



北京城市实验室  
Beijing City Lab

ID of the slides

30



## Slides of BCL

[www.beijingsitylab.com](http://www.beijingsitylab.com)

## How to cite

Author(s), Year, Title, Slides at Beijing City Lab, <http://www.beijingsitylab.com>

E.g. Long Y, 2014, Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest, Slides at Beijing City Lab, <http://www.beijingsitylab.com>



# Finding Public Transportation Community Structure based on Large-Scale Smart Card Records

Song Gao, University of California, Santa Barbara  
Ying Long, Beijing Institute of City Planning



# 1. Introduction

# Citizens as Sensors (Goodchild, 2007)



- Web 2.0 & Geo-tagging
- The increasing pervasiveness of location based services
  - GSM (Global System for Mobile Communications)
- Cell phone data
  - GPS (Global Positioning System)
- Taxi trajectories
  - SNS (Social Networking Services)
- Foursquare Check-ins, Twitter, Facebook, Sina Weibo
  - Wi-Fi (wireless fidelity)
- Rich in spatiotemporal information (Big Geo-data)
- Better for describing and understanding urban structure

# Smart card data (SCD) from public transportation



- SCD collected by automated fare collection systems
  - Bus
  - Subway
- Since 1990's the use of smart card has become significant (Blythe, 2004)
- Overwhelmingly adopted by Chinese cities
  - Over 100 cities in 2007
- With precise bus trip information (both boarding and getting off)
  - A spatial resolution of bus stop
  - A temporal resolution of second

# Existing researches using SCD



- Most researches advocate providing decision making for planning and design of public transportation systems (see Pelletier et al (2011) for a review.
- In South Korea, Joh and Hwang (2010) analyzed the trip trajectories using four million individuals' trips from the bus SCD, and correlated them with land use characteristics in the Seoul Metropolitan Area.
- In the U.K., Roth et al (2011 coauthored with Michael Batty@CASA) used real-time “Oyster” card database of individual person movements in the London subway to reveal the polycentric urban structure.
- Long & Thill (2013) combined SCD and household travel survey to analyze jobs-housing relationships in Beijing.

# This Work



- What is the ground-truth community structure of strongly connected traffic analysis zones (TAZs) by public transportation?
- Do commuting flows exhibit a different community structure pattern, in contrast to that from the overall trips?
- .....



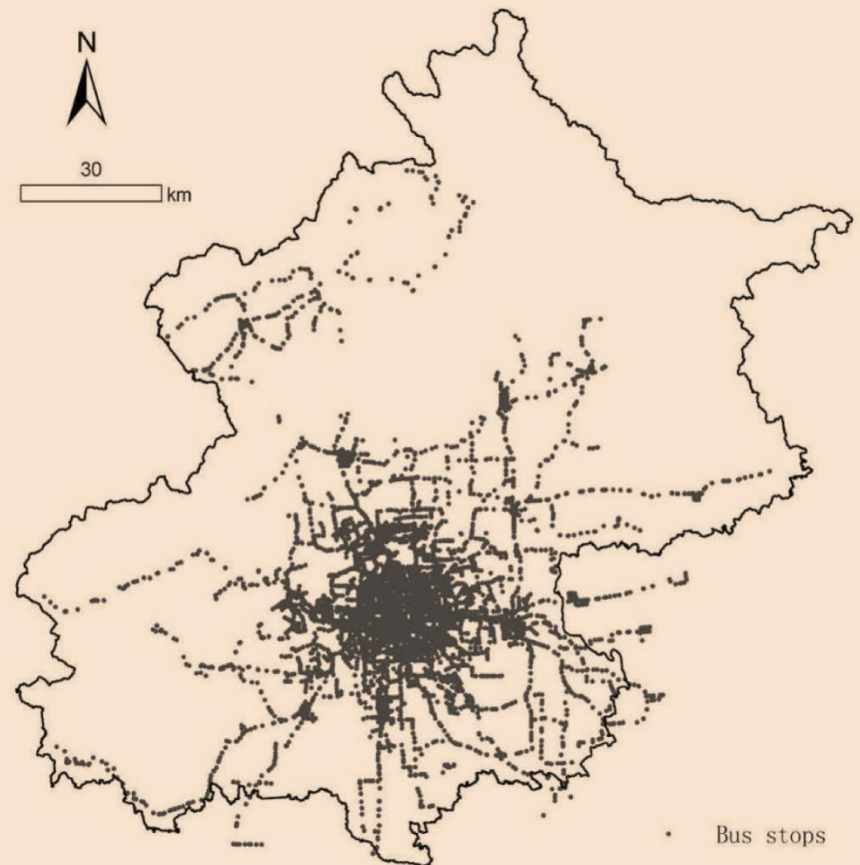
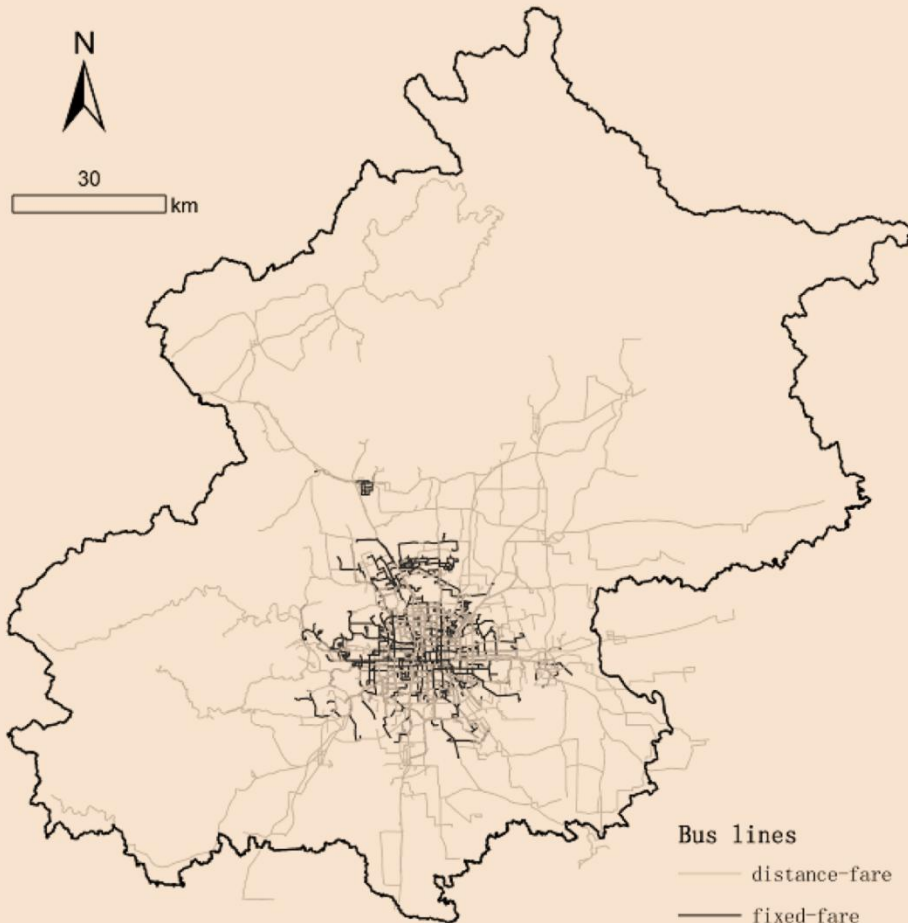
# **2. Public Transportation Data**



# Beijing Bus Systems

- 1287 bus lines

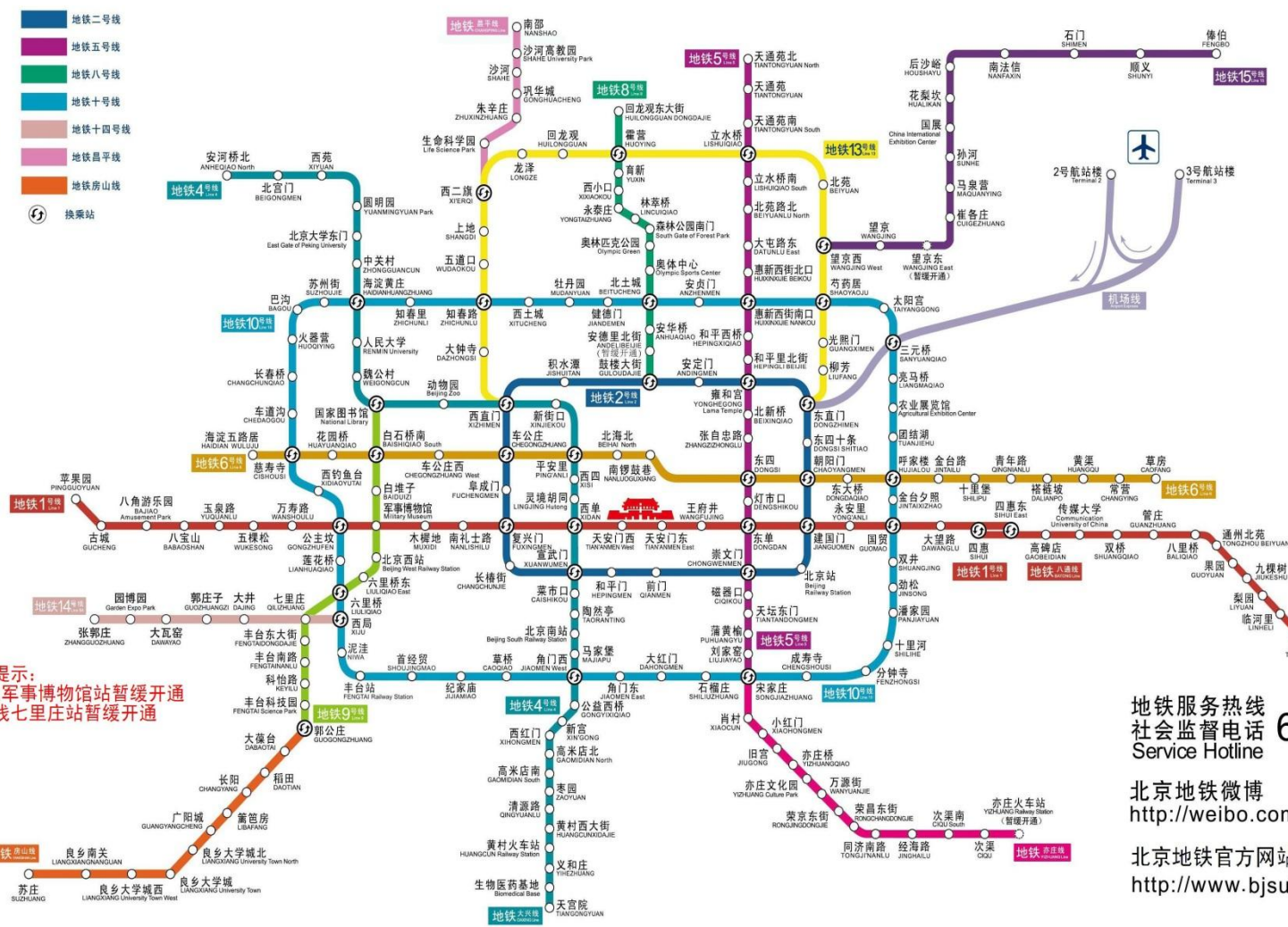
- 8691 bus stops



# 北京地铁线路图

## Beijing Subway Map

- 图例:**  
**Legend**
- 地铁一号线
  - 地铁二号线
  - 地铁八号线
  - 地铁五号线
  - 地铁四号线
  - 地铁六号线
  - 地铁九号线
  - 地铁十号线
  - 地铁十三号线
  - 地铁十四号线
  - 地铁十五号线
  - 地铁昌平线
  - 地铁亦庄线
  - 地铁房山线
  - 机场线
  - 换乘站



**特别提示:**  
9号线军事博物馆站暂缓开通  
14号线七里庄站暂缓开通

地铁服务热线  
社会监督电话 **68345678**  
Service Hotline

北京地铁微博  
<http://weibo.com/bjsubway0528>

北京地铁官方网站  
<http://www.bjsubway.com>

# Beijing Bus & Subway SCD



- 97.9 million trips from anonymized 10.9 million smart card users during a one-week period from April 5 to April 11, 2010.



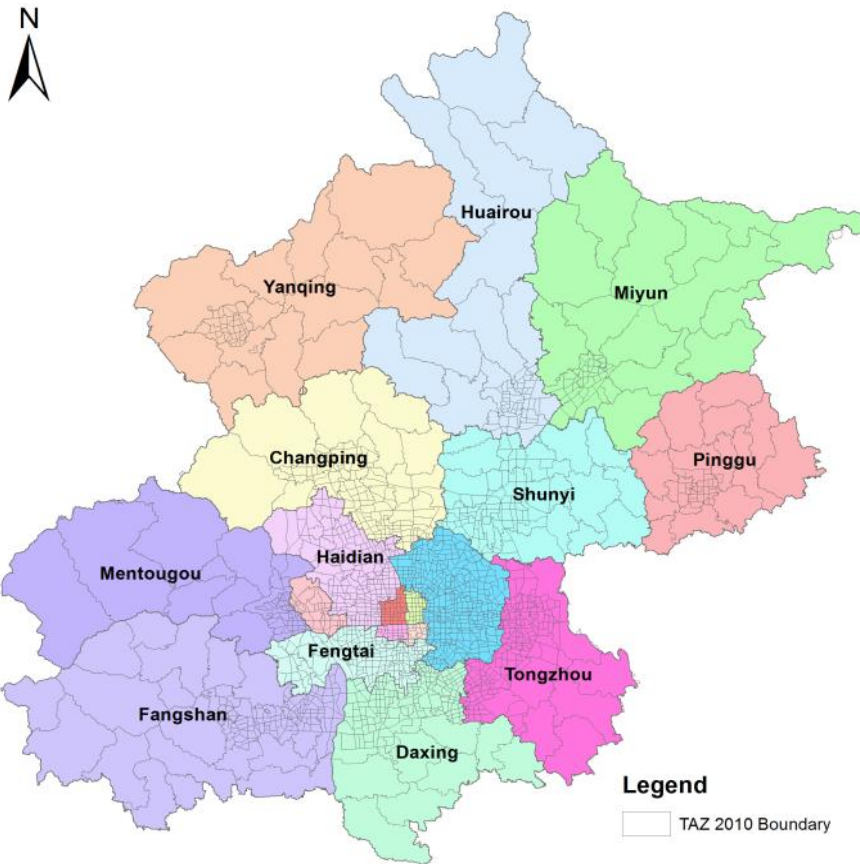
# Data Structure



Variable	Exemplified Values
Card ID	10007510038259911”
Card Type	2
Line ID	602, 40, 102
Line Type	0, 1
Driver ID	11032, 332
Vehicle ID	111223, 89763
Departure Date	2010-04-08
Departure Time	“06-22-30”, “11-12-09”
Departure Stop	11, 5, 14
Arrival Time	“09-52-05”, “19-07-20”
Arrival Stop	3, 14, 9

# Traffic analysis zones

- 1911 TAZs
- both central regions and suburbs
- OD trips extraction



Inner Districts of Beijing

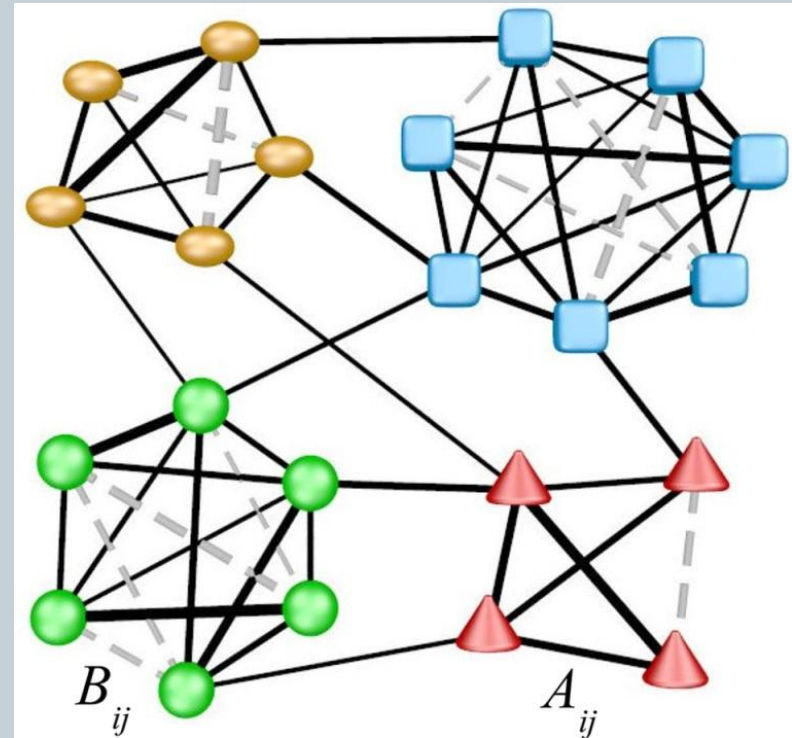


# 3. Community Detection

# Community Detection Algorithm

The nodes of the network can be grouped into sets of nodes so that each community is densely connected internally.

- ❖ **Modularity maximization**
- ❖ **Minimum-cut method**
- ❖ **Hierarchical clustering**
- ❖ **Girvan–Newman algorithm**
- ❖ **Clique based methods**



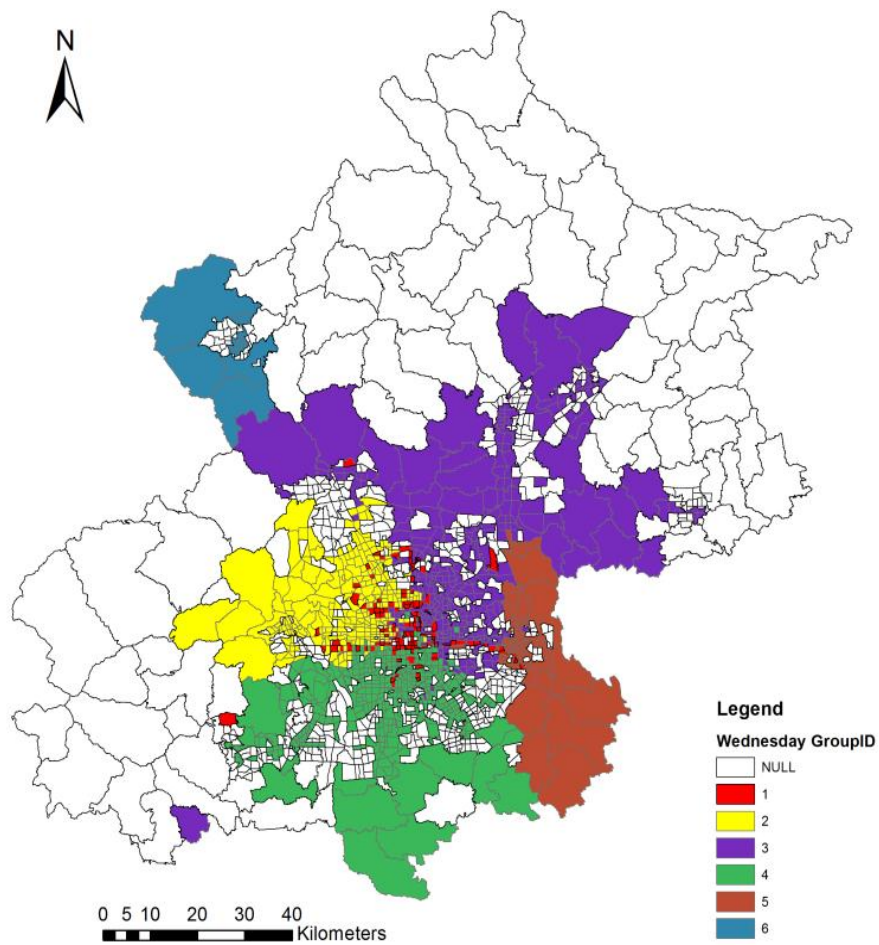
# Modularity Maximization Algorithm



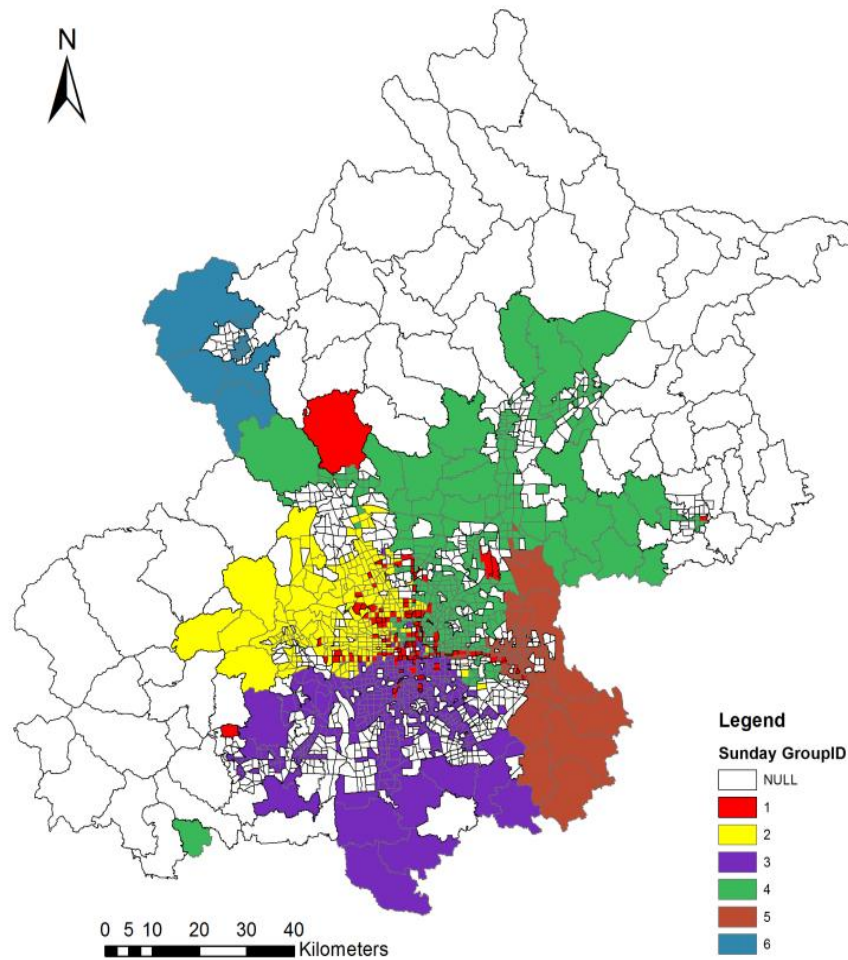
- **Modularity** is defined as the sum of differences between the fraction of edges falling within communities and the expected value of the same quantity under the random null model. (Newman and Girvan, 2004)
- A bottom-up fast greedy algorithm (Clauset et al. 2004, Gao et al. 2013) was adopted for searching an optimized graph partition that maximizes the modularity measure.



# Weekday

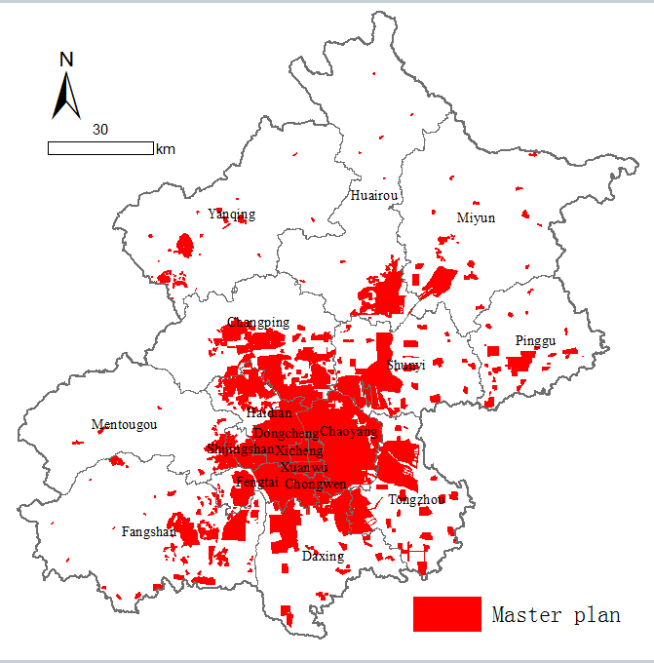
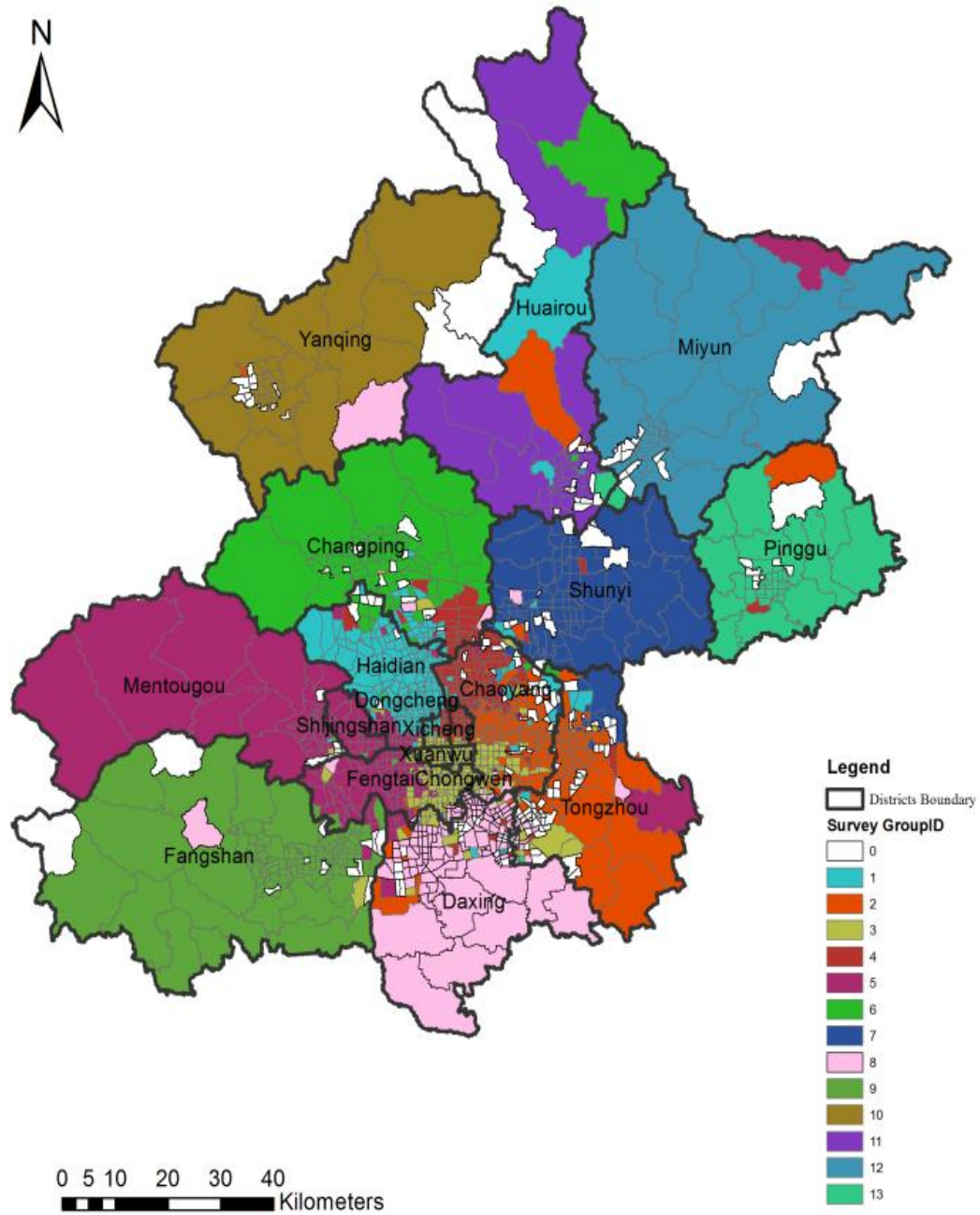


# Weekend



# Compared with Survey Data

The SCD community detection results actually match better with the Beijing planned urban development area boundary.



Master plan

# Available Online to Public

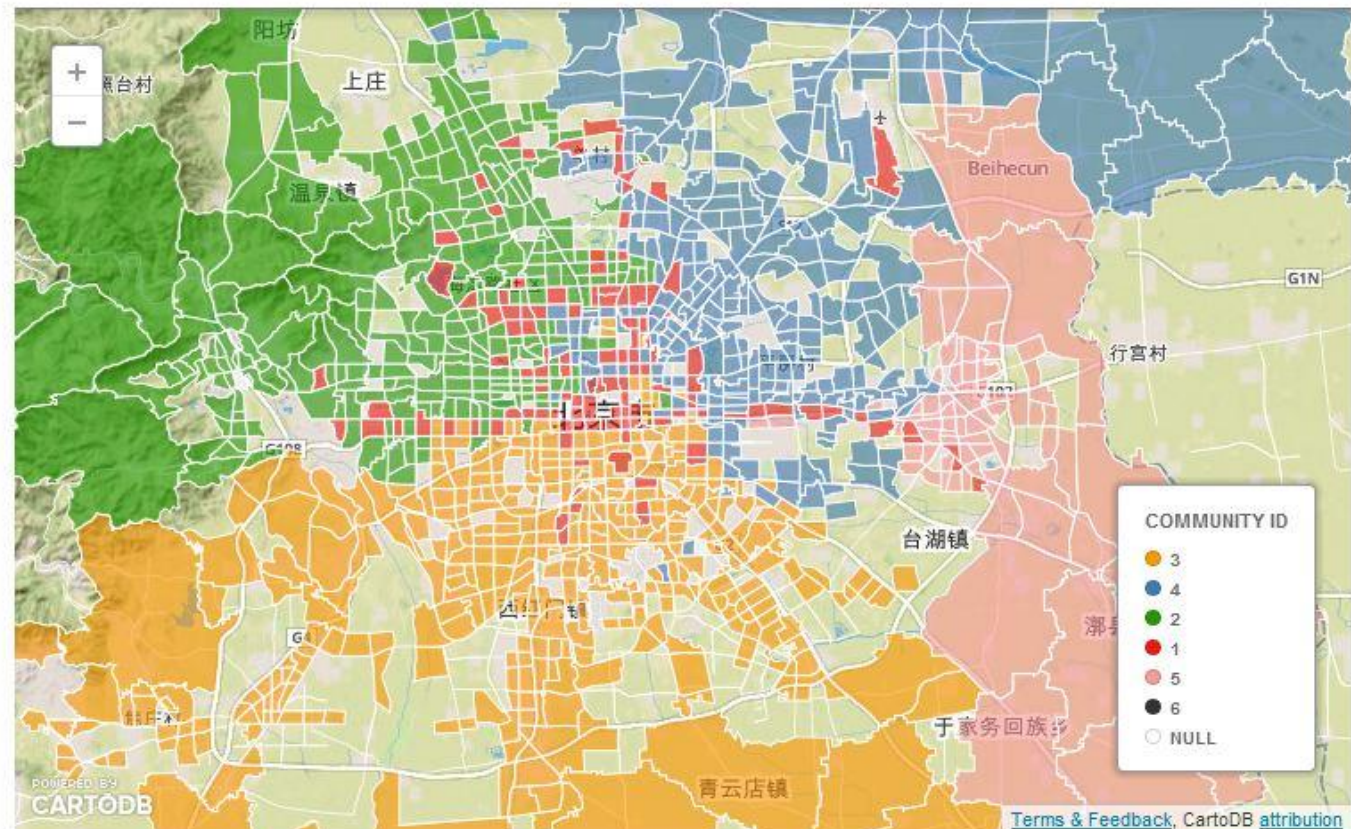


for urban planning and governance, and ultimately producing the science of cities required for sustainable urban development. The lab's current mix of planners, architects, geographers, economists, and policy analysts lends unique research strength.

## Communities of public transportation in Beijing

SHARE

Communities detected by the bus/metro smartcard records of Beijing in 2010 (Wed, a typical weekday). By Song Gao and Ying Long. longying1980@gmail.com



**BCL**  
Beijing City Lab

Home

Models

Members

Working papers

Slides

Data released

Blogs

Publications

Links

Contact

About



# 4. Commuting Patterns

# Identifying House Places



- The departure bus stop of the first trip (TRIP<sub>1</sub>) will be the housing place of a cardholder.
- Grounded on: – In the 2005 household survey, 99.5% person's first trips start from home
- Note that a person is assumed to live around the
- identified stop with a maximum distance of 750m
  - Retrieved from walking trips in the 2005 survey
  - This could be the spatial resolution of our identification results

# Identifying Working Places



- An activity of a non-student (except at home) with a duration longer than 6 hours is assumed to be working.
- Grounded on:
  - In the 2005 survey, 96% persons work over 6 hours per day

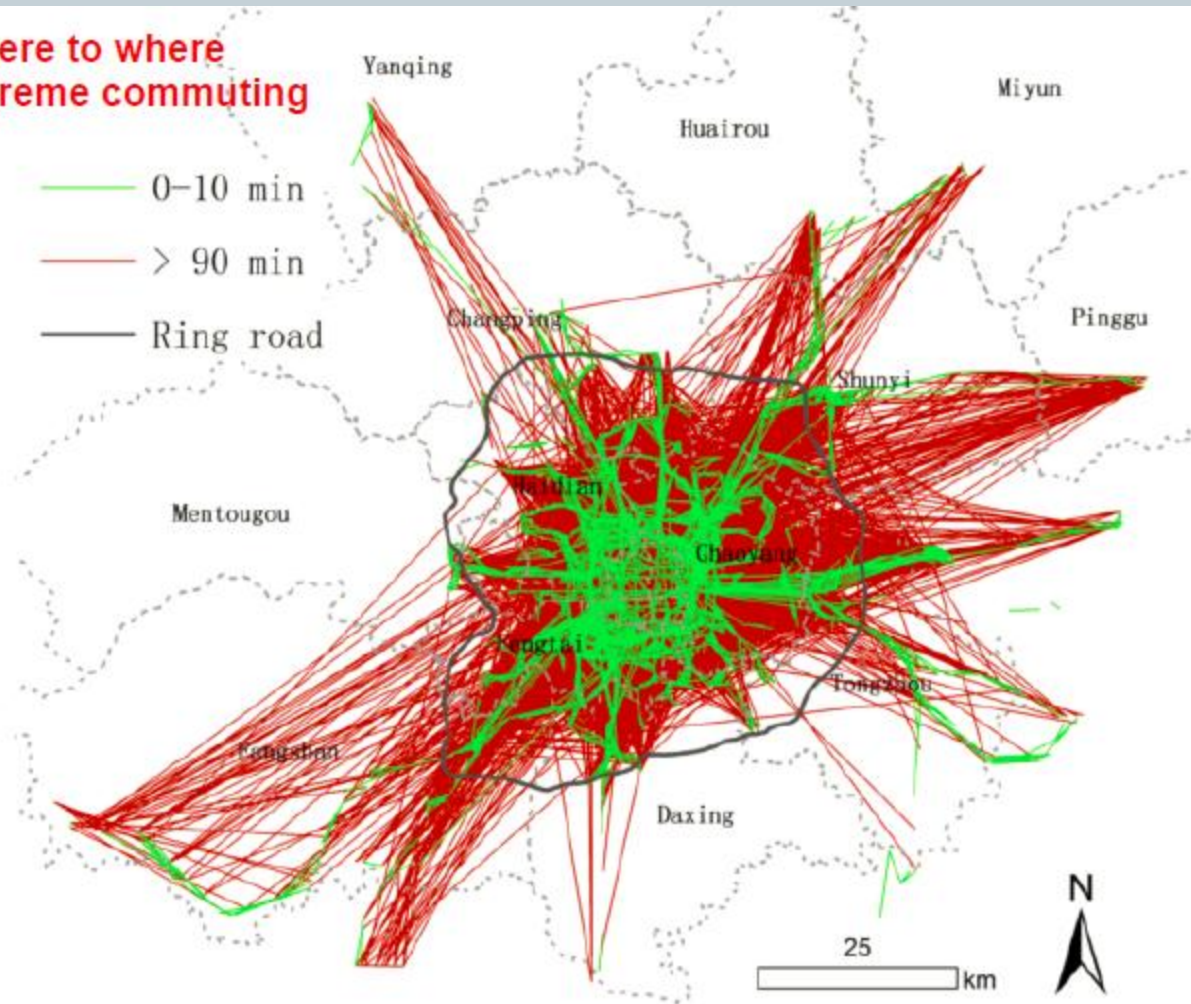
# Identifying Commuting Time

- Where to where
- Extreme commuting

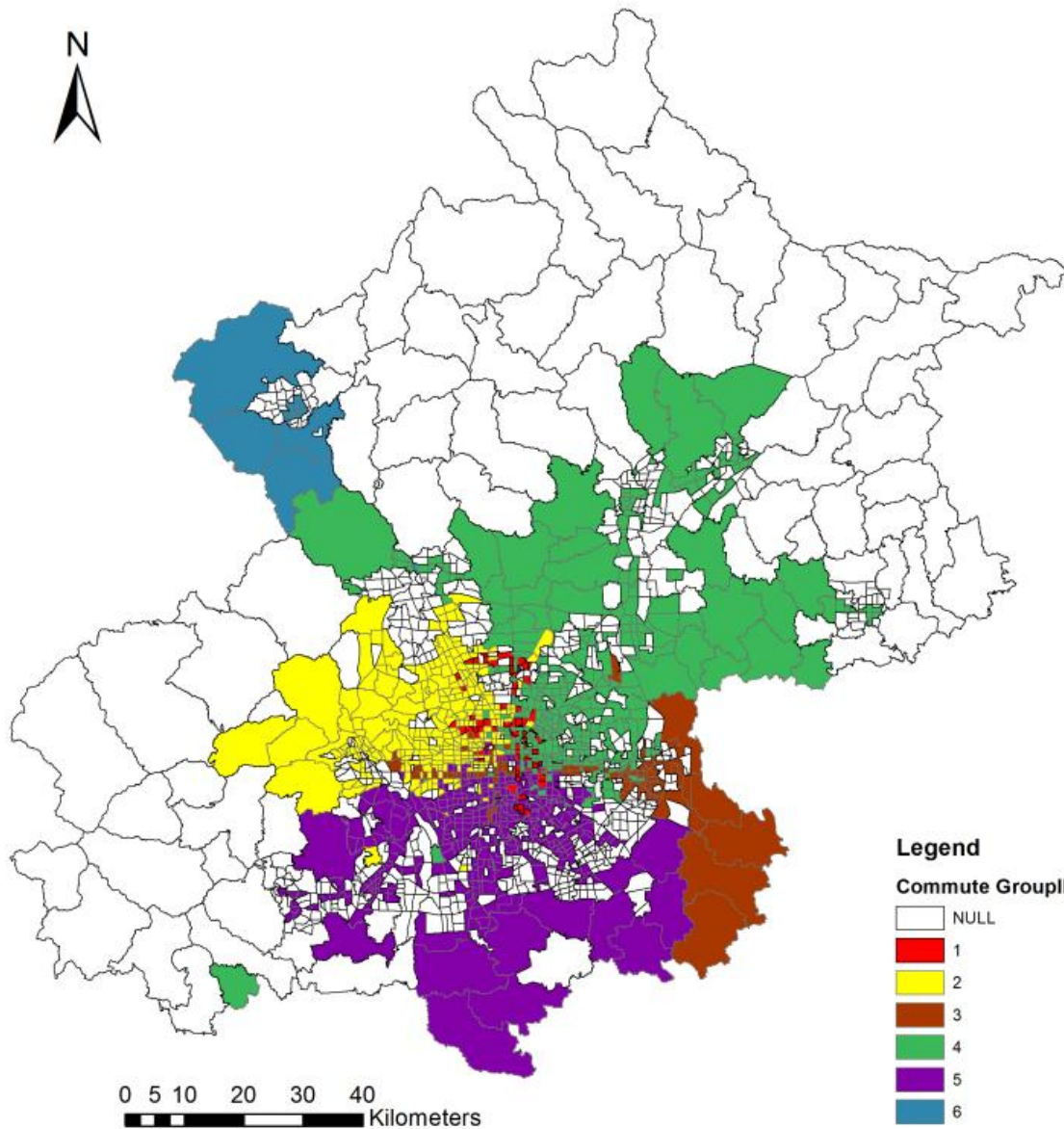
— 0–10 min

— > 90 min

— Ring road



# Commuting Community



- There exist strong public transit connections through the whole day in TAZs that are located along the central west-to-east and north-to-south corridor including the *Chang'an Avenue* in Beijing
- The residential community 'Tiantong Yuan' in Southeast has strong commuting connections to the central regions.



# Conclusions



- First, this research applies a **network-analysis approach** to investigate the ground-truth community structure of strongly connected TAZs via public transportation, which yields insights urban structure in Beijing from the **public transportation functional zone** perspective.
- Second, the daily community detection results using SCD are different from that using **household travel surveys** and the **SCD community** boundary matches better with Beijing urban planned area than the household travel survey.
- In the future, the **map matching** of these OD trips to actual streets and further analysis could be beneficial for the reliability analysis of street networks and emergency transportation management for future strategic planning.

# Thanks for your attention!



<http://www.geog.ucsb.edu/~sgao/>