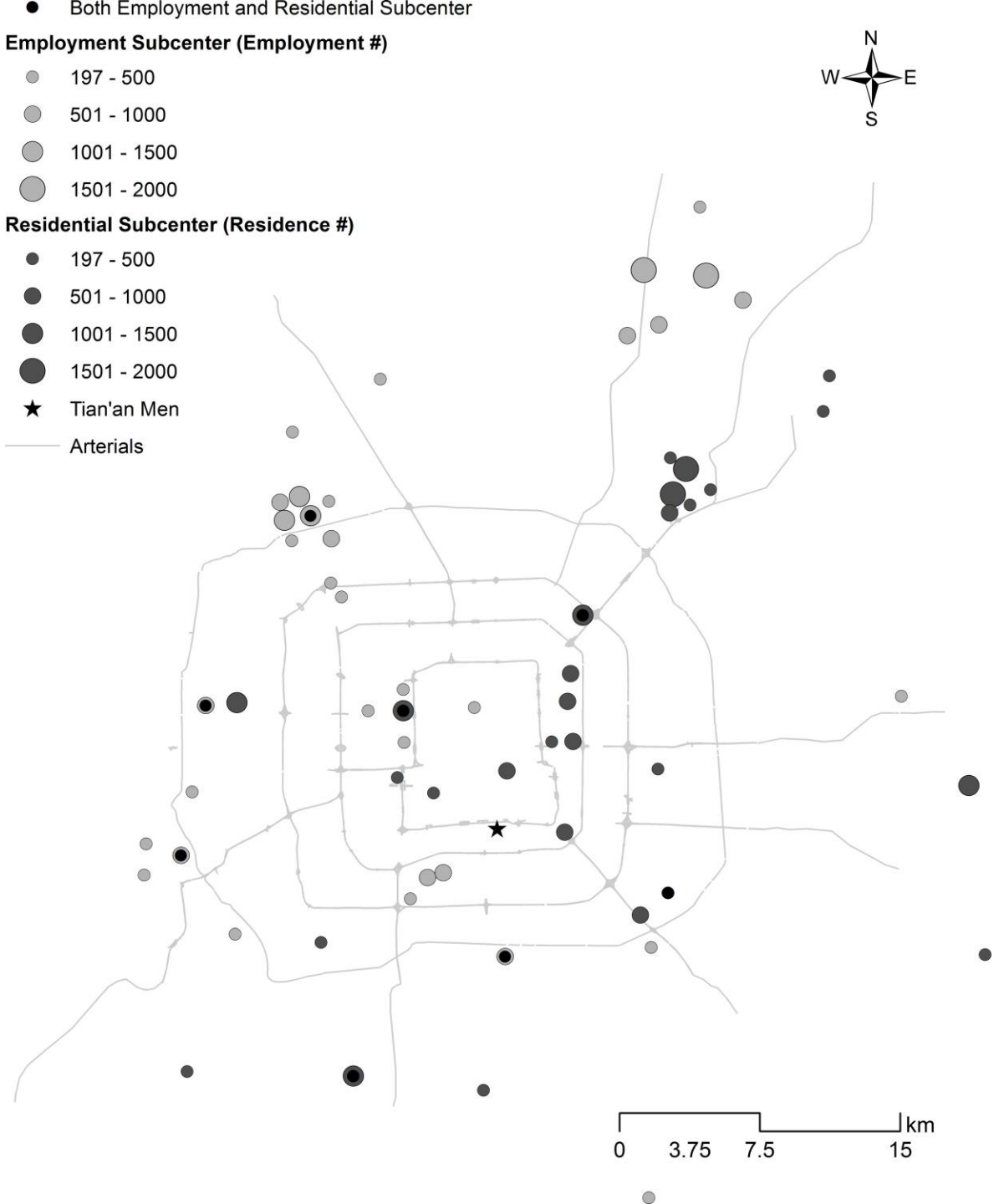
Emergence of big data

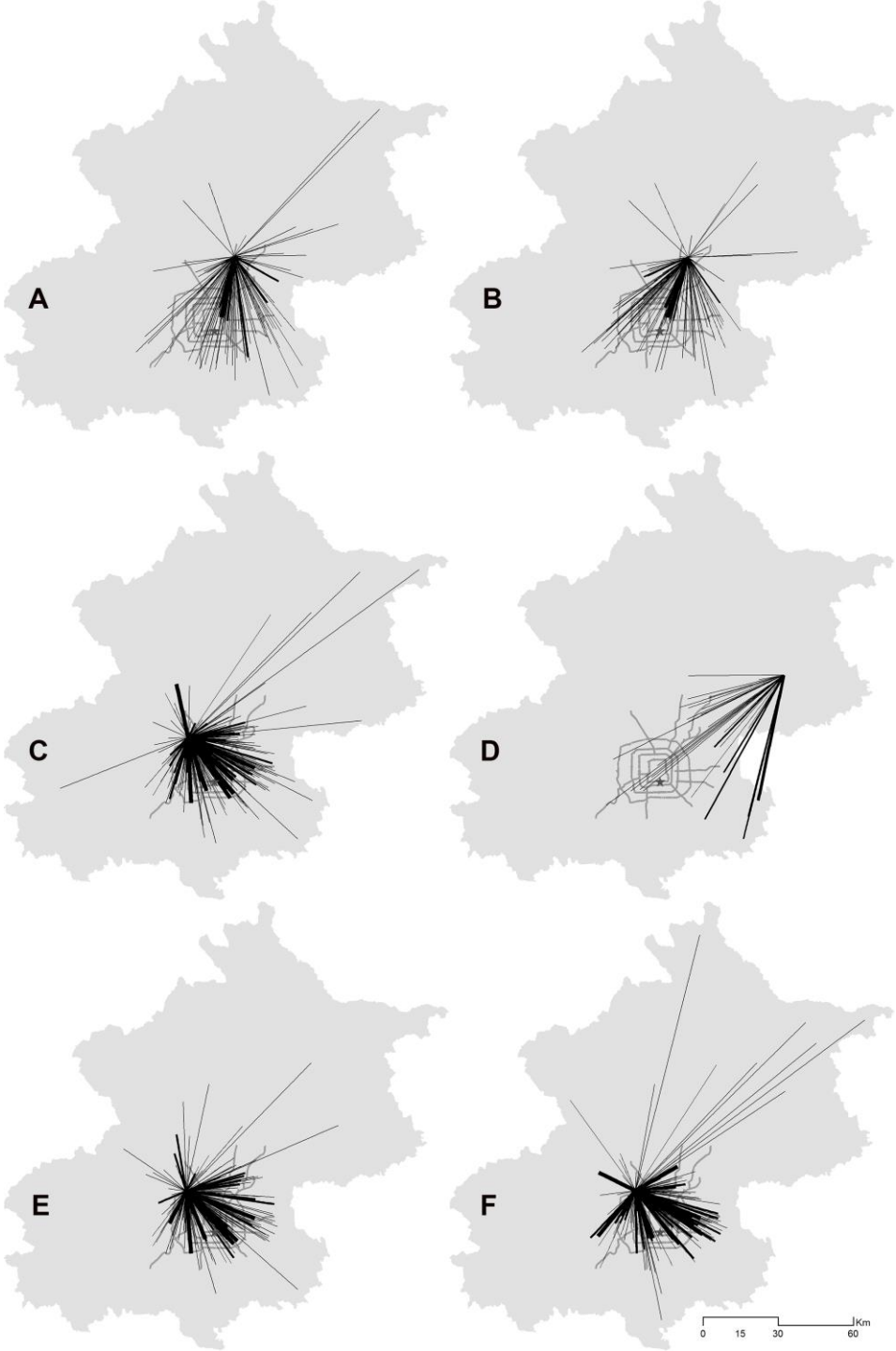
Big data and Urban China

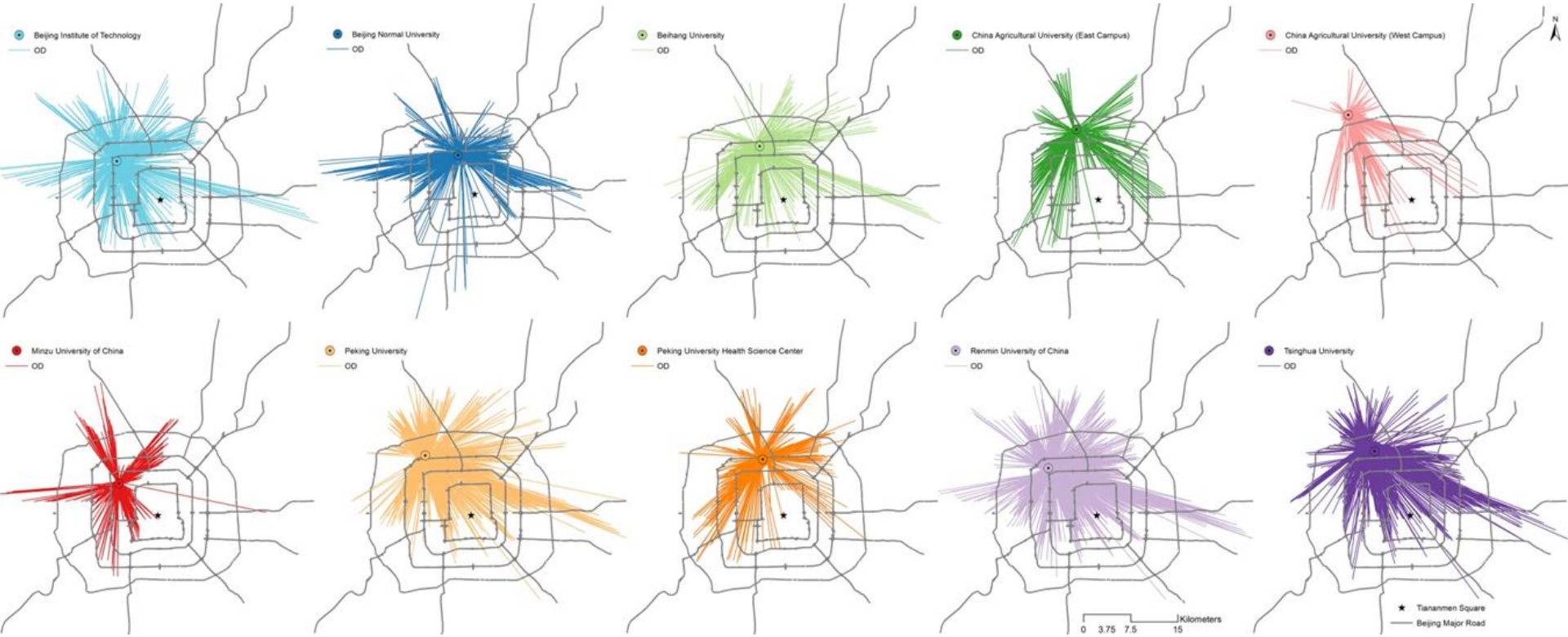
Spatial patterns of human settlements/movements

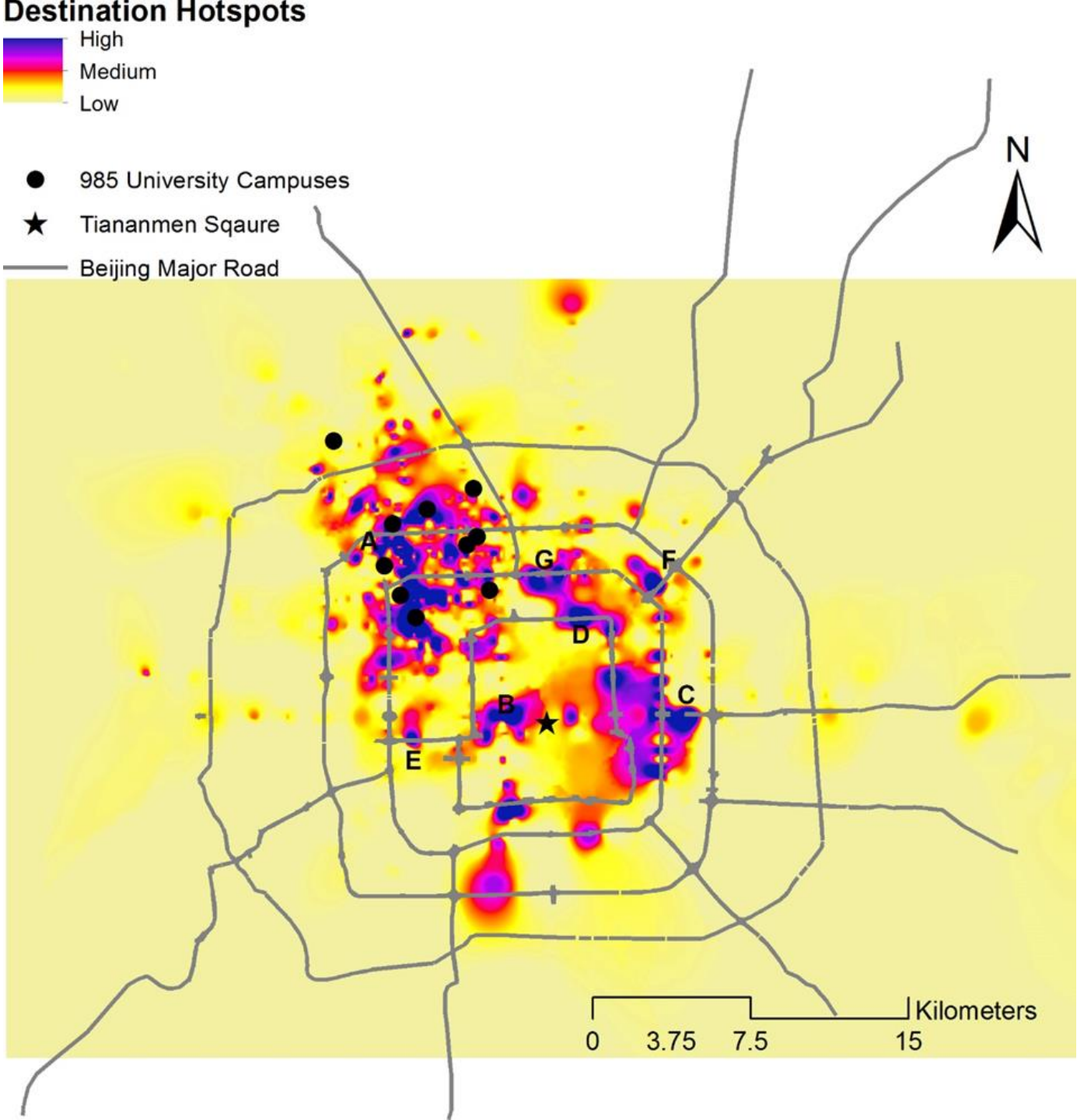

Jobs
Density
Low
High
0 1.75 3.5 7 Kilometers
Residences
N
Tian'anmen

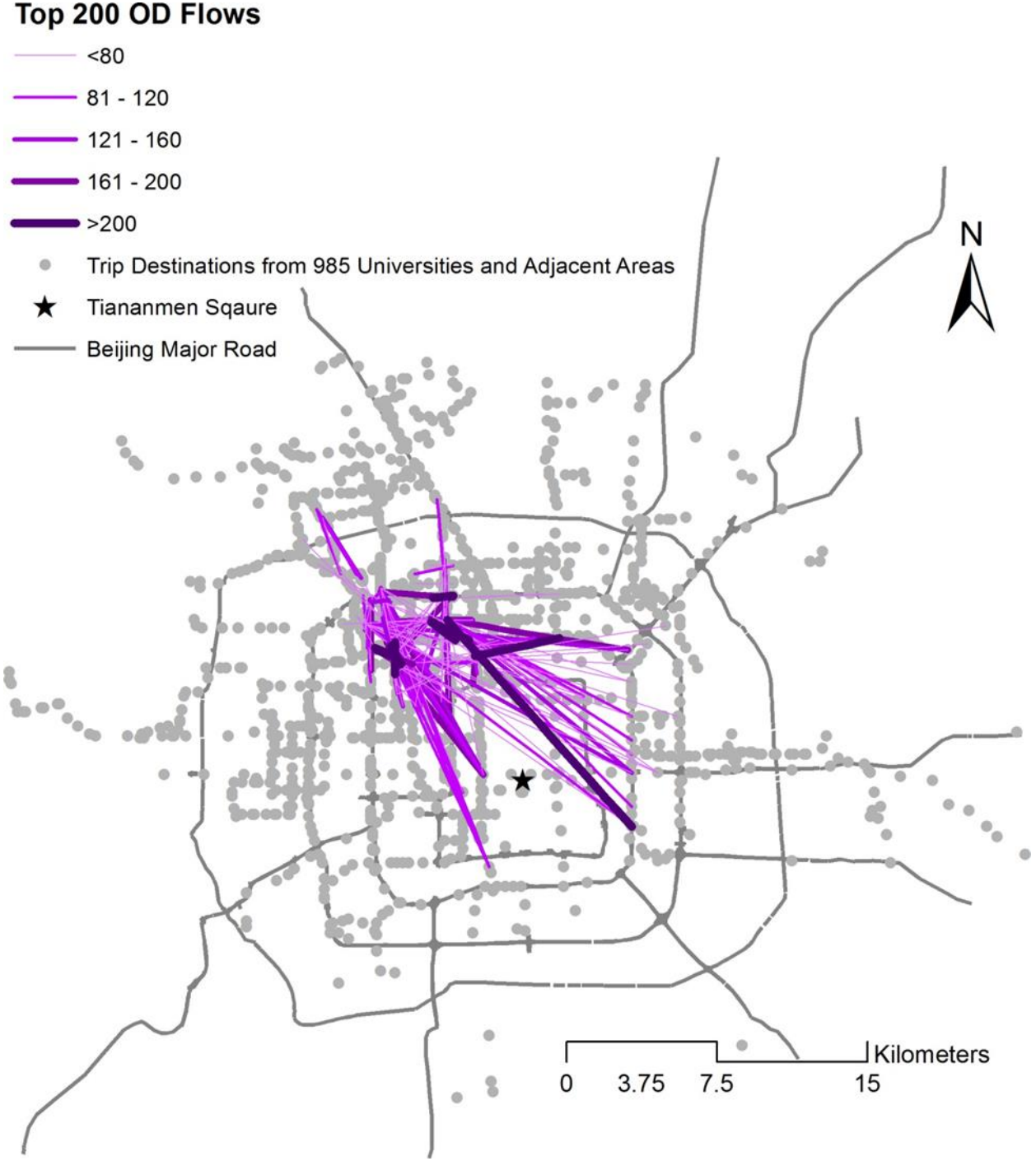

N
Tian'anmen
0 2 4 8 12 16
Kilometers
Bus OD Flows for Tact
(Top 500 OD pairs with the most commuters)
Job Density
Low
High


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
N
W
E
S
km
0
3.75
7.5
15

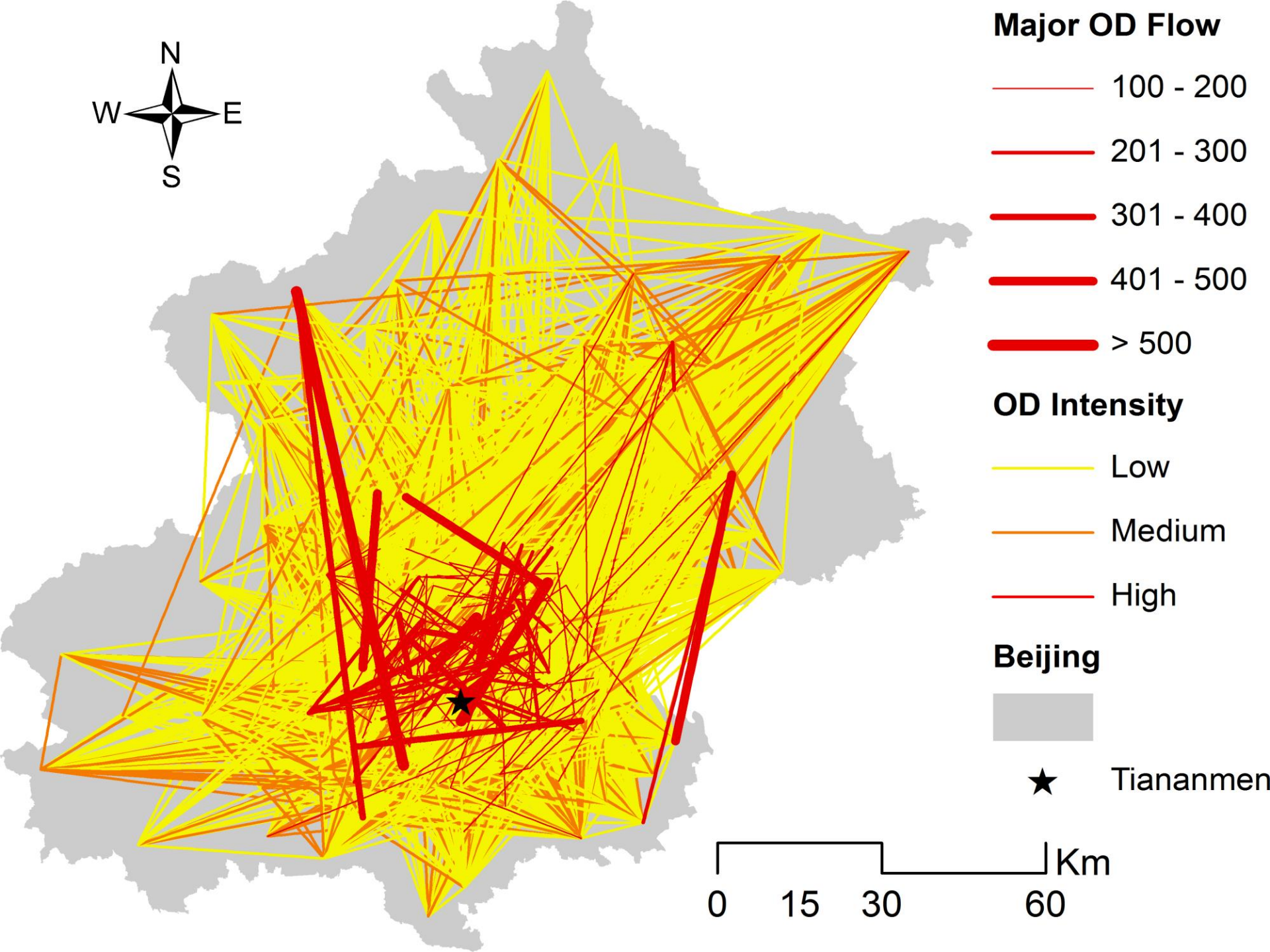

A
B
C
D
E
F
0
15
30
60
Km


Beijing Institute of Technology
OD
Beijing Normal University
OD
Beihang University
OD
China Agricultural University (East Campus)
OD
China Agricultural University (West Campus)
OD
N
Minzu University of China
OD
Peking University
OD
Peking University Health Science Center
OD
Renmin University of China
OD
Tsinghua University
OD
0 3.75 7.5 15
Kilometers
Tiananmen Square
Beijing Major Road


Destination Hotspots
High
Medium
Low
985 University Campuses
Tiananmen Sqaure
Beijing Major Road
N
A
B
C
D
E
F
G
Kilometers
0
3.75
7.5
15

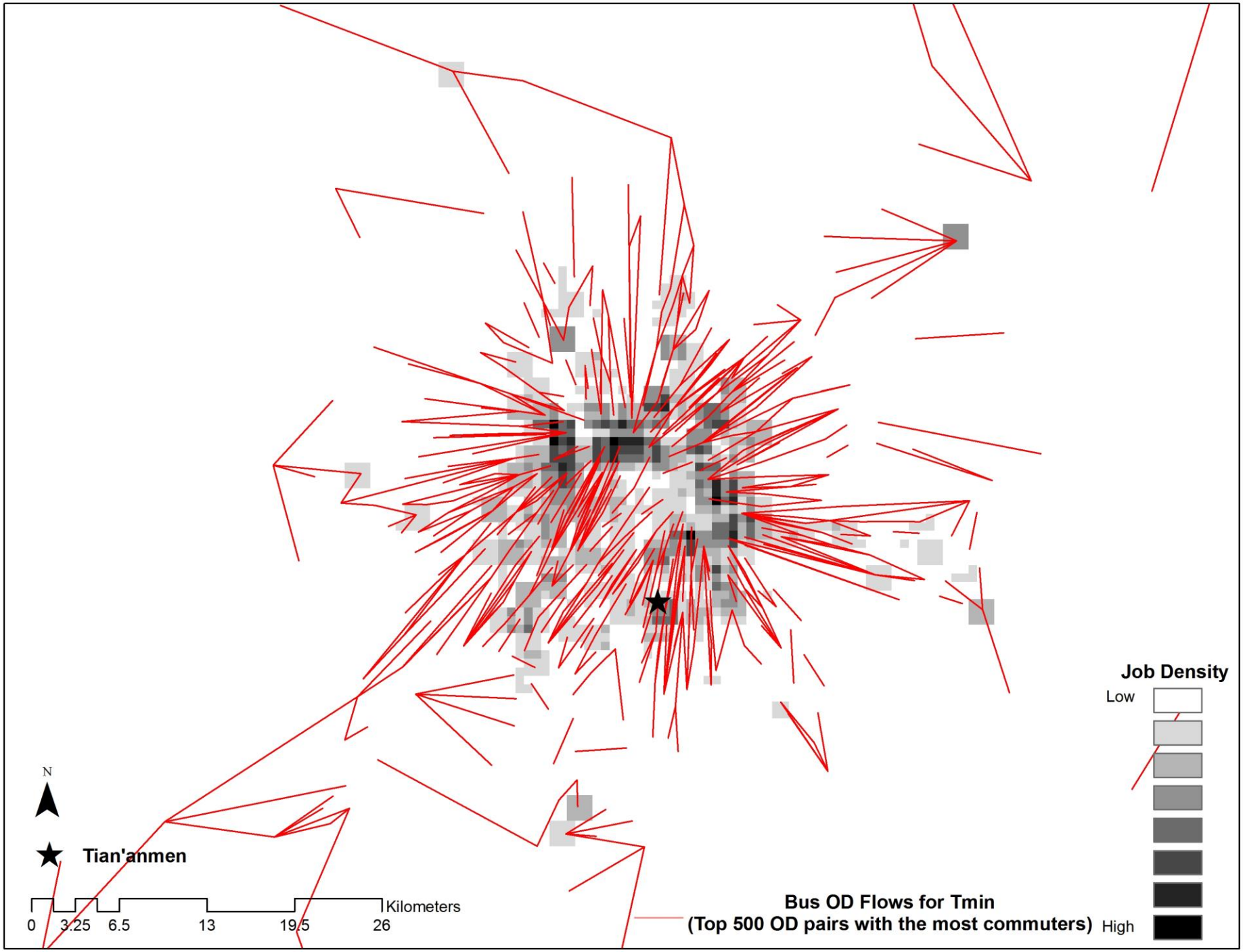

Top 200 OD Flows
<80
81 - 120
121 - 160
161 - 200
>200
Trip Destinations from 985 Universities and Adjacent Areas
Tiananmen Sqaure
Beijing Major Road
N
Kilometers
0
3.75
7.5
15

Hot spots where there are problems

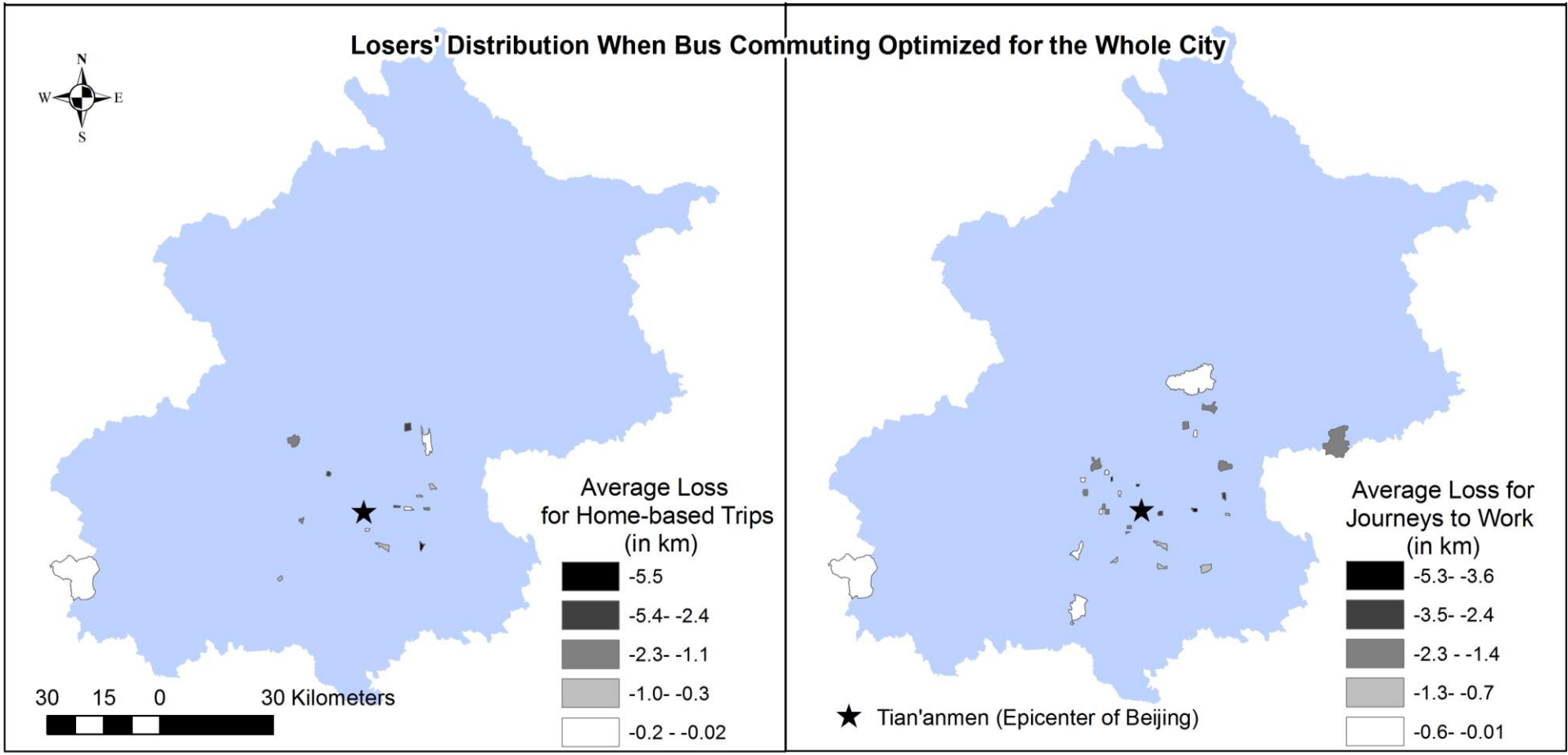

N
W
E
S
Major OD Flow
100 - 200
201 - 300
301 - 400
401 - 500
> 500
OD Intensity
Low
Medium
High
Beijing
Tiananmen
Km
0
15
30
60


A
B
C
D
D
Tian'an Men Square
Commuting Distance (km)
<=8.1
8.2 - 16.2
16.3 - 24.3
24.4 - 32.4
> 32.4
N
0 2.5 5 10 Kilometers

Optimize
commuting and traffic


Job Density
Low
High
N
Tian'anmen
0
3.25
6.5
13
19.5
26
Kilometers
Bus OD Flows for Tmin
(Top 500 OD pairs with the most commuters)

Identify losers and winners in optimization

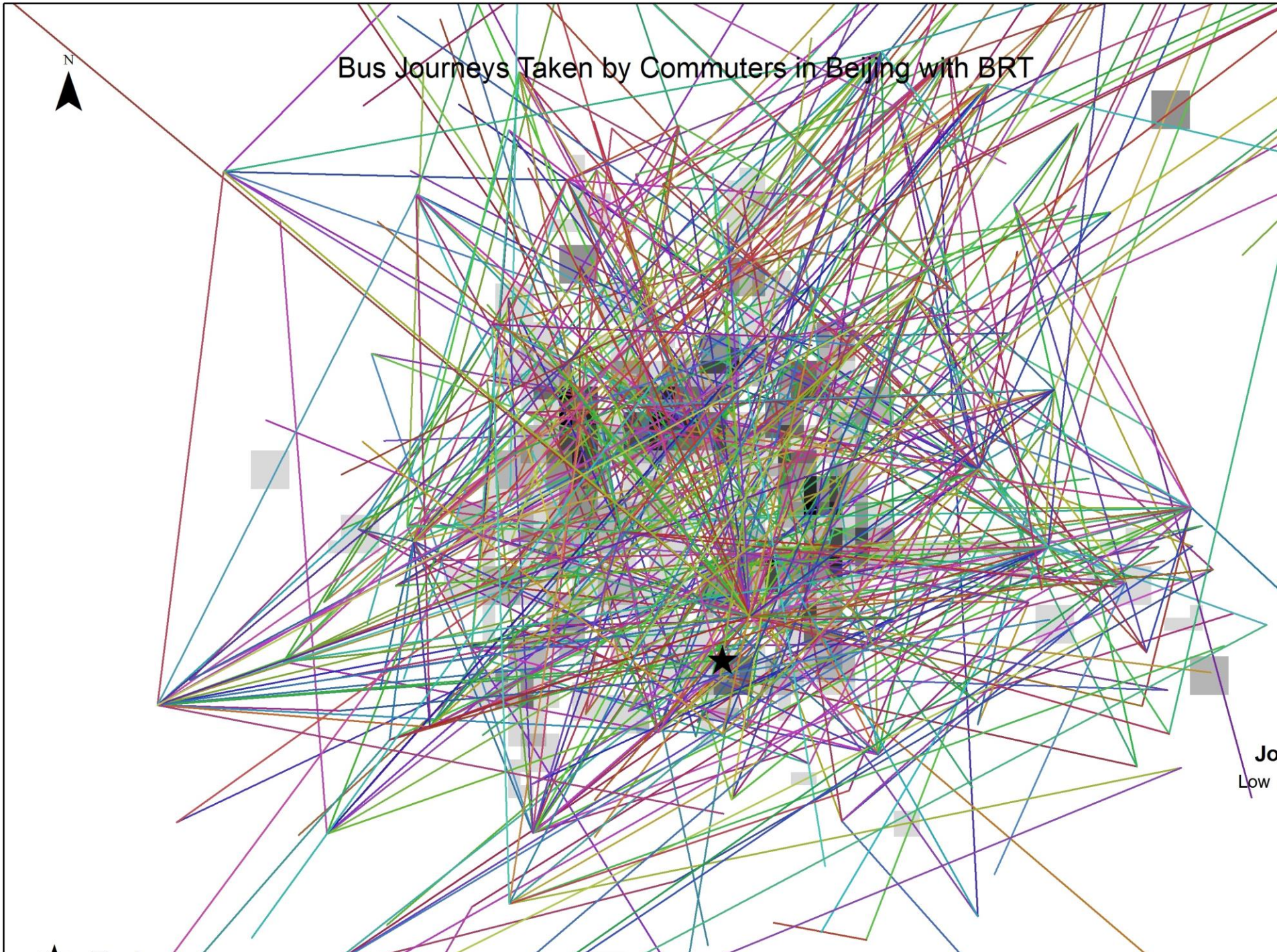

Losers' Distribution When Bus Commuting Optimized for the Whole City
N
W
E
S
30 15 0 30 Kilometers
Average Loss for Home-based Trips (in km)
-5.5
-5.4- -2.4
-2.3- -1.1
-1.0- -0.3
-0.2 - -0.02
Tian'anmen (Epicenter of Beijing)
Average Loss for Journeys to Work (in km)
-5.3- -3.6
-3.5- -2.4
-2.3 - -1.4
-1.3- -0.7
-0.6- -0.01

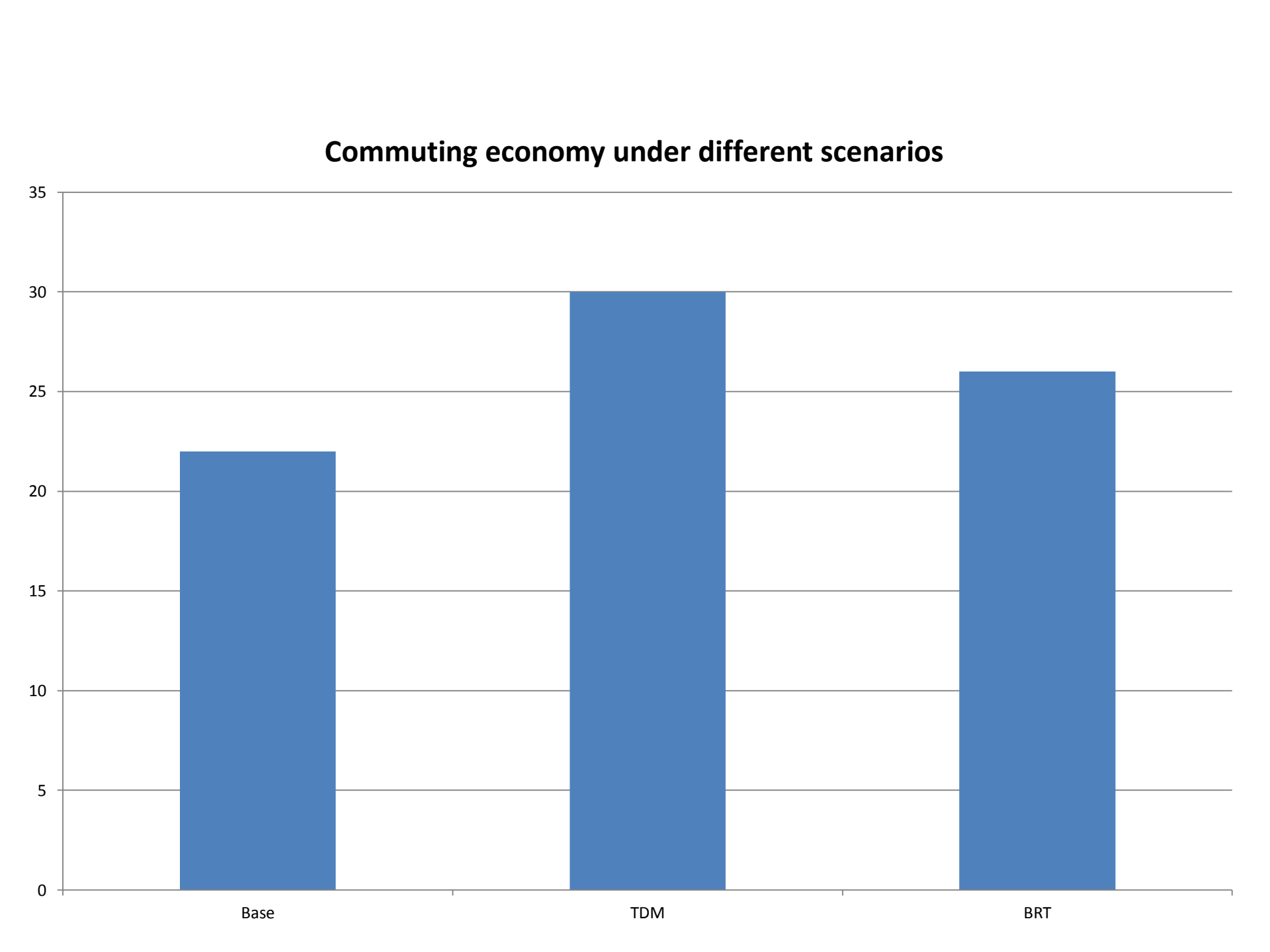
Construct policy scenarios to understand impacts of different policies

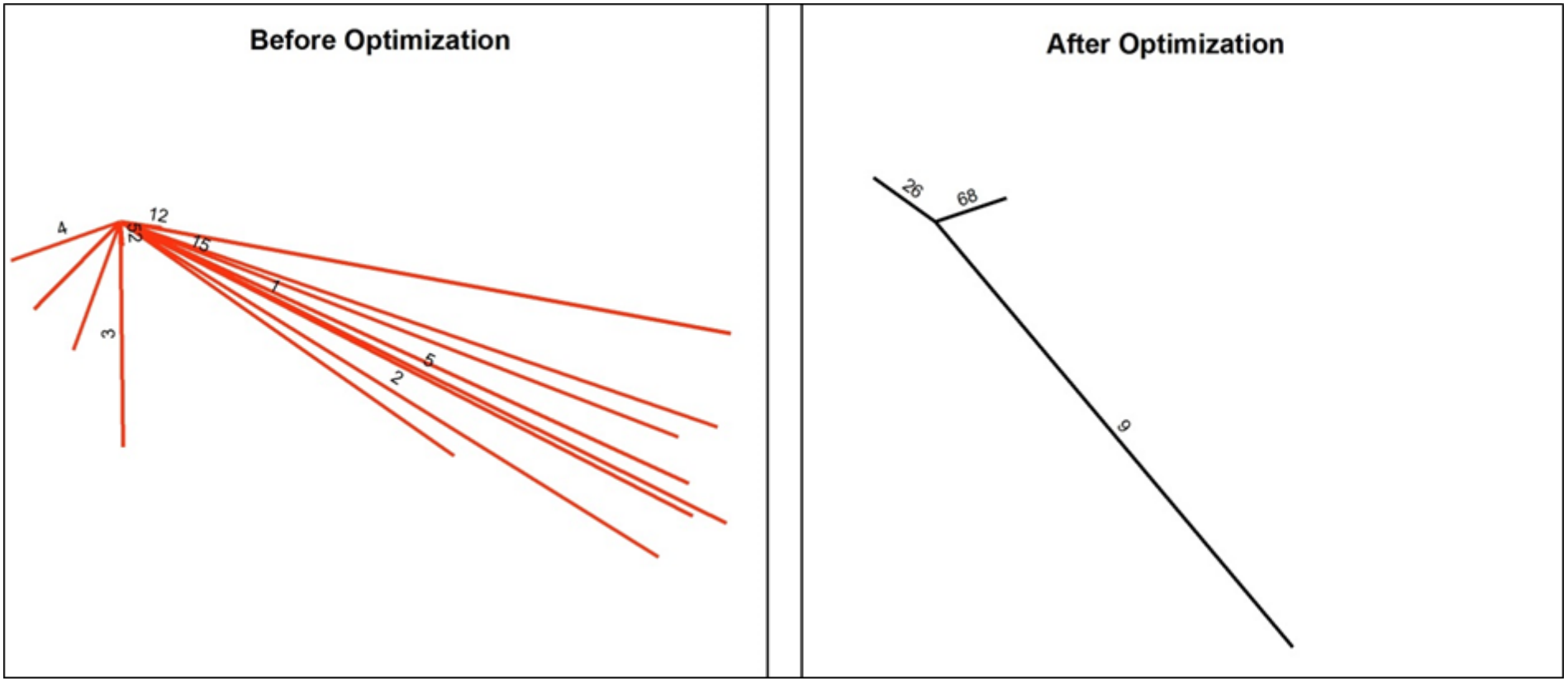
Policy Scenario 1: Doing nothing

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.


N
Bus Journeys Taken by Commuters in Beijing with BRT
Jo
Low


Commuting economy under different scenarios
35
30
25
20
15
10
5
0
Base
TDM
BRT

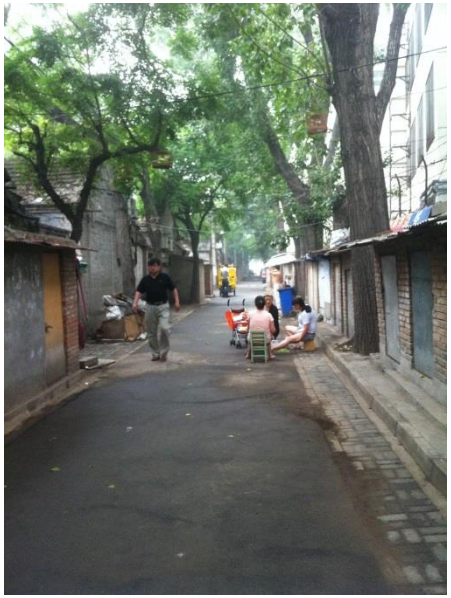

Before Optimization
4
12
9
2
15
1
3
5
2
After Optimization
26
68
9

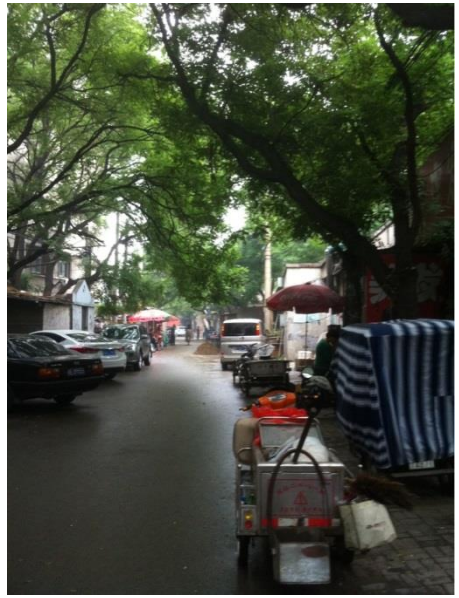
Trip Distance (Km)	Number of Resident Workers	
	Before the Optimization	After the Optimization
<4.05	1	9
4.05-8.1	93	0
8.1-12.2	8	94
12.2-16.3	0	0
>16.3	1	0

Even with big data, we cannot forget traditional and open data

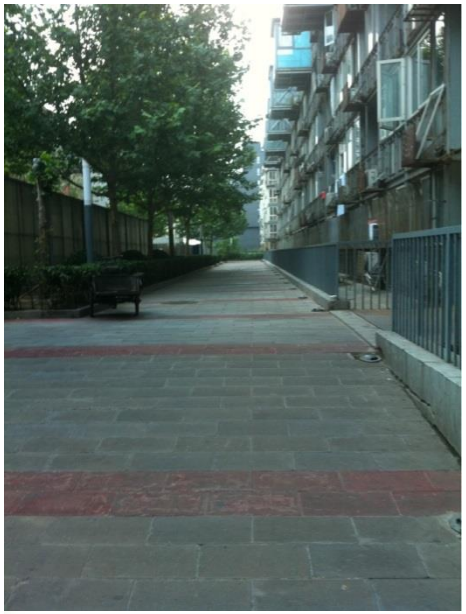
United States Census Bureau


Columbia University
Ivy League school founded in 1754
Barnard College
students study · student computing services
Teachers College, Columbia University
office of admissions...
Columbia Business School
Columbia Law School
greene hall · sexuality and gender
Manhattan School of Music
theory and...
Sakura Park
South Lawn
Morningside Park
Morningside Pond
West 111th Street People's Garden
Harlem Tavern
Riverside Dr
Broadway
Claremont Ave
Henry Hudson Pkwy
New York State Reference Rte 907V
Hudson River
Riverside Dr
West End Ave
Broadway
Manhattan Ave
Morningside Ave
Saint Nicholas Ave
Hancock Pl
La Salle St
W 128th St
W 127th St
W 122nd St
W 120th St
W 123rd St
W 119th St
W 118th St
W 116th St
W 115th St
W 114th St
W 113th St
W 112th St
W 111th St
W 109th St
W 108th St
W 117th
W 115th St
W 114th St
W 113th St
9A



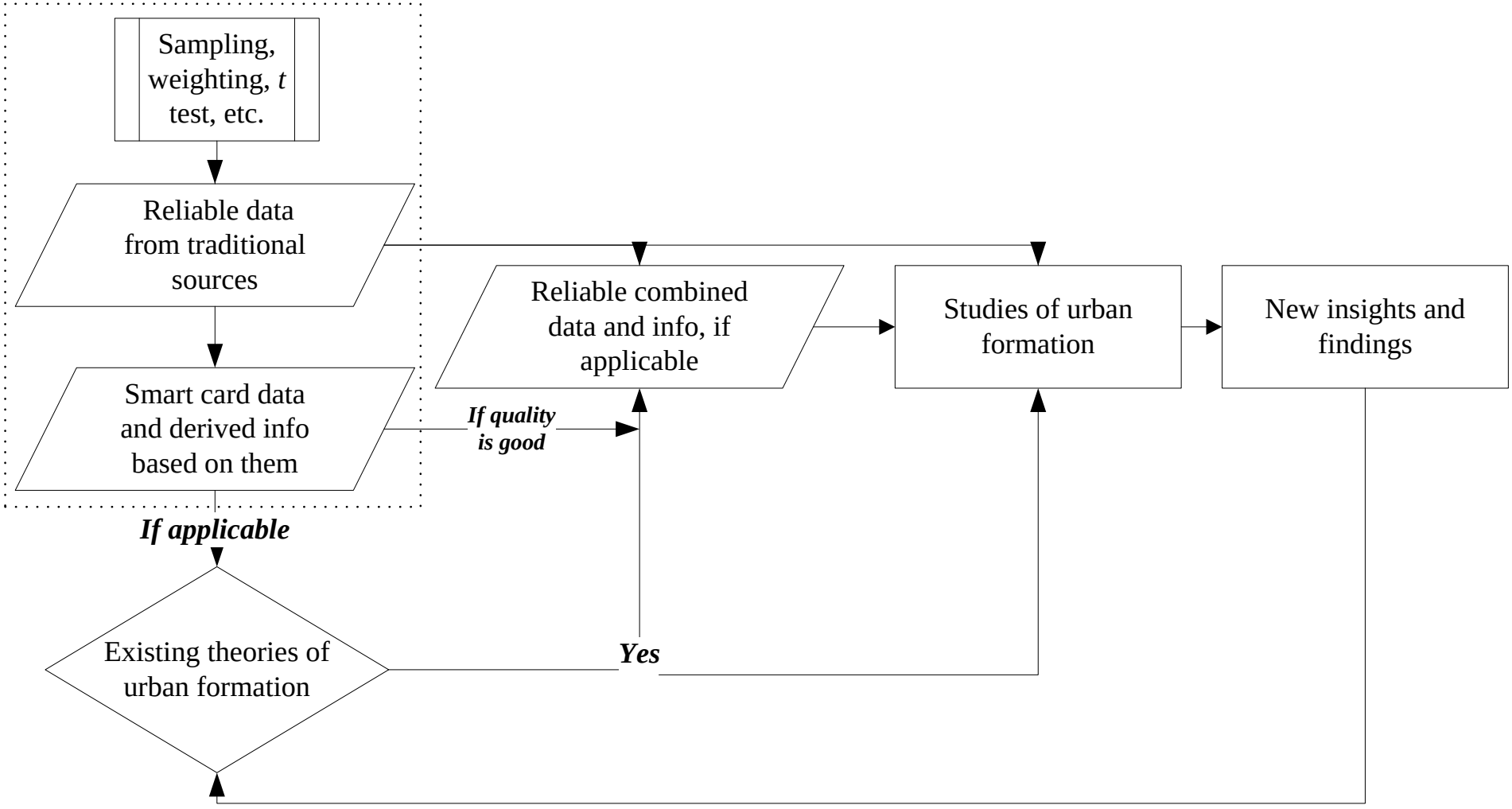




国奥村
DREAMIC WORLD
2008



Prototype	Characteristics	Landmarks and Directions	Spatial Index in Figure
1	University campuses, hotels and old Danwei compounds left with mostly apartment buildings	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	A (Areas around and areas to its northeast)
2	Parks with luxury hotels, high-end apartments, specialized research institutes, hospitals and some mixed-use residential areas	Areas north to Yu Yuan Tan Park; Area adjacent to Tian Tan Park in the east	A (South to A, the darkest area); The U-shaped area south to the star
3	Traditional Hu'tong with old, cheap, small, shared and underserviced rental housing units	Areas in between Qian Men Da Jie and Zhu Shi Kou Da Jie	B
4	Residential areas with mixed-age housing units adjacent to freeway interchanges or arterials, railways within the fifth ring road	Areas near Yong Ding Men and Nan Sha Wo Bridges	C
5	Low density, developing areas with relatively cheap housing units in the suburb	Areas adjacent to the sixth ring road and Jingshi Freeway interchange; Areas adjacent to Yan Chun Railway Station	D

Verify and even extend theories


Sampling, weighting, *t* test, etc.
Reliable data from traditional sources
Smart card data and derived info based on them
If quality is good
If applicable
Existing theories of urban formation
Yes
Reliable combined data and info, if applicable
Studies of urban formation
New insights and findings

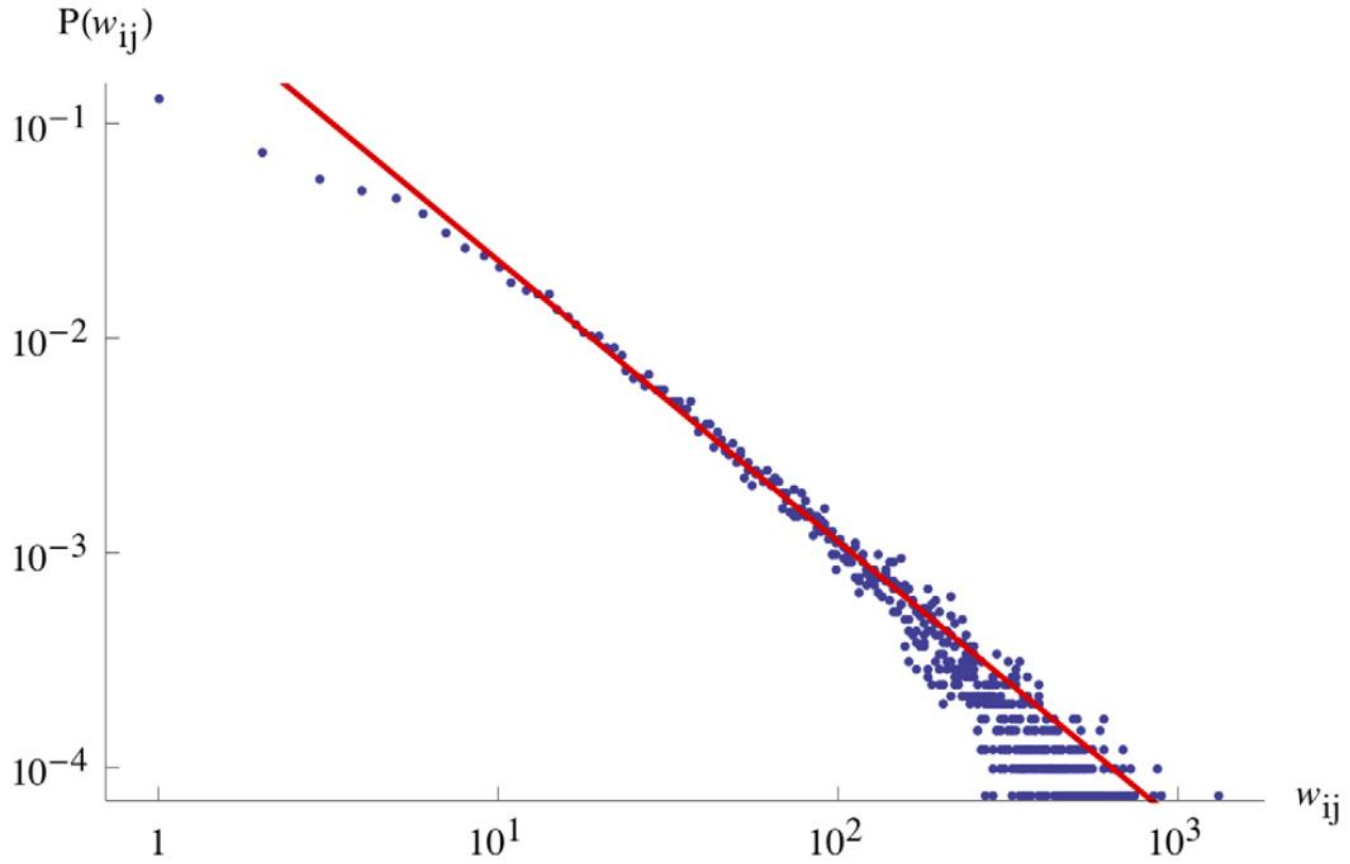
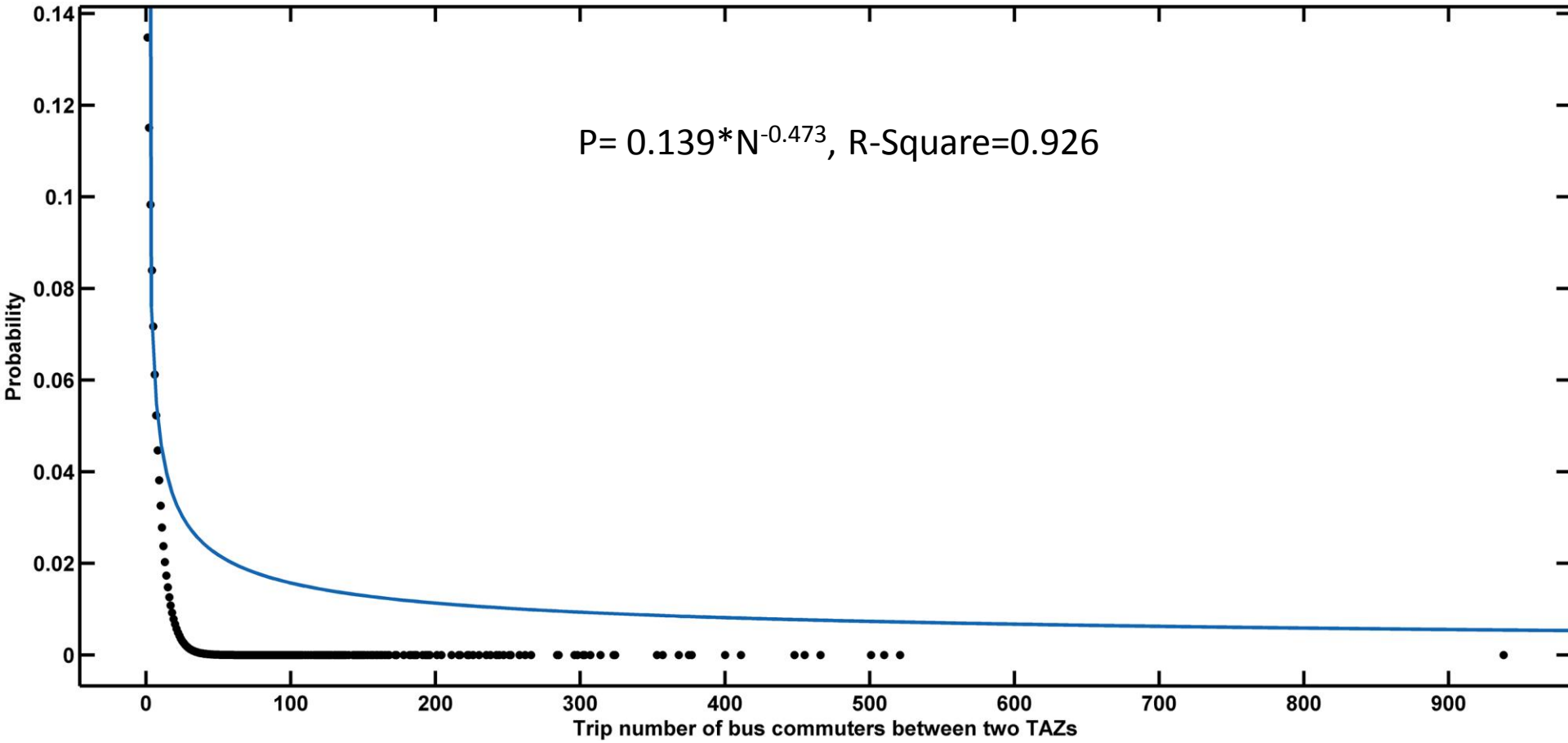
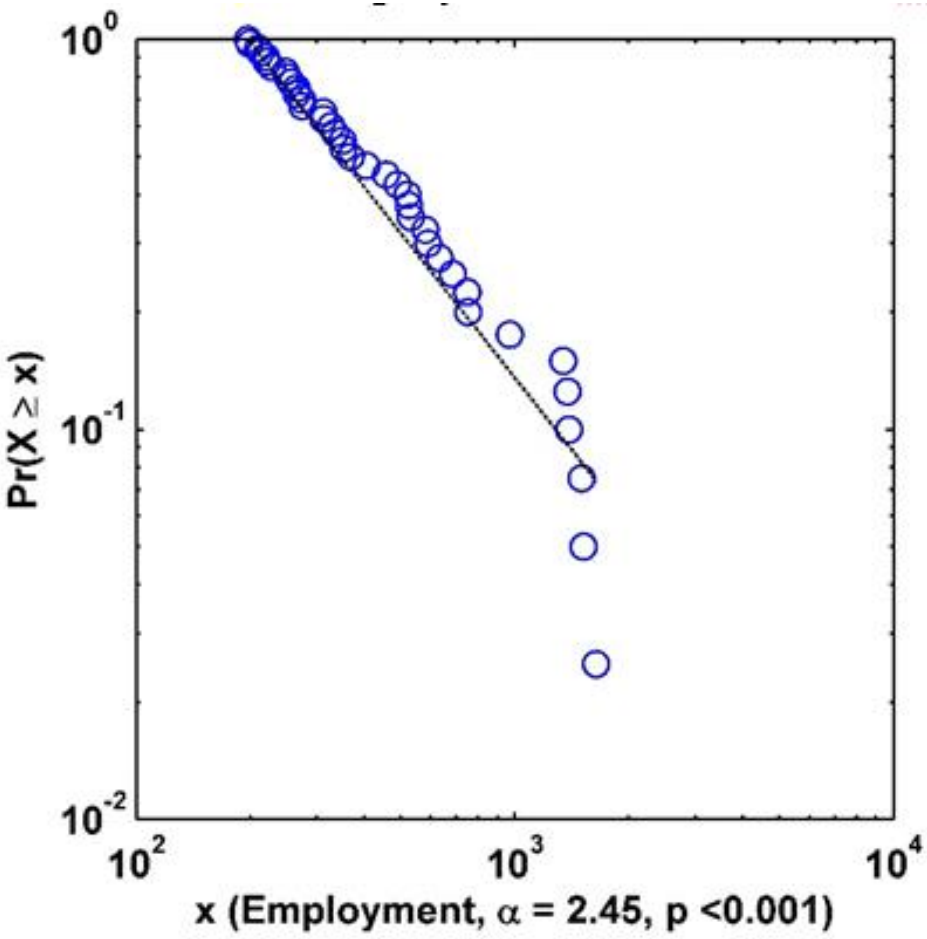
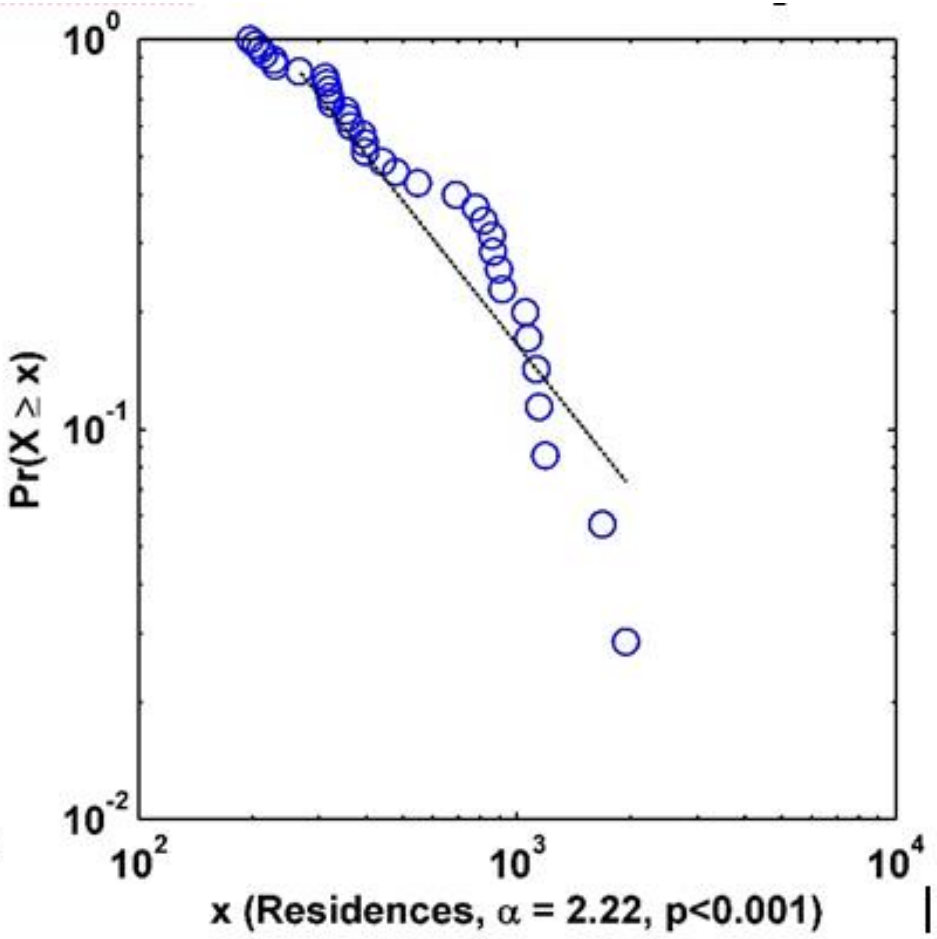


Figure 1. Flow distribution. Loglog plot of the histogram of the number of trips between two stations of the tube system. The line is a power law fit with exponent ≈ 1.3.
doi:10.1371/journal.pone.0015923.g001


P= 0.139*N-0.473, R-Square=0.926
Probability
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
100
200
300
400
500
600
700
800
900
Trip number of bus commuters between two TAZs


10^0
10^{-1}
10^{-2}
Pr(X ≥ x)
10^2
10^3
10^4
x (Employment, α = 2.45, p <0.001)

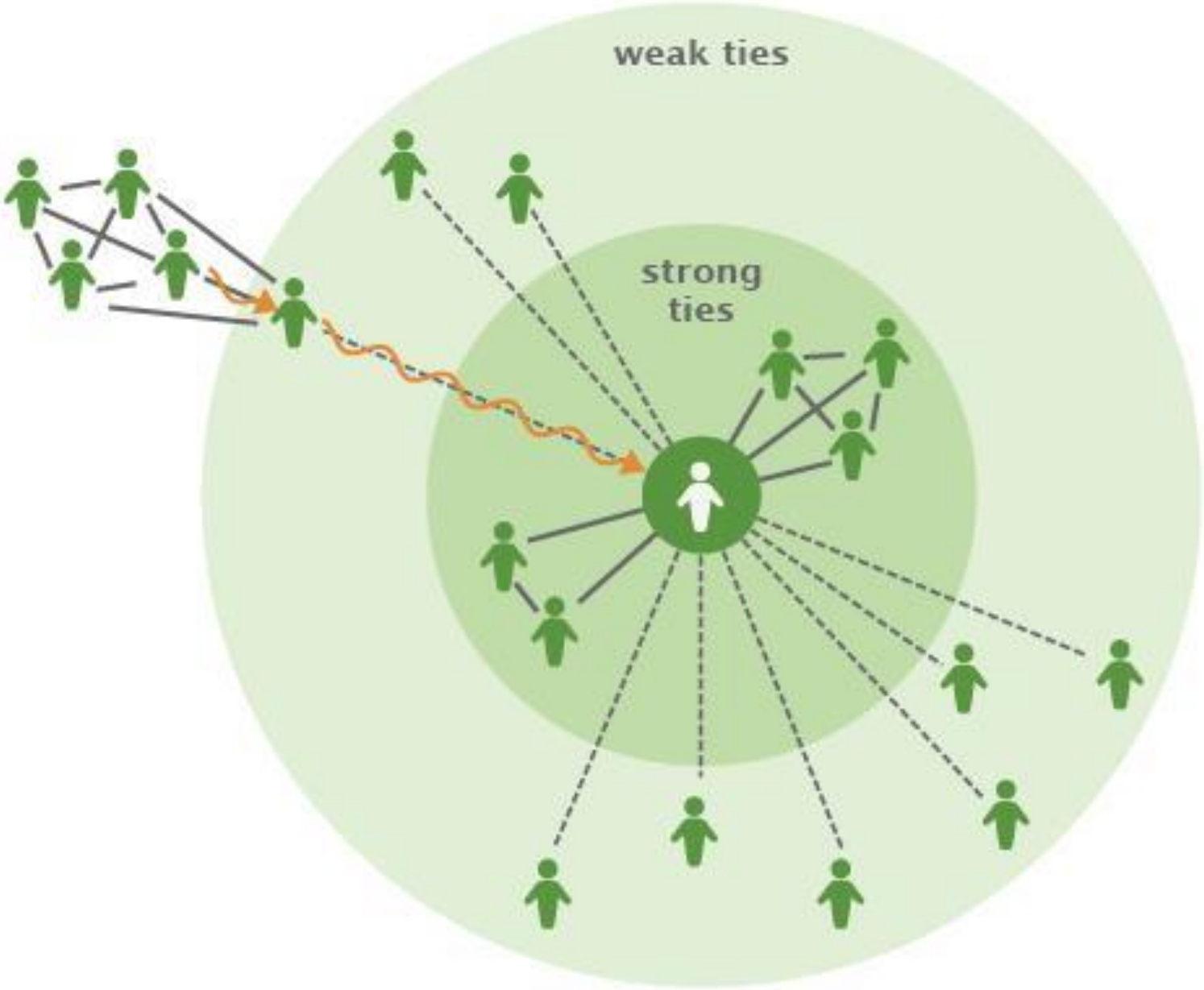

10^0
10^{-1}
10^{-2}
Pr(X ≥ x)
10^2
10^3
10^4
x (Residences, α = 2.22, p<0.001)

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”!)**


weak ties
strong ties


Q&A