

基于公交卡大数据的 城市空间结构分析

龙瀛，博士
清华大学建筑学院
北京城市实验室



北京城市实验室
Beijing City Lab

<http://www.beijingcitylab.com>



新浪微博：龙瀛a1_b2 北京城市实验室BCL
微信公众号：beijingcitylab

Smart card data from public transportation

智能交通卡、一卡通、八达通...

- Smart card data 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



Bus SCD in Beijing



- Anonymous card
- Over 90% bus riders use smart cards till April 2007

- One week in April 2008 from
 - Monday to Sunday
 - Subway not included
- 80 million records for **8.5 million cardholders**
 - One record for a bus riding of a cardholder
 - 1.3 daily bus ridings per person
- One week bus/metro SCD in 2010, 2013 and 2014
 - 100 m records for 10 m cardholders
- The largest scale in the world
 - Singapore, London, Seoul

SCD data structure

Variable	Exemplified Values
Card ID	“10007510038259911”, “10007510150830716”
Card Type	1, 2, 3, 4
Line ID	602, 40, 102
Line Type	0, 1
Driver ID	11032, 332
Vehicle ID	111223, 89763
Departure Data	2008-04-08
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

Bus lines



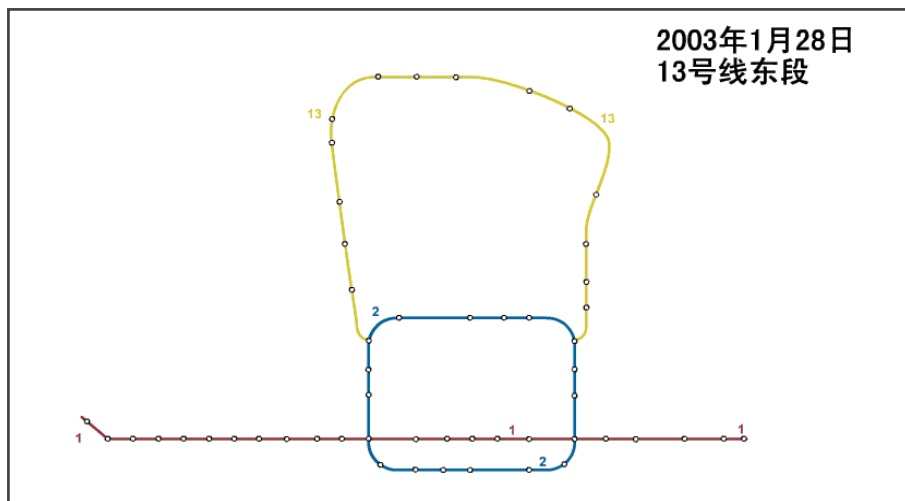
- **The Beijing Metropolitan Area**
 - 16410 sqkm
- **1287 bus lines**
- **Two types**
 - Distance-fare (long dist, inner city to suburb)
 - Fixed-fare (short dist, within inner city)

Bus stops



- 8691 bus stops
- The spatial resolution of identification
- The average distance between two stops
 - 231 m

Subway



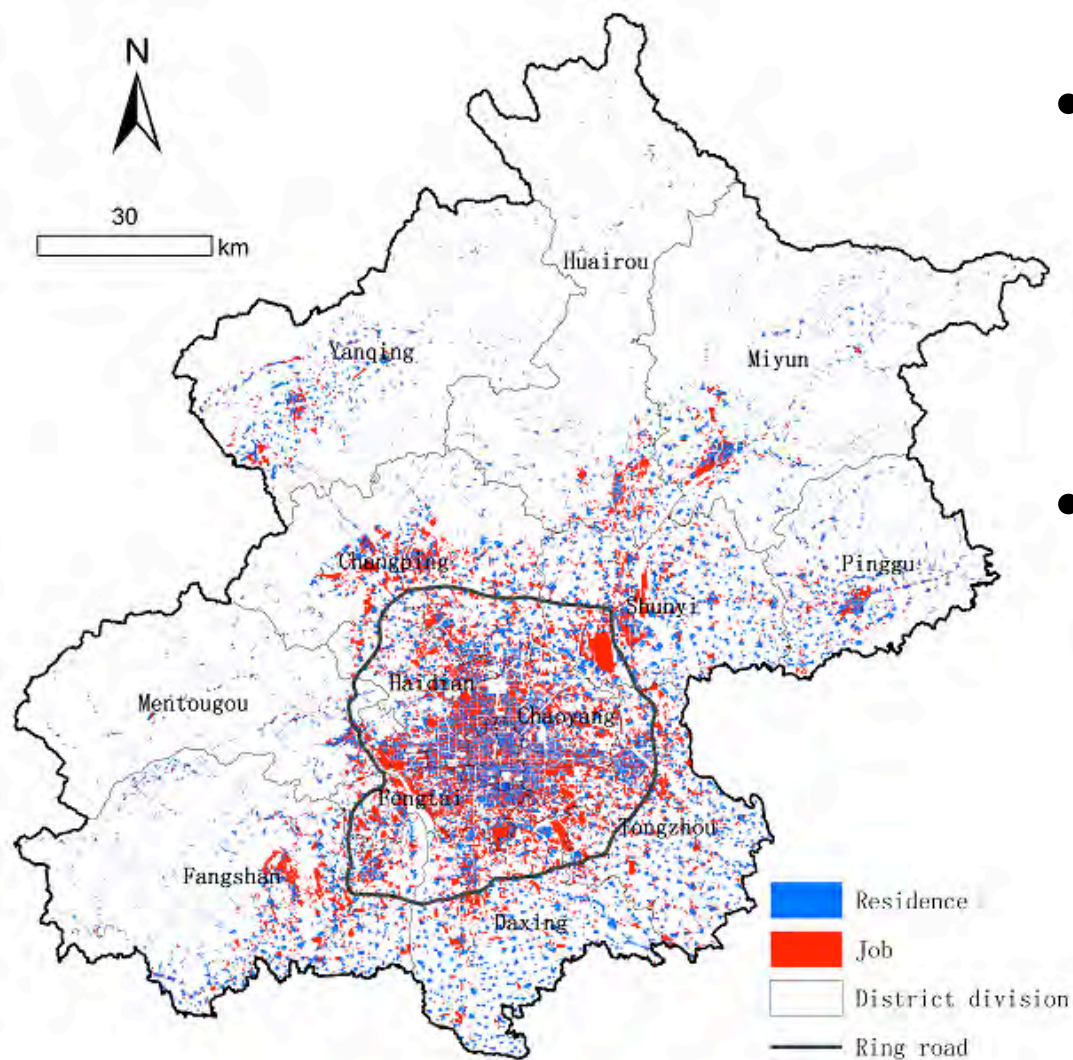
- 1969年10月1日全长23.6公里的北京地铁1号线一期工程（苹果园-北京站）建成通车，到2003年12月27日地铁八通线的开通，北京地铁运营里程突破100公里用了34年
- 截至2014年1月，北京地铁共有17条运营线路。它包含16条地铁线路、1条机场轨道，组成覆盖北京市11个市辖区，拥有273座运营车站、总长465千米运营线路的轨道交通系统。
- 2013年1—11月，北京地铁全网日均客运量达876万人次，同比增长30.5%，最高单日客运量达1106万人次，在公共交通客运总量中所占比重接近40%

Traffic analysis zones



- **1118 TAZs**
- **For aggregating identification results**
 - Originally in the bus stop level

Land use pattern



- **133503 parcels in the BMA**
 - 29112 residential parcels
 - 57285 parcels with job positions
 - Job parcel
- **For calculating residential or job potential of each stop**
 - In case of periodic identification results using one-week data

The Beijing Household Travel Survey

(The 2005 survey)

- Conducted in 2005 for the whole BMA
- 81,760 households/208,290 persons, with 1.36% sampling rate
 - 800 thousand trips
- Information included:
 - Socio-economic attributes of household and persons
 - One-day travel diary of each person
 - Mode, purpose (e.g. **commute**), OD, departure and arrival time, etc.
- For setting rules for identifying jobs-housing places and commuting trips

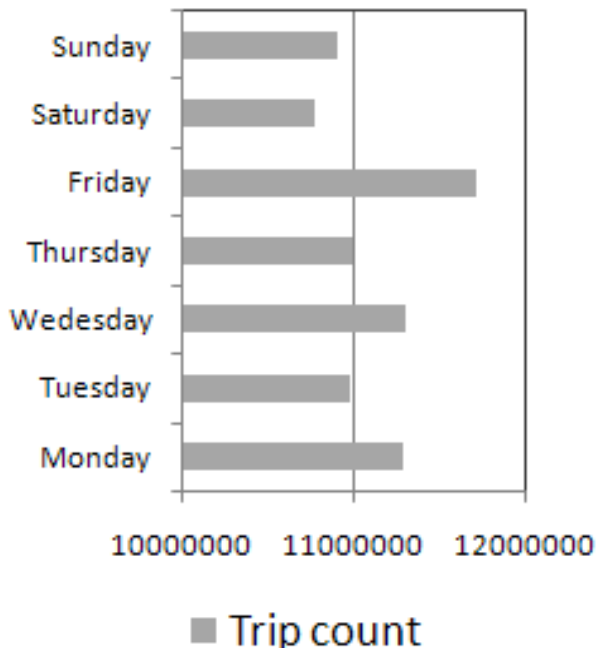
数据处理和可视化

Data preprocessing

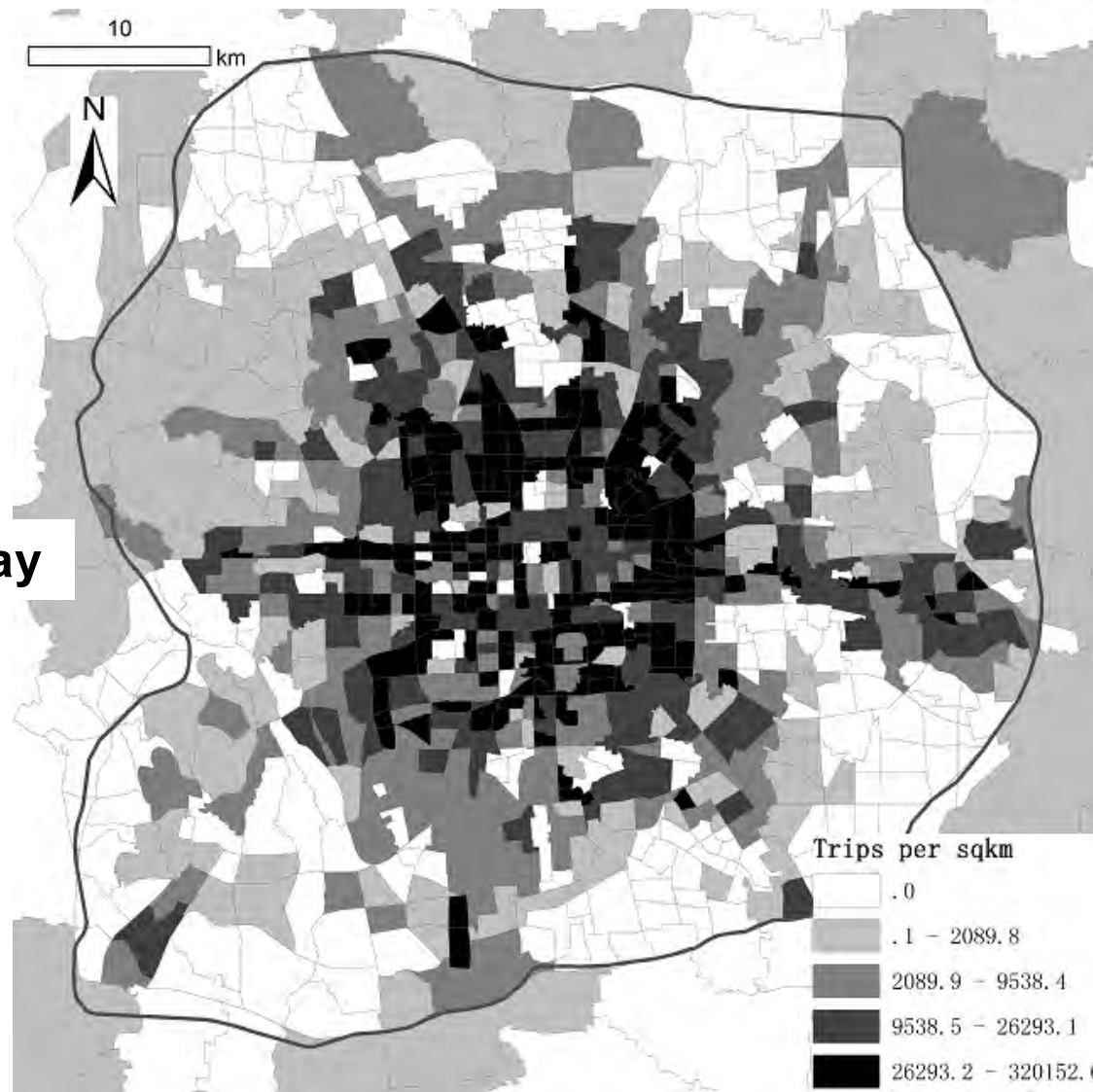
- **Data cleaning**
 - Abnormal records due to inappropriate operation on fare machines
- **Merge adjacent trips for each cardholder**
 - Converting segments into real trips
- **Geocode the SCD**
 - by linking the bus stop ID in the SCD with the bus stop layer.
- **Combine trips of each cardholder**
 - to retrieve cardholder's full bus travel diary
- **Two data models proposed for identifying commuting pattern**
 - **PTD: Position-start Time-Duration**
 - For identifying housing-jobs places
 - **{H0, 18:00 (-1), 13 h} and {J0, 8:00, 9h}**
 - **TRIP: departure / arrival time and space**
 - For identifying commuting trips
 - **{H0, 7:00, J0, 8:00} and {J0, 17:00, H0, 18:00}**

Special attention paid to fixed-fare trips

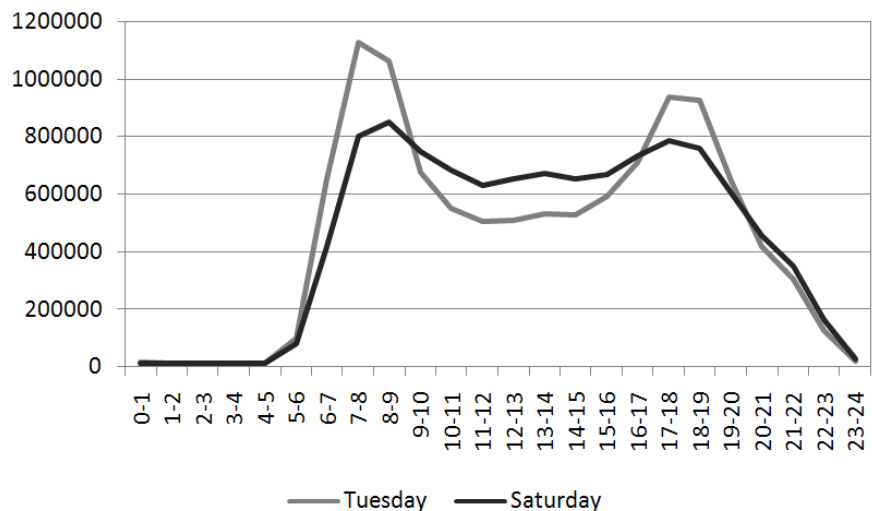
- **Distance-fare**
 - With full spatiotemporal information
- **Fixed-fare**
 - Only departure time



Trip count (bus riding) for each day



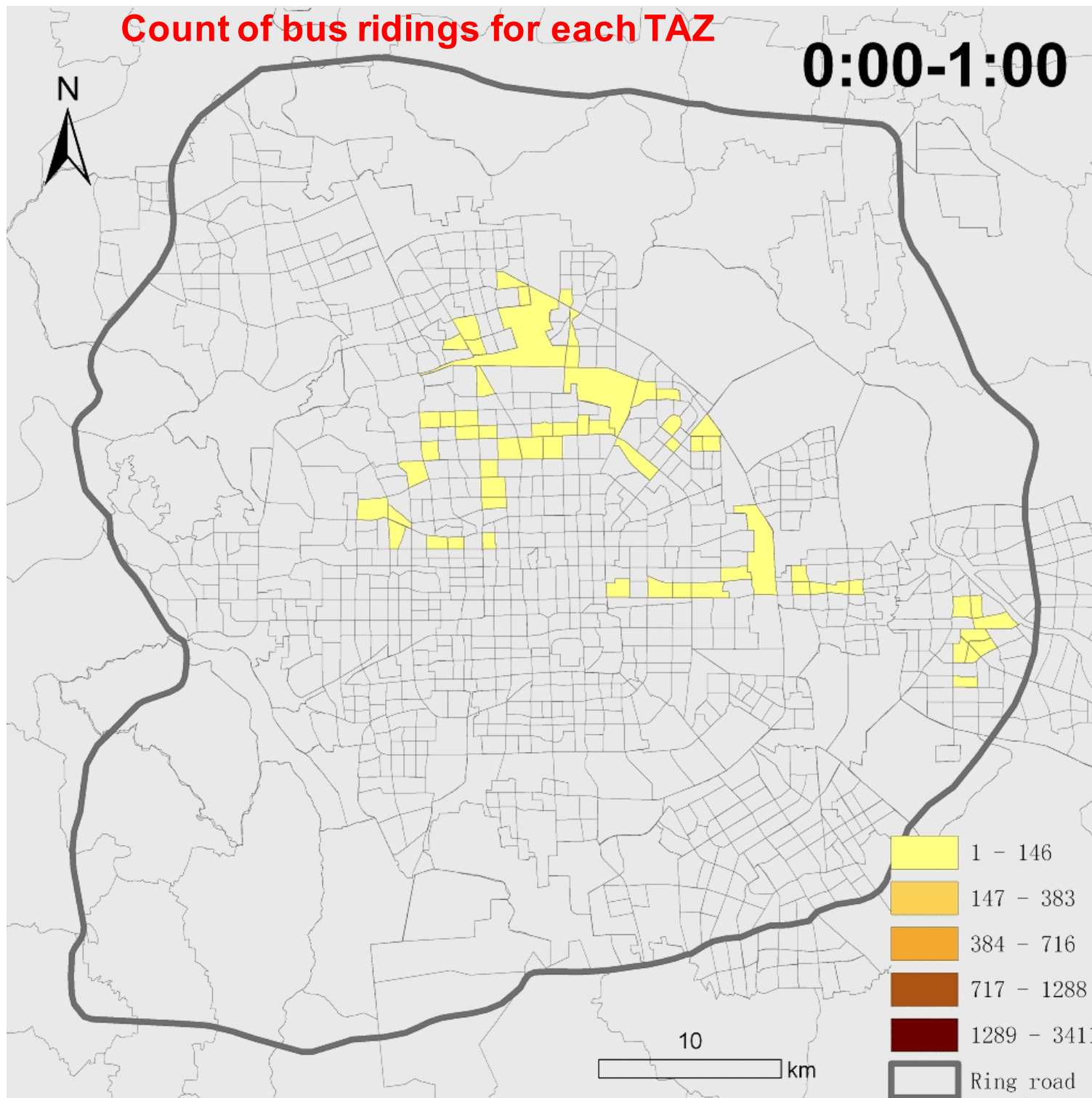
Trip density in the TAZ level in the whole week



Trip count in each hour on Tue and Sat

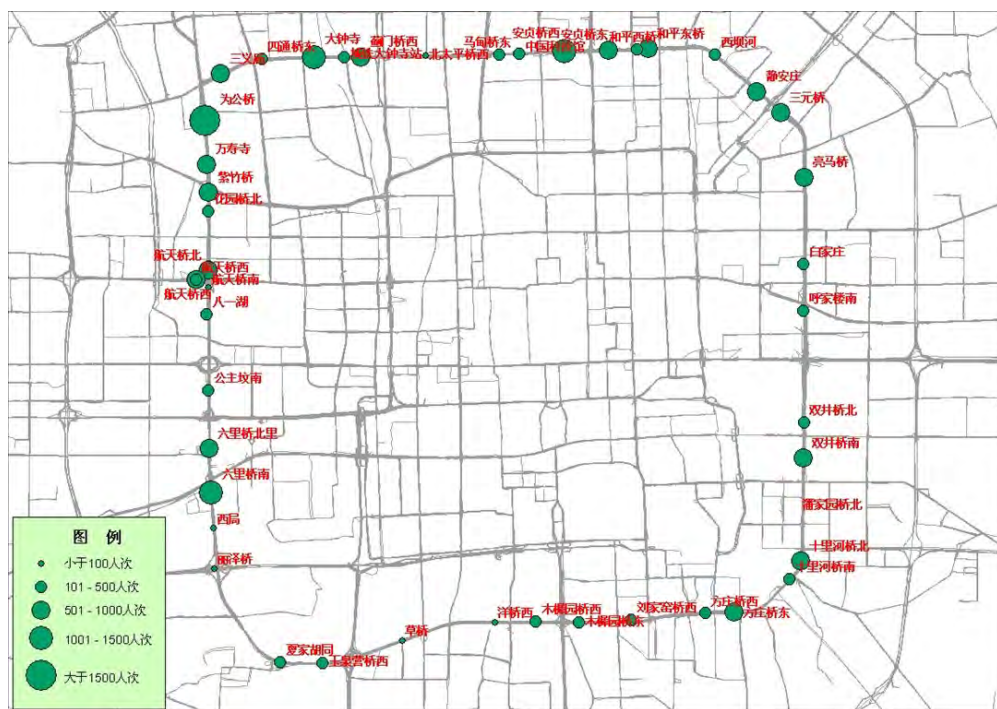
Count of bus ridings for each TAZ

0:00-1:00

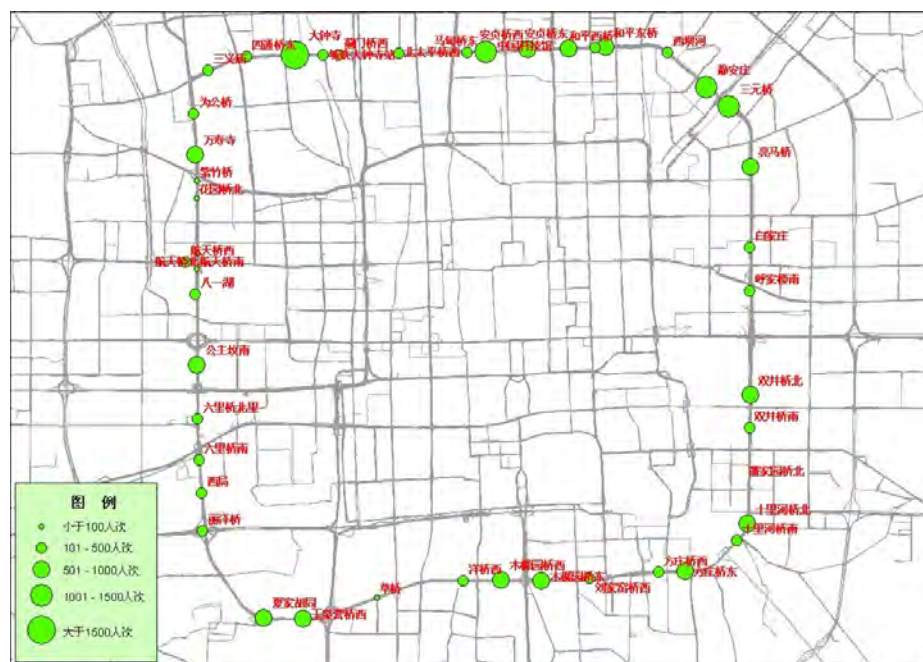


单线路客流分析

早高峰时段7点至9点上下车人次约2.4万人次，占全日上车人次的23.3%。



特8内早高峰时段各站上车人次



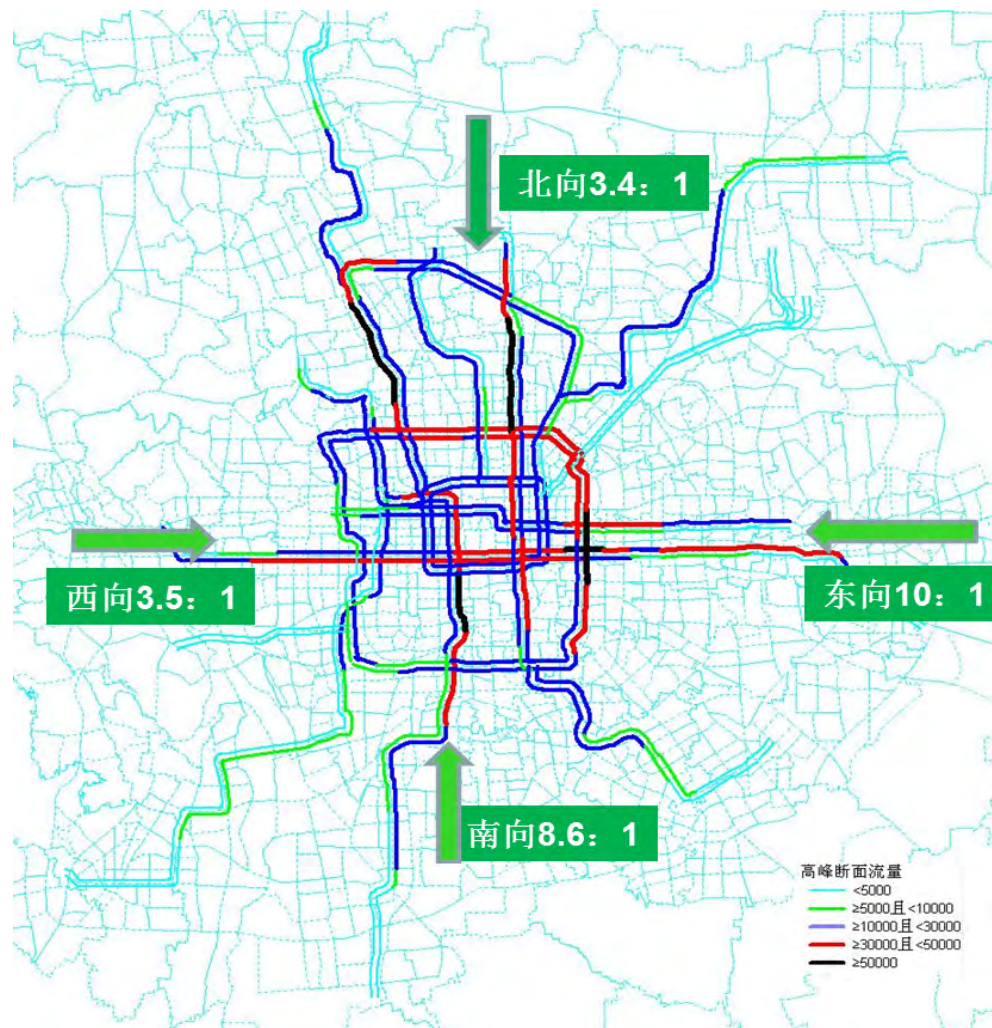
特8内早高峰时段各站下车人次

2013轨道交通早高峰客流状态

客流分布

由于近年来产业的集聚和居住的扩散及公共服务资源配置的不均衡，轨道交通客流呈现较明显的潮汐性。

早高峰时段，进程方向高峰小时断面流量压力明显，轨道交通进出四环的断面流量比平均为**4.4:1**。



2013年全网早高峰小时轨道交通客流分布

地理学报(首篇)、Computers, Environment and Urban Systems

通勤出行

Identification of housing places using one-day data

- The departure bus stop of the first trip (TRIP1) will be the housing place of a cardholder.
- Grounded on:
 - In the 2005 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 750 m
 - Retrieved from walking trips in the 2005 survey
 - This could be the spatial resolution of our identification results

Identification of job places using one-day data

- An activity of a non-student (except at home) with a duration longer than 6 hours is assumed to be working.
 - *If:*
 - *Condition 1:* The card type is not the student card
 - *Condition 2:* $D_k > 360$
 - *Condition 3:* $k \neq 1$
 - he/she is sleeping at home when $k=1$
 - *Then:*
 - the k_{th} place P_k will be regarded as the job place of this cardholder
- Grounded on:
 - In the 2005 survey, **96%** persons work over 6 hours per day.

Combining each day's result to get more solid housing-jobs places

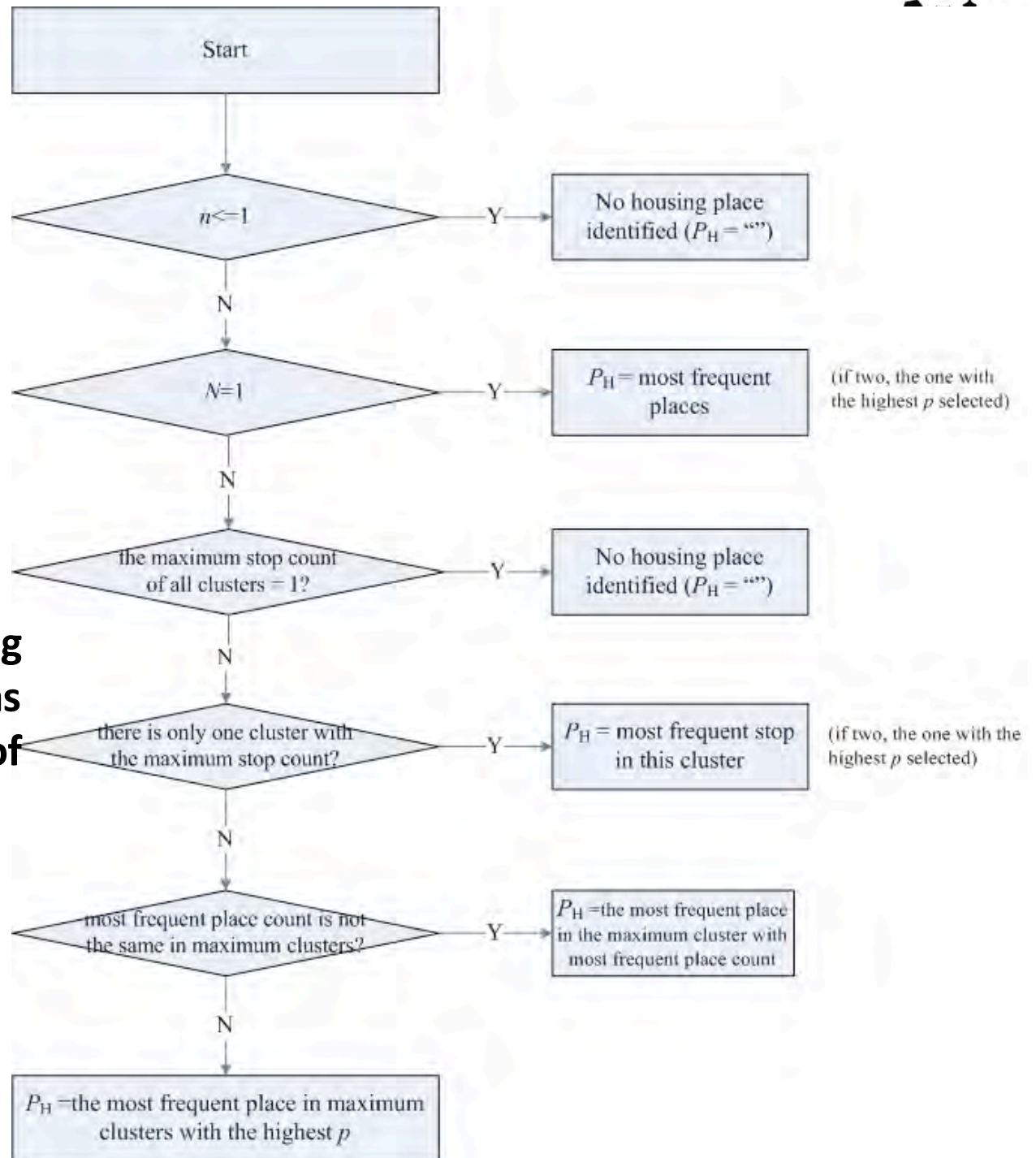
- One day result might not be stable, e.g.
 - A shopping, hospital, or creation activity longer than 6 h
 - One night sleeping at a friend's home
- Periodic pattern of cardholders with identified housing-jobs results like
 - **M T W T F S S**
 - **a a a b a a a (Person 1)**
 - **a a a b b b c (Person 2)**
 - **a a b b c c d (Person 3)**

- A decision tree for combining seven daily result

- A complex procedure

- For **aaabbbc** pattern:

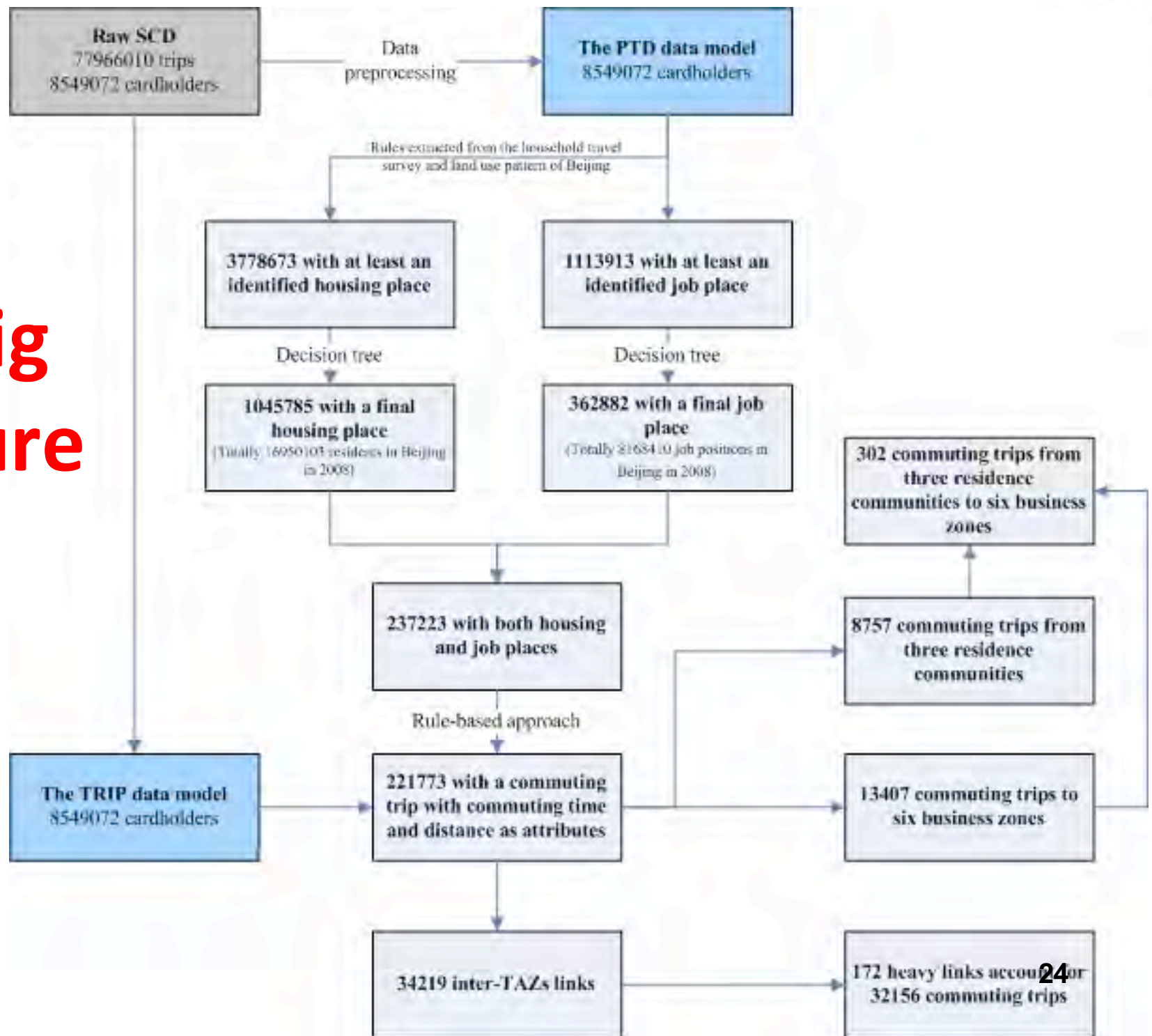
- Housing or job potential of each stop
 - “a” with a higher housing potential was selected as the final housing place of Person 2



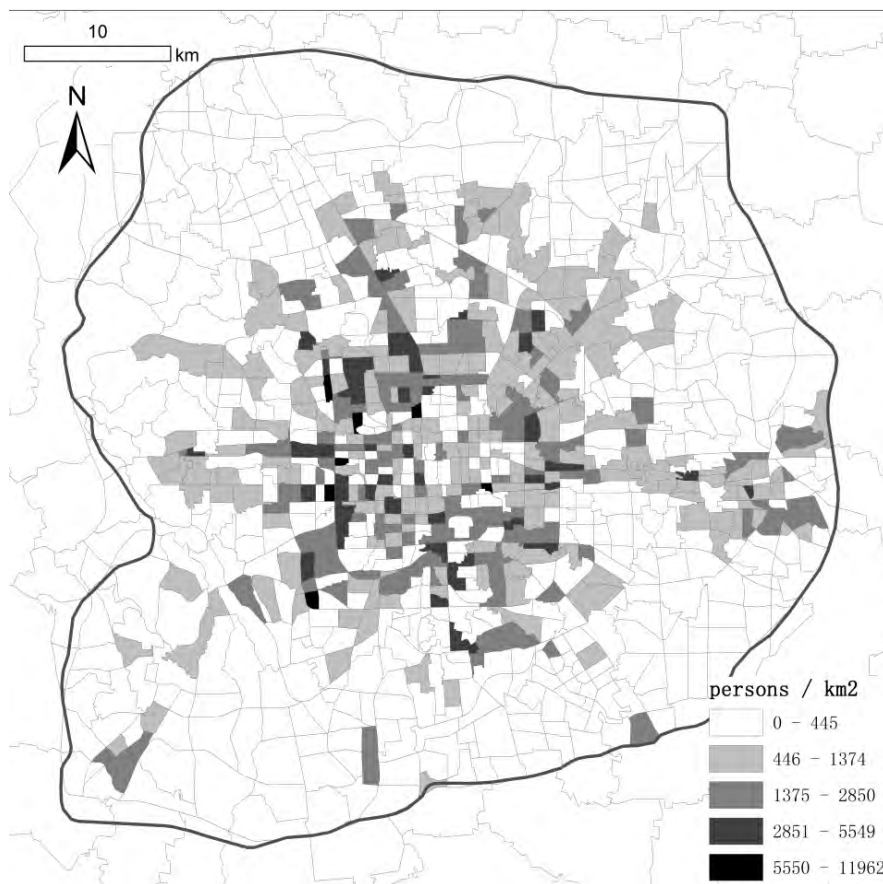
Identification of commuting trips

- Limited to those cardholders with both identified housing and job places.
- Rules used:
 - The boarding bus stop of the first trip in a day is the identified housing place.
 - The job place appears in trips of a day.
 - Both the housing and job places are identified in the same day.
- Commuting time and distance are calculated for each identified commuting trip.

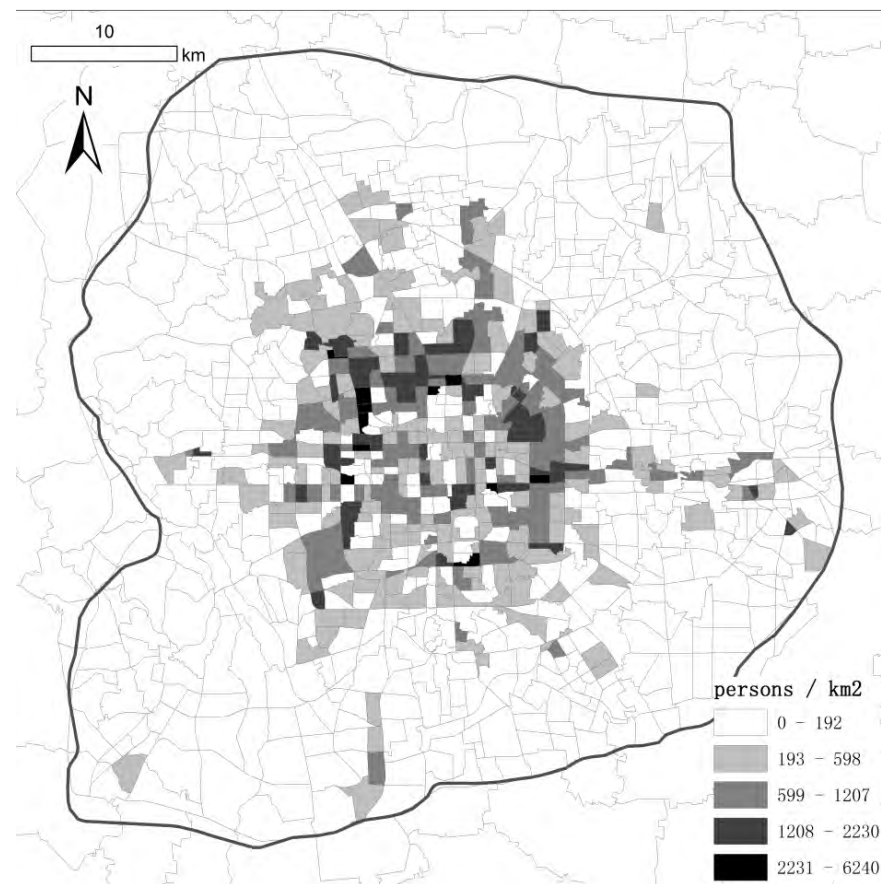
A big picture



1,045,785 cardholders (12.2% of all 8549072 cardholders) are identified with housing places, and 362,882 cardholders (4.2% of all) are identified with job places.



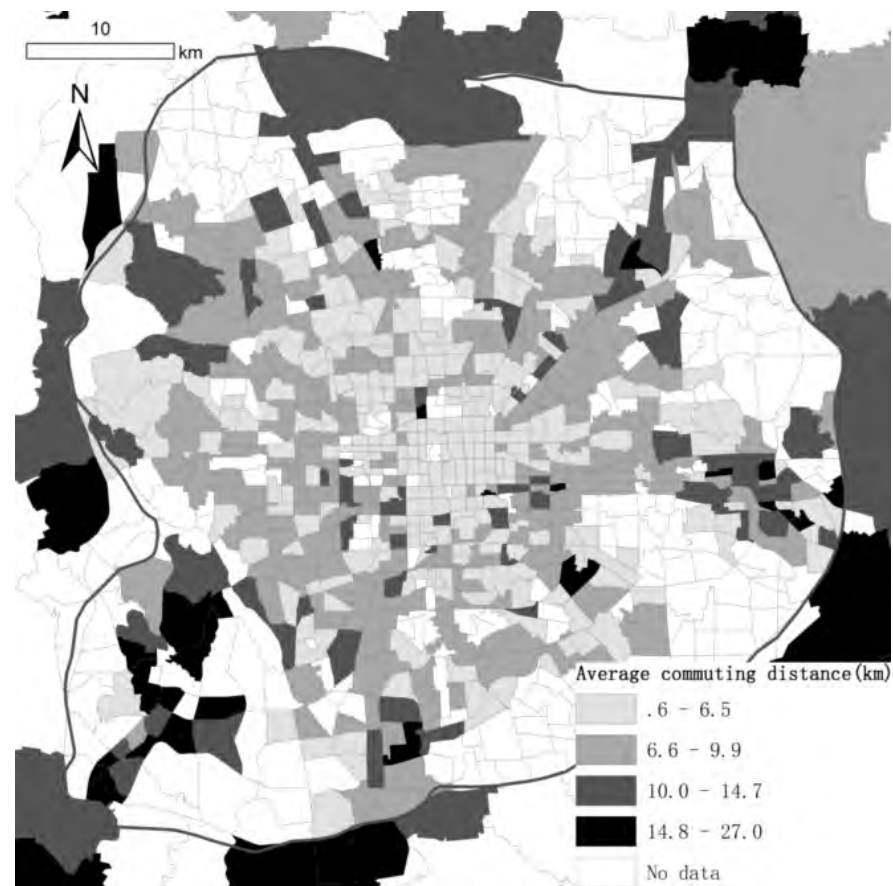
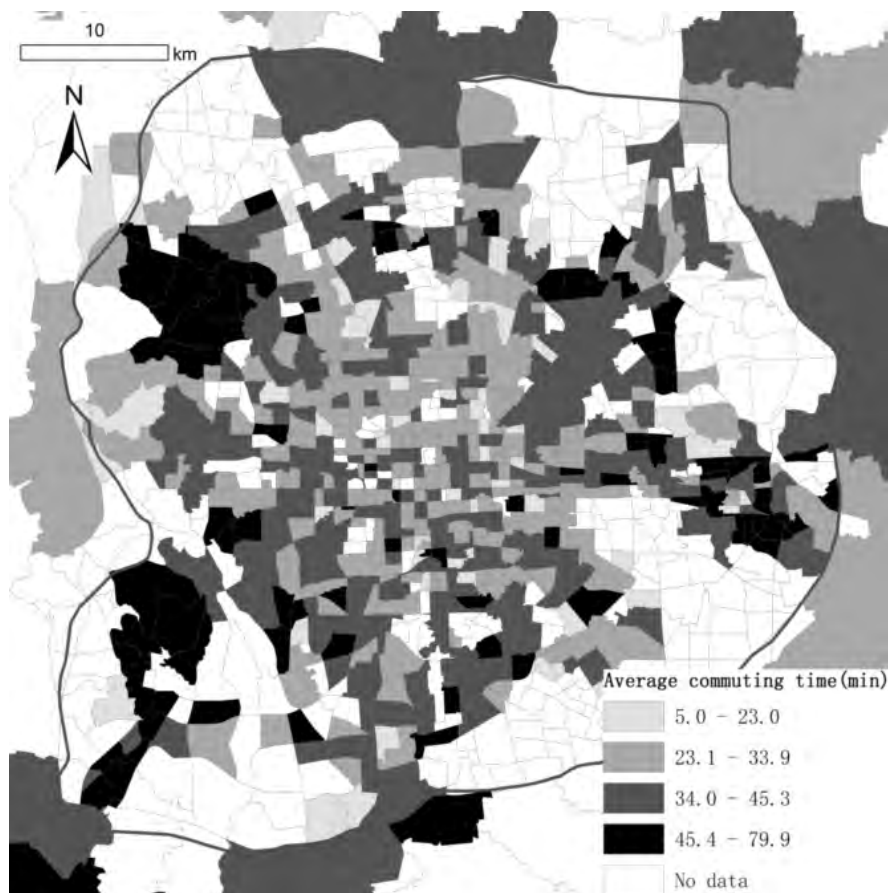
Cardholder density with an identified **housing** place



Cardholder density with an identified a **job** place

Commuting trips identification

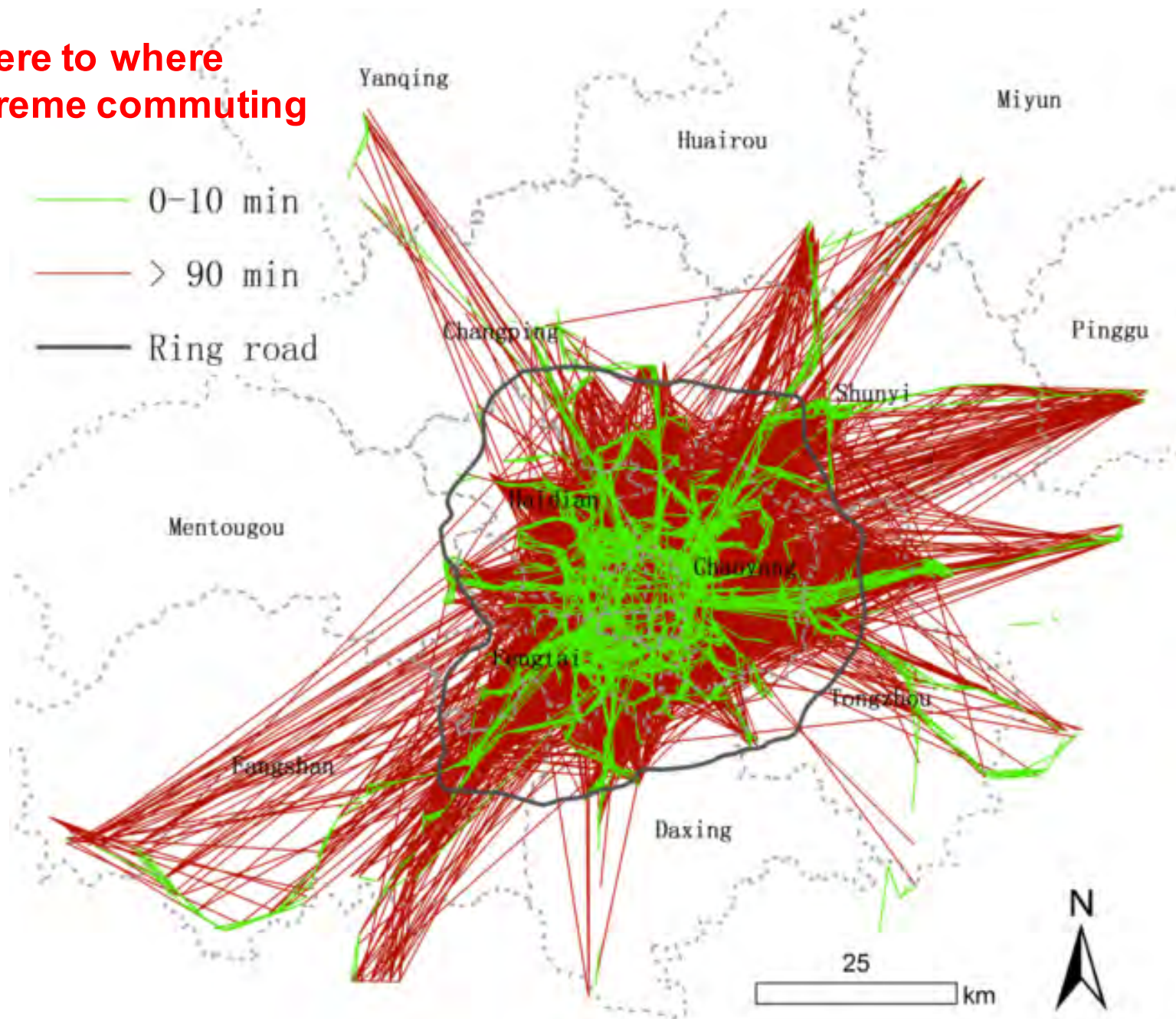
221,773 cardholders identified with commuting trips



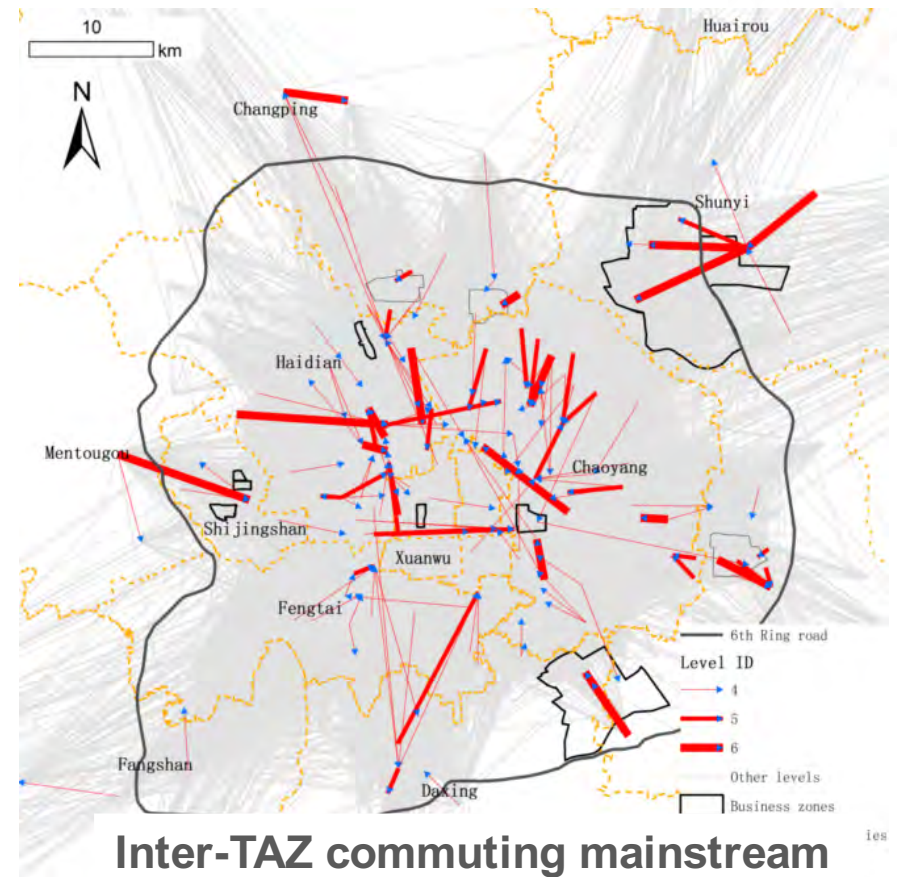
Average commuting **time** and **distance** for each TAZ 26

Bus commuting pattern

- Where to where
- Extreme commuting



- Aggregated at the TAZ level
- The head/tail division rule used
- Policy suggestion on BRTs



2013 年9月开通以来

开通45个方向

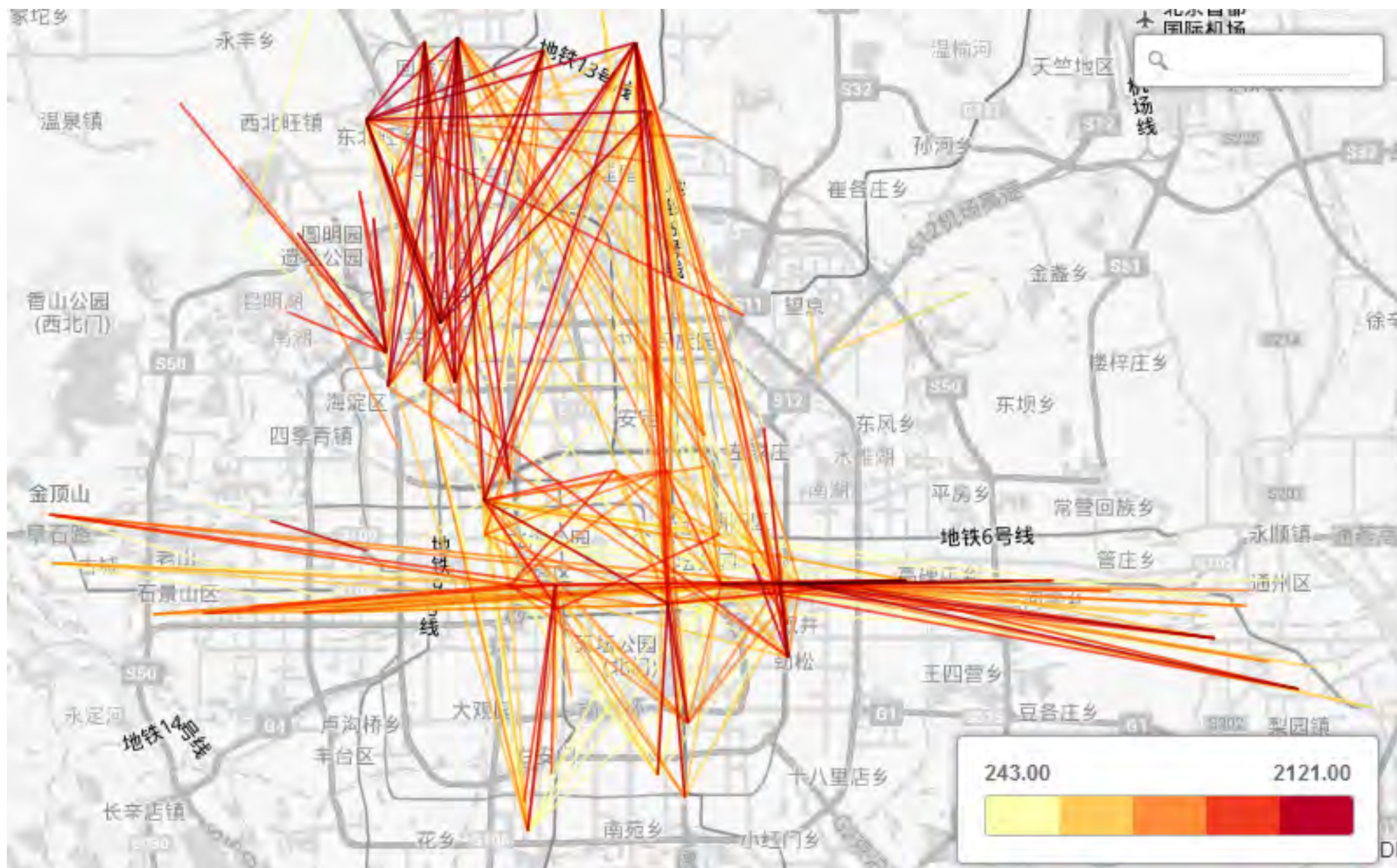
日发车77个班次

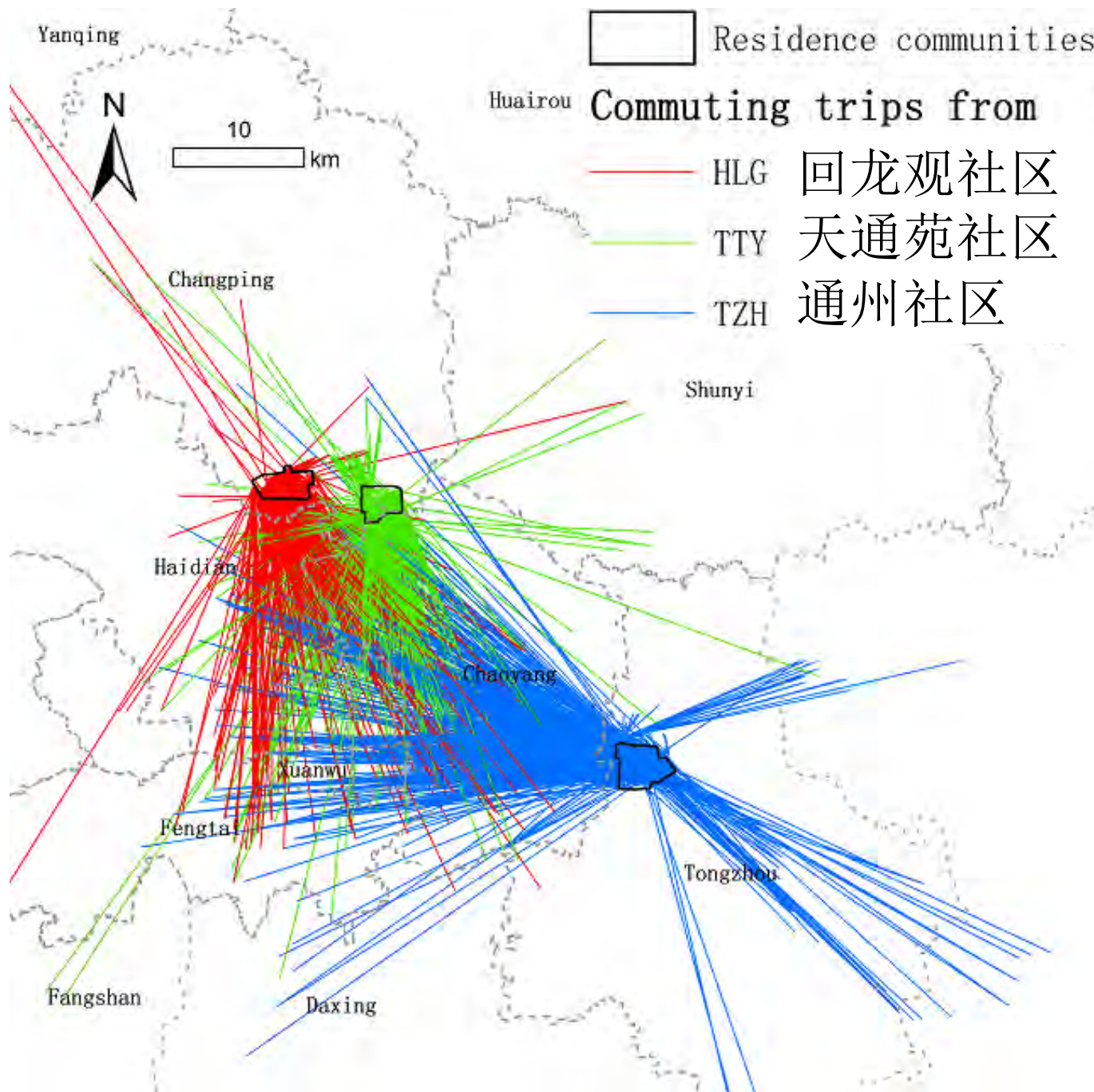
运送乘客10.33万人次

日运送乘客近2000人次

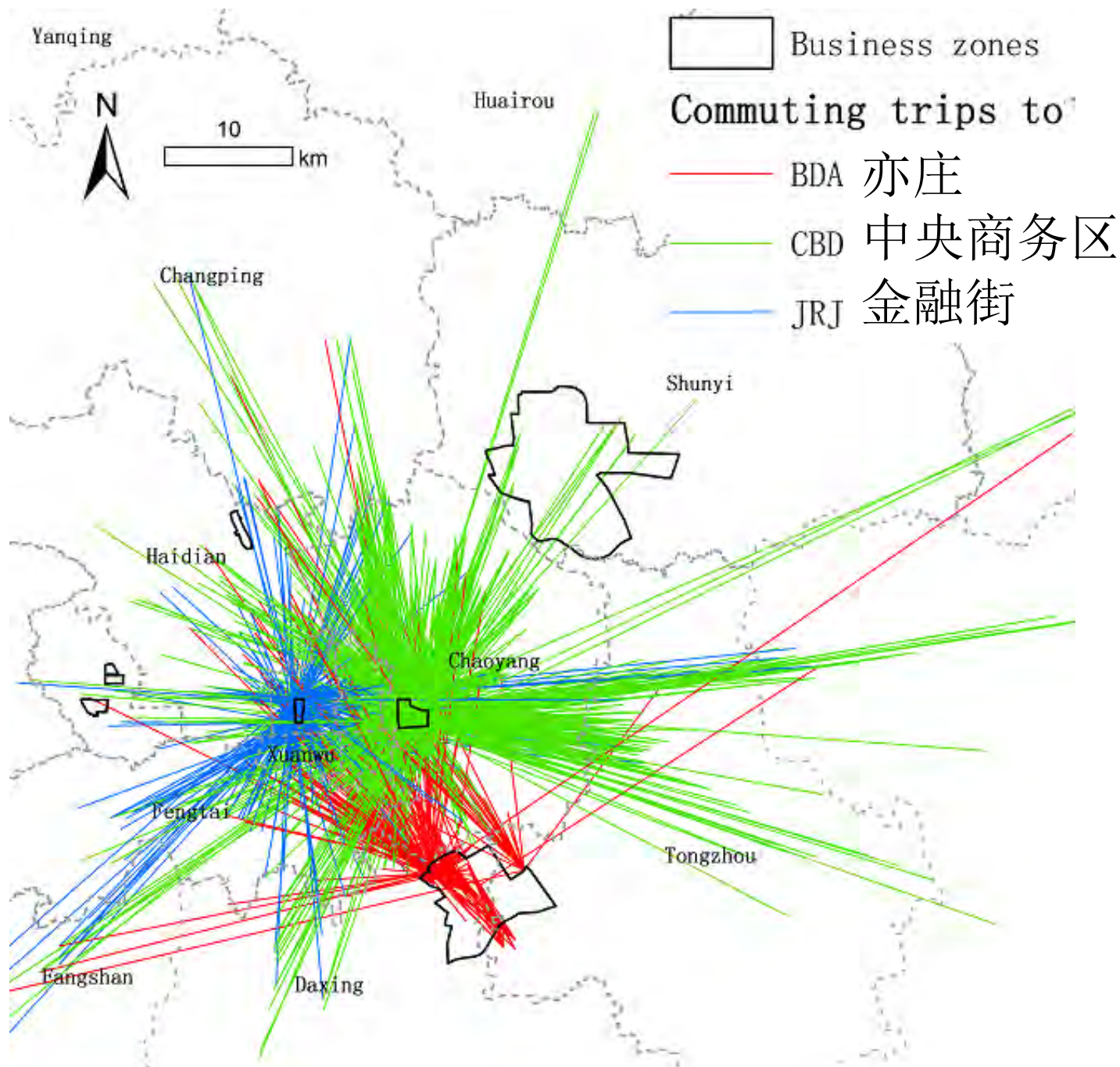
摘自：《2013北京公交社会责任报告》

基于2010年公交和地铁SCD的热点通勤线路

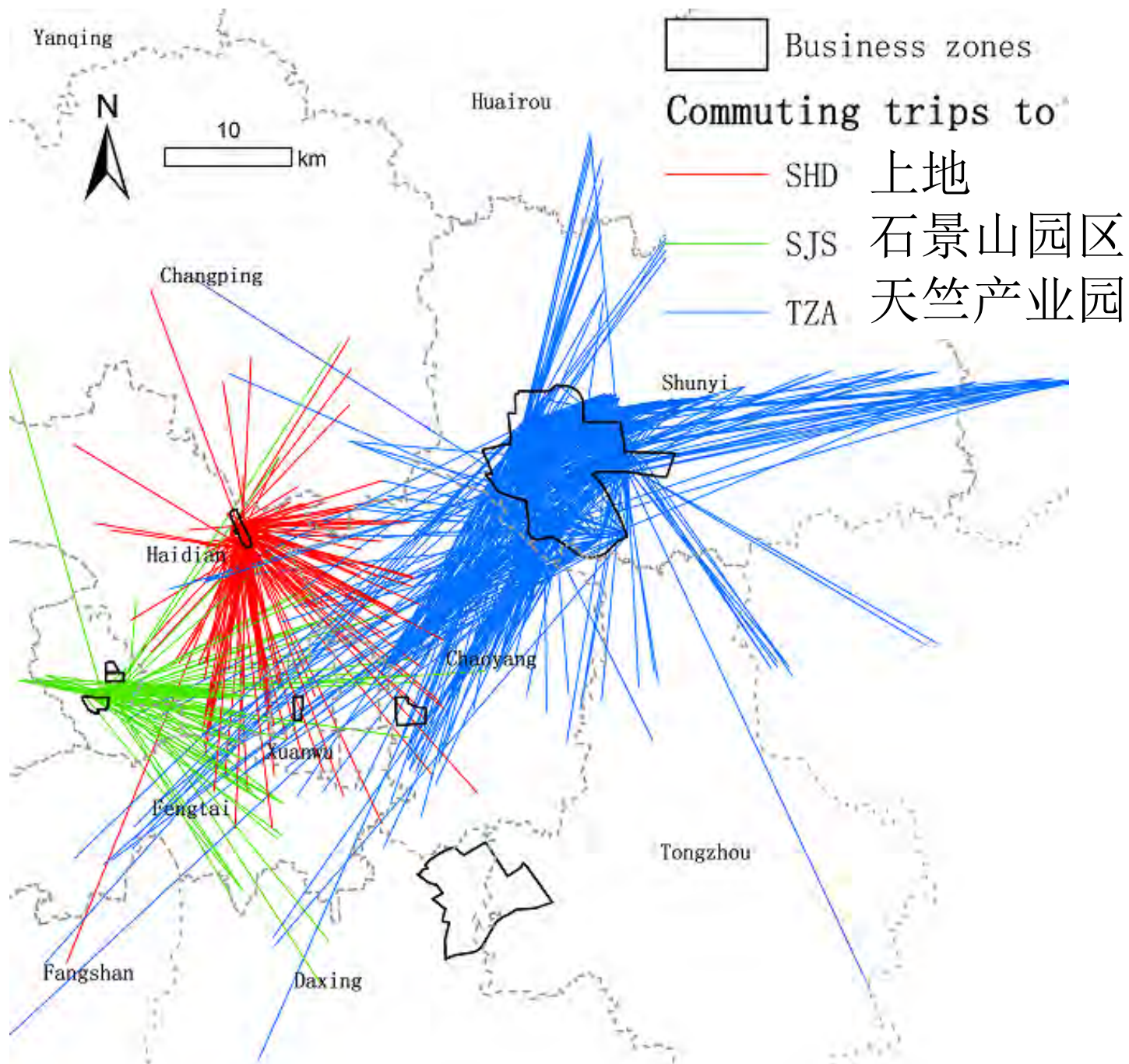




Commutes from main residence communities



Commutes to main business zones (1)



Commutes to main business zones (2)

Commuting time and distance of various residence communities and business zones of Beijing

Zone name	Commuting time(min)	Commuting distance(km)	% of all identified commuting trips
Trips from residence communities			3.9
TZH	45.1	10.0	1.4
HLG	39.4	7.0	1.0
TTY	36.2	6.1	1.5
Trips to business zones			6.0
CBD	41.4	9.4	2.7
SHD	40.4	6.7	0.3
JRJ	34.9	7.1	0.5
TZA	31.6	10.0	1.3
SJS	28.4	6.9	0.3
BDA	26.6	6.4	0.8

? Only 302 commuting trips (0.14% of all identified trips) are from 3 residence communities to 6 business zones in Beijing.

Summary

1. Propose an **algorithm** for identifying housing-jobs places as well as commuting trips using rules extracted from
 - the 2005 household travel survey
 - land use pattern of Beijing
2. Identification results are **acceptable** via comparing with the 2005 survey and existing researches
3. Analyse and map Beijing bus commuting **pattern**
 - Extreme commutes
 - Mainstream of bus commutes
 - Commutes from typical residential communities and to business zones
4. SCD could be a substitute of conventional travel surveys, at least a **complement**.

Our contribution: A promising solution for analyzing urban dynamics

- Mining LBS data using conventional conventional surveys and urban GIS layers with sound validation results
- **A decision tree** for determining the final one-week result using periodic information and spatial distribution of one-day result
- Retrieve Beijing commuting pattern with **more accurate spatial info and more samples** in contrast to existing researches using surveys
 - Although limited to bus riders

Springer book chapter, AAAG

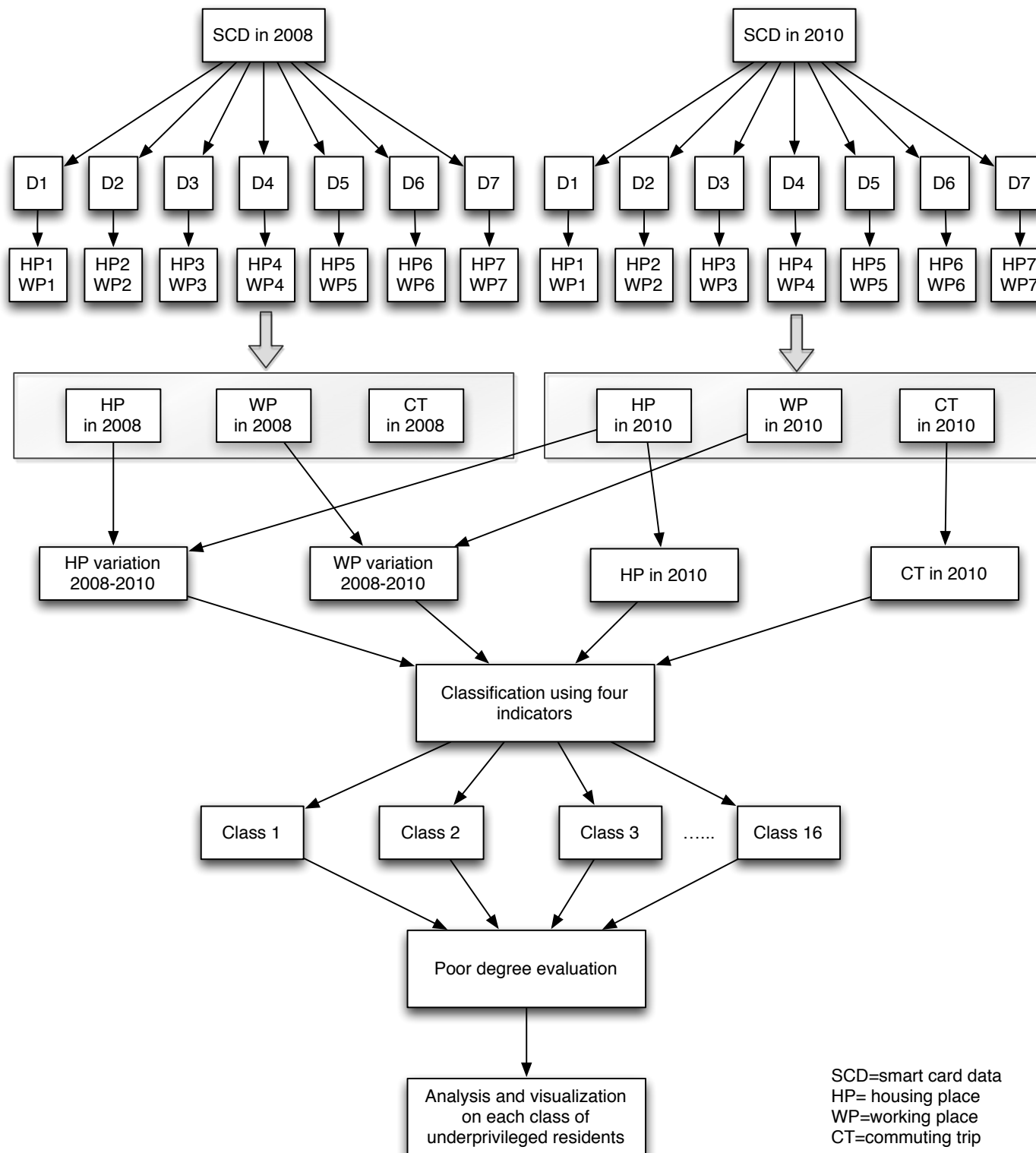
城市贫困

中国人针对社会经济问题的敏感性 以及已有城市贫困问题的研究方法

频繁公交出行者多为经济上的准（底层）



- 2010交通出行调查
 - For income of 618 households, 417 are in the level 1, 166 are in the level 2, 191 are in the level 3, and 8 are in level 4.
- 柴彦威小组调查
 - Among 125 identified frequent riders, 80.8% residents are with the month income less than 6000 CNY, and most of them (50% of all) only in the 2001-4000 CNY level.
- 当地居民访谈



SCD=smart card data
 HP= housing place
 WP=working place
 CT=commuting trip

识别结果



- 11.2万研究对象中，77.4%持卡人居住地点变化，仅13.3%未换工作（其他可能包括找到工作、失业、工作地变化）。**动荡的群体！**

Table 6 Housing place dynamics of FRs during 2008-2010.

Housing place		# cardholders	Ratio (%)	
Not changed		25,492	22.6	
Changed		87,082	77.4	
Inward (km)		42,013	37.3	
	2-5	9,211	8.2	
	5-10	9,651	8.6	
	10-20	13,150	11.7	
	>=20	10,001	8.9	
	Outward (km)		45,069	40.1
	2-5	7,990	7.1	
	5-10	10,139	9.0	
10-20	16,400	14.6		
>=20				
Sum				

Table 7 Workplace dynamics of FRs during 2008-2010.

Workplace		# cardholders	Ratio (%)	
Not changed		14916	13.3	
Changed		2203	2.0	
Working	Inward (km)	12713	11.3	
		6142	5.5	
		2-5	1444	1.3
		5-10	1893	1.7
		10-20	2071	1.8
	>=20	734	0.7	
	Outward (km)	6571	5.8	
		2-5	1371	1.2
		5-10	2018	1.8
		10-20	2156	1.9
>=20		1026	0.9	
Losing job		15909	14.1	
Finding a job		26919	23.9	
Jobless		54830	48.7	
Sum		112,574	100.00	

<http://arxiv.org/abs/1409.5839>

Table 8 Commuting distance variation of FRs (with commuting trips both in 2008 and 2010)

Commuting distance in 2010 – that in 2008 (km)	# cardholders
≥ 20	436
10-20	1,885
5-10	2,266
2-5	2,419
0-2	2,647
-2-0	1,984
-5-(-2)	1,416
-10-(-5)	1,069
-20-(-10)	622
≤ -20	172
Sum	14,916

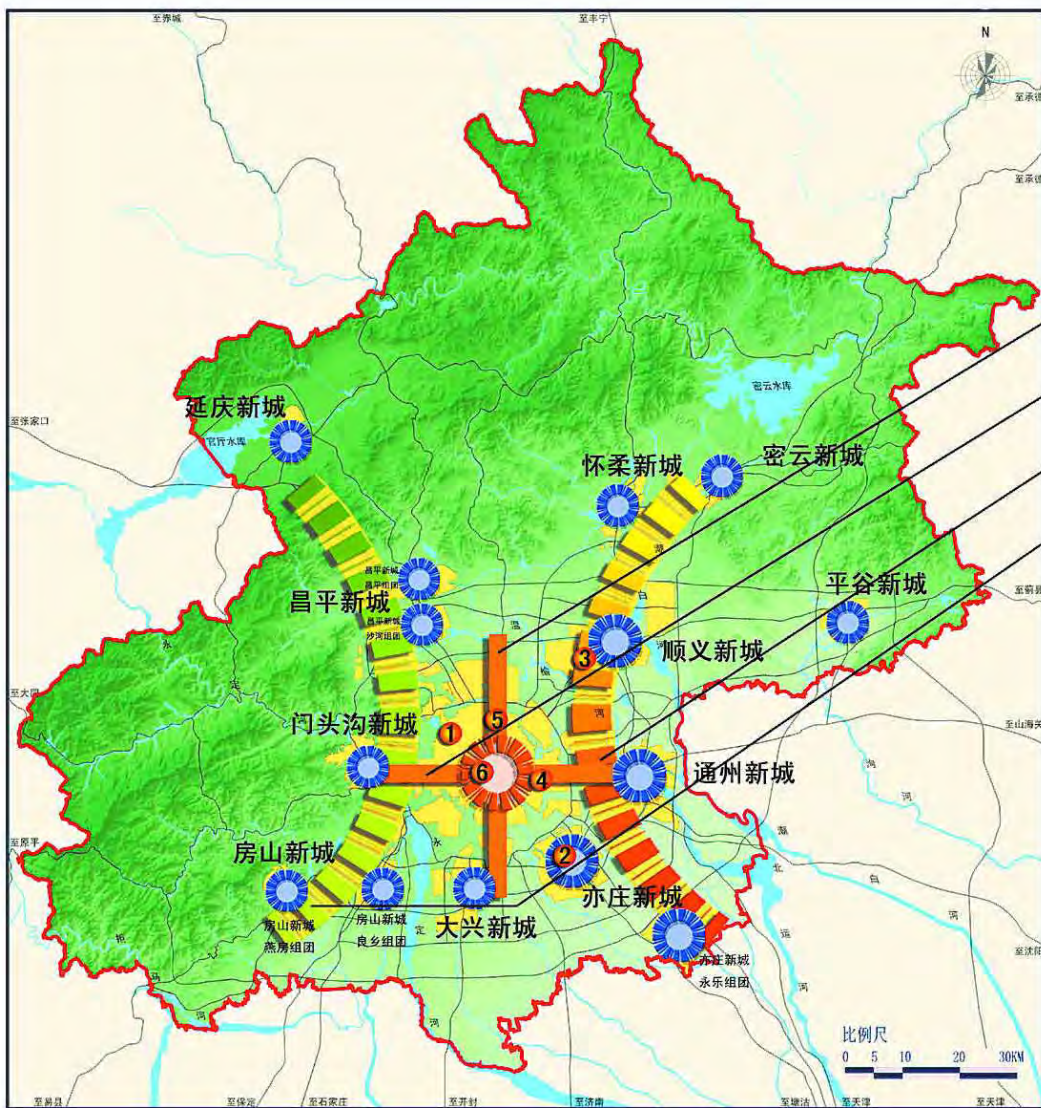
- 后续研究：城市开发对居住区位改变的影响、结合典型调研建立移动性与贫困的关系

Cities, 近期登出

评价城市总体规划实施

北京城市总体规划

北京城市总体规划 (二〇〇四年—二〇二〇年)



北京城市空间结构规划图 两轴—两带—多中心

- 传统中轴线的南北轴
- 沿长安街的东西轴
- 东部发展带
- 西部发展带

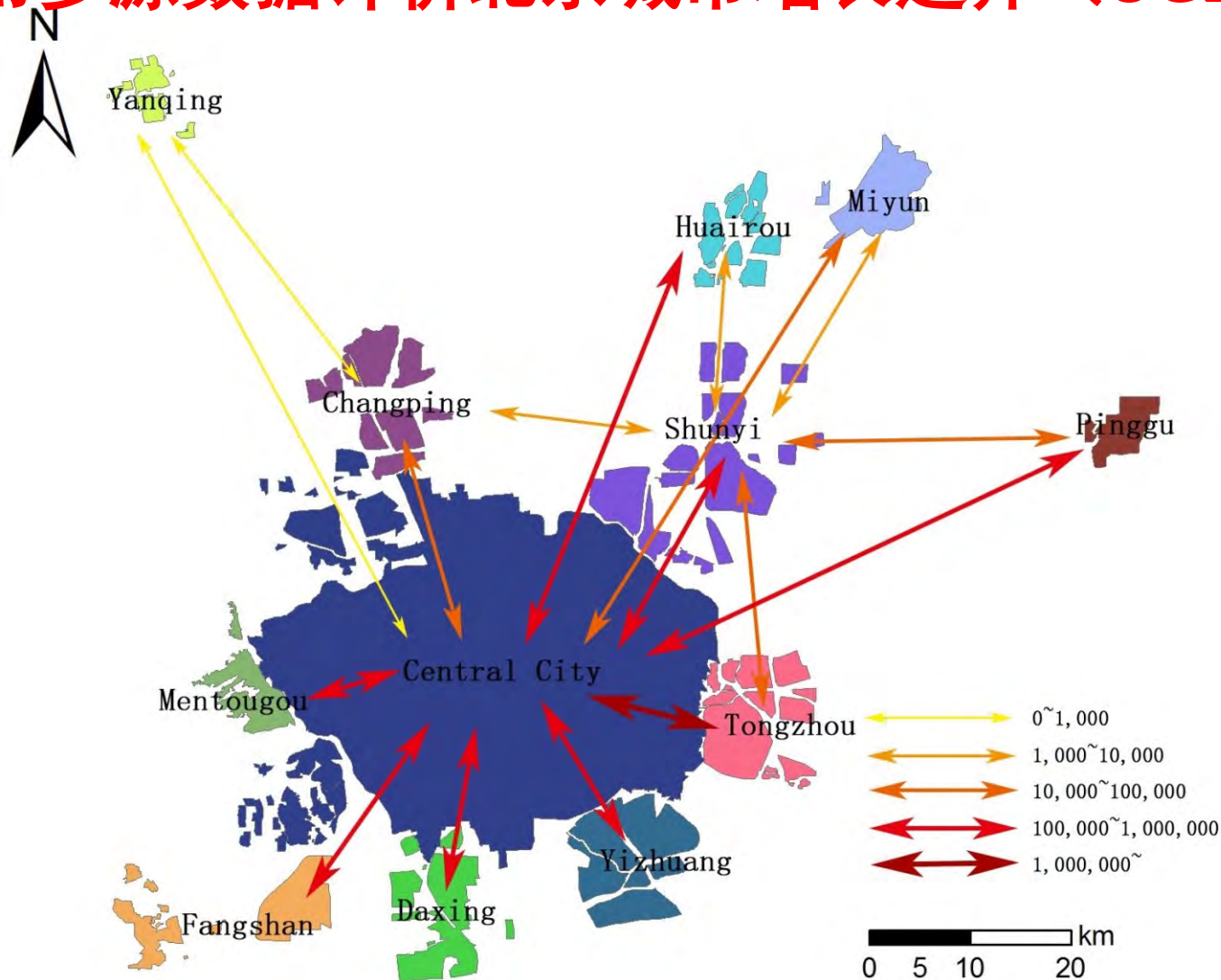
北京市“十一五”规划
确定的六大高端产业功能区：

1. 中关村科技园区
2. 北京经济技术开发区
3. 临空经济区
4. 商务中心区 (CBD)
5. 奥林匹克中心区
6. 金融街



- 两轴两带多中心
- 1800万人、1650平方公里城镇建设用

利用多源数据评价北京城市增长边界 (UGBs)



- 虽然有不可忽视的非正式开发，**95%以上的人类活动和移动位于规划城市增长边界内**（出租车轨迹、公交刷卡记录、位置微博/照片等）
- 各个功能组团的活动强度和之间的联系也可以用于评价规划目标（**两轴两带多中心还是单中心？顺义副中心还是通州副中心？**）

Urban Studies准备中

大学生汲取社会资本

数据情况

- We use the weekday smartcard data from April 6, 2010 to April 9, 2010 (totally 158,262 transit trips, including bus-only, bus plus subway and subway-only trips) to identify and profile the most popular destinations of the student riders from the “985 universities” and associated transit trip flows in Beijing.
- There are eight of 39 “985” universities located in Beijing. We define “popular destinations” as bus and subway stations where a student transit rider stays longer than one hour before he starts a second transit trip.

The top destinations of the students

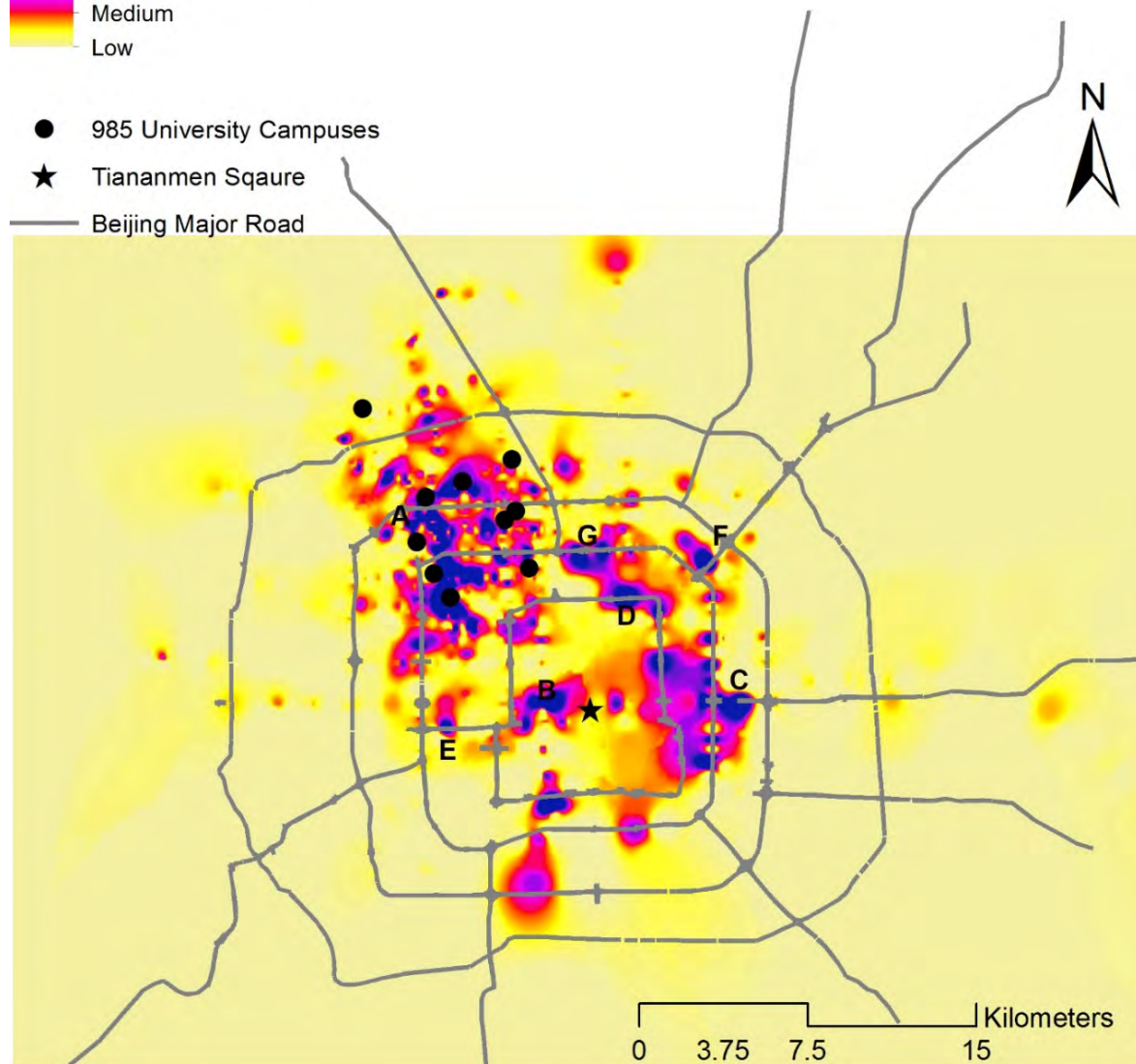
Destination Hotspots



● 985 University Campuses

★ Tiananmen Square

— Beijing Major Road



- 距离大学比较近的区域，例如中关村（A）片区包含了最多的“最受欢迎的地方”的热点。而西单（B）和国贸（C）这两个金融片区则站在了“最受欢迎的地方”的第二梯队里，雍和宫（D）、六里桥（E）、三元桥（F）和亚运村（G）同样被包含在“最受欢迎的地方”里。共同特征是这些地方都有密度较高的建筑，商场和餐饮店。

Top 200 OD Flows

<80

81 - 120

121 - 160

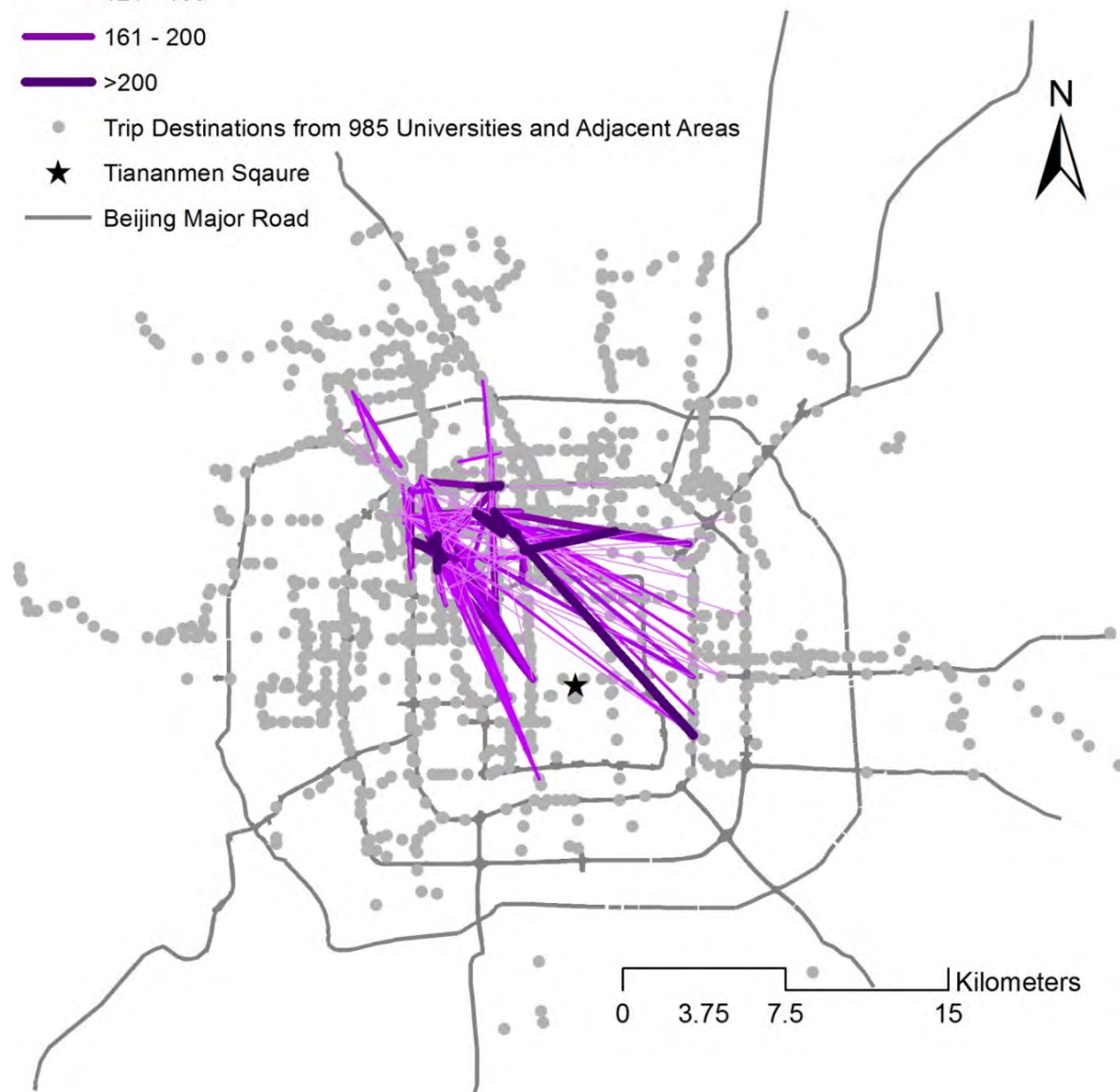
161 - 200

>200

● Trip Destinations from 985 Universities and Adjacent Areas

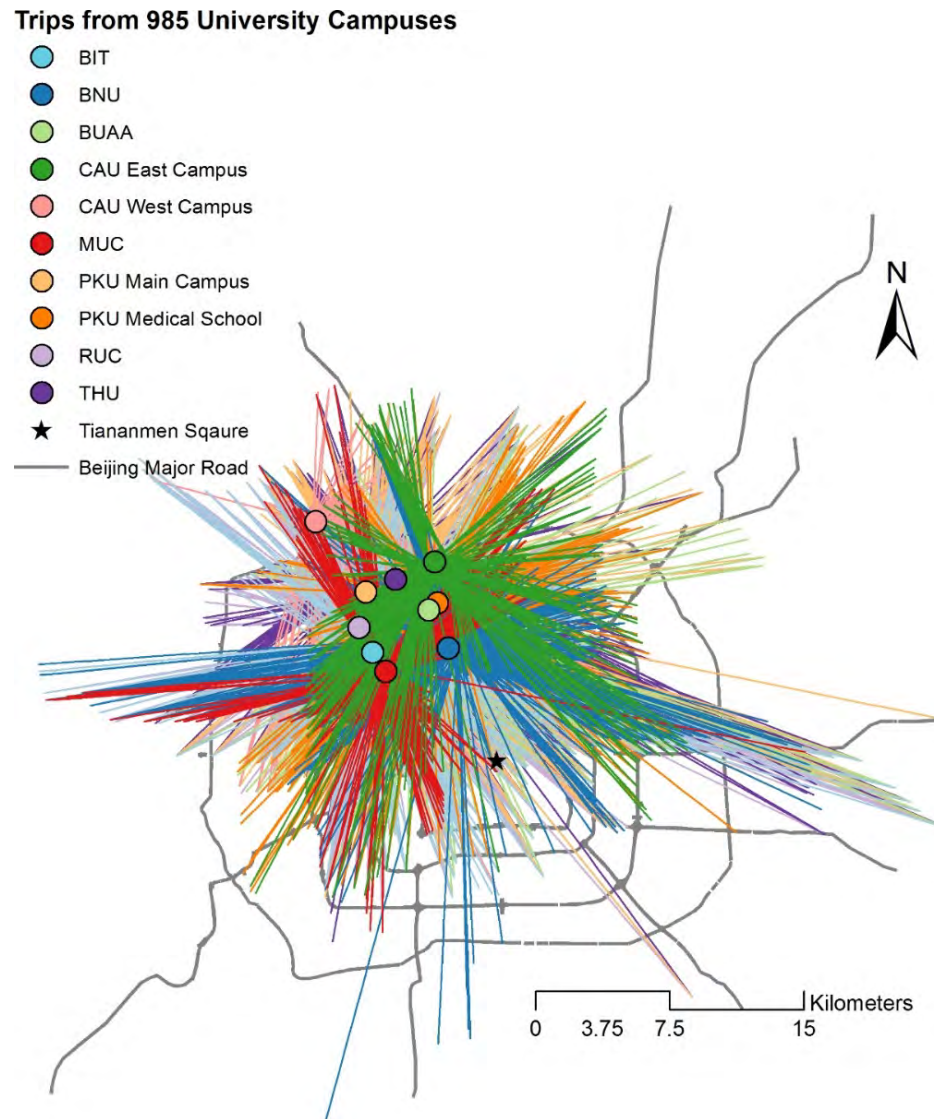
★ Tiananmen Square

— Beijing Major Road

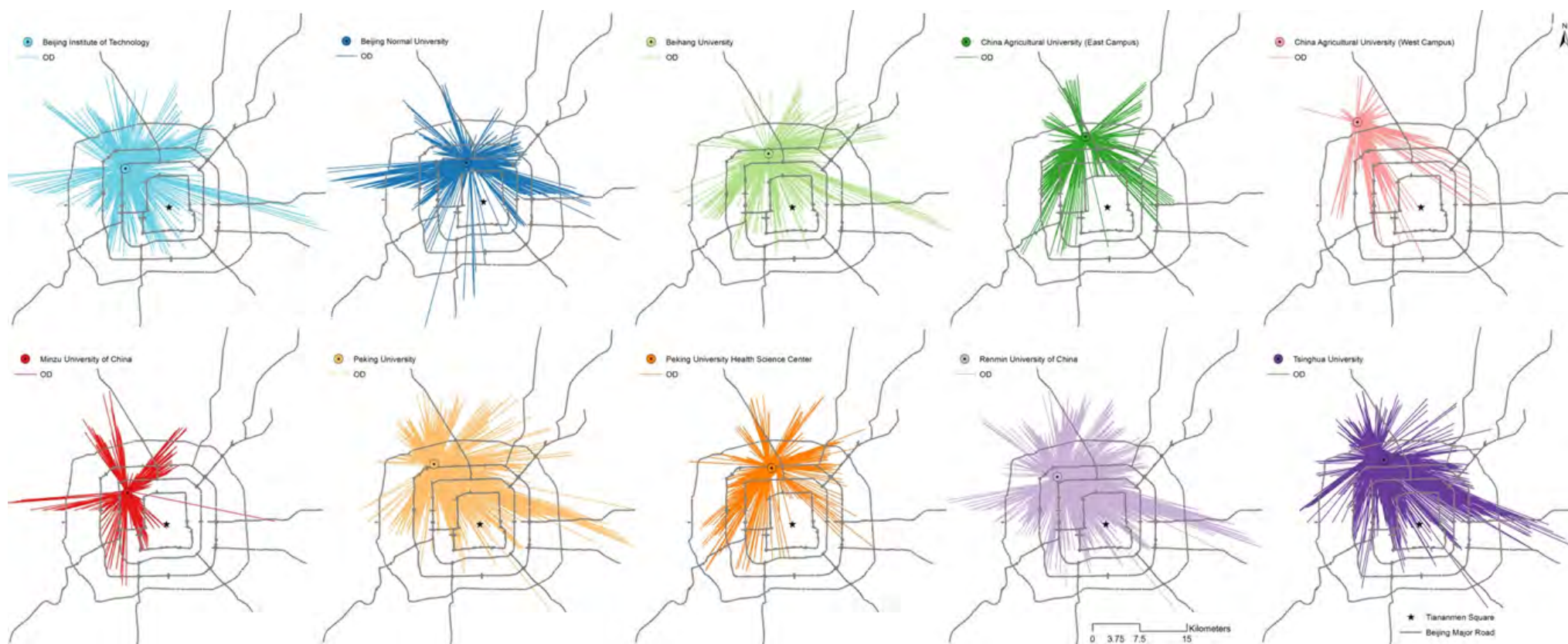


- 显示了大学校园和“最受欢迎的地方”的联系强度。大多数比较强的联系都在三环以内，同时也是北京高收入人群、高知名度团体和高收入公司的所在地。但令人惊讶的是各个大学之间的联系却不怎么强。

All the transit trips between the campuses and destinations

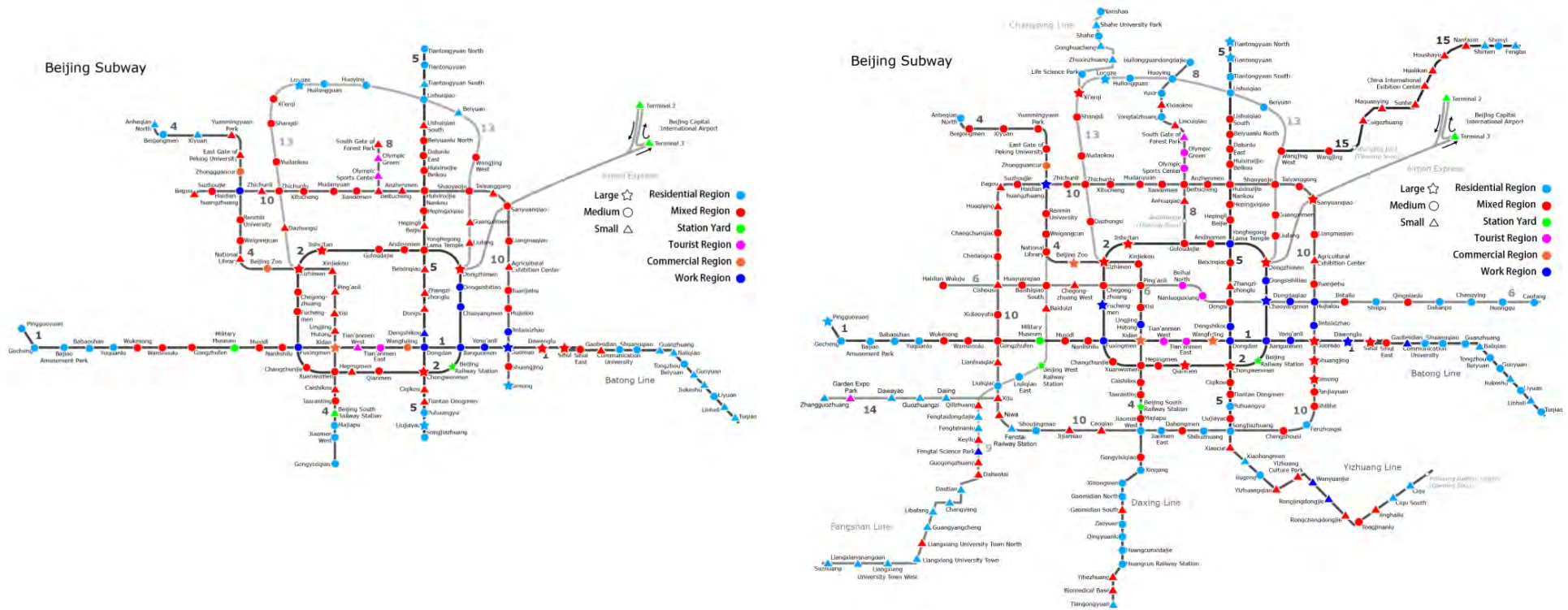


- 表示了各985大学的学生课余时间都去哪。很明显的是大学生去到的地方很多，但是他很少去南三环。985高校和北京经济比较弱的地区联系不强。



城市功能识别及其时空演化

利用轨道交通刷卡数据研究北京城市功能的演进 (2010年5月 vs 2013年7月)



- 利用刷卡数据识别每个站点的城市功能和客流量
- 通过对比2010和2013年的识别结果（城市功能和客流量），可以评价轨道交通基础设施建设，对城市功能的塑造
- 交通导向开发 **TOD OR NOT?**

Both size and function change

Station Name	Scale	2010	2011	2012	2013.3	2013.7
阜成门	● ★ ★ ★ ★	Red	Red	Red	Blue	Blue
刘家窑	★ ★ ★ ● ●	Blue	Blue	Blue	Blue	Red
北苑	▲ ▲ ● ● ●	Blue	Red	Blue	Blue	Blue
西苑	▲ ● ● ● ●	Blue	Blue	Blue	Blue	Red
灵镜胡同	▲ ● ● ● ●	Red	Blue	Blue	Blue	Blue
后沙峪	∅ ● ▲ ▲ ▲	White	Red	Red	Blue	Blue
六里桥东	∅ ∅ ▲ ● ●	White	White	Red	Blue	Blue
六里桥	∅ ∅ ▲ ▲ ●	White	White	Green	Blue	Blue
北海北	∅ ∅ ∅ ▲ ●	White	White	White	Red	Pink
南锣鼓巷	∅ ∅ ∅ ▲ ●	White	White	White	Red	Pink

Large ☆
 Medium ○
 Small ▲

Residential Region ●
 Mixed Region ●
 Station Yard ●
 Tourist Region ●
 Commercial Region ●
 Work Region ●

Explorative Analysis on Extreme Travel Behavior in Beijing

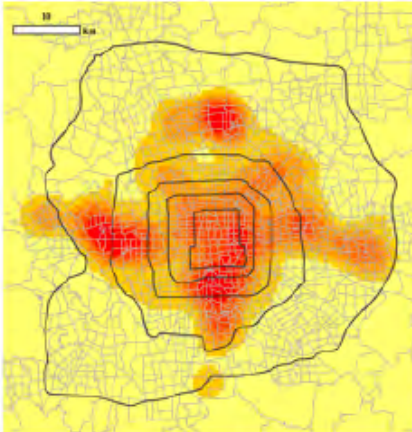
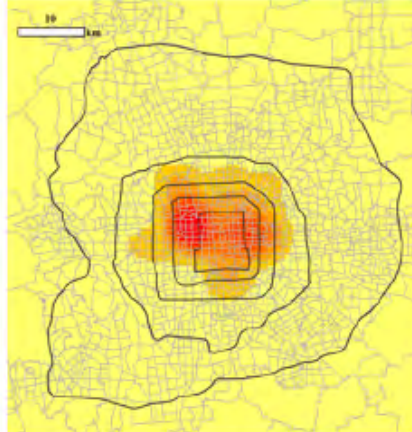

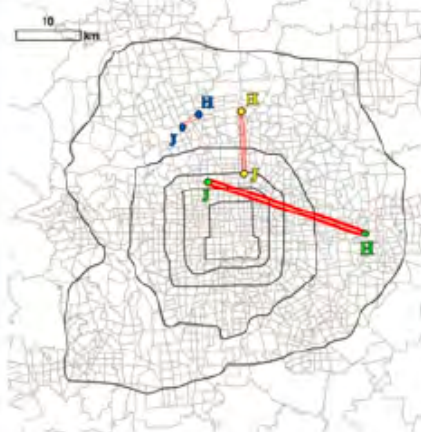
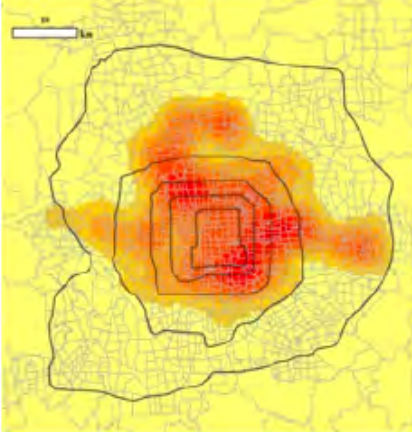
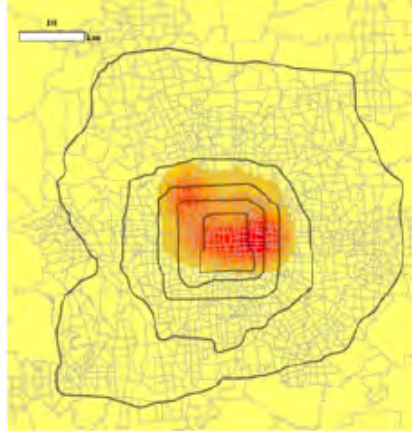




EARLY BIRDS, NIGHT OWLS, AND TIRELESS ITINERANTS

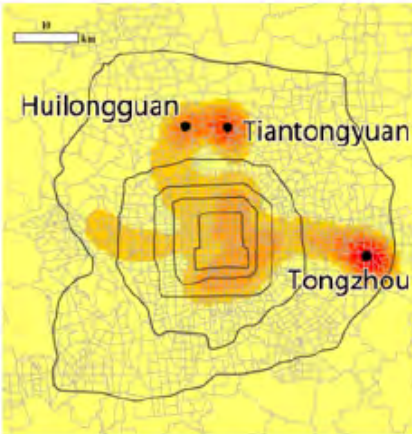
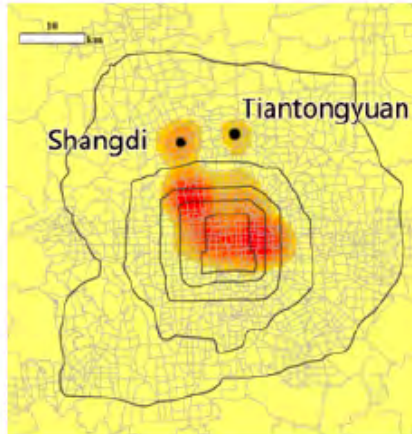


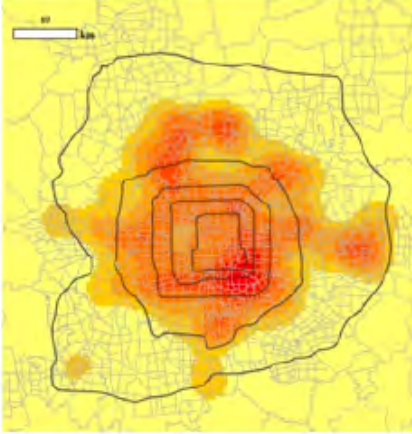
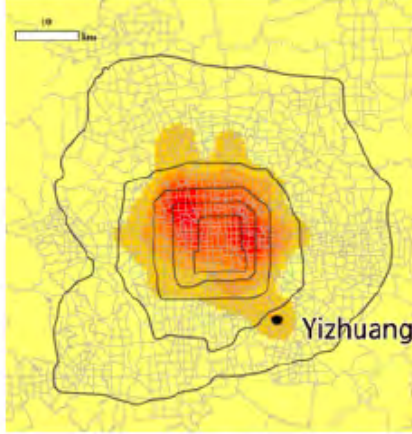
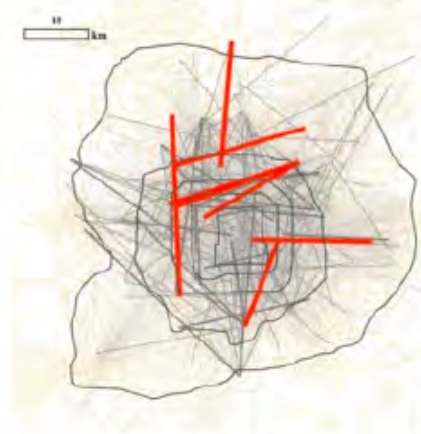



利用一周公共交通刷卡数据研究北京极端出行

利用刷卡和出行调查数据分析北京极端出行行为

- 充分利用大数据在样本大、时空信息完善的特点，勾画特定人群的时空分布及移动性，回答在哪里、如何活动和移动（说格局）？
- 同时利用小数据的社会经济等信息，回答是谁、在做什么、为什么（讲故事）？

Type	Definition
Early Birds (EBs)	First trip < 6AM, more than two days in a week (60% of weekdays)
Night Owls (NOs)	last trip (boarding time) > 10PM, more than two days in a week (60% weekdays)
Tireless Itinerants (TIs)	>= one and a half hours commuting, more than two days in a week
Recurring Itinerants (RIs)	>= 30 trips in weekdays of a week (>= 6 trips per day)

Extreme travelers	Kernel density of housing	Kernel density of jobs	Commuting trips	Typical trips
EBs	 <p>(10.3 k)</p>	 <p>(9.4 k)</p>	 <p>(4.9 k)</p>	
NOs	 <p>(31.6 k)</p>	 <p>(25.0 k)</p>	 <p>(17.5 k)</p>	
Legend	 <p>Low High</p>		 <p>Low High</p>	<ul style="list-style-type: none"> ● Person A ● Person B ● Person C — Ring roads — Routes

Extreme travelers	Kernel density of housing	Kernel density of jobs	Commuting trips	Typical trips
<p>TIs</p>	 <p>Huilongguan • Tiantongyuan • Tongzhou</p> <p>(6.7 k)</p>	 <p>Shangdi • Tiantongyuan</p> <p>(6.7 k)</p>	 <p>Tongzhou</p> <p>(6.7 k)</p>	
<p>RIs</p>	 <p>(25.4 k)</p>	 <p>Yizhuang</p> <p>(7.8 k)</p>	 <p>(2.7 k)</p>	
<p>Legend</p>	 <p>Low High</p>		 <p>Low High</p>	<ul style="list-style-type: none"> ● Person A ● Person B ● Person C — Ring roads — Routes

Type	EBs	NOs	TIs	RIs	AB
Sleep/Rest	0.4%	0.4%	0.0%	0.2%	0.1%
Shopping	1.0%	0.4%	0.0%	6.9%	1.4%
Pick-up or Drop-off Others	1.0%	0.0%	0.0%	11.7%	4.6%
Accompany Others	0.0%	0.0%	0.0%	0.4%	0.2%
Taking Delivery of Goods	0.6%	0.0%	0.0%	1.3%	0.4%
Go Home	1.3%	96.2%	0.0%	0.7%	44.4%
Have Meals	1.3%	0.0%	0.0%	31.2%	2.4%
Work	60.2%	2.1%	100.0%	2.4%	21.9%
Official Travel	0.9%	0.0%	0.0%	10.0%	0.9%
Go to Class/Study	11.8%	0.0%	0.0%	5.5%	3.9%
Personal Affairs	6.8%	0.0%	0.0%	0.5%	2.6%
Homework/Take Care of the Family	0.6%	0.0%	0.0%	10.8%	0.1%
Recreation, Entertainment and Fitness	10.9%	0.4%	0.0%	0.2%	6.4%
Visit Relatives and Friends	1.8%	0.0%	0.0%	3.1%	10.1%
Others	1.2%	0.4%	0.0%	15.1%	0.6%

Table 4 Selected socioeconomic characteristics of extreme travelers

ID	Extreme travelers	EBs (676)	NOs (236)	TIs (627)	RIs (100)	ABs (116,142)
1	% annual household income \geq 100 k CNY	4.9	4.2	6.7	5.0	7.4
2	% renting house	11.0	17.8	20.4	16.0	16.1
3	# average household car ownership	0.22	0.21	0.25	0.22	0.31
4	% higher education (undergraduate and graduate)	14.2	18.2	33.5	25.0	21.1
5	% Beijing <u>Hukou</u>	87.0	82.2	74.8	83.0	82.4
6	% public-sector employees	13.5	7.6	15.8	7.0	10.4
7	% fulltime workers	60.9	84.7	94.4	42.0	45.9
	% fulltime students	12.7	2.1	1.3	1.0	7.3
	% retirees	20.9	5.9	0.8	38.0	29.1

Note that numbers in brackets are the total count of extreme travelers.

谢谢!

- 基于公共交通刷卡数据开展了诸多定量城市研究
 1. 通勤出行、职住平衡和过度通勤
 2. 城市贫困
 3. 评价城市总体规划实施
 4. 大学生汲取社会资本
 5. 城市功能识别及其时空演进
- 大数据时代，传统数据同样重要!
- 本报告所介绍的诸多实践，**是科学地理解城市的一次尝试**
(不是规划信息化、规划新技术)
 - The New Science of Cities (Michael Batty)



北京城市实验室
Beijing City Lab

<http://www.beijingcitylab.com>

我们从来也没有如此清晰地观察社会系统