Bus Landscapes:
Analyzing Commuting Pattern
using Bus Smart Card Data in Beijing

Ying Long, Beijing Institute of City Planning 龙 瀛

Jean-Claude Thill, The University of North Carolina at Charlotte

1 INTRODUCTION

Human as sensors

- The increasing pervasiveness of location based services (LBS), including

  - GSM (Global System for Mobile Communications)

    - Namely cell phone data

  - GPS (Global Positioning System)

    - Trajectories

  - SNS (Social Networking Services)

    - Check-in, Twitter, Facebook, Sina Weibo

  - Wi-Fi (wireless fidelity)

- Rich in spatiotemporal information (big data)

- Better for describing and understanding urban structure

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

Existing researches using SCD

- Most researches advocate providing decision making for planning and design of public transportation system (see Pelletier et al (2011) for a review).
  - This is also the focus of Chinese papers
- In South Korea, Joh and Hwang (2010) analysed 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.
- Less attention was paid on analyzing jobs-housing relationships as well as commuting pattern of a metropolitan city.

This talk is regarded with

- Using Beijing bus SCD for identifying cardholders' housing and job places

- Identifying commuting trips of cardholders

- Analyzing and mapping bus commuting pattern of Beijing

2 DATA

Bus lines

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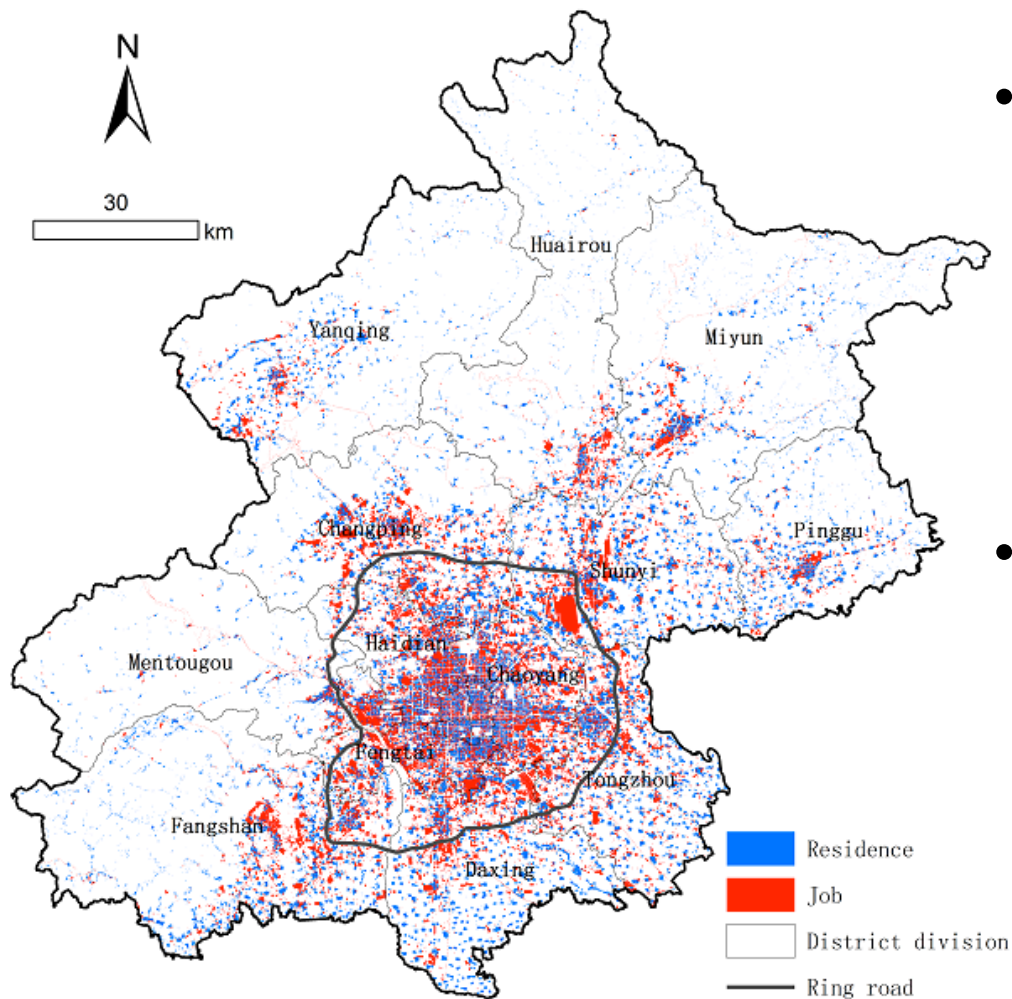 - Fixed-fare (short dist, within inner city)

Bus stops

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

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

r11

Bus SCD in Beijing

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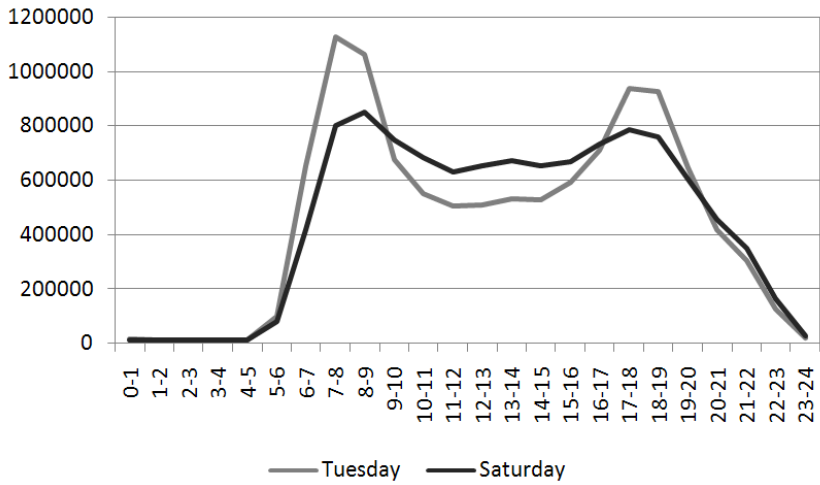
•Over 90% bus riders use smart cards till April 2007

- One week in April 2008 from
  - Monday to Sunday
  - Subway not included
- 80 million records for 8.5 million cards/persons
  - One record for a bus riding of a cardholder
  - 1.3 daily bus ridings per person

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
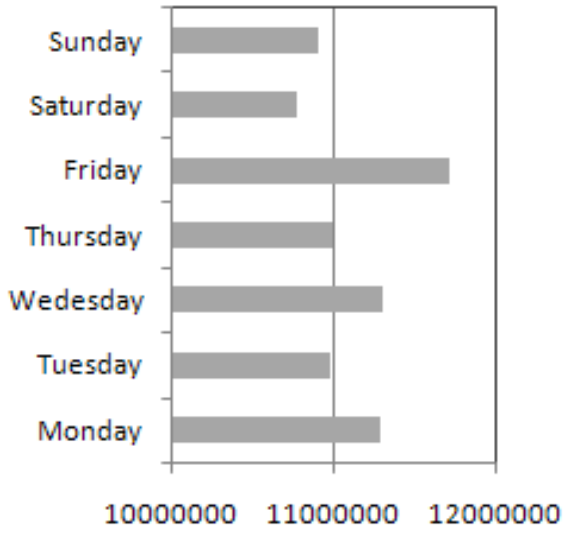
Trip count (bus riding) for each day

Trip count in each hour on Tue and Sat

Trip density in the TAZ level in the whole week

14

Count of bus ridings for each TAZ

0:00-1:00

	1 - 146
	147 - 383
	384 - 716
	717 - 1288
	1289 - 3411
	Ring road

10 km

15
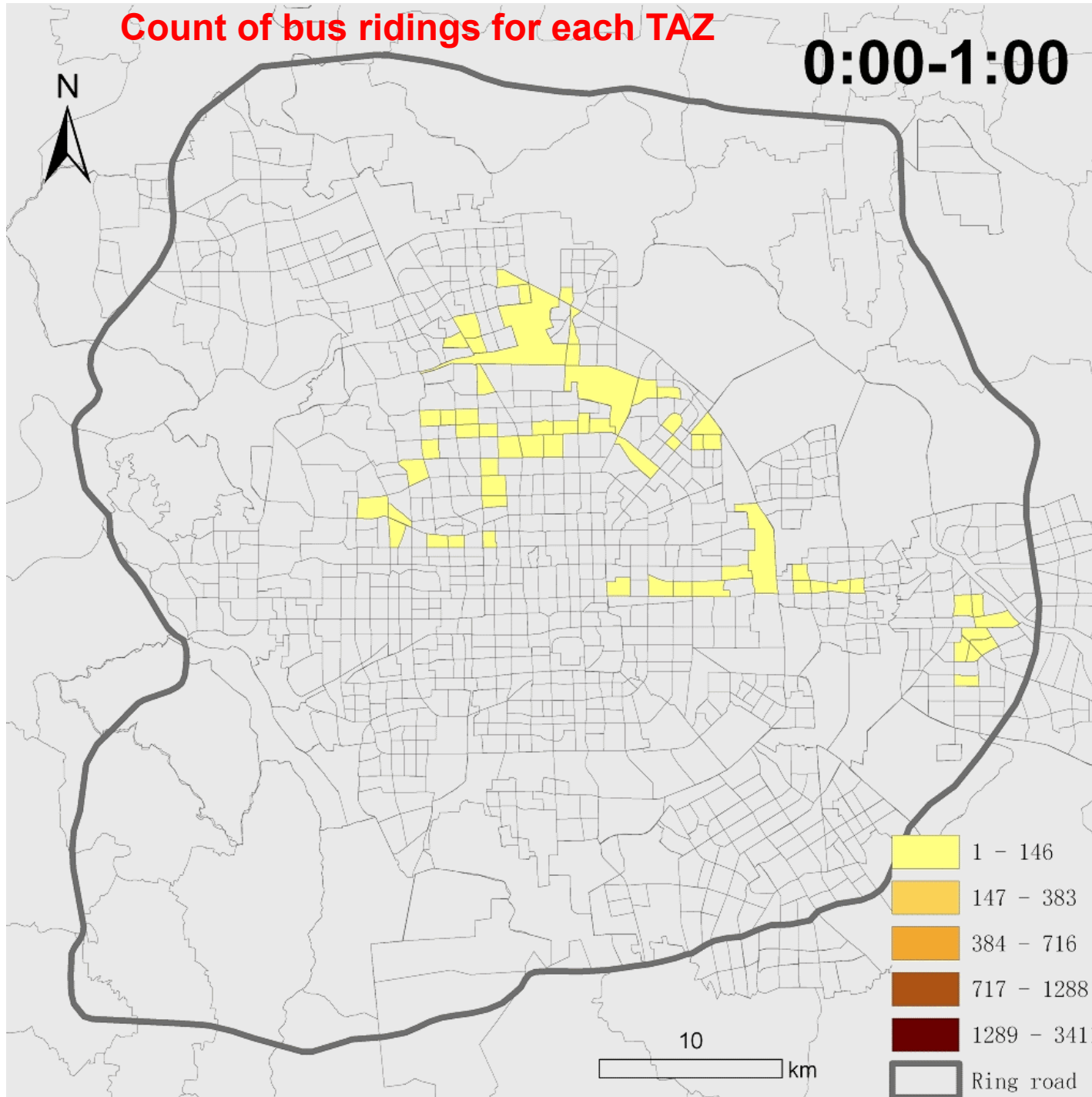
Special attention paid to fixed-fare trips

- Distance-fare
    - With full spatiotemporal information
- Fixed-fare
    - Only arrival time and stop ID
    - No departure time and stop ID

The Beijing Household Travel Survey

(The 2005 survey)

- Conducted in 2010 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

3 APPROACH

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}

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*: $k <> 1$
      - 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.

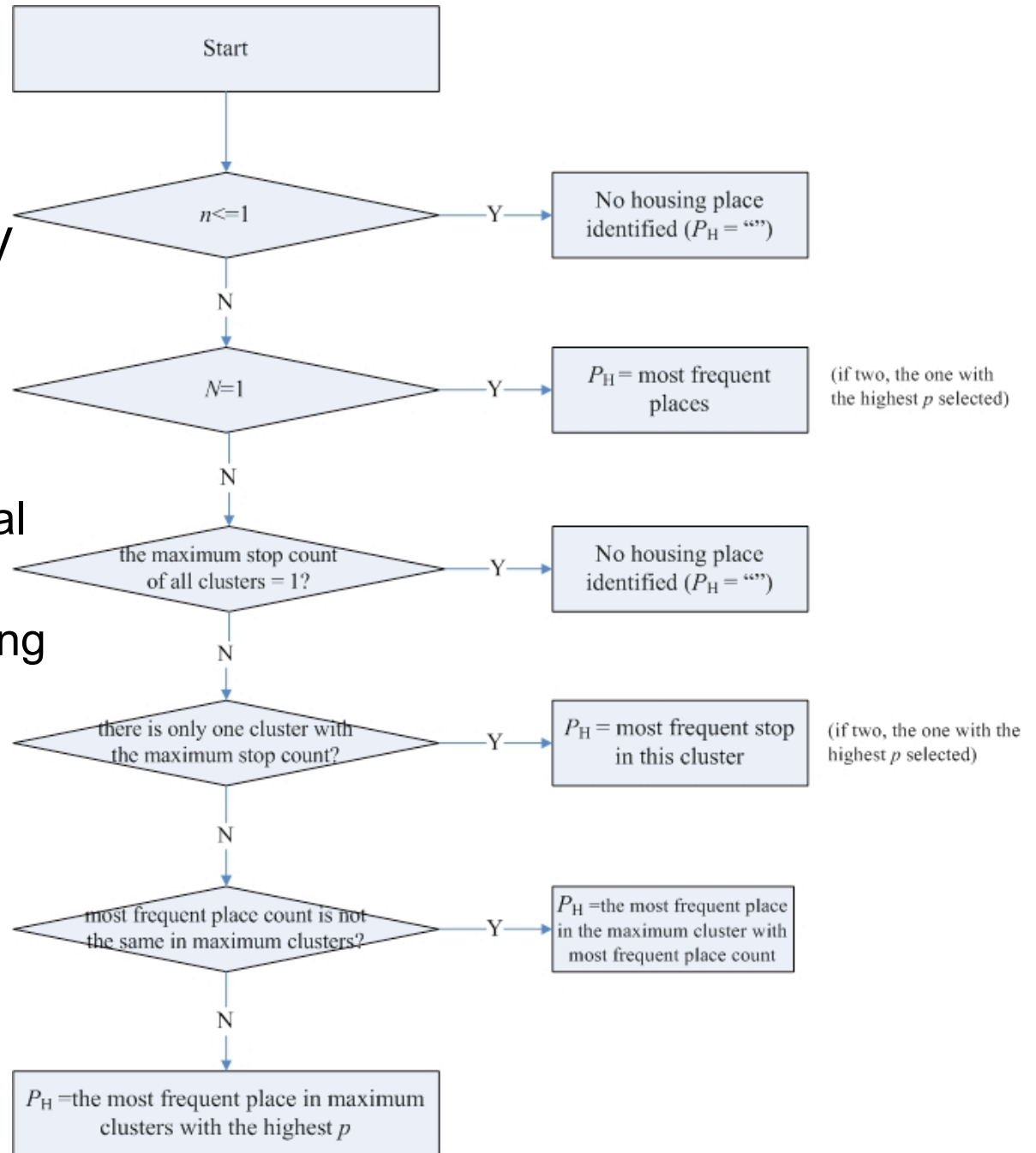# Combining each day's result to get more solid housing-jobs places

- 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

Start

$n <= 1$ — Y → No housing place identified (P_H = "")

N

$N = 1$ — Y → P_H = most frequent places (if two, the one with the highest p selected)

N

the maximum stop count of all clusters = 1? — Y → No housing place identified (P_H = "")

N

there is only one cluster with the maximum stop count? — Y → P_H = most frequent stop in this cluster (if two, the one with the highest p selected)

N

most frequent place count is not the same in maximum clusters? — Y → P_H = the most frequent place in the maximum cluster with most frequent place count

N

P_H = the most frequent place in maximum clusters with the highest p
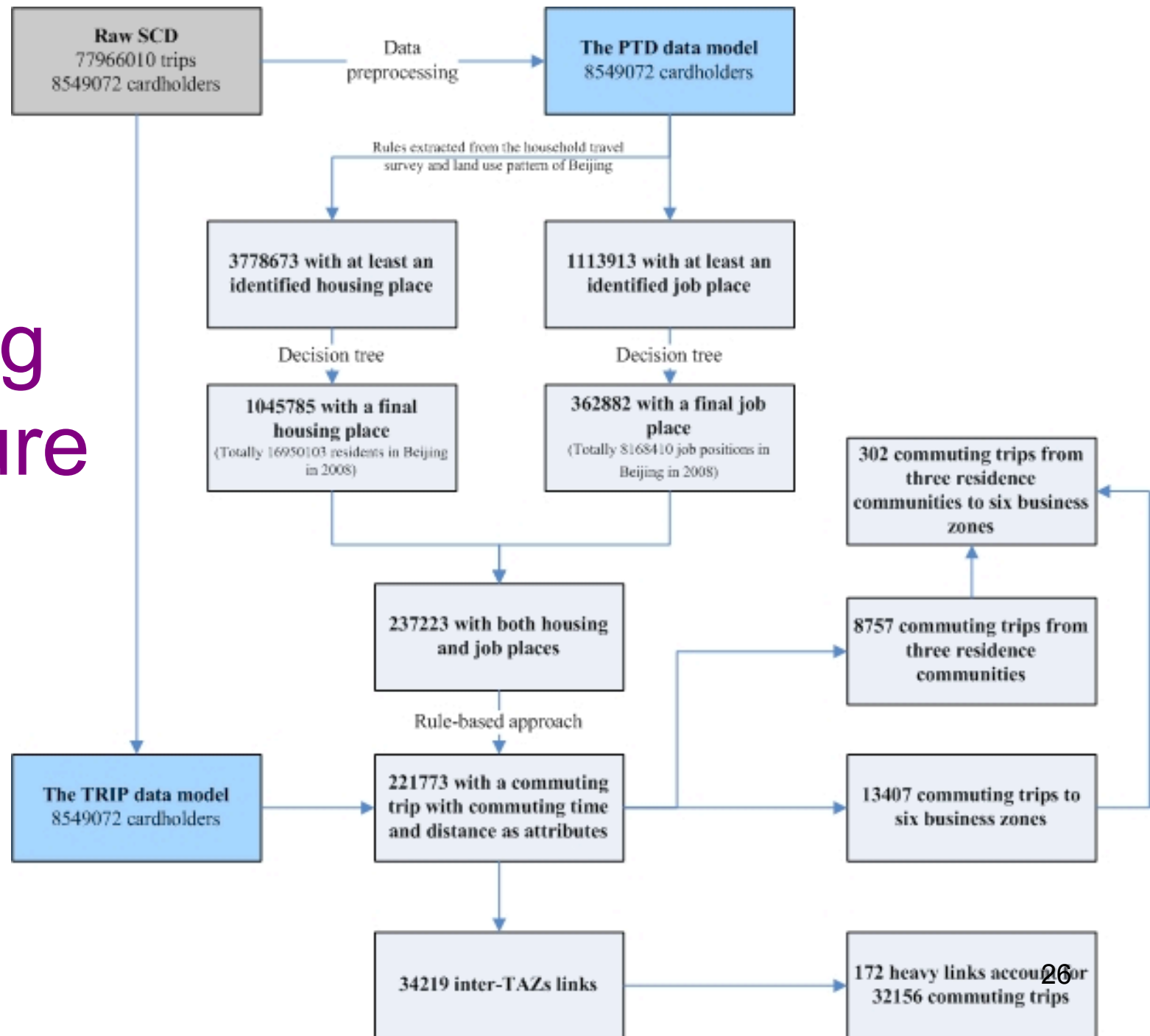
Identification of commuting trips

- 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.

4 RESULTS

A big picture

Raw SCD
77966010 trips
8549072 cardholders

Data preprocessing →

The PTD data model
8549072 cardholders

Rules extracted from the household travel survey and land use pattern of Beijing

3778673 with at least an identified housing place

1113913 with at least an identified job place

Decision tree

Decision tree

1045785 with a final housing place
(Totally 16950103 residents in Beijing in 2008)

362882 with a final job place
(Totally 8168410 job positions in Beijing in 2008)

237223 with both housing and job places

302 commuting trips from three residence communities to six business zones

8757 commuting trips from three residence communities

Rule-based approach

The TRIP data model
8549072 cardholders

221773 with a commuting trip with commuting time and distance as attributes

13407 commuting trips to six business zones

34219 inter-TAZs links

172 heavy links account for 32156 commuting trips

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.
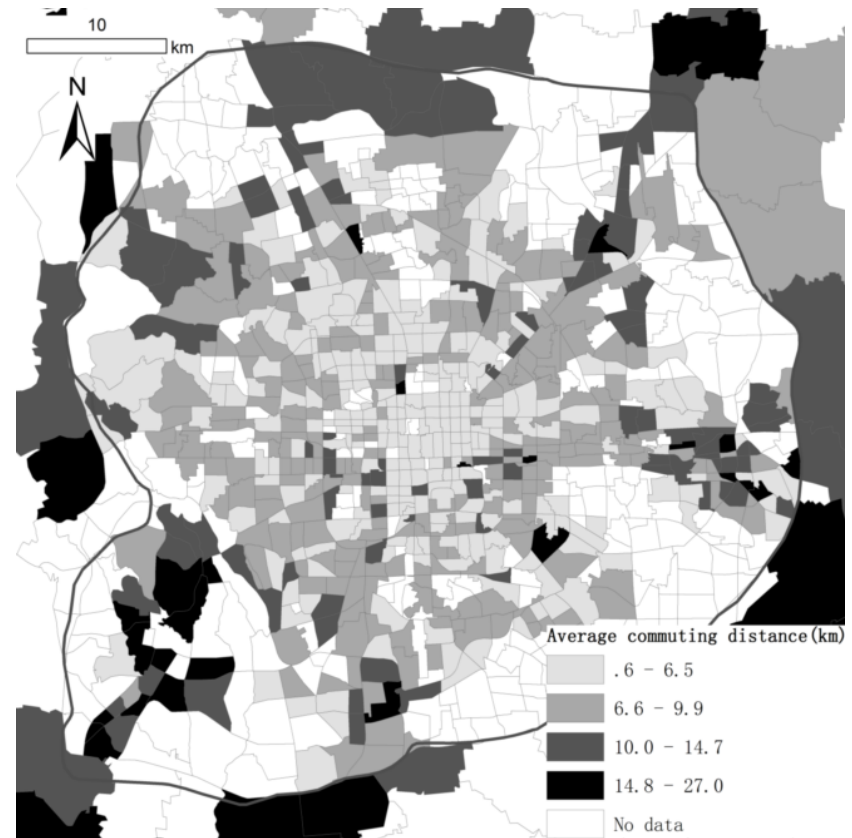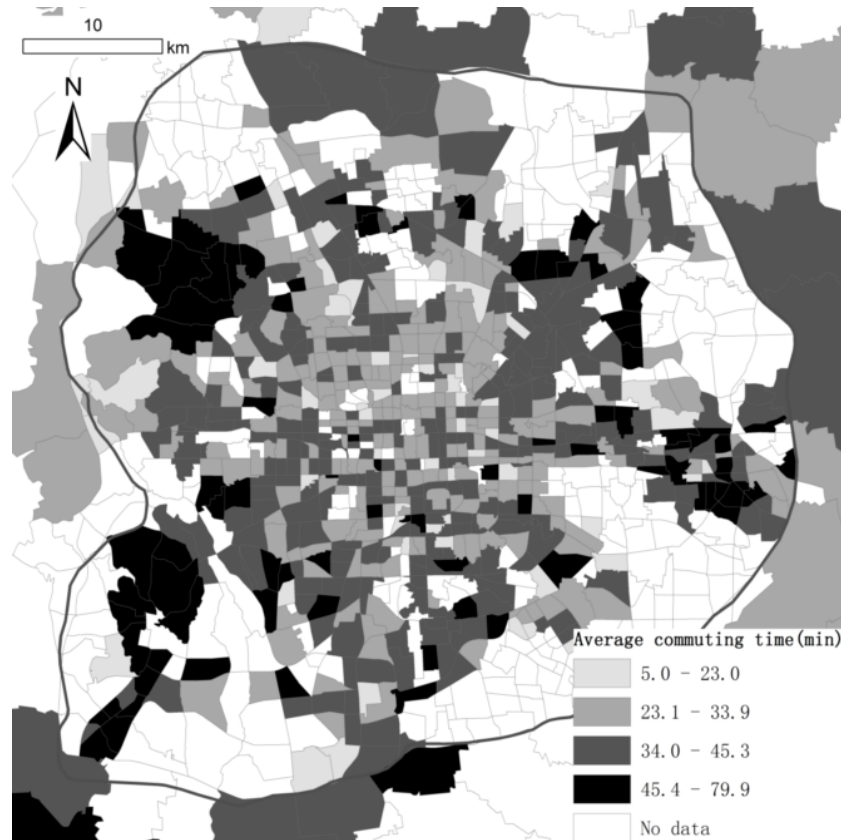

Cardholder density with an identified housing place
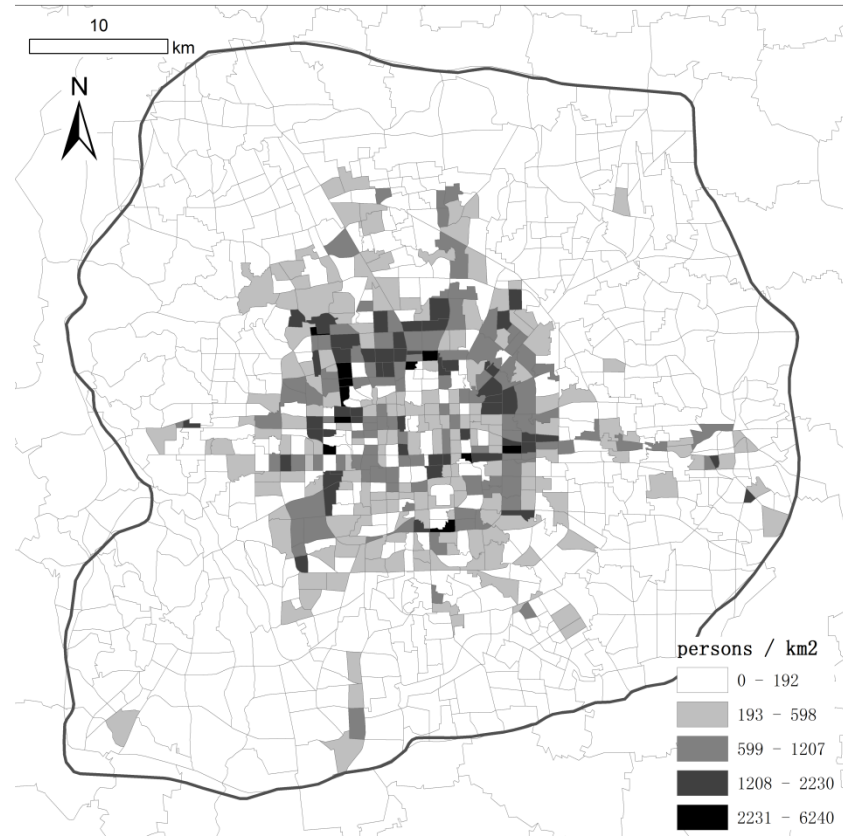
Cardholder density with an identified a job place

Commuting trips identification

221,773 cardholders identified with commuting trips


Average commuting time and distance for each TAZ 28

Comparing with existing researches

| Name	Travelling modes and year	Sample size	Average commuting time (min)	Average commuting distance (km)
Our commuting trips	**Bus, 2008**	**221773**	**36.0 (24.2)**	**8.2 (7.0)**
The 2005 survey	Bus, 2005	6651	40.5 (23.1)	8.4 (8.3)
Liu and Wang, 2011	Bus, 2007	307	46.3 (N/A)	N/A
Wang and Chai, 2009	Bus, 2001	227	55.1 (30.4)	N/A
Zhao et al, 2011	Bus and metro, 2001	220	52.4 (26.6)	N/A

Note that the number in the bracket is the standard deviation of the average commuting time and distance. Except our commuting trips, bus samples in other researches are extracted from the whole survey with all travelling modes.
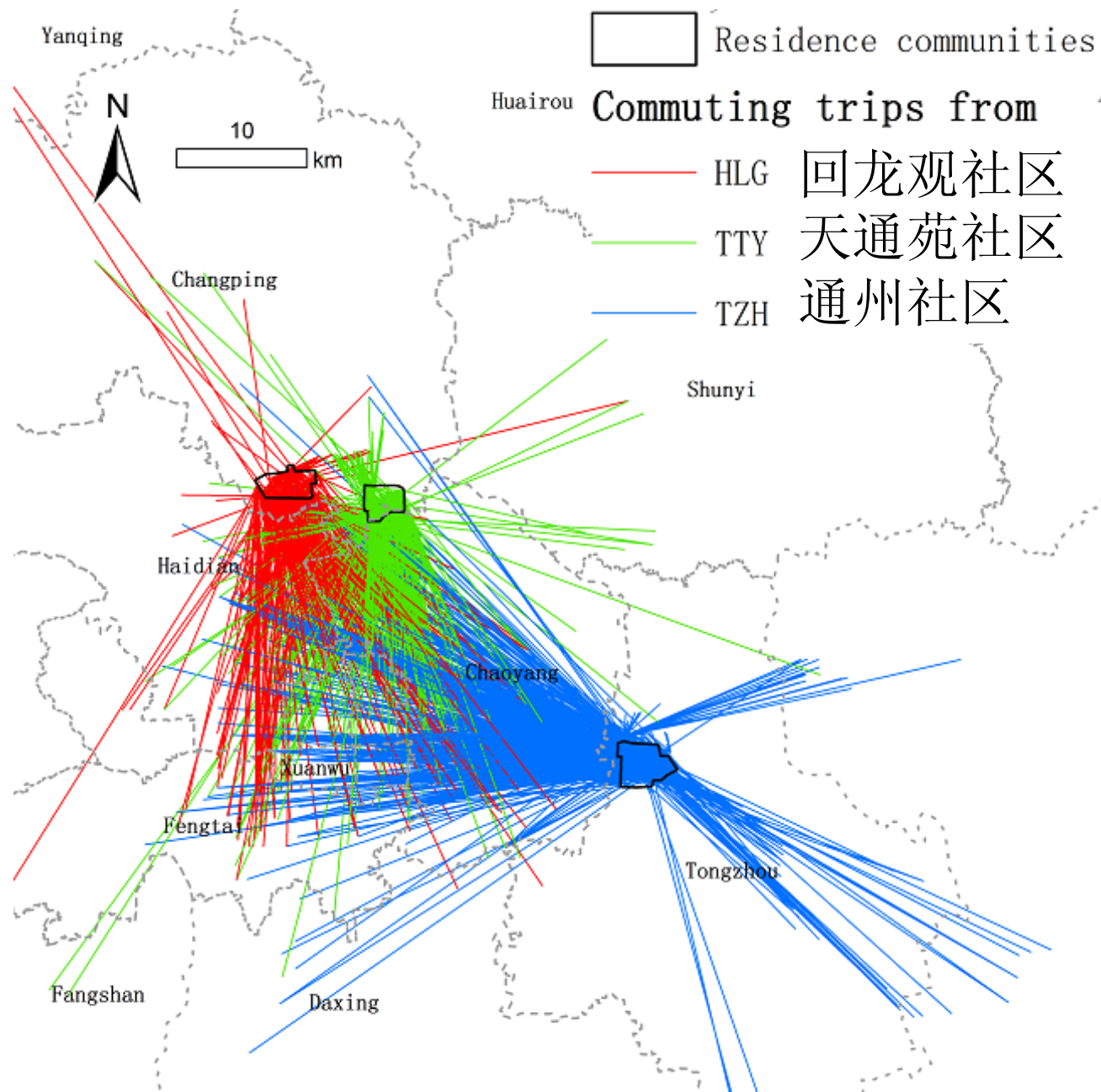
Bus commuting pattern
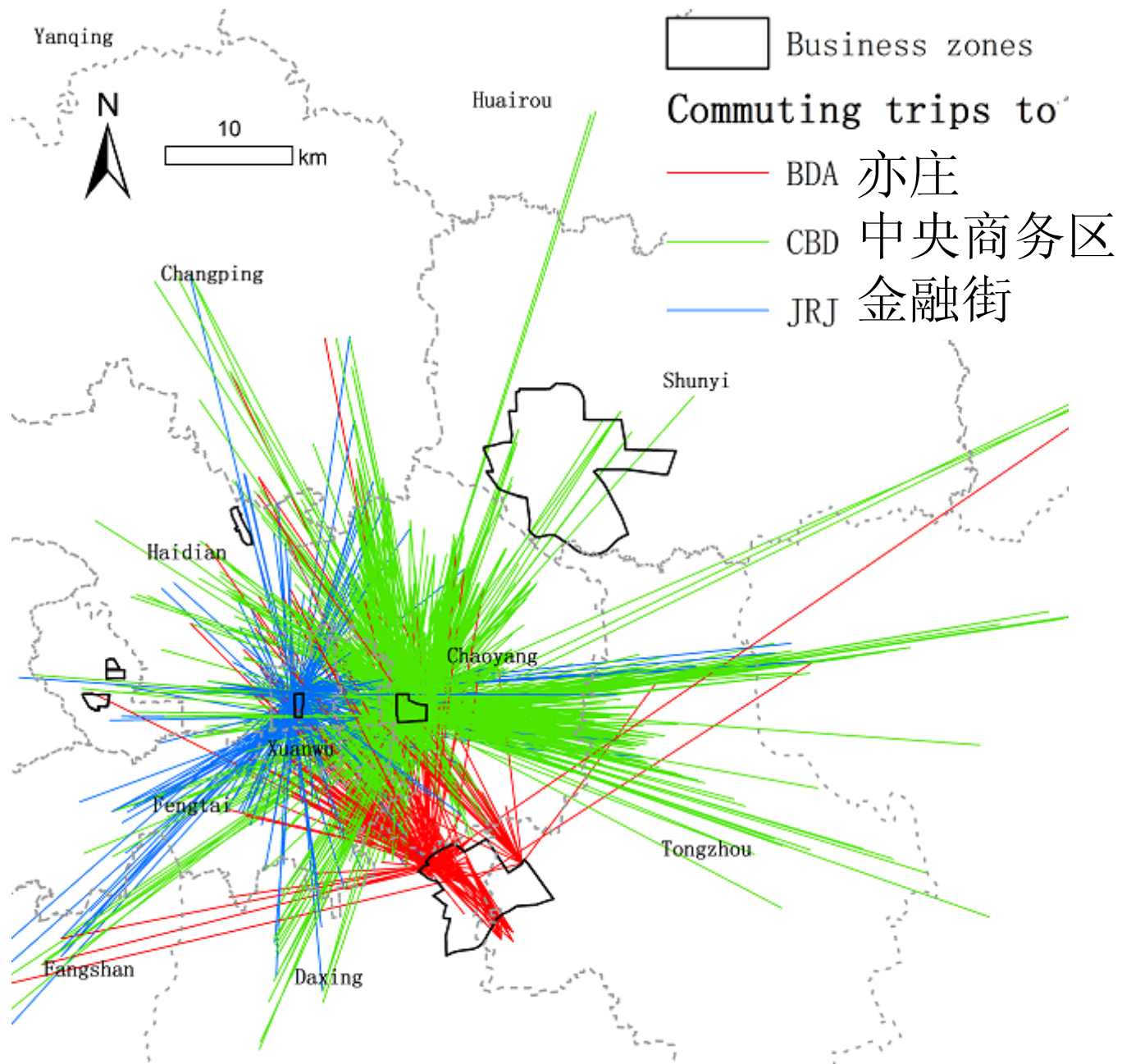
- **Where to where**
- **Extreme commuting**
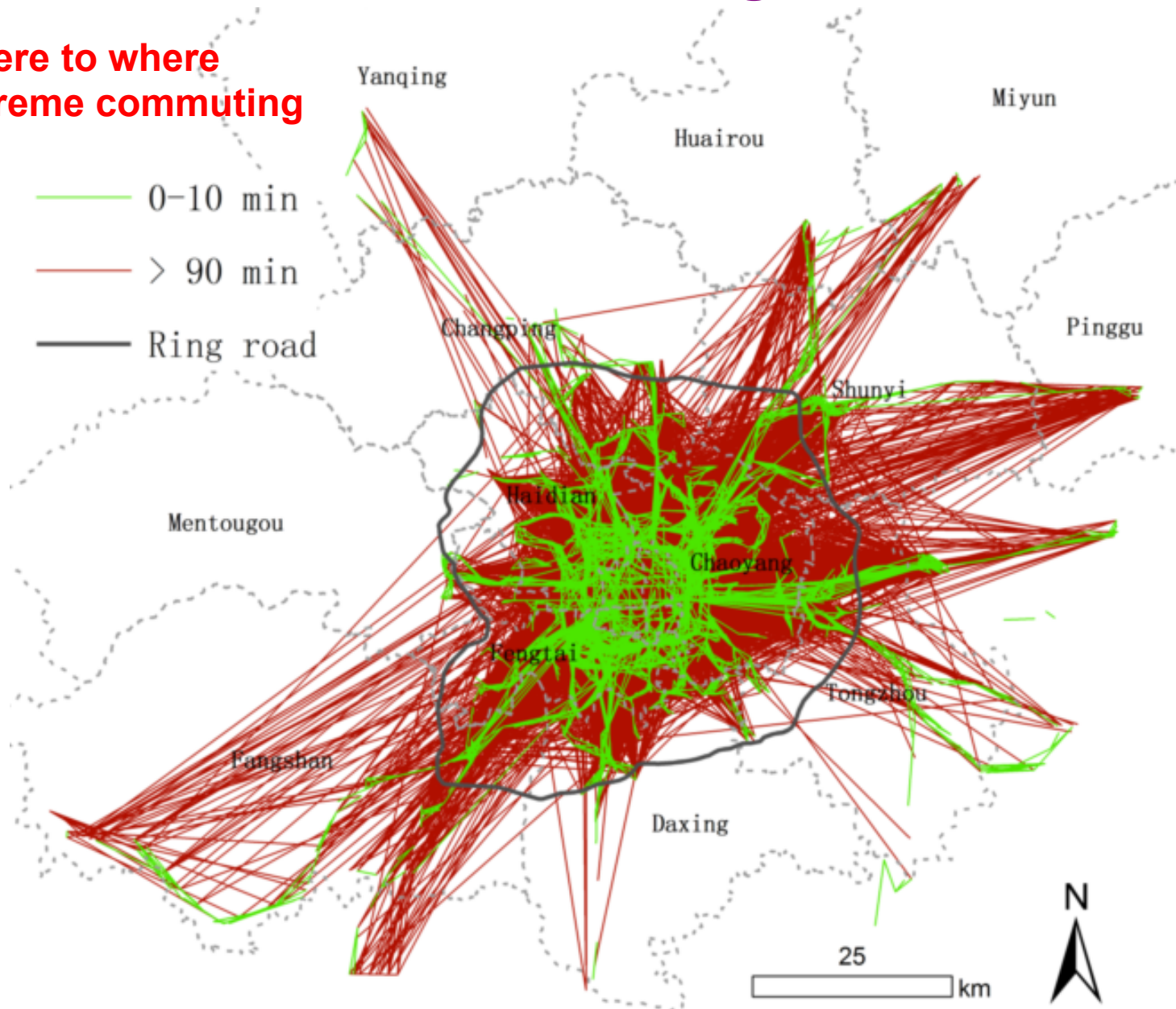
- **Aggregated in the TAZ level**
- **The head/tail division rule used**
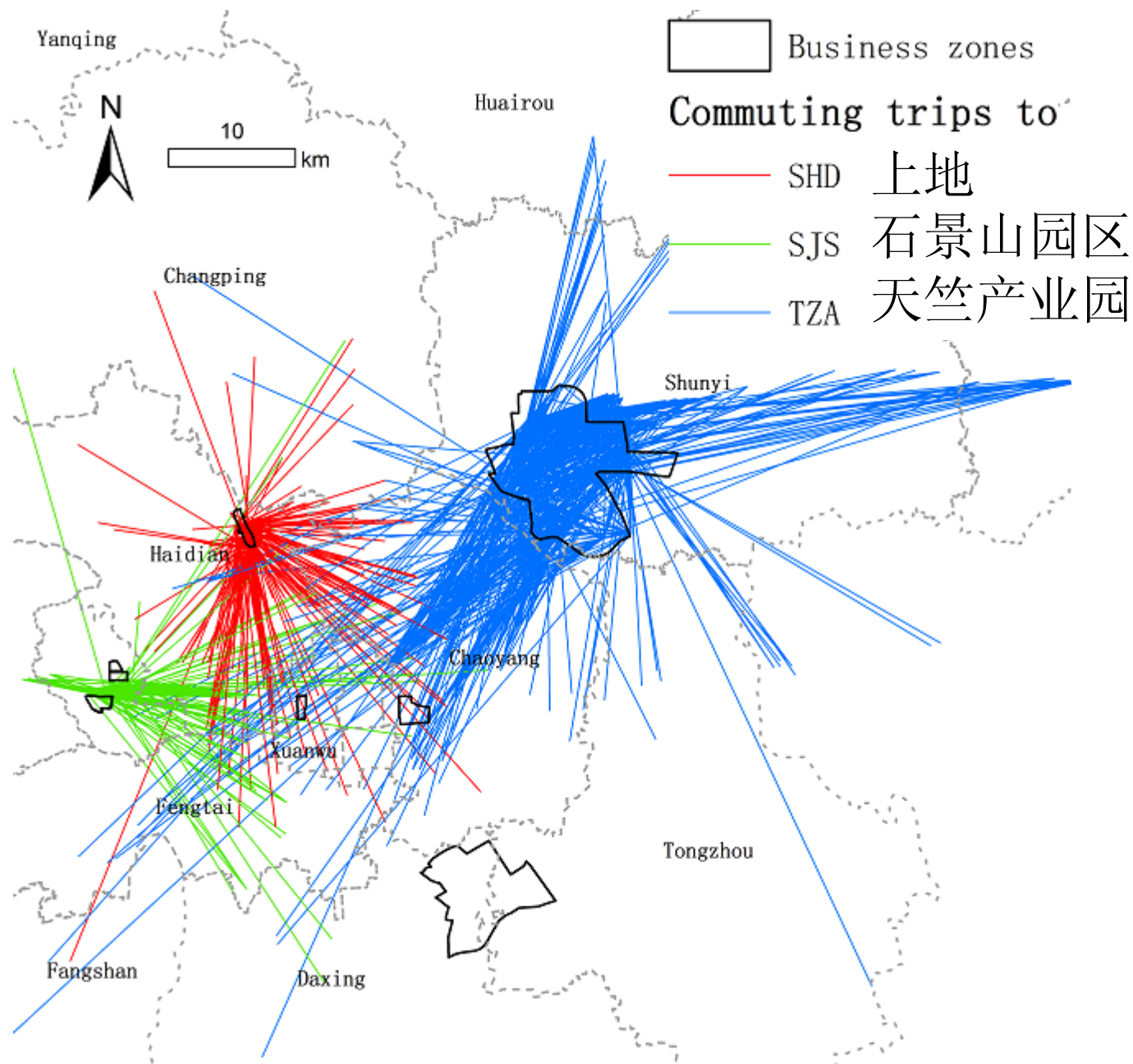- **Policy suggestion on BRTs**
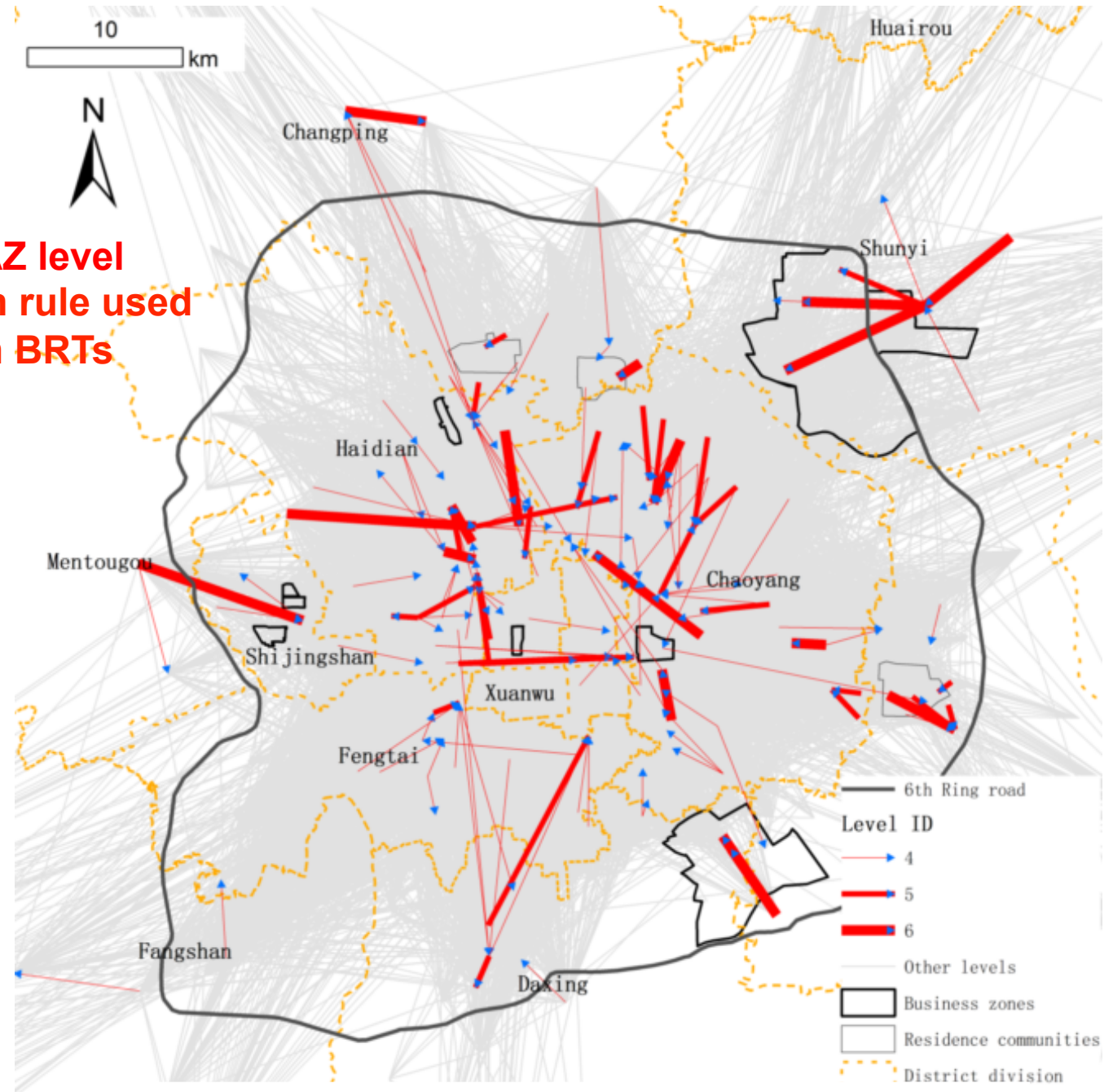
Inter-TAZ commuting mainstream

Commutes from main residence communities

Commutes to main business zones (1)

33

Commutes to main business zones (2)

34

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.

5 CONCLUSIONS AND DISCUSSION

Concluding remarks

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

Next steps

- Analyze SCD in 2010 with extra subway ridings
  - One week in April, 2010
  - 11 million cardholders

  - Share: 11.5% subway + 21.2% bus = 40%

- Move focus on non-commuting trips
  - Identify other activities combining POIs
    - Shopping
    - Recreation

Questions?

longying1980@gmail.com