Urban Computing With City Dynamics

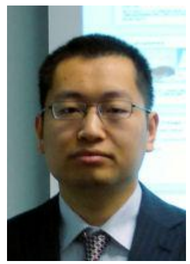
Yu Zheng

Ph.D., Researcher

Microsoft Research Asia

Students who have worked with me in urban computing

Yin Lou
@ Cornell



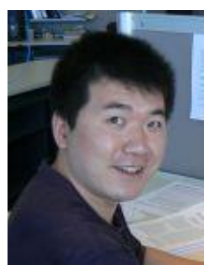
C. Y. Zhang
@ UNT

Jing Yuan
@ USTC

Ling-Yin Wei
@ NCTU

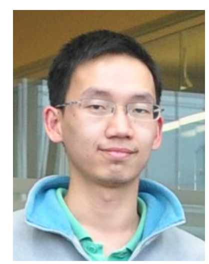
Kevin Zheng
@ UQ

Darshan
@ ETH



Wei Liu
@ U. Sydney

Liuhang Zhang
@ USTC



Yanchi Liu
@ USTB

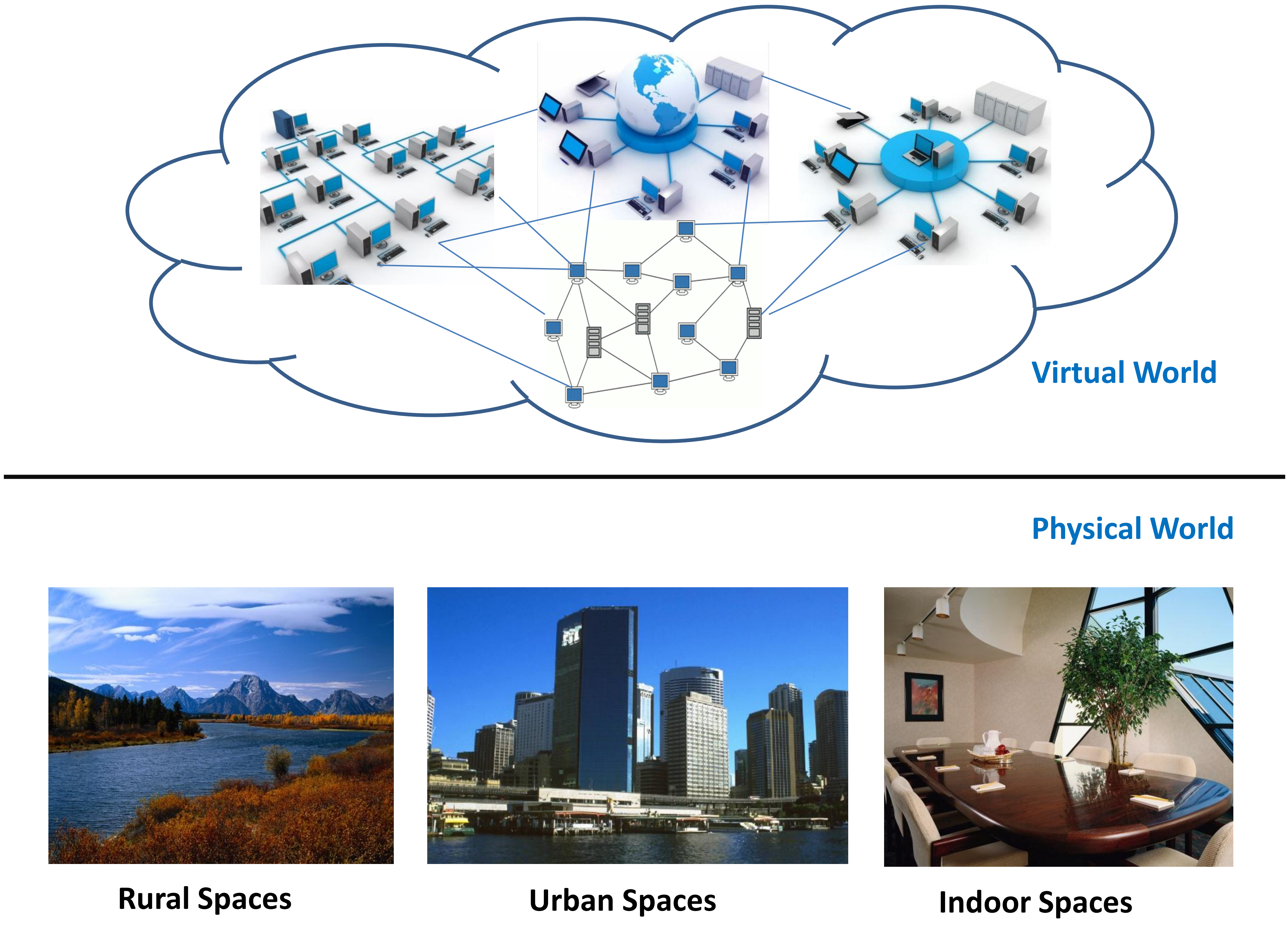
Hechen Liu
@ U. Florida

Outline

- Background
- Fundamental algorithms
- Application scenarios for end users
  - Driving direction service
  - Taxi recommendations
  - Travel itinerary suggestion
  - Other social applications
- Application scenarios for governments
  - Anomaly detection
  - Glean the problematic urban planning
  - Discover regions of different functions

Outline

- Background
- Fundamental algorithms
- Application scenarios for end users
  - Driving direction service
  - Taxi recommendations
  - Travel itinerary suggestion
  - Other social applications
- Application scenarios for governments
  - Anomaly detection
  - Glean the problematic urban planning
  - Discover regions of different functions


Virtual World
Physical World
Rural Spaces
Urban Spaces
Indoor Spaces

Why Urban Computing

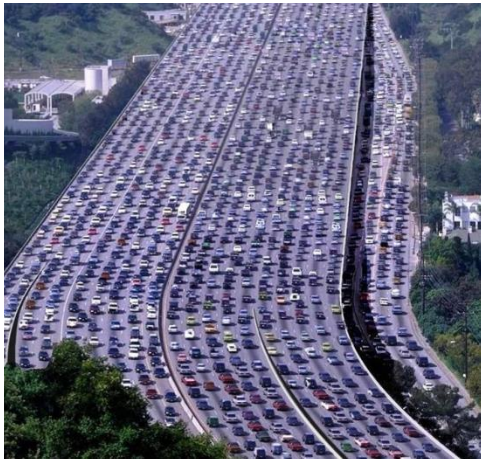
- 50% of people live in urban areas (just 0.4% of earth surface)
- The greatest wave of urbanization is coming

Years ago

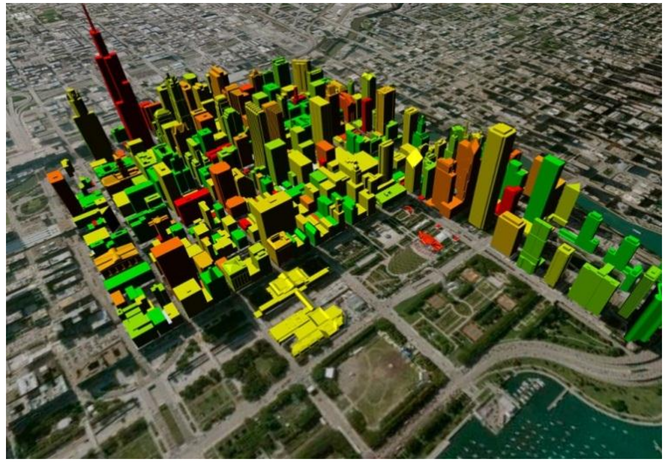
2010

Beijing

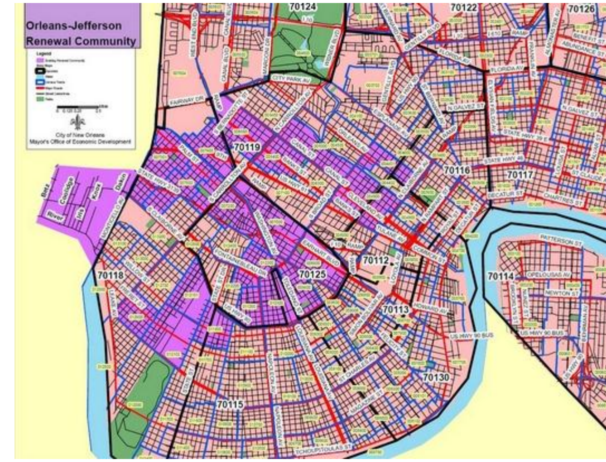
Shanghai



Traffic jams



Energy consumption



City renewal

Population

City reconstruction

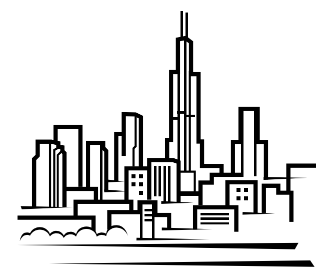
Pollution

Why Urban Computing

- Bigger cities do more with less – from Scientific American
  - Productive: economy, sciences, and technology
  - Greener: less energy consumption per person

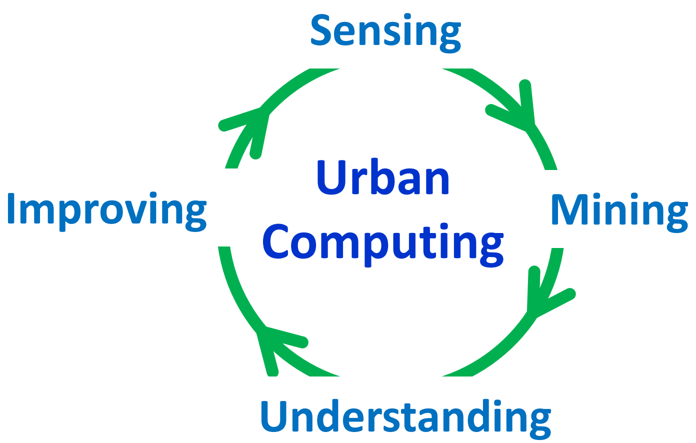
2M People

1M people

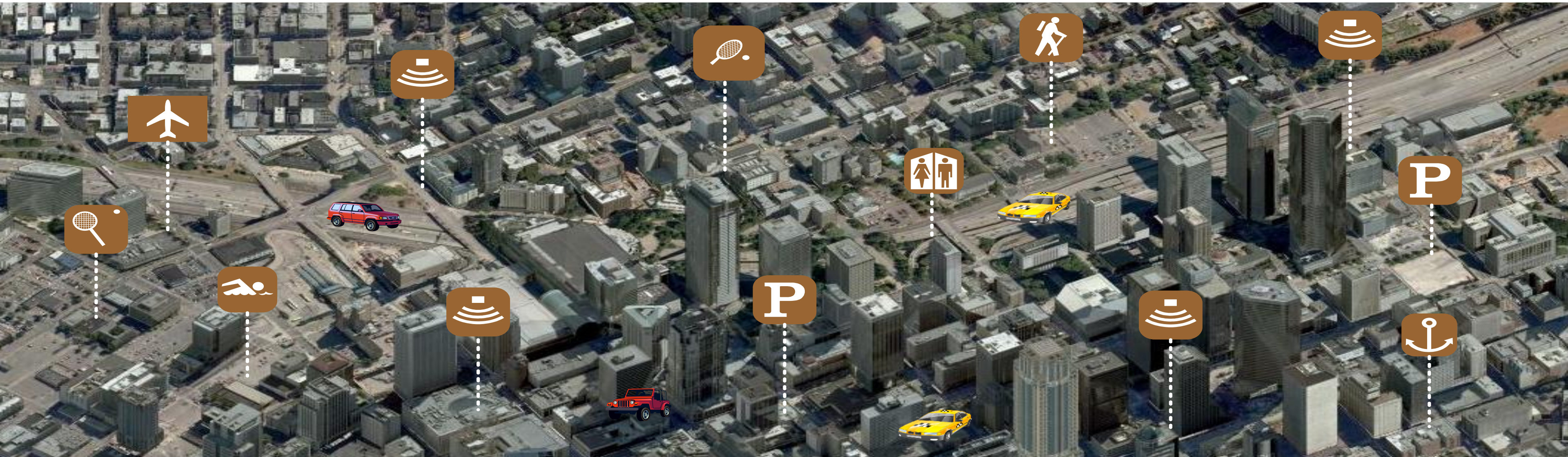


1M people

What's Urban Computing



Urban computing is emerging as a concept where every sensor, device, person, vehicle, building, and street in the urban areas can be used as a component to sense city dynamics to enable a city-wide computing to tackle the challenges in urban areas as so to serve people and cities.

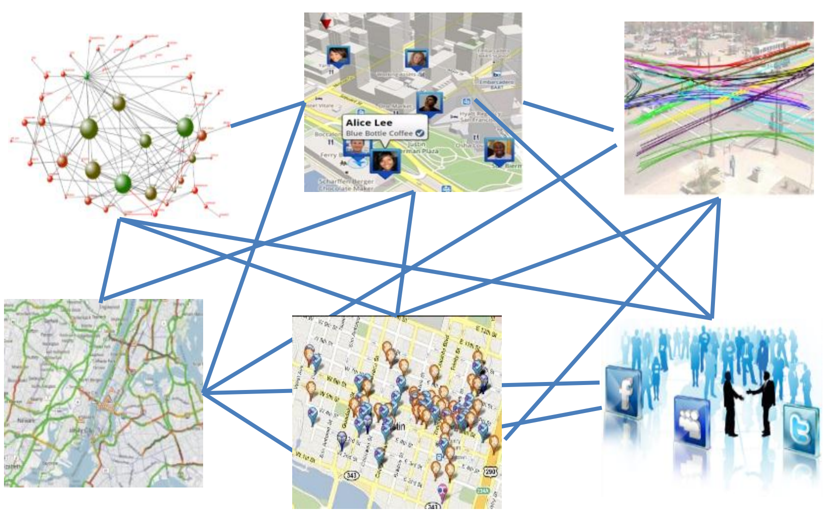


Differences and Relations

- Smart Cities
  - Current cities →Urban computing →Smart cities
  - Unobtrusively sensing (Leveraging what we already have)
- Internet of IoT
  - Infrastructure connecting objects
  - Lack of **human** and **social**
- Cloud Computing
  - Technology and platform
  - Many urban computing scenarios can be built on the Cloud

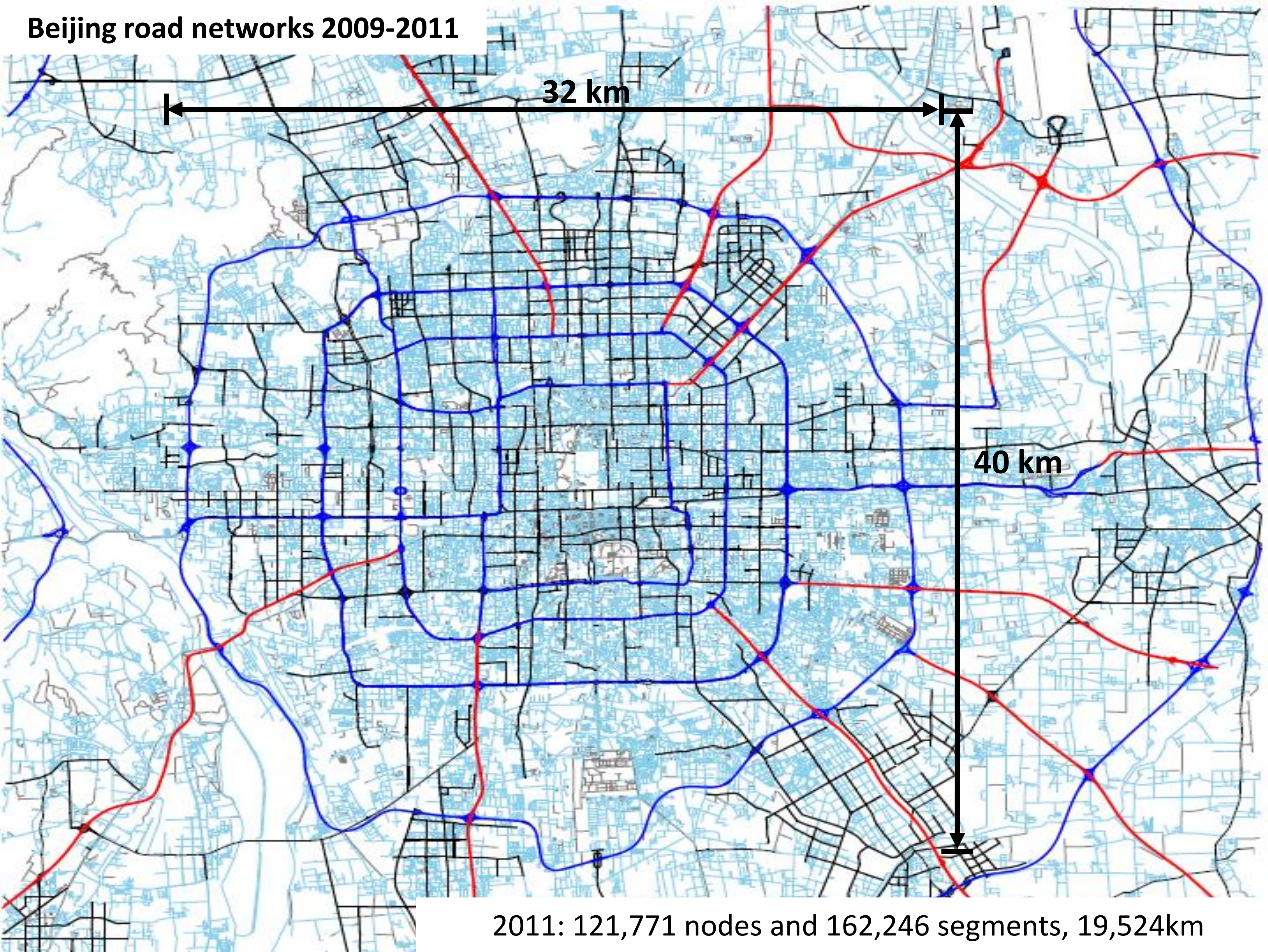
City Dynamics

- Scope
  - Traffic flow
  - Human mobility
  - Energy consumption
  - Environment
  - Economic
  - Populations
  - ……



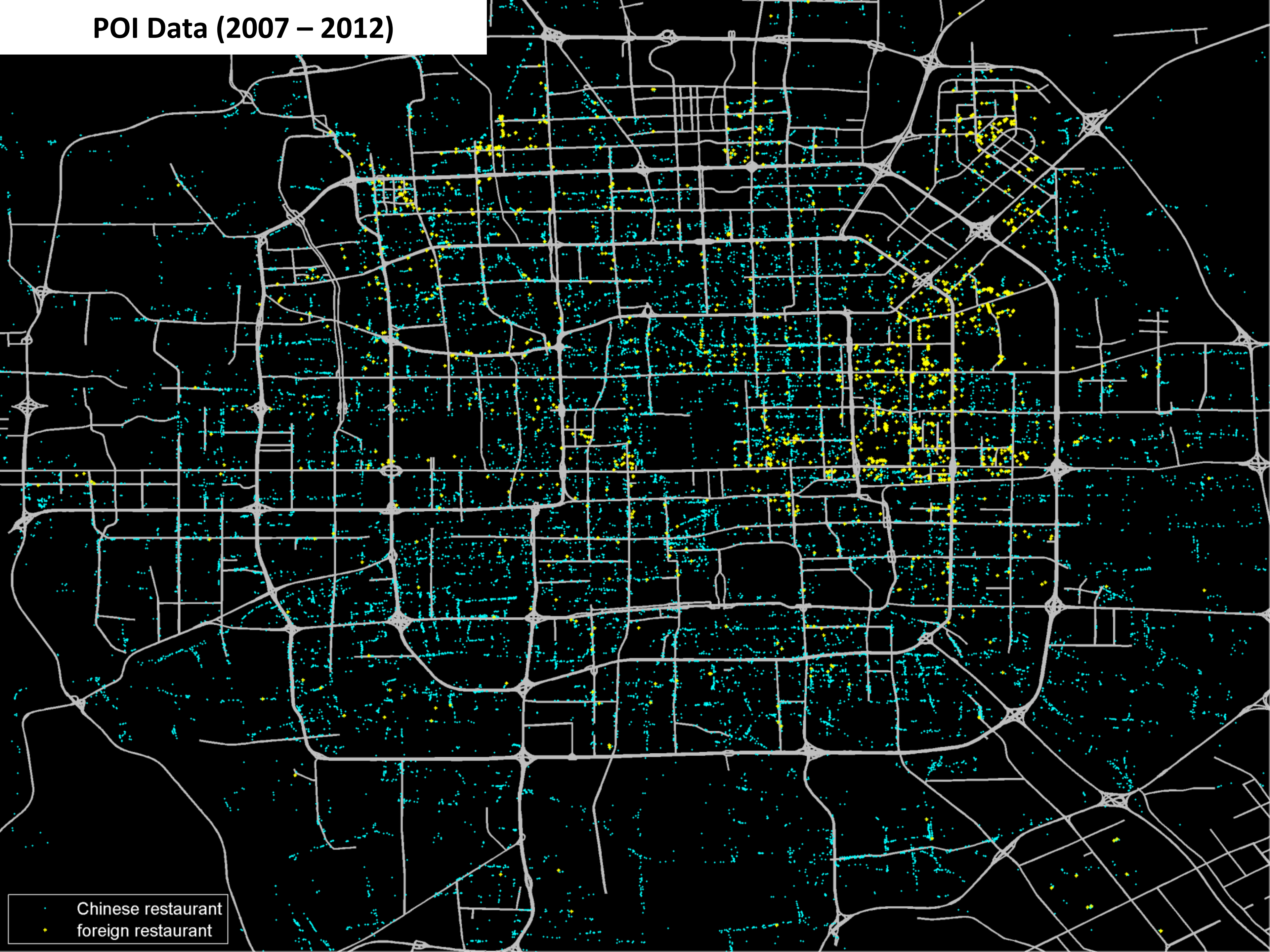
- Data available
  - Mobile phone signal
  - GPS traces of vehicles and people
  - Ticketing data in public transportation systems
  - User-generated content
  - Transportation sensor networks
    - Camera and loop sensors
    - Parking lots
  - Environmental sensor network
    - Air quality
    - Temperature
    - Radiation
  - Transaction records of credit cards
  - ….

Beijing road networks 2009-2011



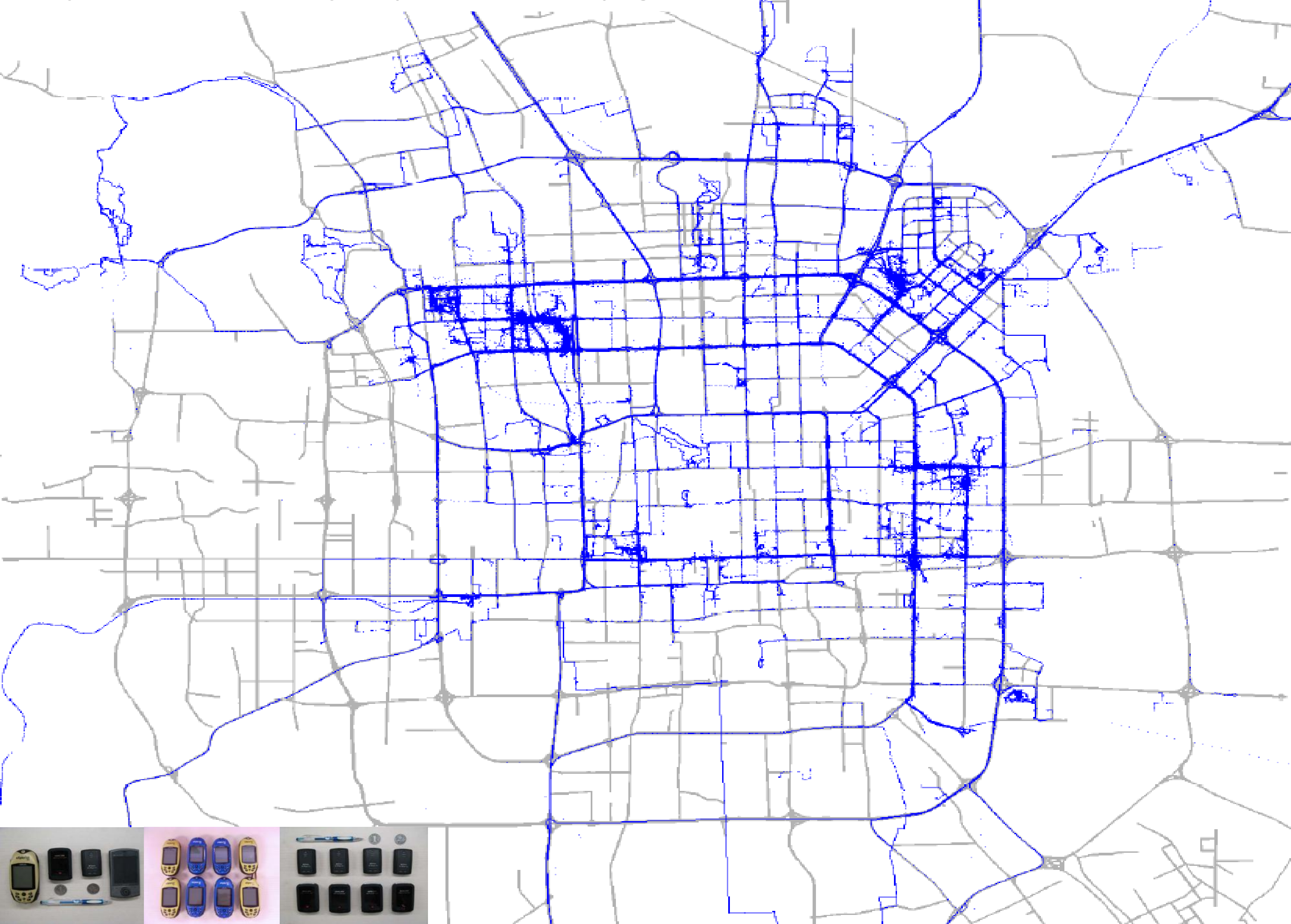
2011: 121,771 nodes and 162,246 segments, 19,524km

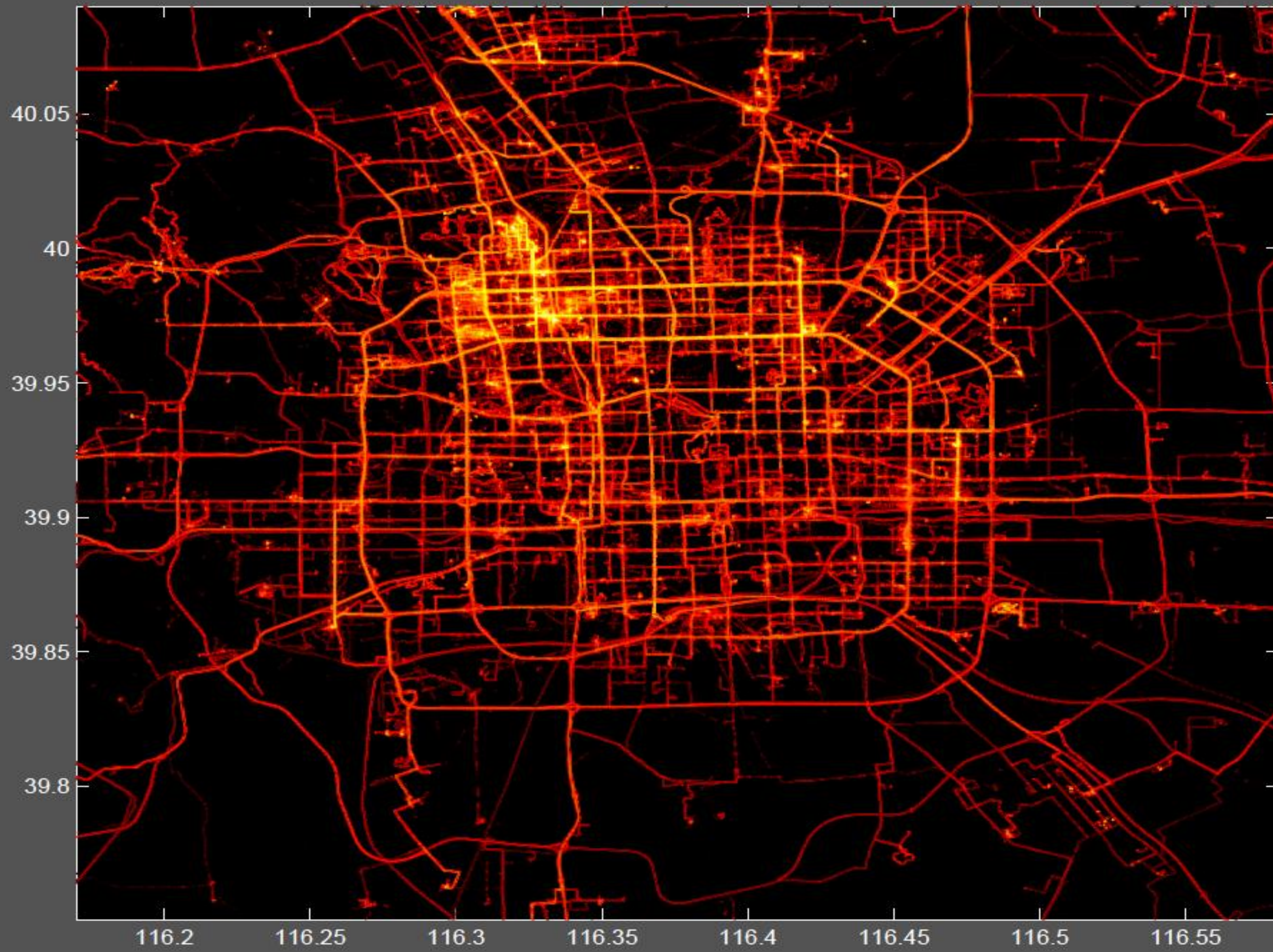
POI Data (2007 – 2012)



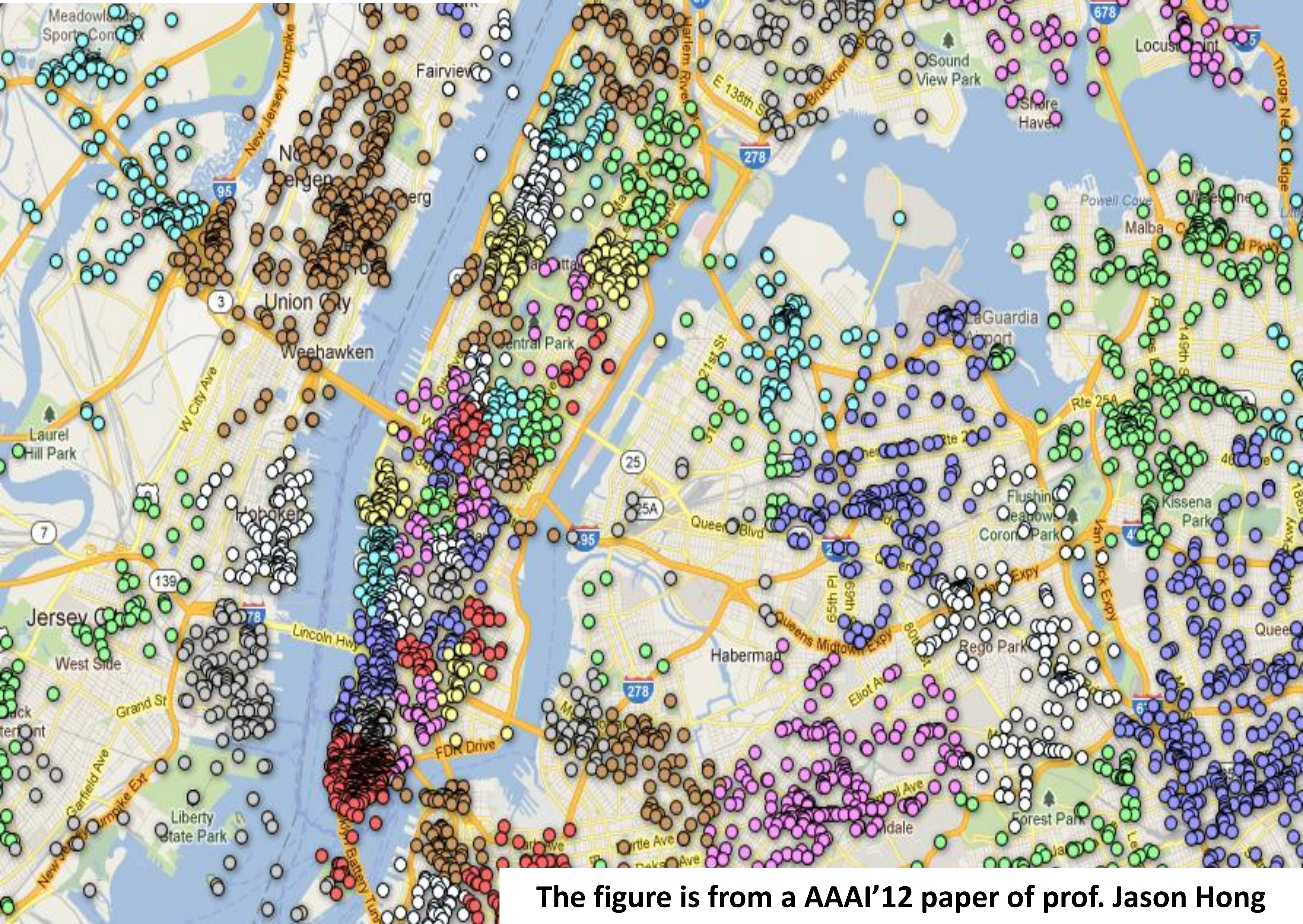
pub/bar
theaters

People’s location history (trajectories) in Beijing



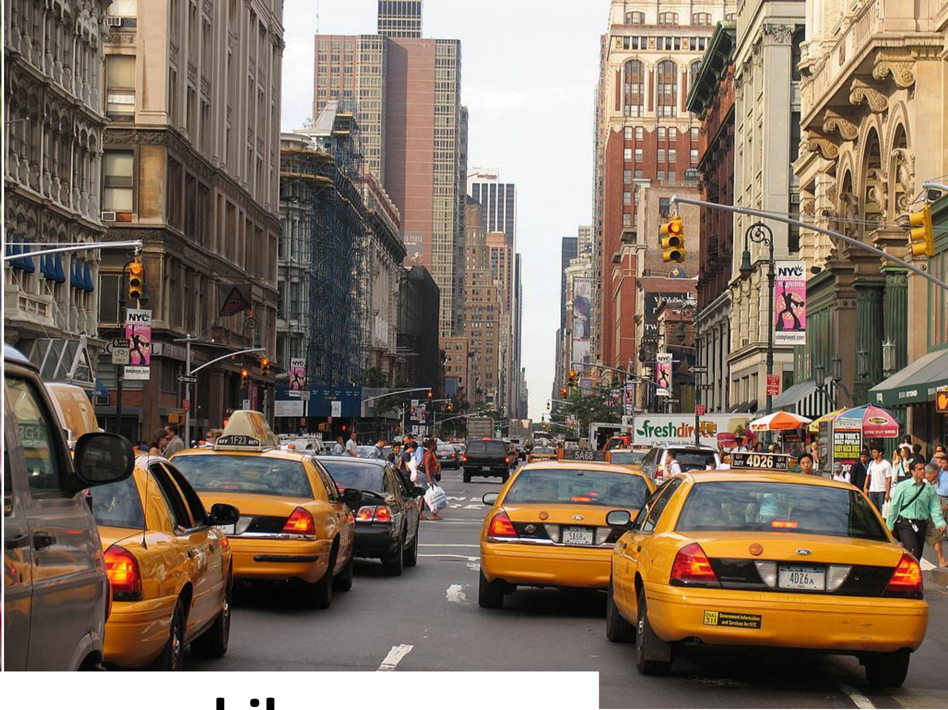

40.05
40
39.95
39.9
39.85
39.8
116.2
116.25
116.3
116.35
116.4
116.45
116.5
116.55

People’s check-in data in LBSNs

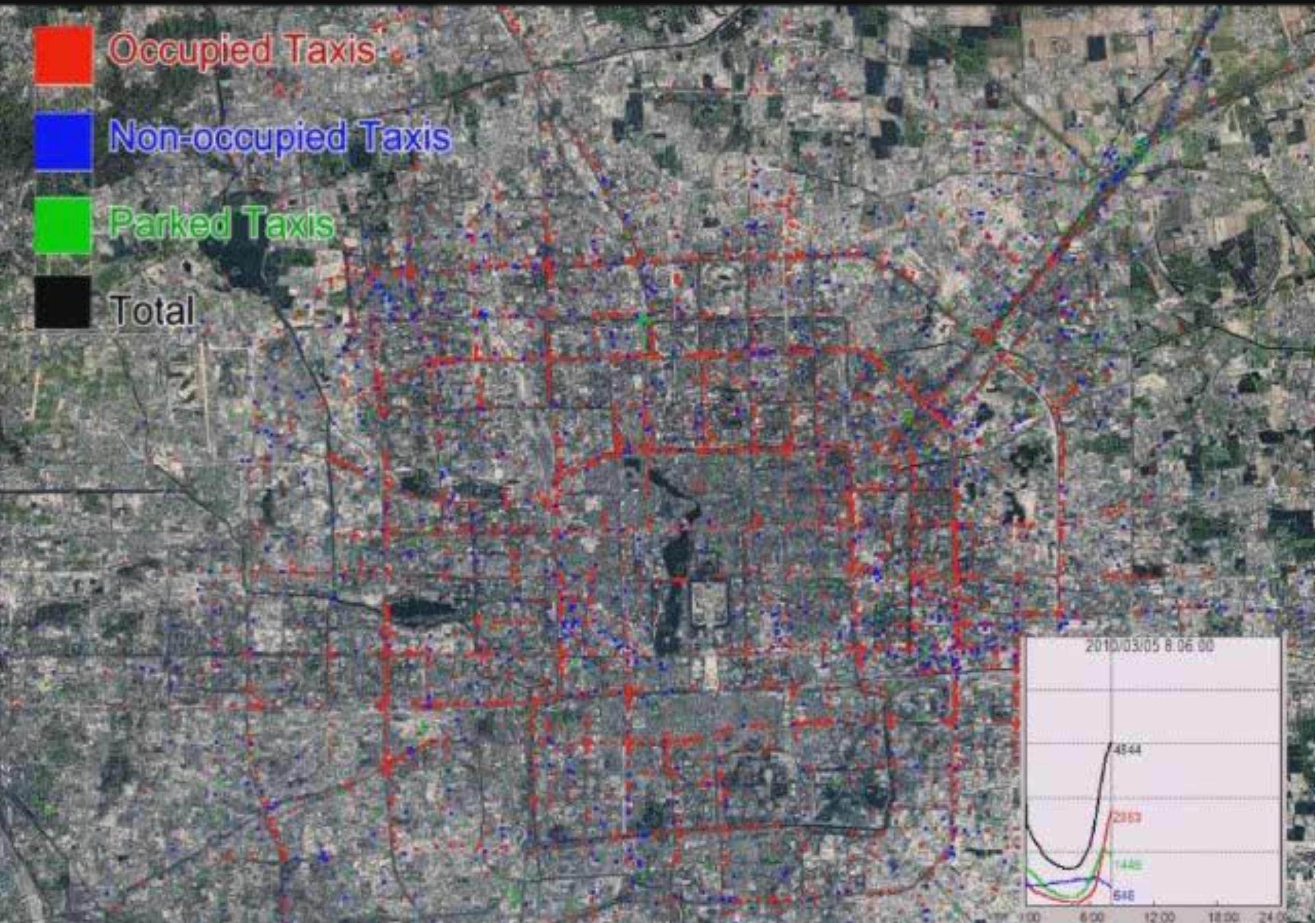


The figure is from a AAAI’12 paper of prof. Jason Hong

GPS trajectories of 33,000 taxis in 2009, 2010, and 2011

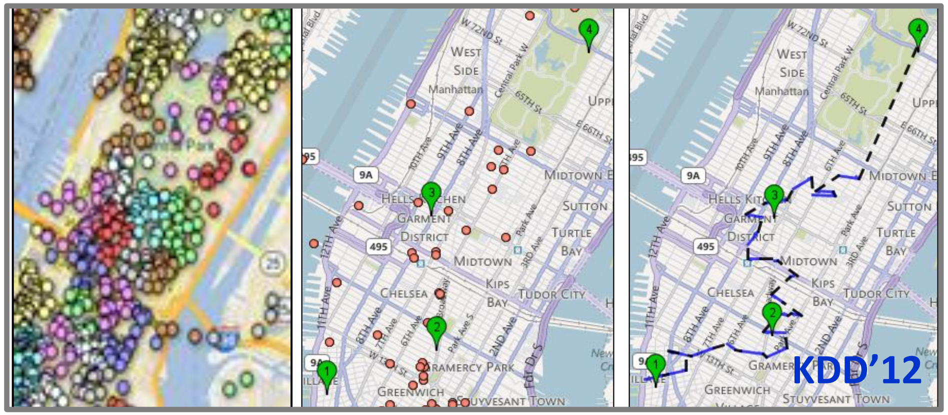


GPS-equipped taxis are **mobile sensors**

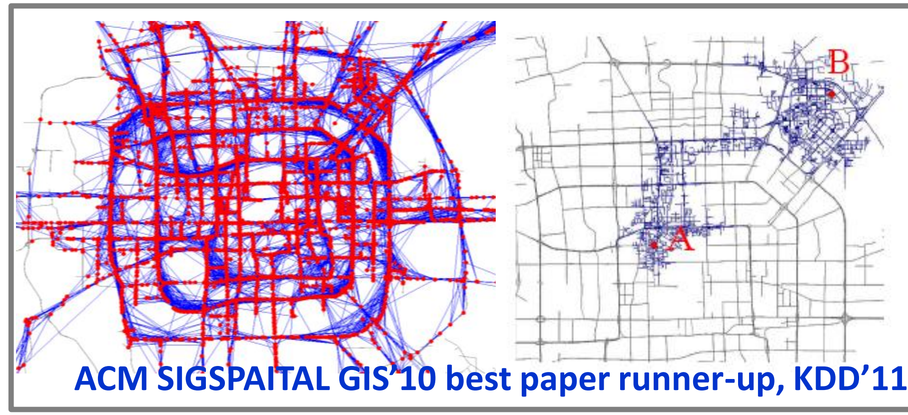


Occupied Taxis
Non-occupied Taxis
Parked Taxis
Total
2010/03/05 8:06:00
4844
646
6:00
12:00
18:00

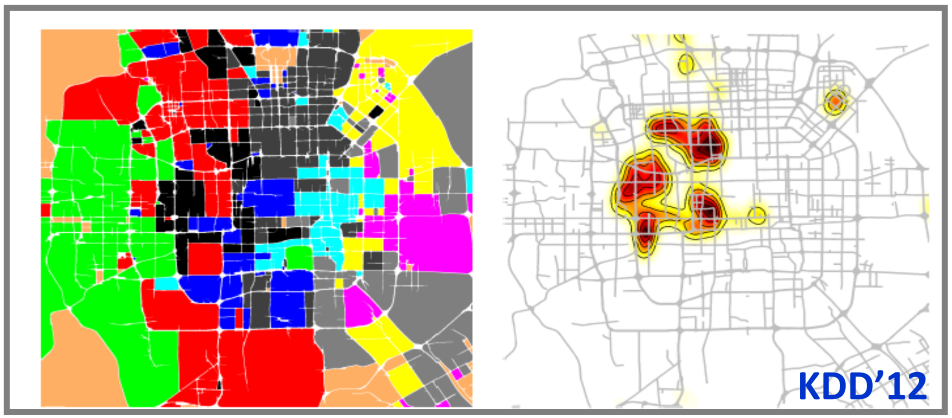
Rank	Cities	Countries/Regions	Taxicabs
1	The Mexico city	Mexico	103,000
2	Bangkok	Thailand	80,000
3	Seoul	South Korea	73,000
4	**Beijing**	**China**	**67,000**
5	Tokyo	Japan	60,000
6	Shanghai	China	50,000
7	New York City	USA	48,300
8	Buenos Aires	Argentina	45,000
9	Moscow	Russia	40,000
10	St.Paul	Brazil	37,000
11	Tianjin	China	35,000
12	Taipei	Taiwan	31,000
13	New Taipei City	Taiwan	23,500
14	Singapore	Singapore	23,000
15	Osaka	Japan	20,000
16	Hong Kong	China	18,000
17	Wuhan	China	18,000
18	London	England	17,000
19	Harbin	China	17,000
20	Guangzhou	China	16,000
21	Shenyang	China	15,000
22	Paris	France	15,000



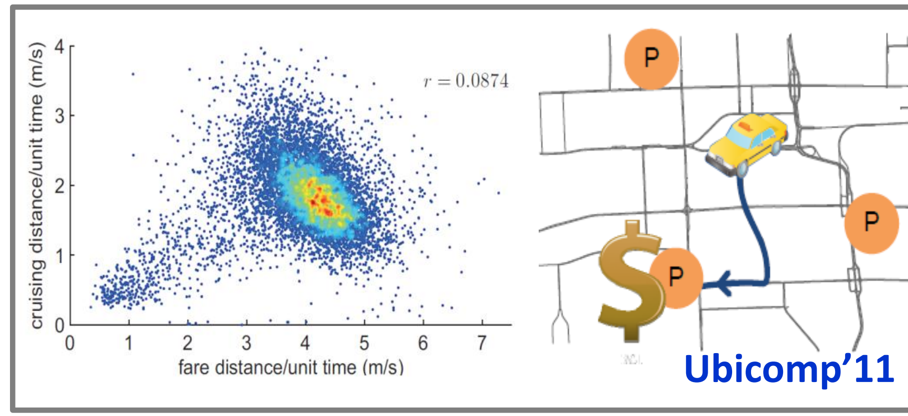
Route Construction from Uncertain Trajectories



Finding Smart Driving Directions

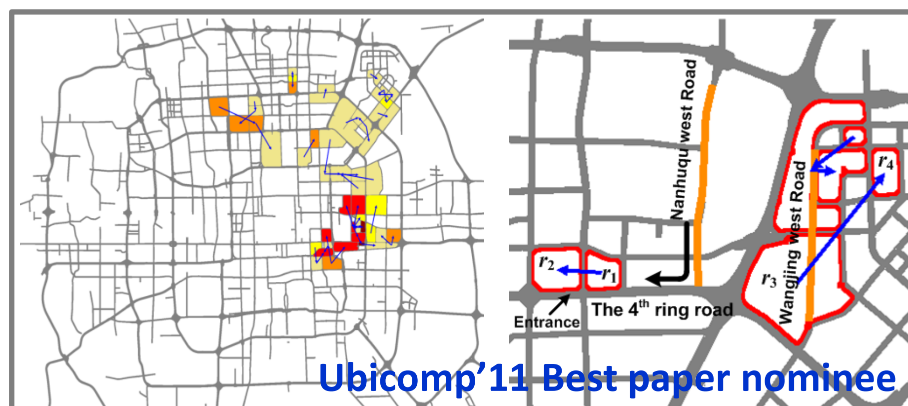


Discovery of Functional Regions



Passengers-Cabbie Recommender system

Anomalous Events Detection KDD’11



Urban Computing for Urban Planning

Outline

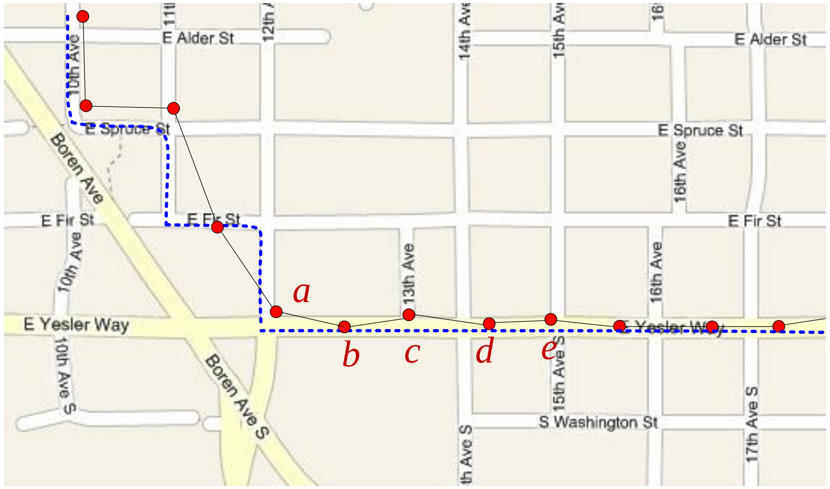
- Background
- Fundamental algorithms
- Application scenarios for end users
  - Driving direction service
  - Taxi recommendations
  - Travel itinerary suggestion
- Application scenarios for governments
  - Anomaly detection
  - Glean the problematic urban planning
  - Discover regions of different functions

Fundamental Algorithms

- Map-matching
- Map segmentation
- Stay point detection

Map-matching

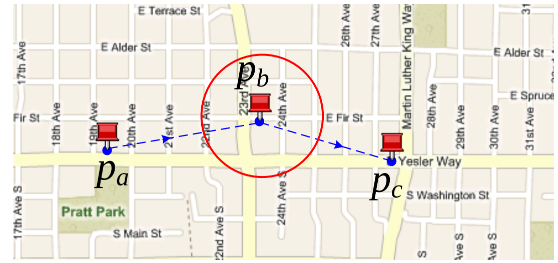
- Project a trajectory onto a road network
- A fundamental step in many transportation applications
  - Navigation and driving
  - Traffic analysis
  - Taxi dispatching and recommendations

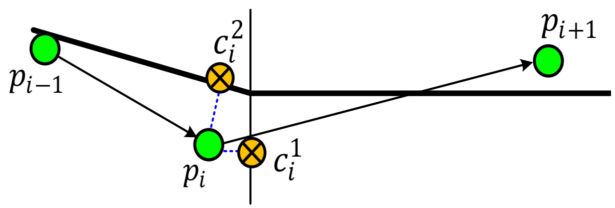


Yin Lou, Chengyang Zhang, Yu Zheng. [Map-Matching for Low-Sampling-Rate GPS Trajectories](). ACM SIGSPATIAL GIS 2009

Map-matching

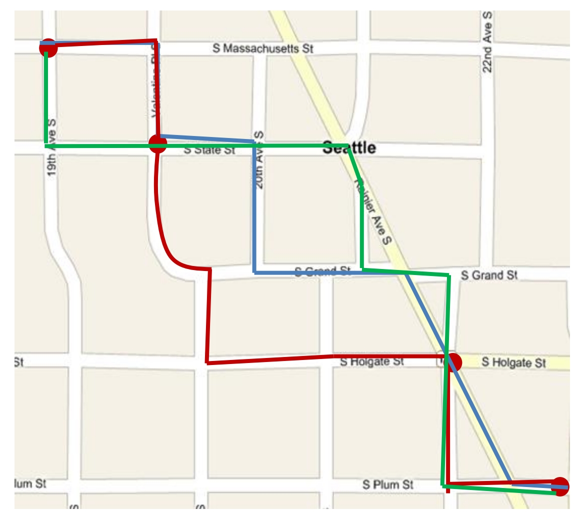
- Challenges (low-sampling rate)
- Solution
  - Consider both local and global information
  - Incorporating both spatial and temporal features





$$N(c_i^j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i^j-\mu)^2}{2\sigma^2}} \qquad V(c_{i-1}^t \rightarrow c_i^s) = \frac{d_{i-1\rightarrow i}}{w_{(i-1,t)\rightarrow(i,s)}}$$

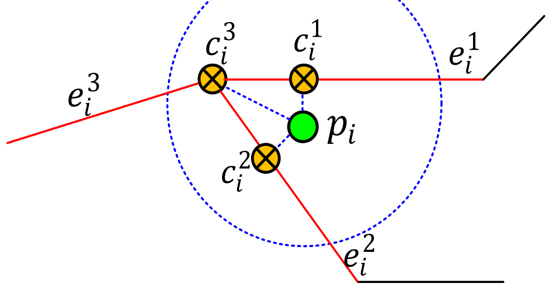
$$F_t(c_{i-1}^t \rightarrow c_i^s) = \frac{\sum_{u=1}^{k}(e'_u.v \times \bar{v}_{(i-1,t)\rightarrow(i,s)})}{\sqrt{\sum_{u=1}^{k}(e'_u.v)^2} \times \sqrt{\sum_{u=1}^{k}\bar{v}_{(i-1,t)\rightarrow(i,s)}^2}}$$



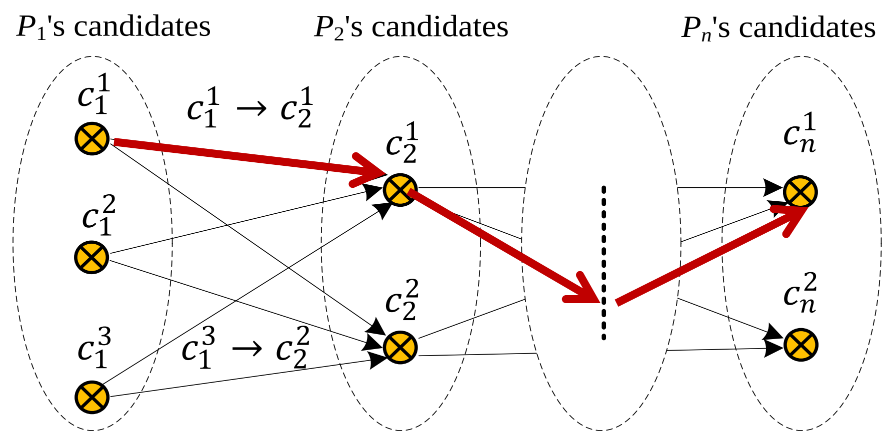
Yin Lou, Chengyang Zhang, Yu Zheng. [Map-Matching for Low-Sampling-Rate GPS Trajectories]. ACM SIGSPATIAL GIS 2009

Map-matching

- Basic Solution
  - Find candidate road segments for each GPS point
  - Calculate local and global features
  - Dynamic programing



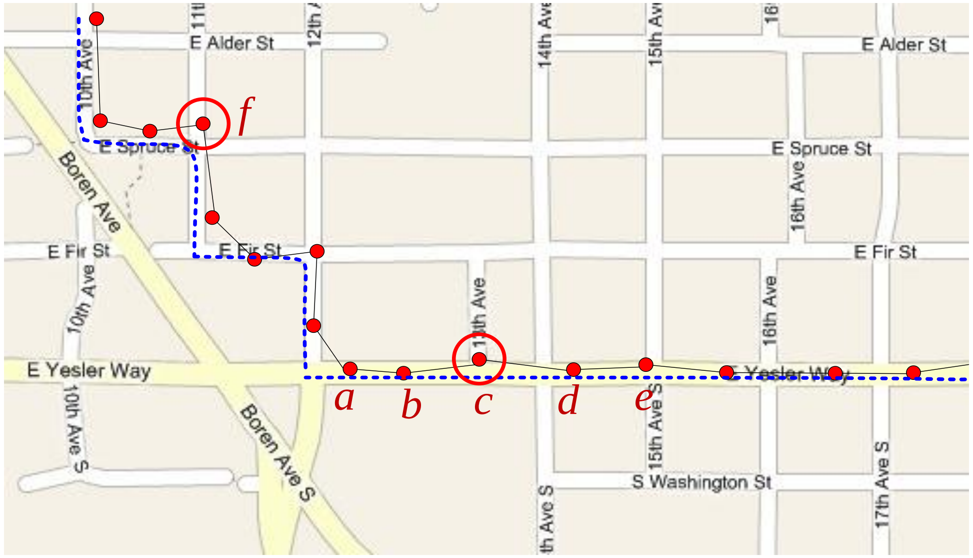
$$N(c_i^j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i^j-\mu)^2}{2\sigma^2}}$$



$$F(c_{i-1}^t \to c_i^s) = F_s(c_{i-1}^t \to c_i^s) * V(c_{i-1}^t \to c_i^s)$$

Map-matching

- Advanced solution
  - Mutual influence
  - Weighted influence



Jing Yuan, Yu Zheng, et al. [An Interactive-Voting based Map Matching Algorithm](). In MDM 2010.

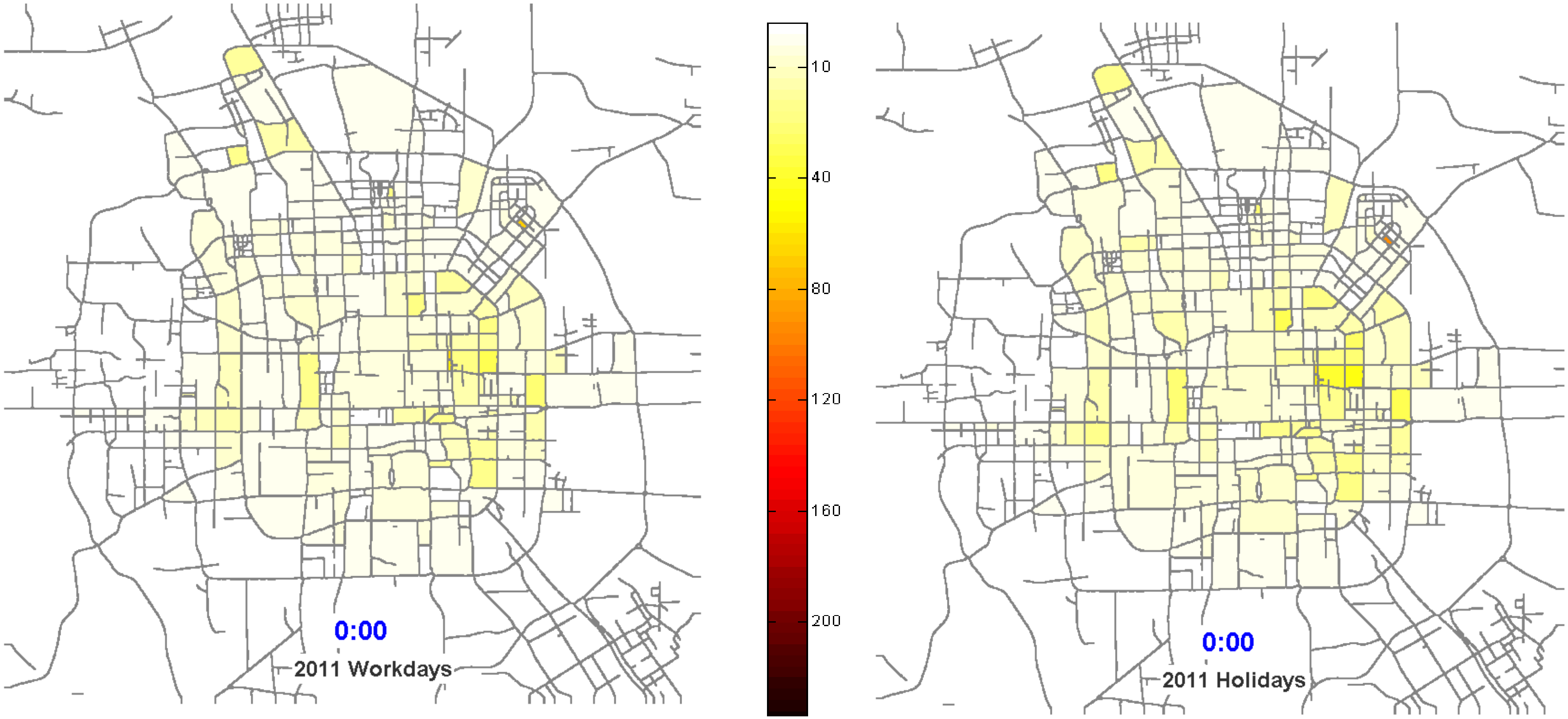
Map Segmentation

- Partition a map into disjoint regions
- Three possible methods

(a) Grid-based (b) Hierarchy-based (c) Morphology-based

Nicholas Jing Yuan, Yu Zheng, Xing Xie. Segmentation of Urban Areas Using Road Networks. MSR-TR-2012-65. 2012.

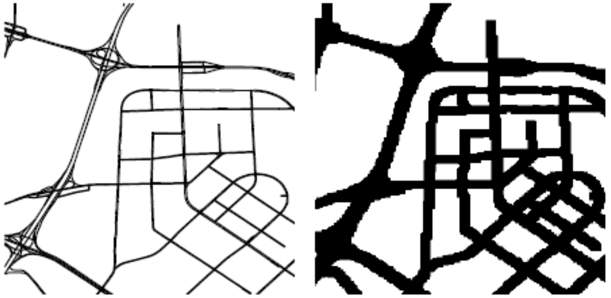
Heat Maps of Beijing (2011)



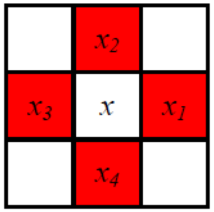
Nicholas Jing Yuan, Yu Zheng, Xing Xie. [Segmentation of Urban Areas Using Road Networks](). MSR-TR-2012-65. 2012.

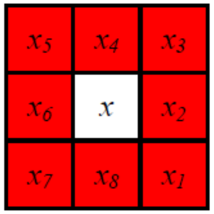
Map Segmentation

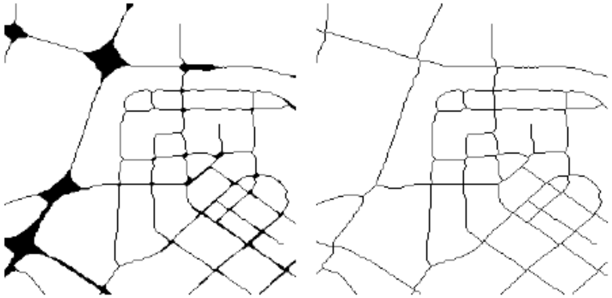
- Morphology-based segmentation method
  - Represent a road network with a raster model
  - Dilation
  - Thing
  - Connected component labeling



(a) Before dilation (b) After dilation





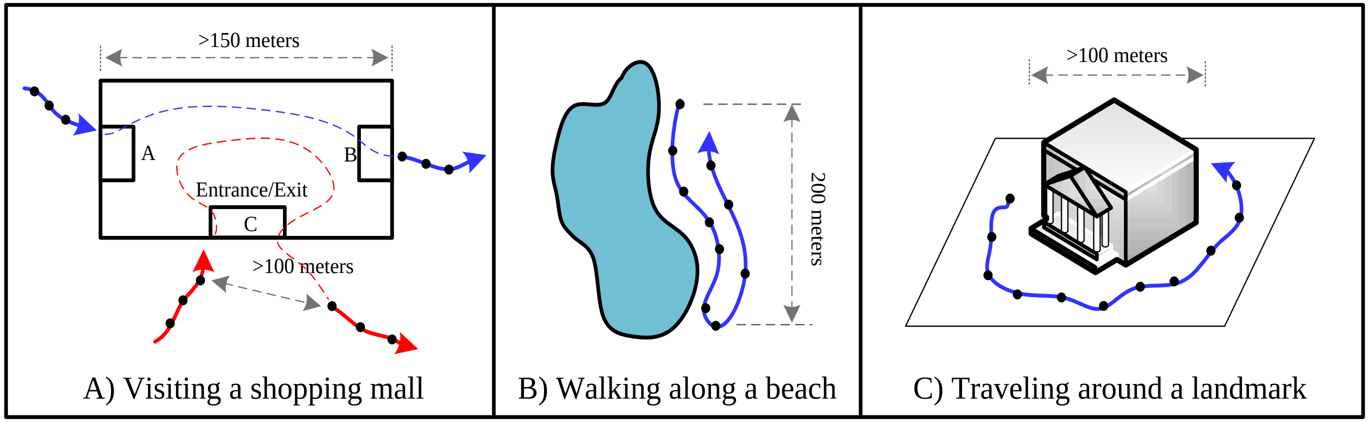


(a) After 7 iterations (b) Until convergence

Nicholas Jing Yuan, Yu Zheng, Xing Xie. [Segmentation of Urban Areas Using Road Networks](). MSR-TR-2012-65. 2012.

Stay Point Detection

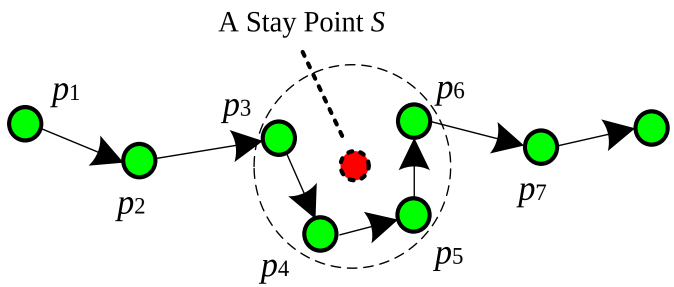
- Stay point:
  - A location where an individual has stayed for a while (t) within a distance threshold (d)
  - Carry semantic meanings than other points
  - Does not only mean remaining stationary



A) Visiting a shopping mall B) Walking along a beach C) Traveling around a landmark

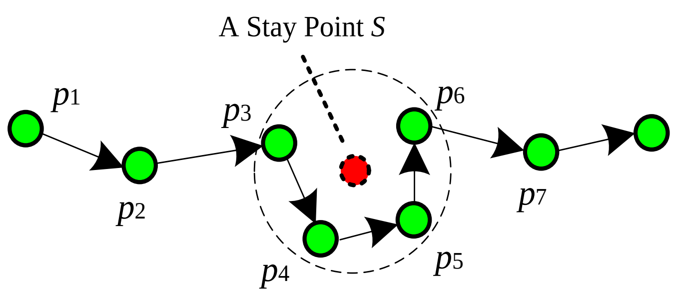
Stay Point Detection

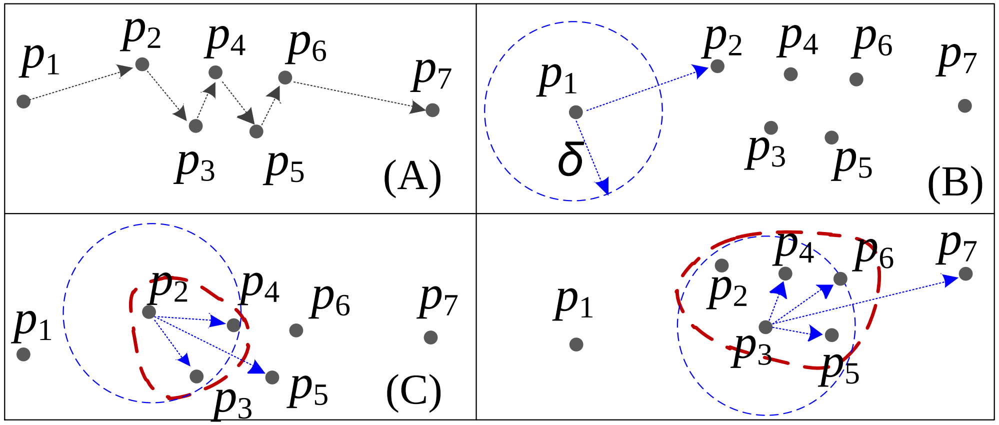
- Applications
  - Modeling human location history or human mobility
  - Parking place detection



Stay Point Detection

- Solution
  - A sort of density based clustering
  - Two thresholds: t and d
  - Two version presented in two papers





Quannan Li, Yu Zheng, et al. Mining user similarity based on location history. ACM SIGSPATIAL GIS 2008

Jing Yuan, Yu Zheng, et al, Where to Find My Next Passenger? , UbiComp 2011

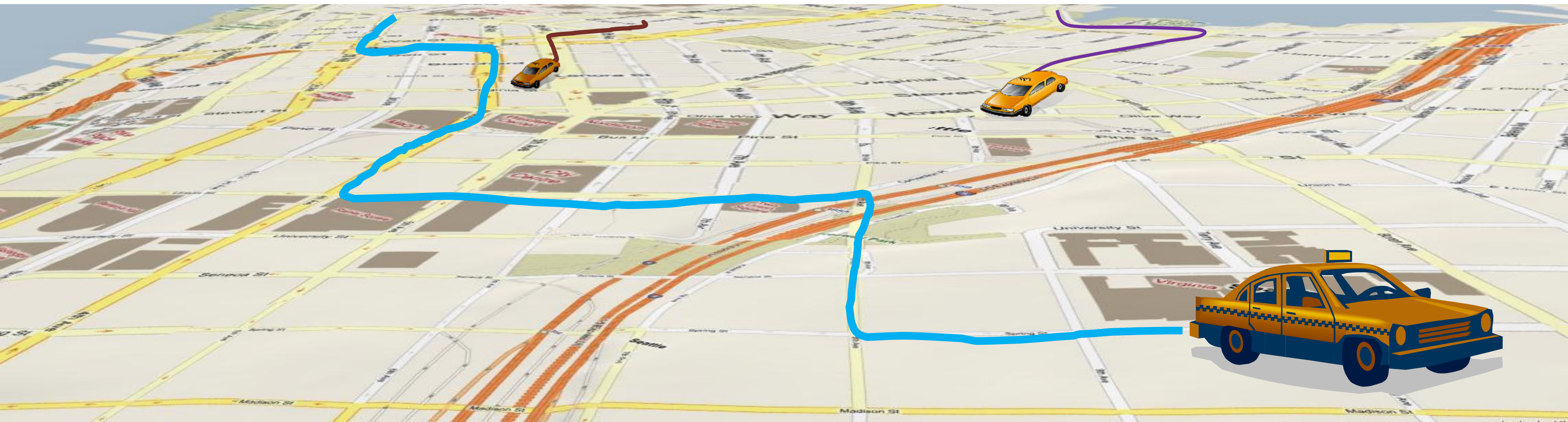
References

- Map matching
  - Yin Lou, Chengyang Zhang, Yu Zheng, et al. Map-Matching for Low-Sampling-Rate GPS Trajectories. ACM SIGSPATIAL GIS 2009.
  - Jing Yuan, Yu Zheng, et al. An Interactive-Voting based Map Matching Algorithm. In MDM 2010.
- Map segmentation
  - Nicholas Jing Yuan, Yu Zheng, Xing Xie. Segmentation of Urban Areas Using Road Networks. MSR-TR-2012-65. 2012
- Stay point detection
  - Quannan Li, Yu Zheng, et al. Mining user similarity based on location history. ACM SIGSPATIAL GIS 2008
  - Jing Yuan, Yu Zheng, et al, Where to Find My Next Passenger? , UbiComp 2011

Outline

- Background
- Fundamental algorithms
- Application scenarios for end users
  - Driving direction service
  - Taxi recommendations
  - Travel itinerary suggestion
  - Other social applications
- Application scenarios for governments
  - Anomaly detection
  - Glean the problematic urban planning
  - Discover regions of different functions

Finding Smart Driving Directions



Driving Direction Based on Taxi Trajectories

- A *time-dependent*, user-specific, and self-adaptive driving directions service using
  - GPS trajectories of a large number of taxicabs
  - GPS log of an end user

Physical Routes

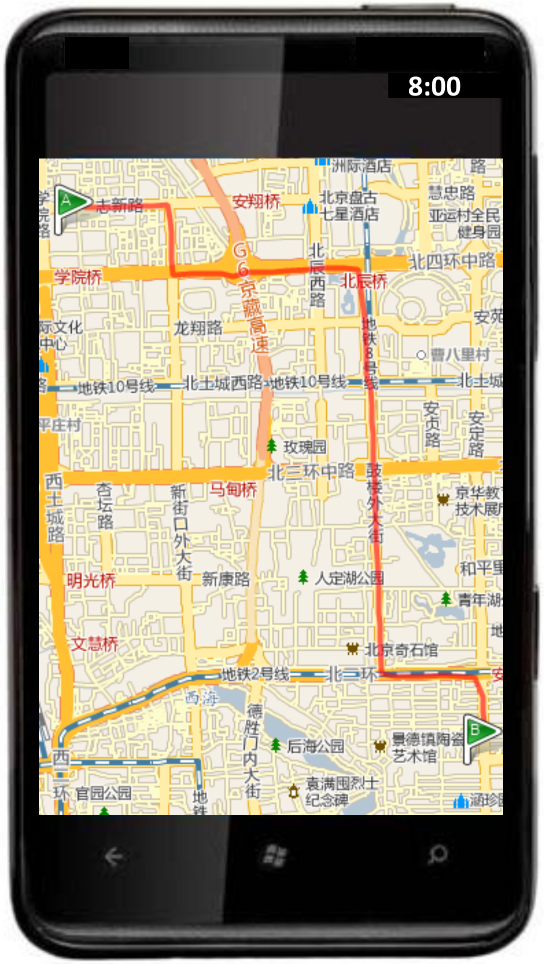
Traffic flows

Drive behavior

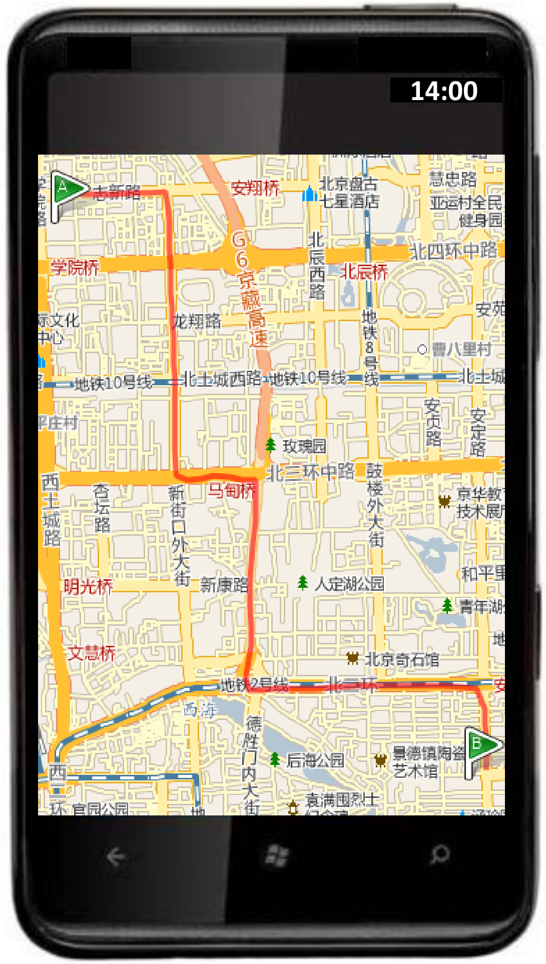
ACM SIGSPATIAL GIS 2010 best paper runner-up award and a publication on KDD 2011

Driving Direction Based on Taxi Trajectories

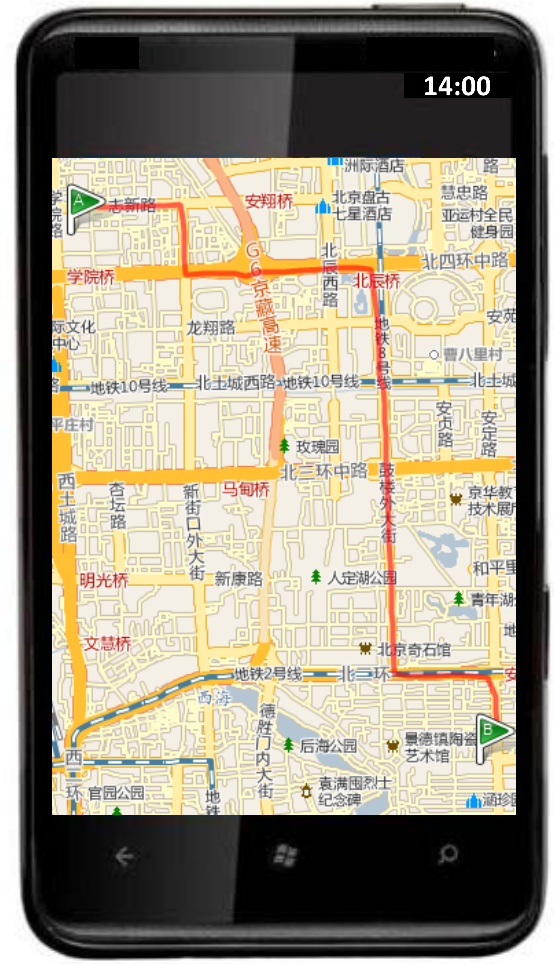
Driver A



Driver A

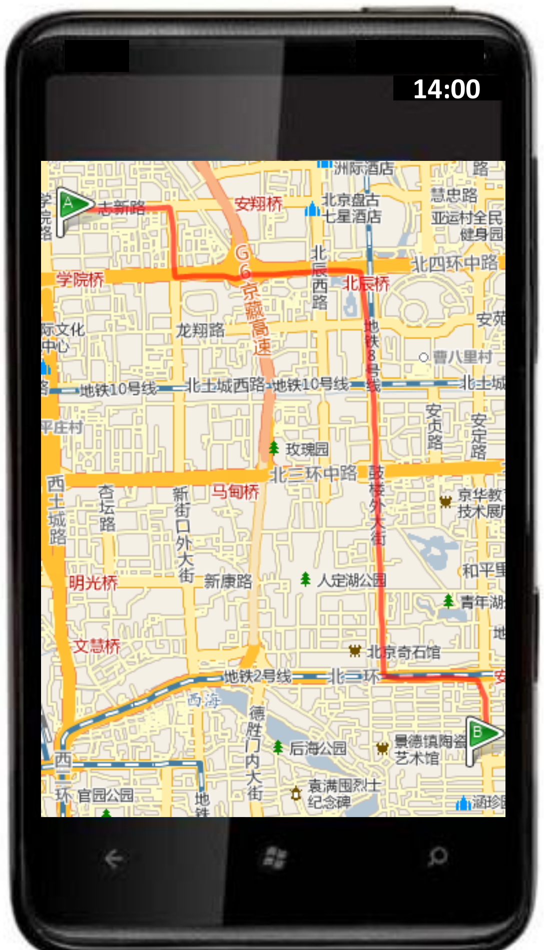


Driver B



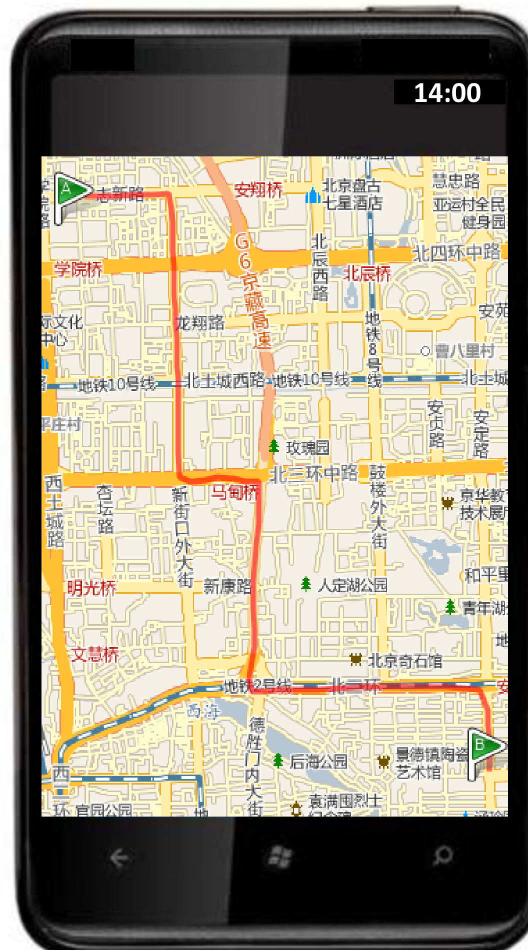
Driving Direction Based on Taxi Trajectories

Driver B



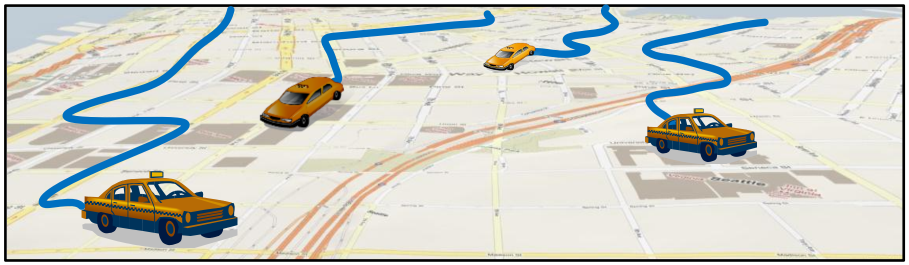
Log user B's driving routes for 1 month

Driver B



Motivation

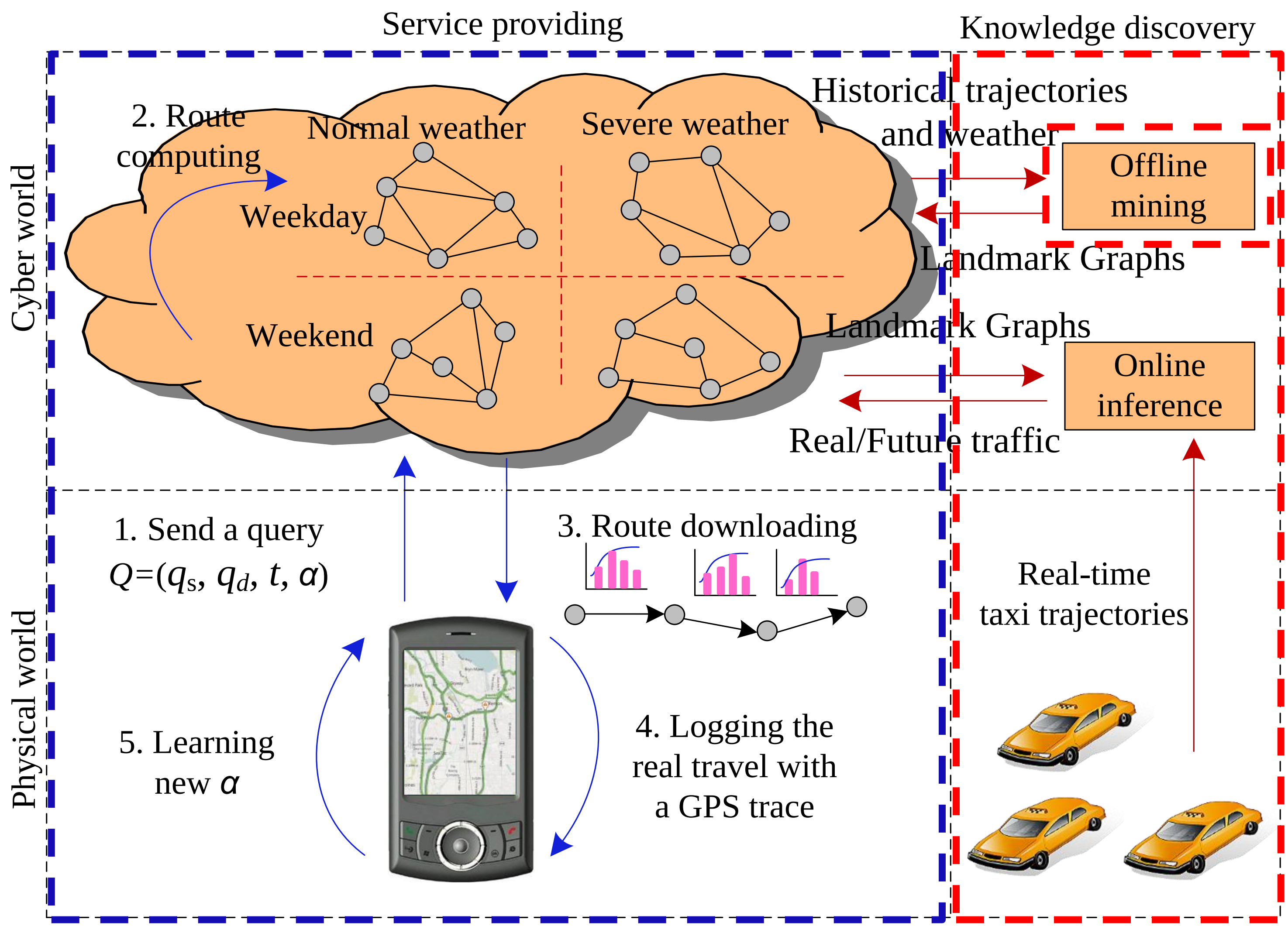
- Taxi drivers are **experienced** drivers
- GPS-equipped taxis are **mobile sensors**
- GPS logs imply the **drive behavior** of a user



Human Intelligence + Traffic patterns

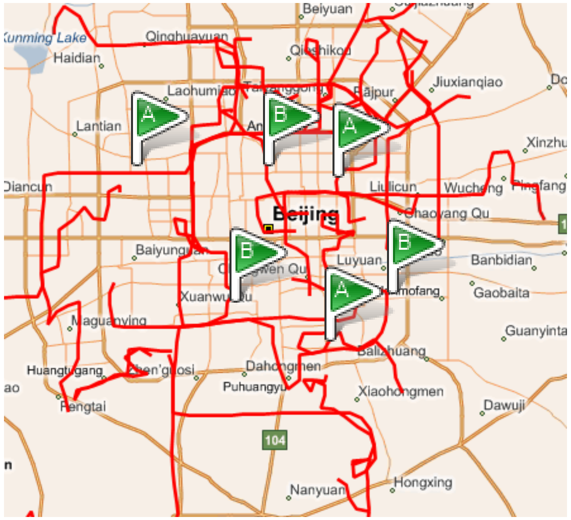
Drive behavior

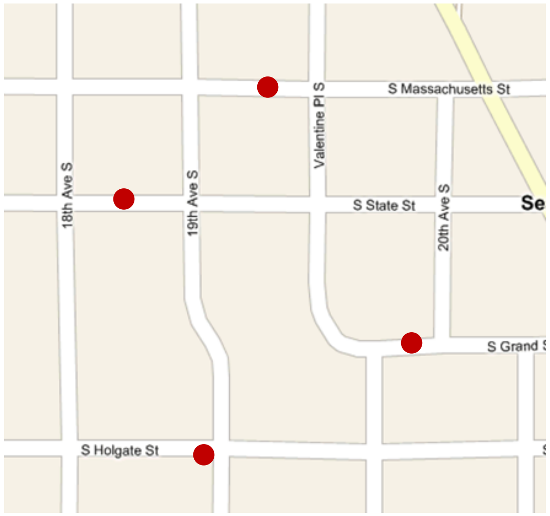
System Overview

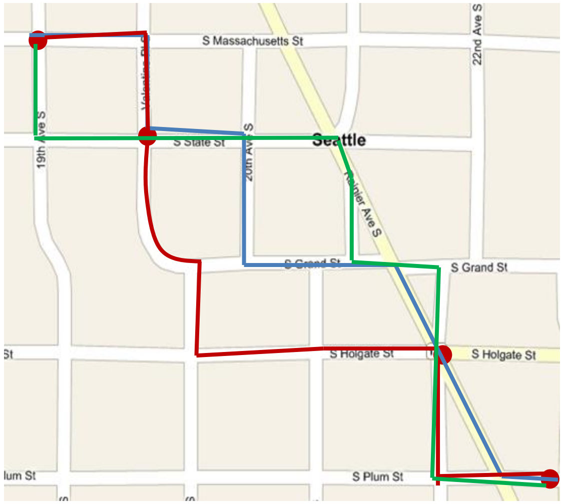


Offline Mining

- Challenges
  - Intelligence modeling
  - Data sparseness
  - Low-sampling-rate



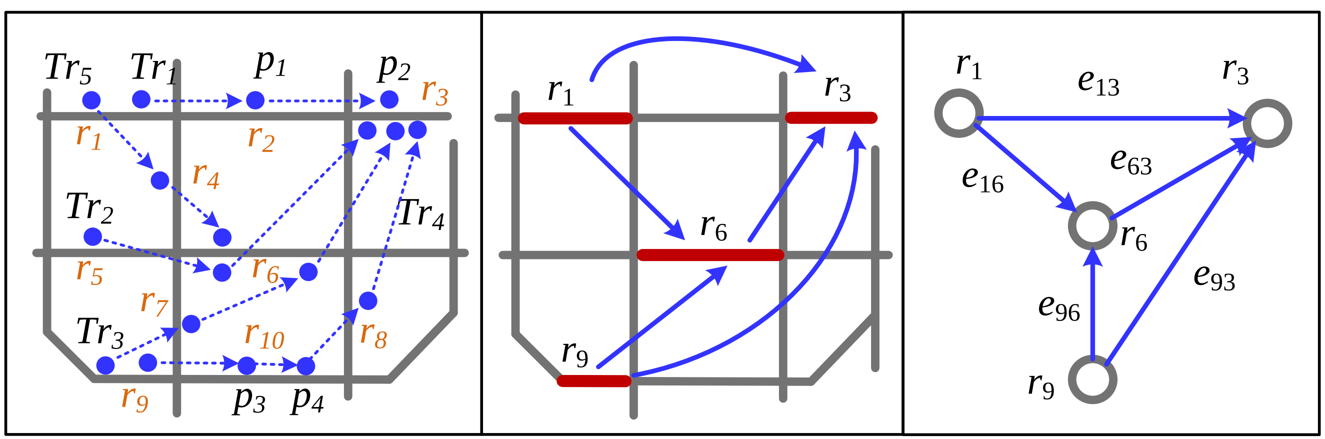




T-Drive
Driving Directions Based on Taxi Traces

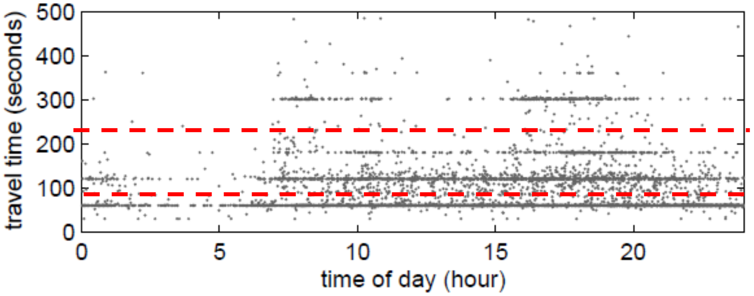
Offline Mining

- Detecting landmarks
  - A landmark is a frequently-traversed road segment
  - Top k road segments, e.g. k=4
- Building landmark edges
  - Number of transitions between two landmark edges > δ
  - E.g., $\delta = 1$

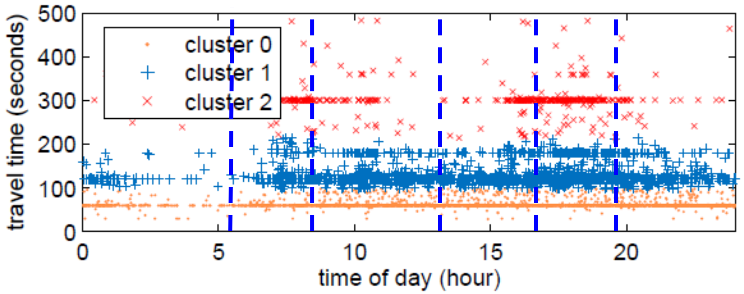


A) Matched taxi trajectories B) Detected landmarks C) A landmark graph

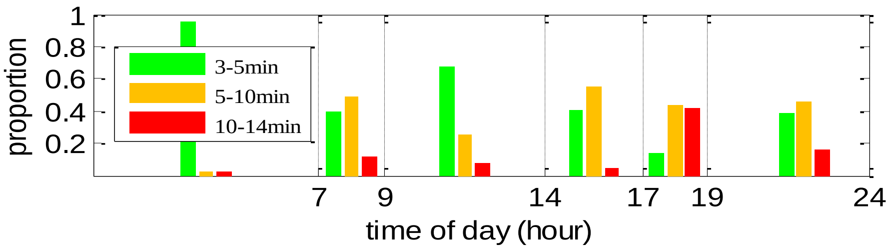
Mining Taxi Drivers' Knowledge



- Learning travel time distributions for each landmark edge
  - Traffic patterns vary in time on an edge
  - Different edges have different distributions



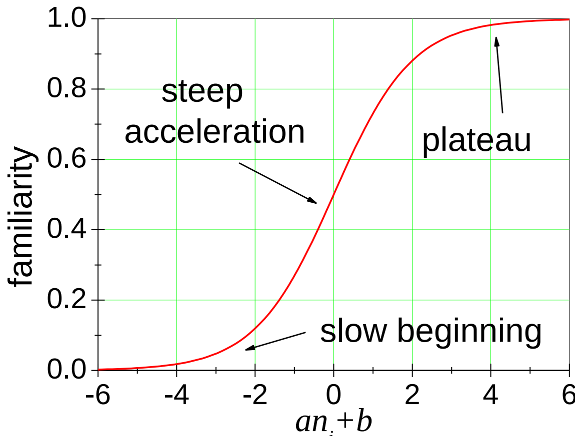
- Differentiate taxi drivers' experiences in different regions



C) Distributions of travel time

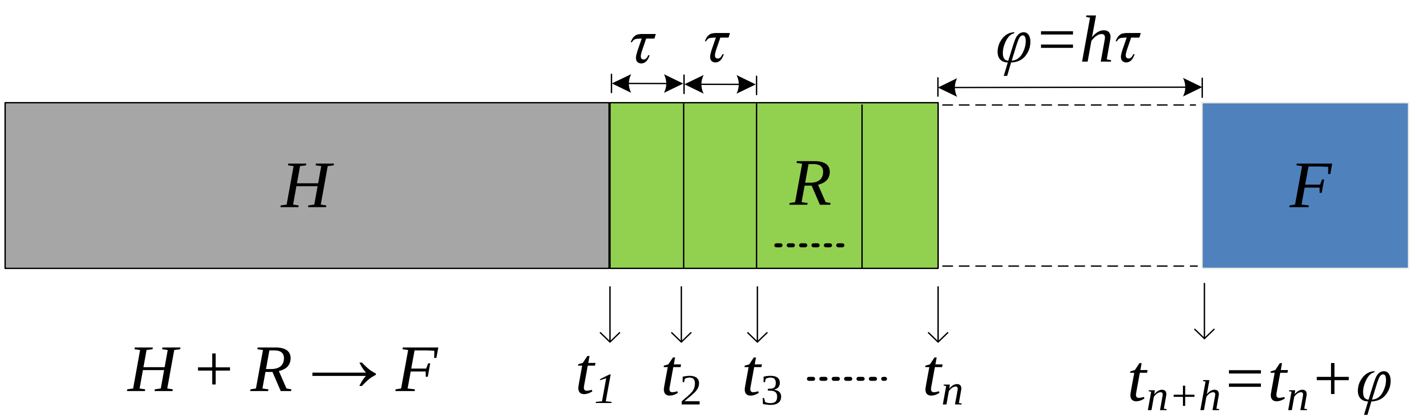
Sigmoid learning curve

$$f(n_i) = \frac{1}{1 + e^{-(an_i+b)}}$$



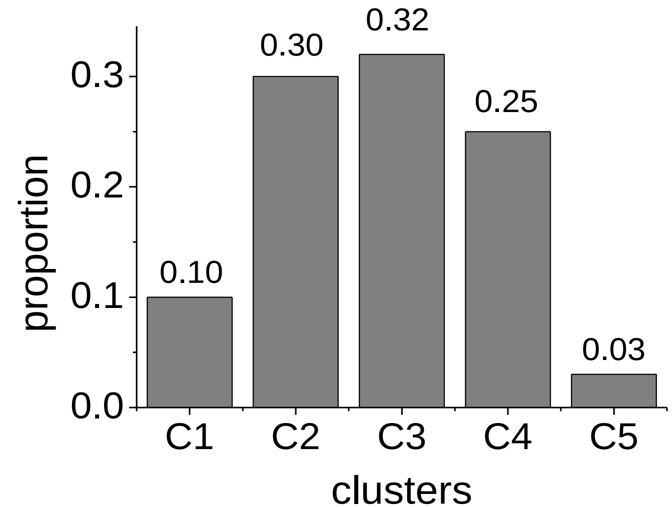
Online Inference

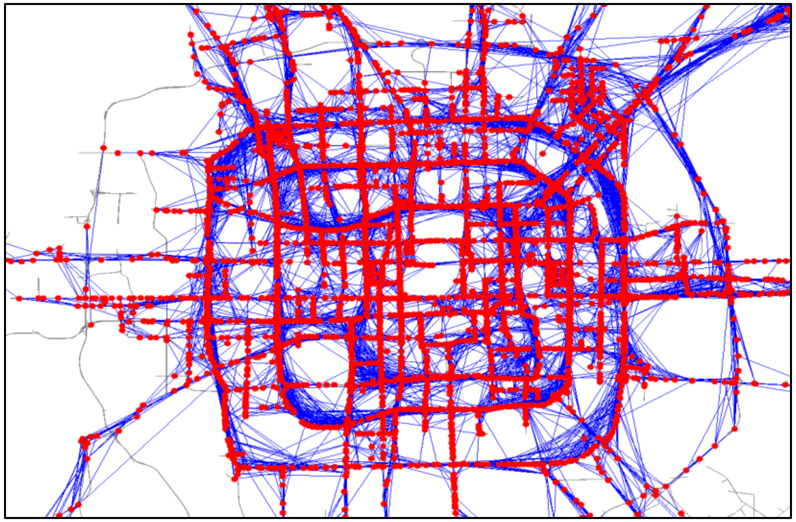
- Predict feature traffic conditions (F) on each landmark edge
  - based on the historical landmark graph (H) and
  - the recent GPS trajectories of taxis (R)
  - using a mth-order Markov chain



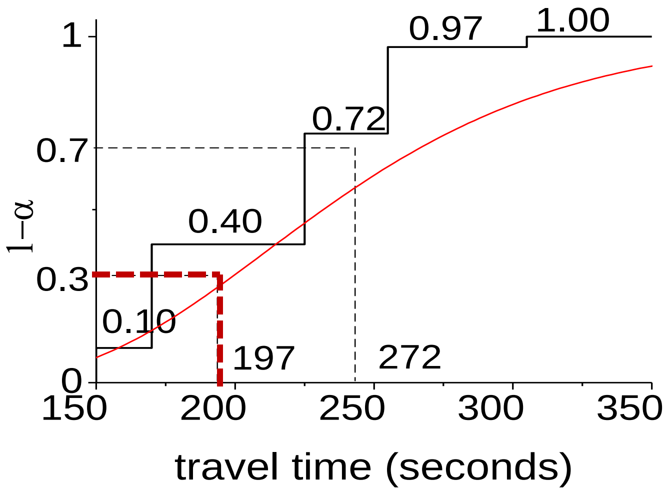
Route Computing

- Rough routing
  - Given a user query (q_s, q_d, t, α)
  - Search a landmark graph for a rough route: a sequence of landmarks
  - Using a time-dependent routing algorithm



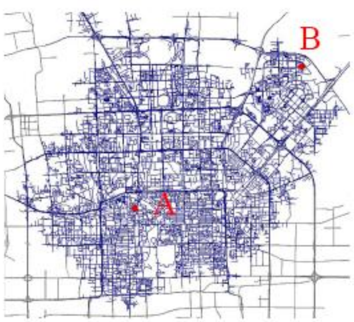


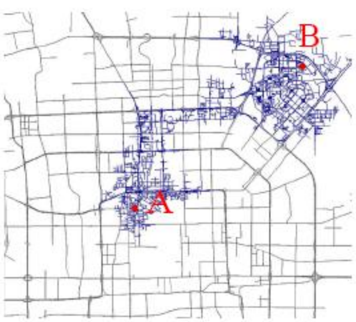
A landmark graph

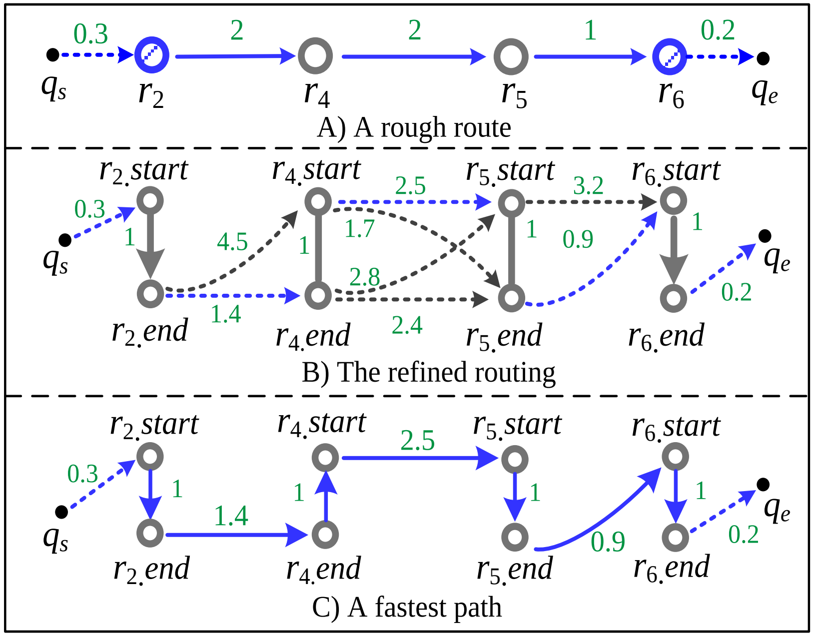


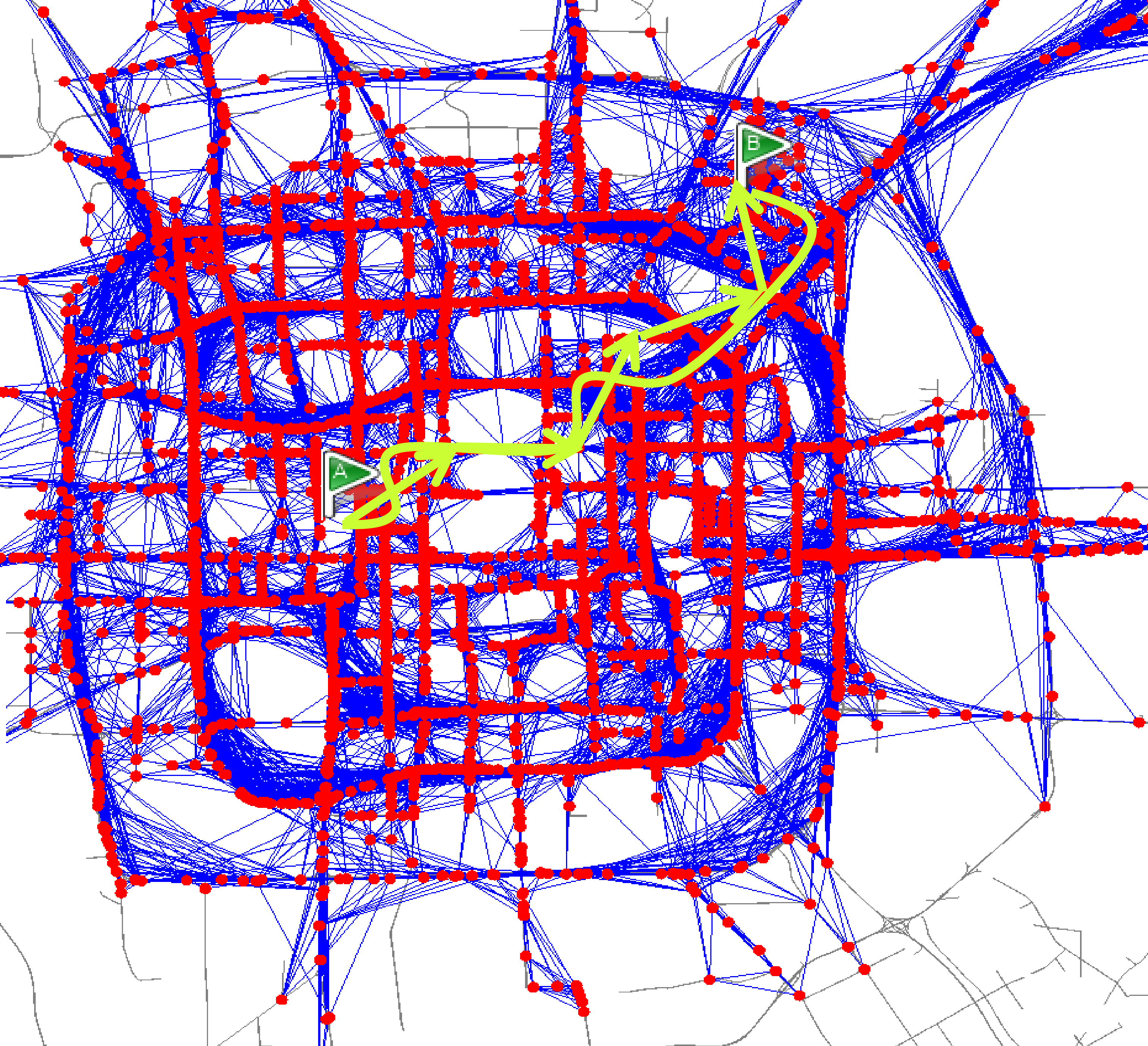
Route Computing

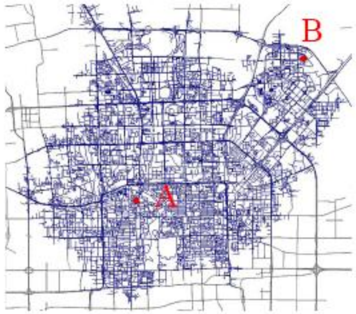
- Refined routing
  - Find out the fastest path connecting the consecutive landmarks
  - Can use speed constraints
  - Dynamic programming
- Very efficient
  - Smaller search spaces
  - Computed in parallel

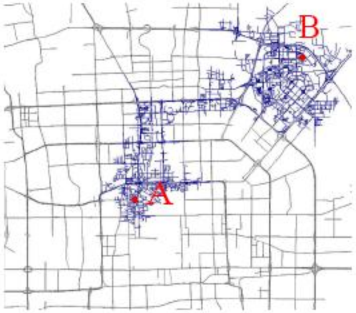








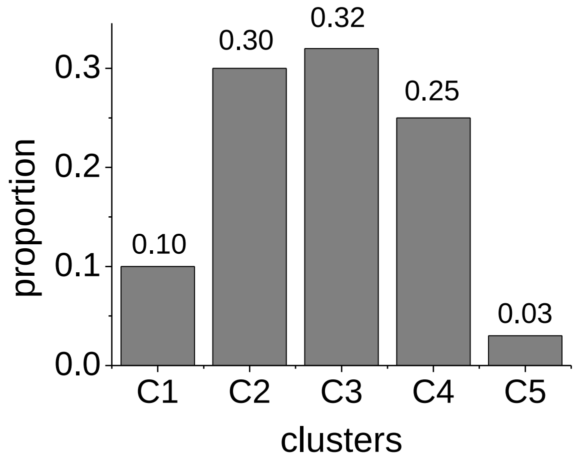




Reported by MIT Technology Review Twice, featured once

Learning an end user's drive behavior

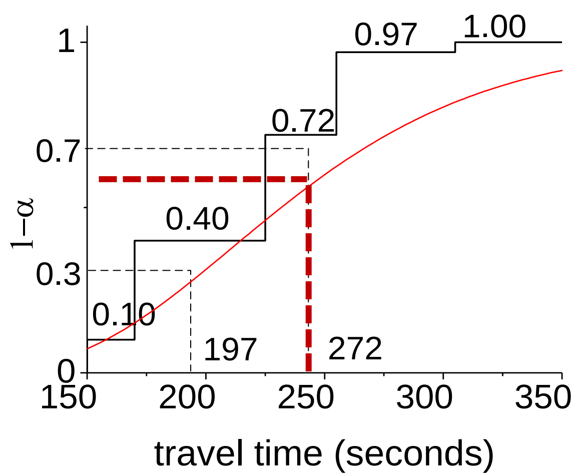
- Drive behavior
  - Vary in persons and places
  - Vary in progressing driving experiences
  - Custom factor: $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$



$$\tilde{\alpha}_i^{(M)} = CDF_i\left(T_i^{(M)}\right)$$

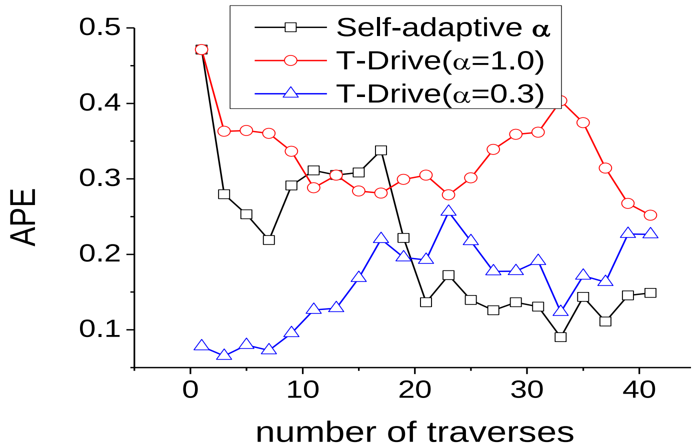
Weighted Moving Average:

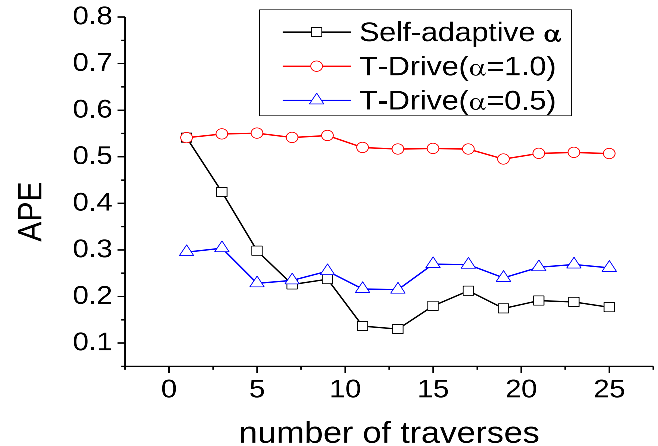
$$\alpha_i^{(M+1)} = \frac{\sum_{j=1}^{n} j \tilde{\alpha}_i^{(M-n+i)}}{\sum_{j=1}^{n} j} = \frac{2}{n(n+1)} \sum_{j=1}^{n} j \tilde{\alpha}_i^{(M-n+j)}$$



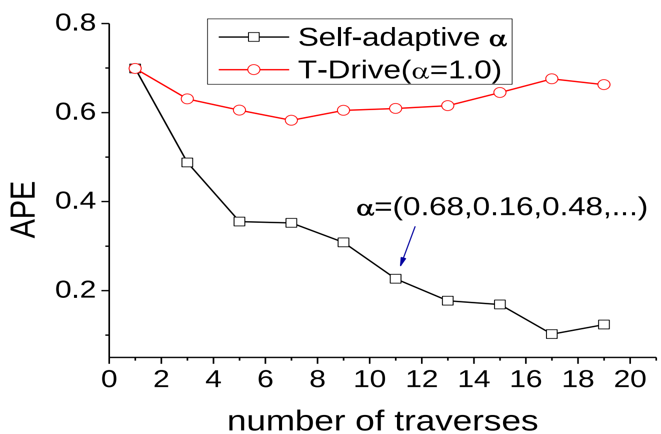
Evaluation on Routing

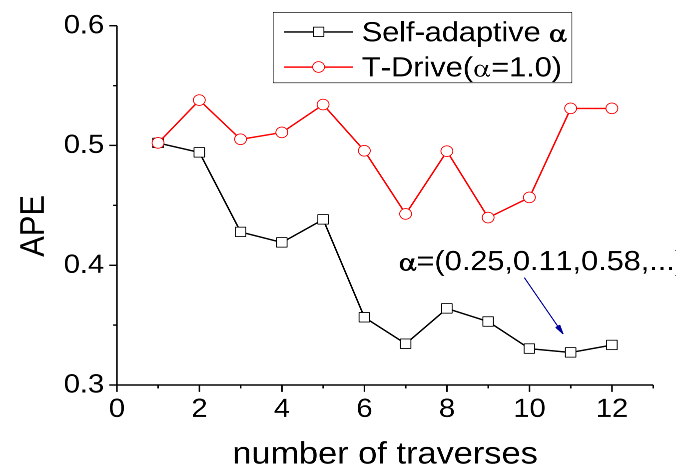
- User A on different routes





- Two users

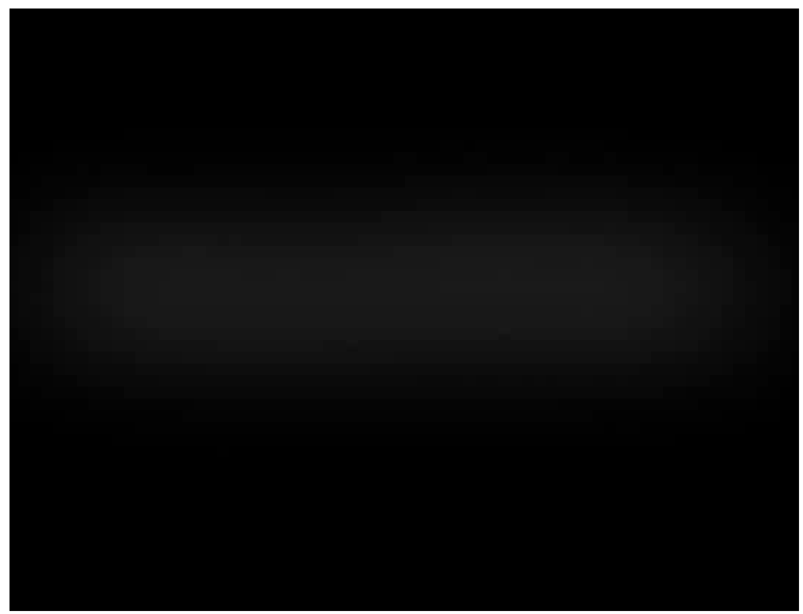




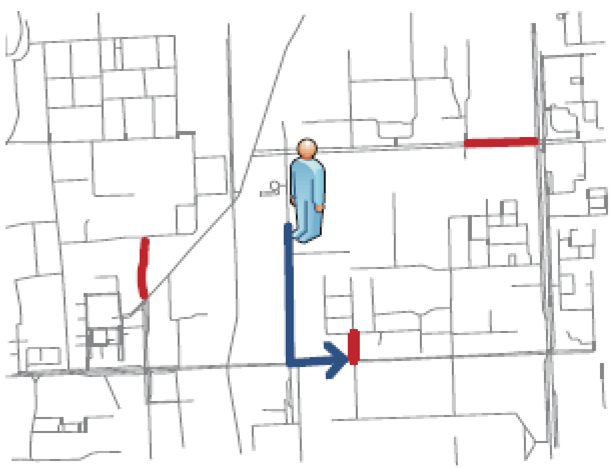
T-Drive
Driving Directions Based on Taxi Traces

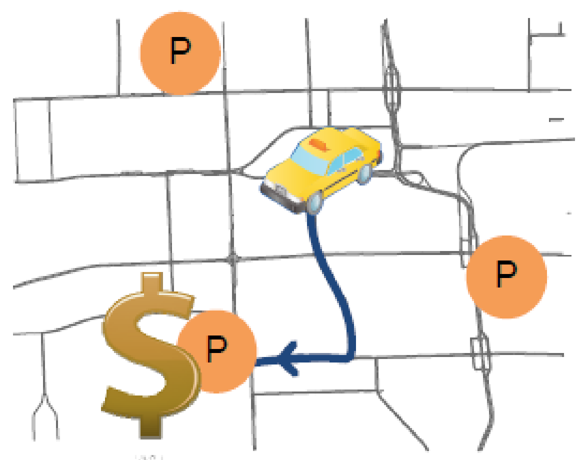
Results

- **More effective**
  - **60-70%** of the routes suggested by our method are faster than Bing and Google Maps.
  - Over **50%** of the routes are **20+%** faster than Bing and Google.
  - On average, we save **5** minutes per 30 minutes driving trip.
- More efficient



Taxi Recommender Systems

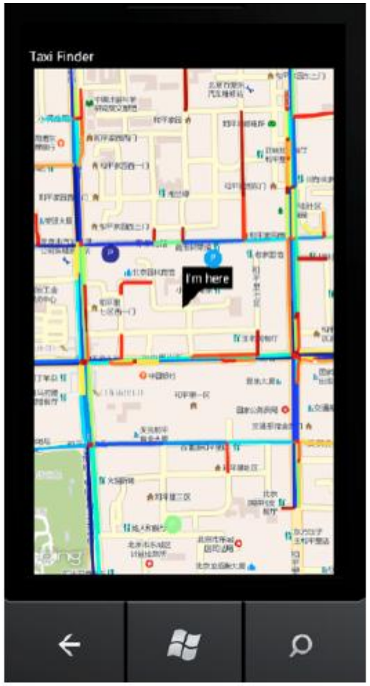


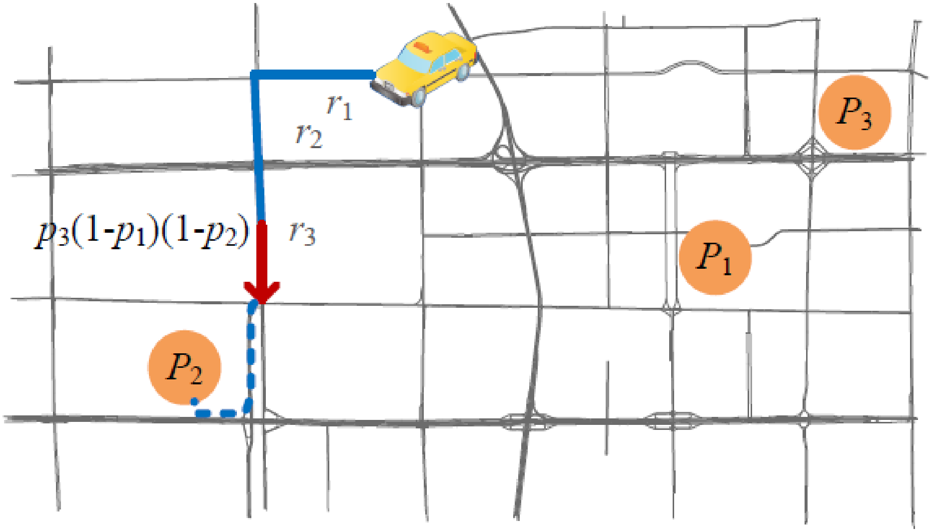


Jing Yuan, Yu Zheng, et al. [Where to Find My Next Passenger?](#) , UbiComp 2011

Taxi Recommender Systems

- Two-folder recommendations
  - Users: some road segments or parking places around them
  - Taxi drivers: top-k parking places and the trips

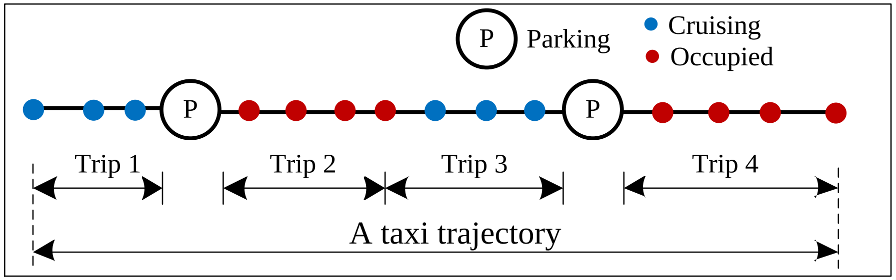




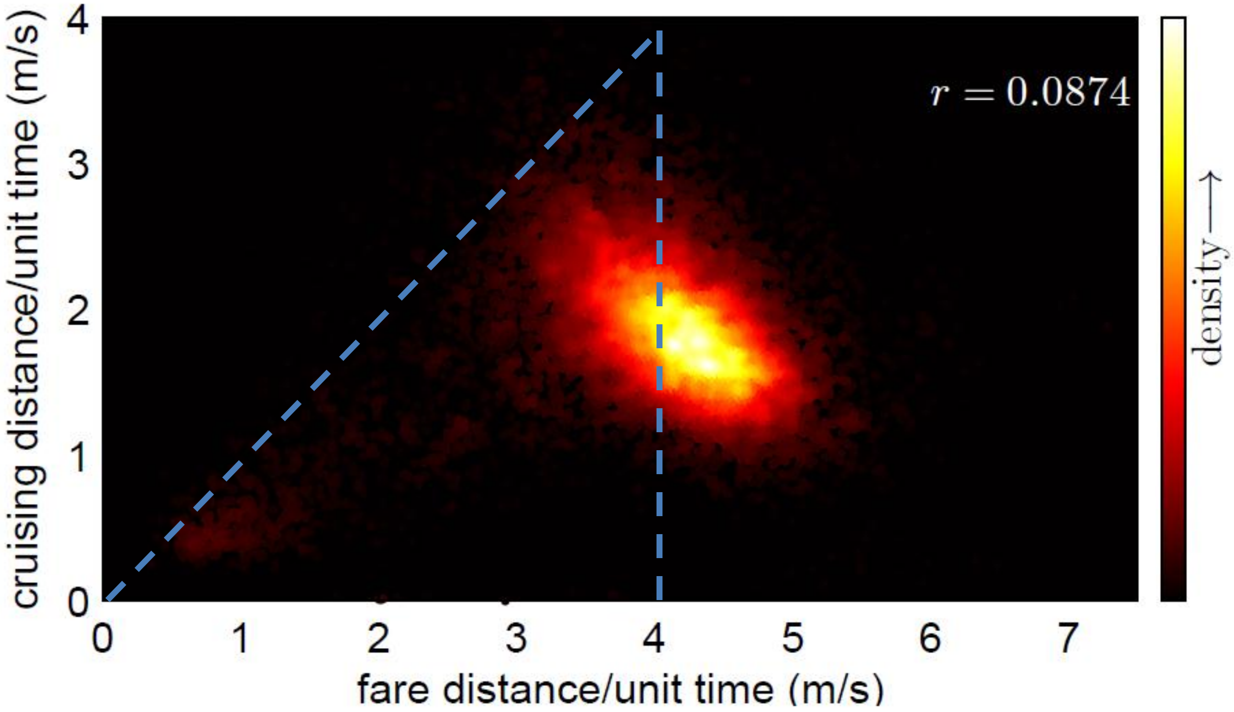
Jing Yuan, Yu Zheng, et al. [Where to Find My Next Passenger?](), UbiComp 2011

Taxi Recommender Systems

- Concepts
- Observations

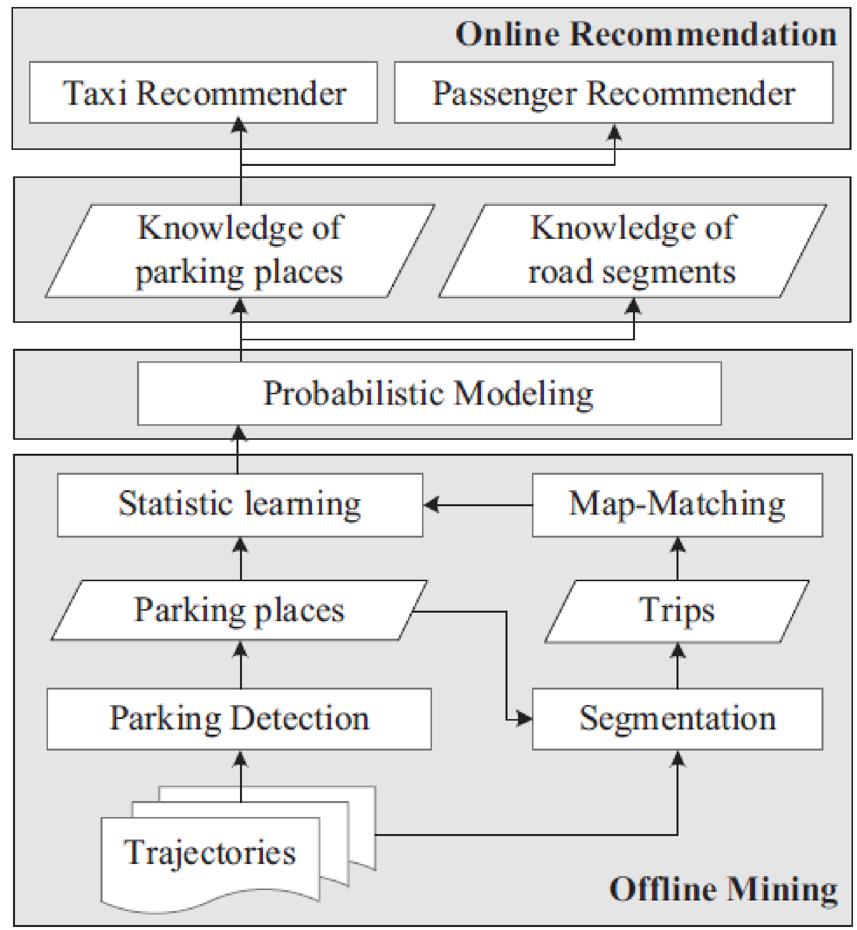


"Taxi drivers prefer to park at somewhere rather than cruising on streets when having no passengers"



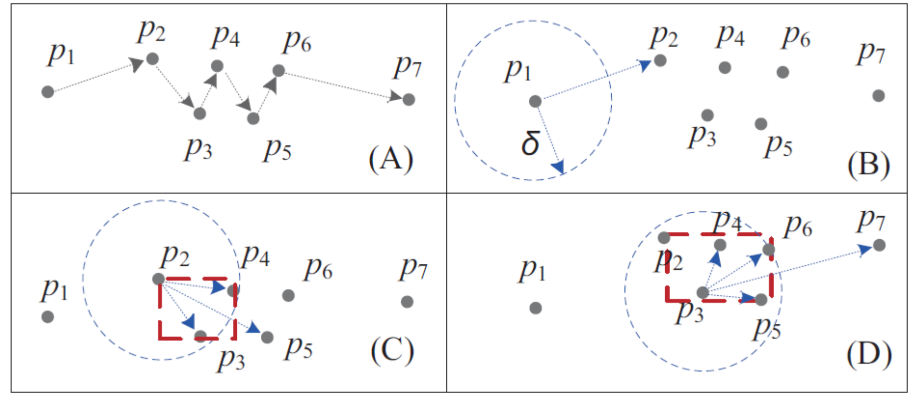
Taxi Recommender Systems

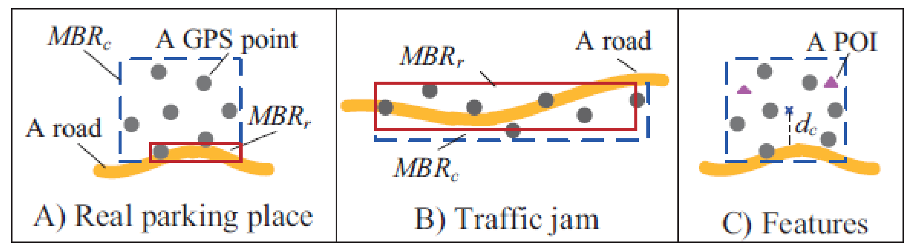
- Framework

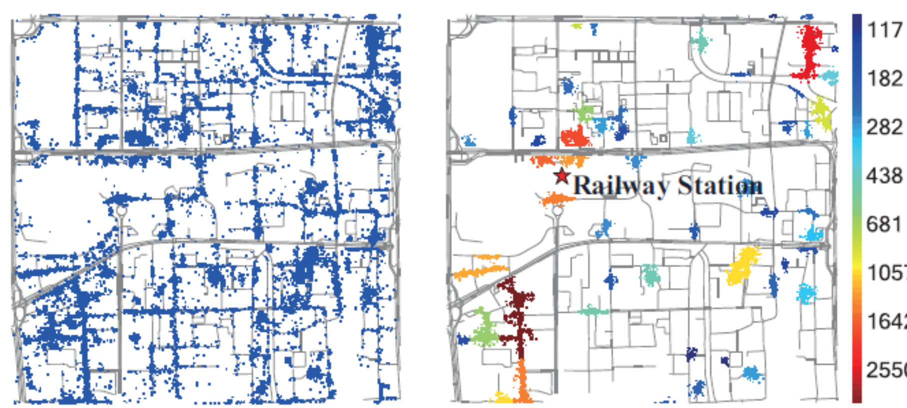


Taxi Recommender Systems

- Parking place detection
  - Stay points detection
    - Just candidates
    - Could be traffic jams/lights
  - Filtering
    - Supervised classification
    - Features
      - Spatio-temporal features
      - POIs
      - Collaborative features
  - Density-based clustering







Taxi Recommender Systems

- Statistic learning
  - Knowledge on road segments

$$\Pr(\mathcal{C}; r|t) \qquad \Pr(\mathcal{O}; r|t) \qquad \Pr(\mathcal{C} \rightsquigarrow \mathcal{O}|r, t)$$

$$\Pr(d_a < D_N \leq d_b|r, t)$$

  - Knowledge in a parking place

$$\Pr(\mathcal{P} \xrightarrow{(t_a, t_b]} \mathcal{O}|T_P > 0, t)$$

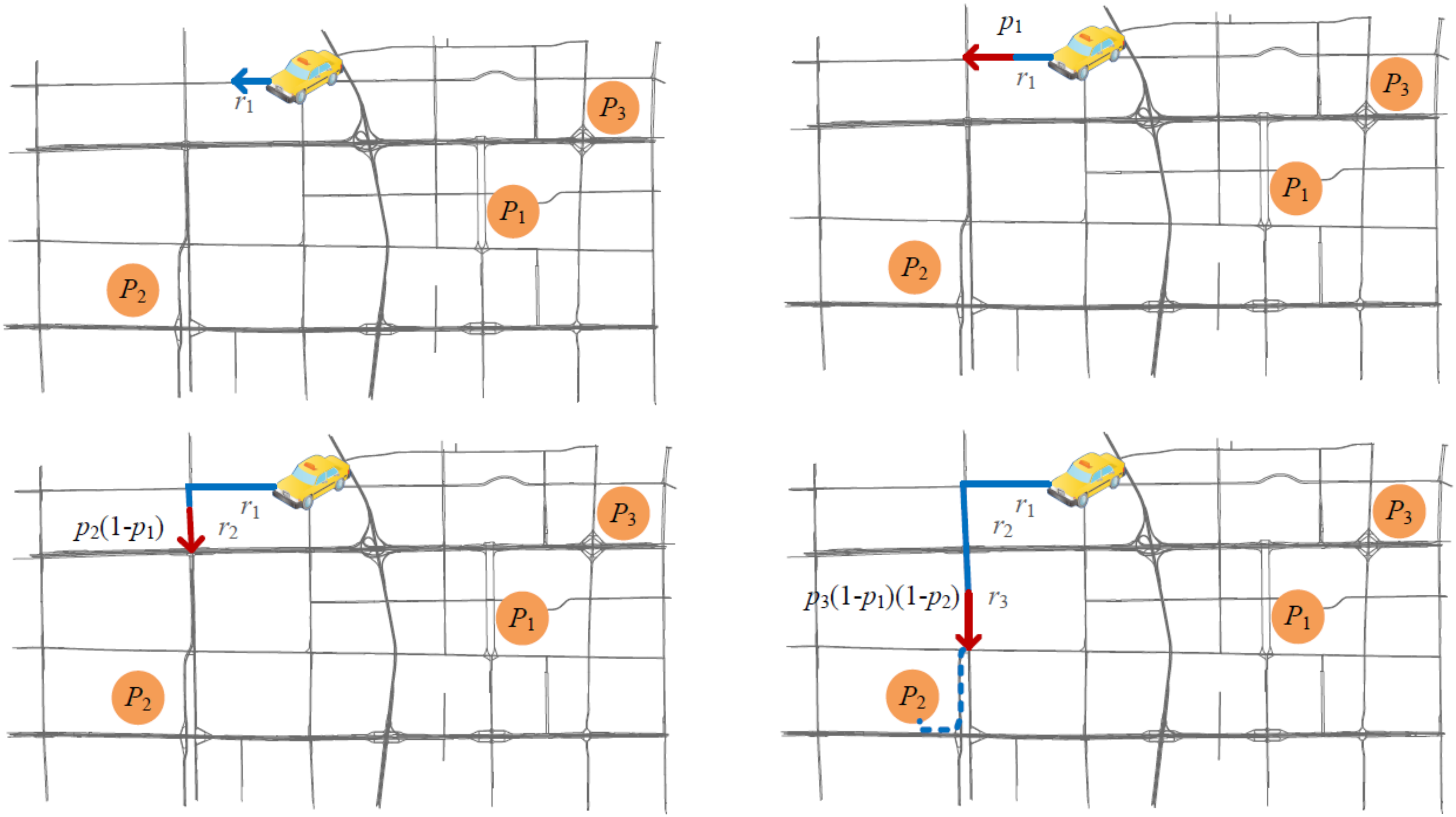
Jing Yuan, Yu Zheng, et al. Where to Find My Next Passenger? , UbiComp 2011

Taxi Recommender Systems

- Probability modeling
  - Taxi drivers
    - Maximum the profit of a taxi driver in a unit time
    - What is a good parking place
      - High probability to take a passenger in it and on the way towards it
      - Pick up the next passenger quickly
      - The fare distance/duration is big
  - Users
    - High probability to find a vacant taxi
    - A short waiting time

Taxi Recommender Systems

- Probability modeling for the taxi recommender



Taxi Recommender Systems

- Probability model for taxi recommender

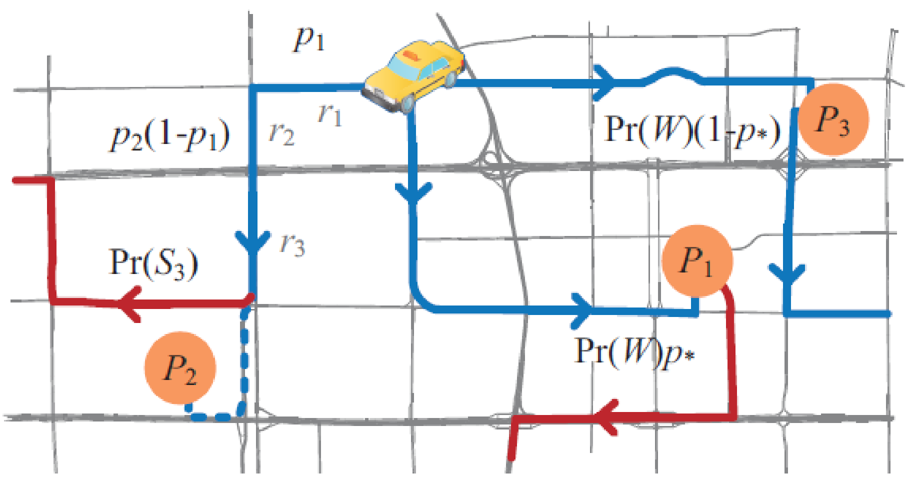
$$S = \bigcup_{i=1}^{n+1} S_i,$$

$$\Pr(S_i) = \begin{cases} p_1, & i = 1 \\ p_i \prod_{j=1}^{i-1} (1 - p_j), & i = 2, 3, \ldots, n, \\ p_* \prod_{j=1}^{n} (1 - p_j), & i = n + 1. \end{cases}$$

$$p_i = \Pr(\mathcal{C} \rightsquigarrow \mathcal{O} | r_i, T_0 + t_i)$$

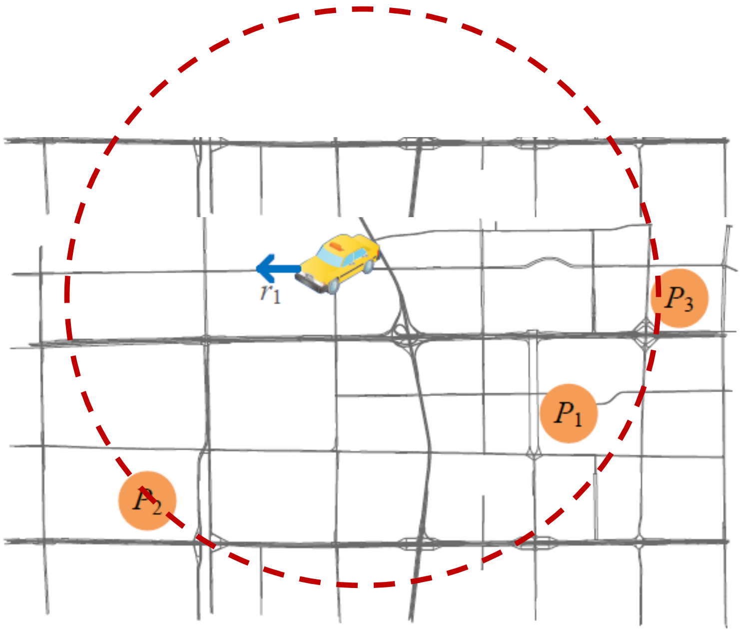
$$p_* = \Pr(\mathcal{P} \xrightarrow{(0, t_{max}]} \mathcal{O} | T_0 + t_n)$$

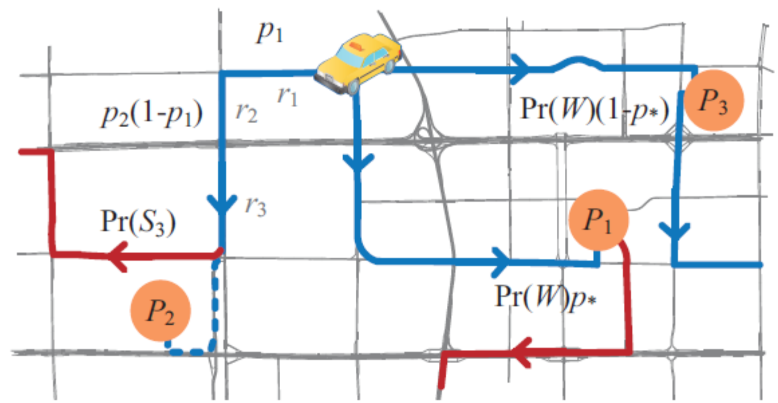
$$\boxed{\begin{aligned} \Pr(S) &= 1 - \Pr(\overline{\bigcup_{i=1}^{n+1} S_i}) \\ &= 1 - (1 - p_*) \prod_{j=1}^{n} (1 - p_j). \end{aligned}}$$



Taxi Recommender Systems

- NP hard problem (Approximation)
  - Select top k parking places close to a taxi
  - Find the shortest path to each parking place
  - Compute the probability of taking a passenger for each choice





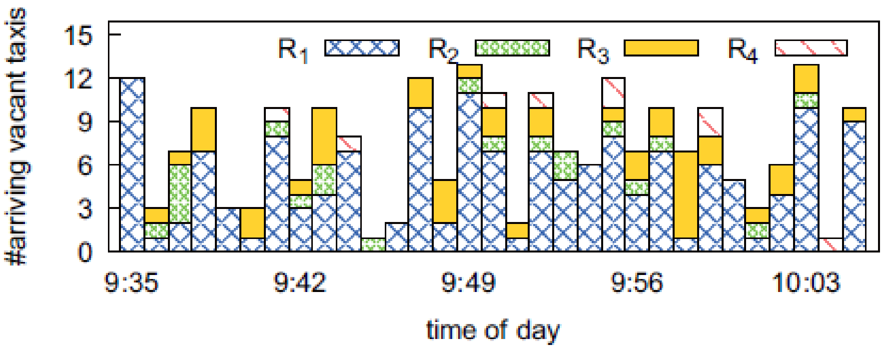
Taxi Recommender Systems

- More challenges
  - Estimate the waiting time a passenger on a road segment
  - Duration that a taxi would wait in a parking place
- Main ideas
  - Suppose the arrival of taxis on a road segment follows a poison distribution
  - Estimate the average time interval between two arrivals

Nicholas Jing Yuan, Yu Zheng, et al. T-Finder: A Recommender System for Finding Passengers and Vacant Taxis. TKDE.

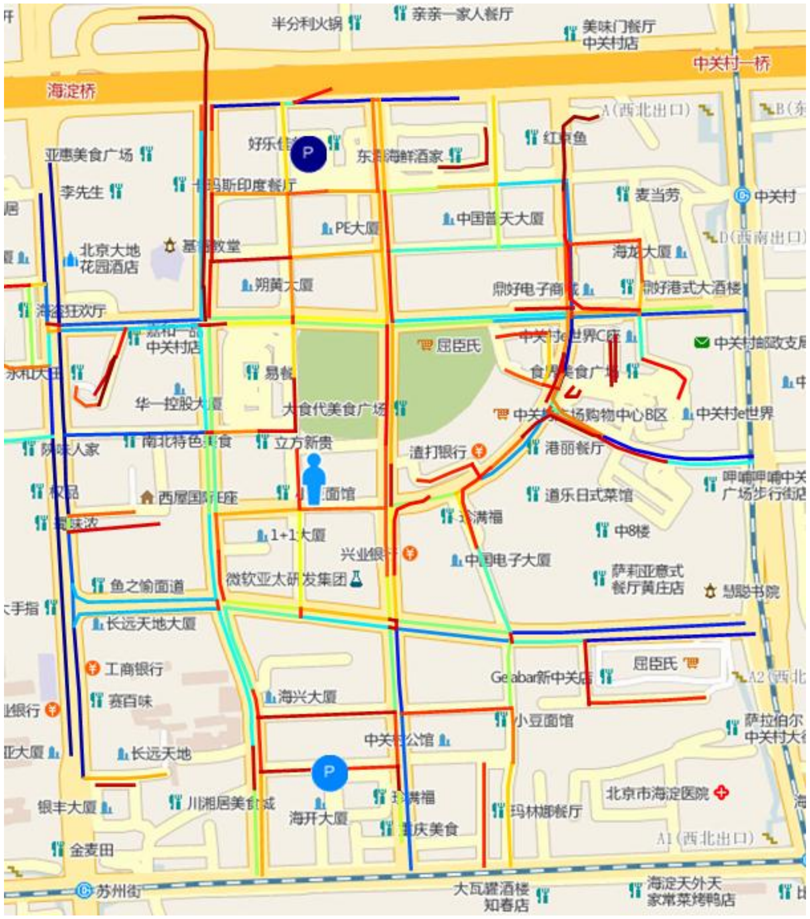
Taxi Recommender Systems

- Evaluation



(a) weekday, Suzhou Street

area	Zhichun Rd.				Suzhou St.			
time	8:40-9:10				9:35-10:05			
road	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
$\#$	20.8	3.3	0.8	0.8	128.3	35.0	16.7	7.5
$Rank$	1	2	3	3	1	2	3	4
$Rank_p^d$	2	1	4	3	1	3	2	4
$Rank_t^d$	2	1	3	4	1	2	4	3
$Rank_p^w$	1	3	2	4	1	3	2	4
$Rank_t^w$	1	2	4	3	2	1	3	4
$Rank_p^{d,w}$	1	2	3	4	1	3	2	4
$Rank_t^{d,w}$	1	2	4	3	1	2	3	4



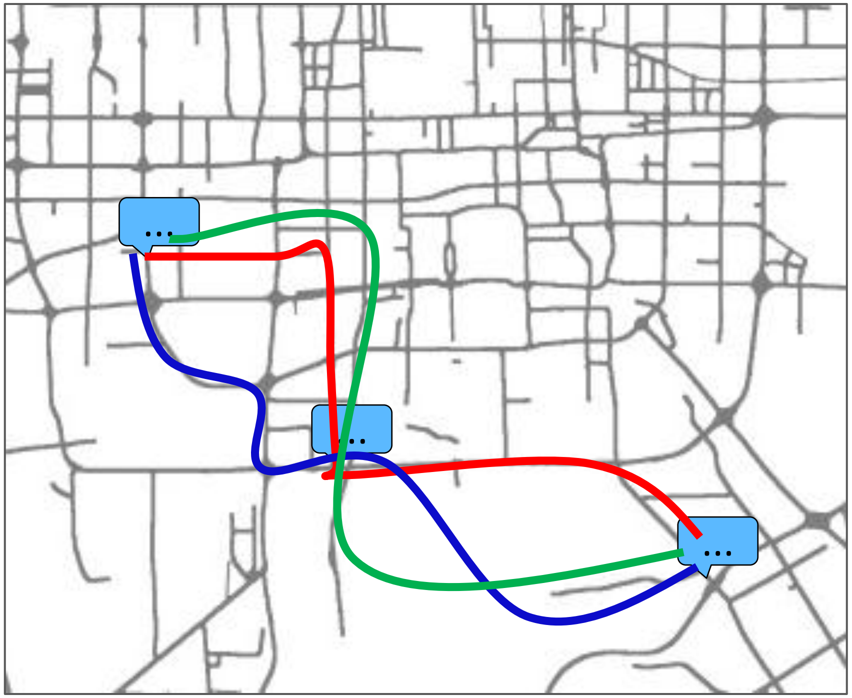
Taxi Recommender Systems

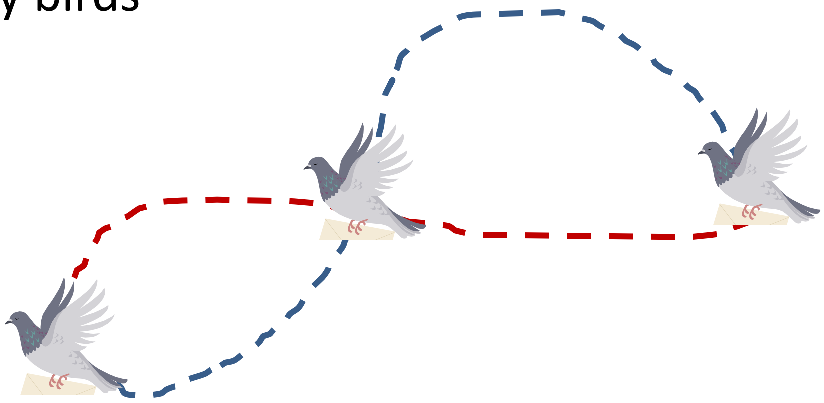
- What’s the next
  - Large-scale real-time taxi ridesharing
    - Constraints: users, taxi drivers, government
    - NP complete problem
    - Approximation: search and optimization problem
  - On-site Discussion

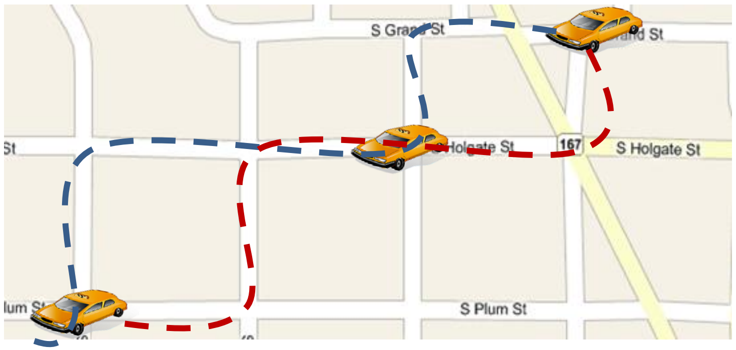
Constructing Popular Routes from Uncertain Trajectories

Constructing Popular Routes from Uncertain Trajectories

- Uncertain trajectories
  - check-ins or geo-tagged photos
  - Taxi trajectories, trails of migratory birds

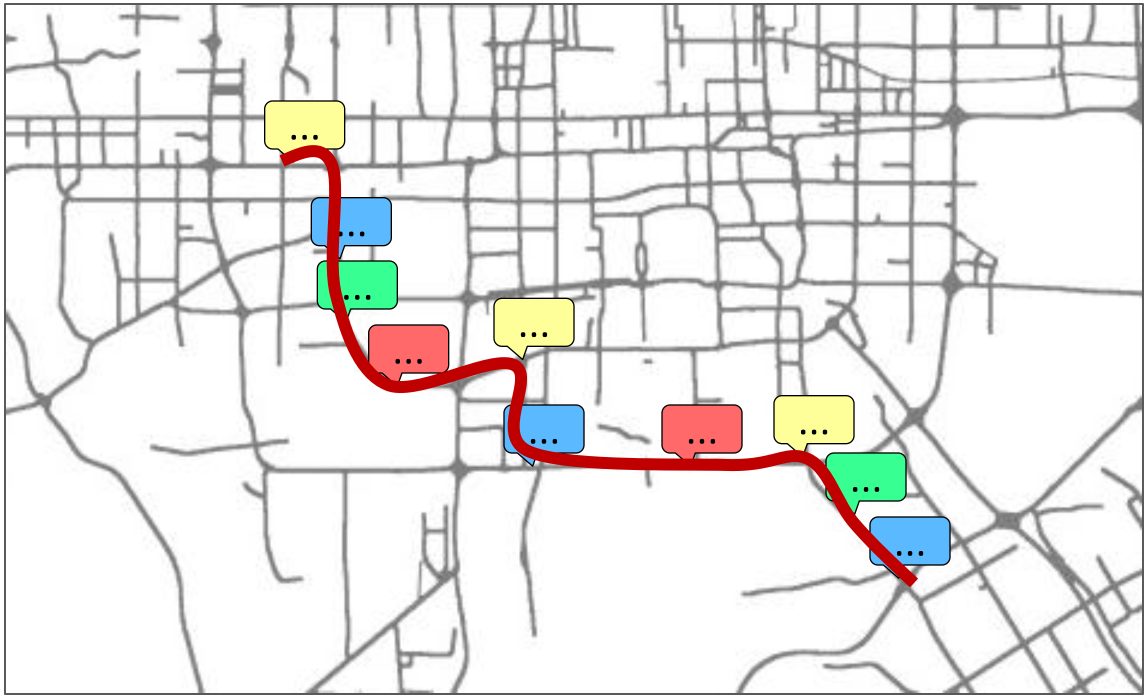




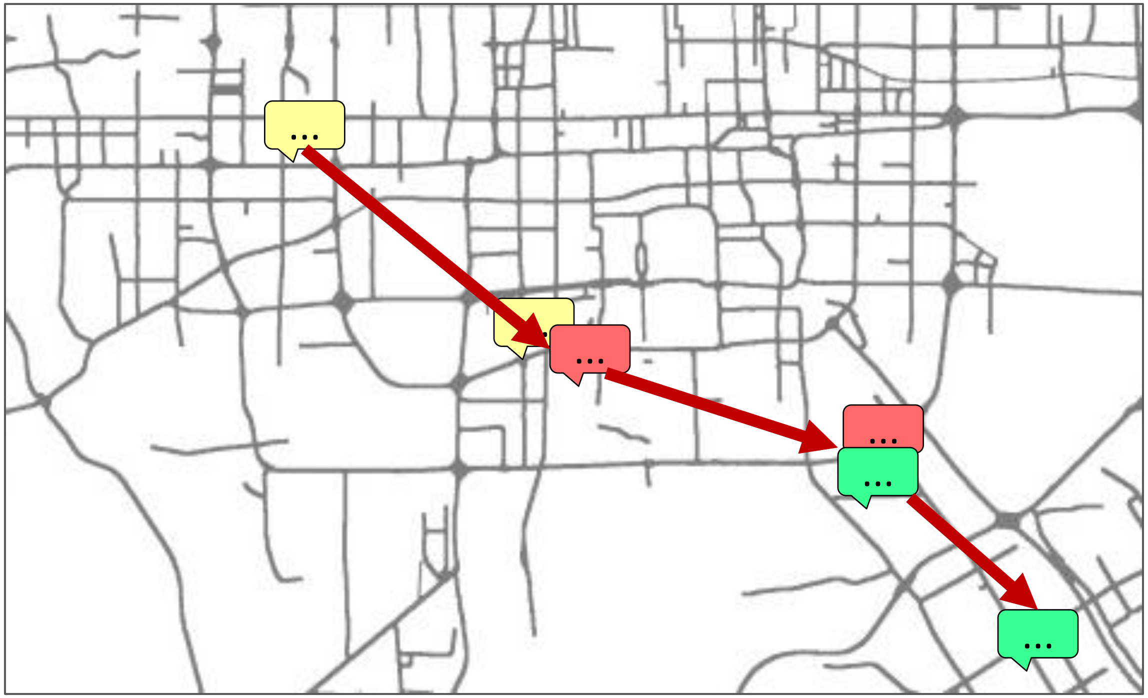


Constructing Popular Routes from Uncertain Trajectories

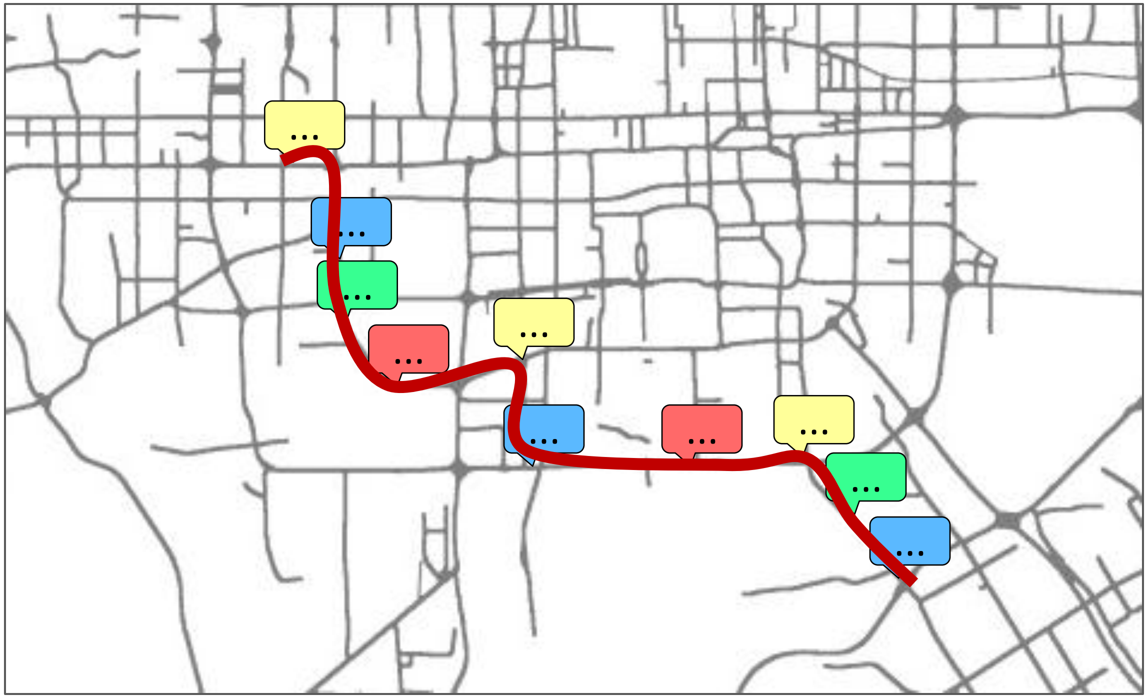
- Goal: Using collective knowledge: The route may not exist in the dataset
  - Mutual reinforcement learning (***uncertain + uncertain → certain***)



Concatenation

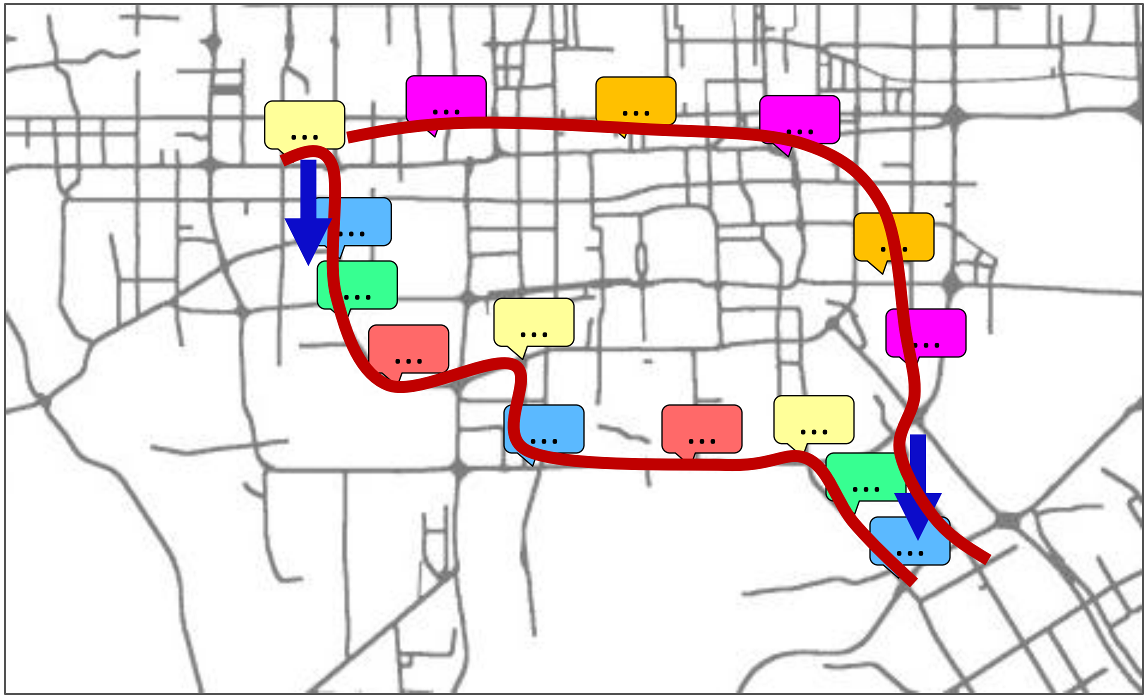


Mutual reinforcement construction



Constructing Popular Routes from Uncertain Trajectories

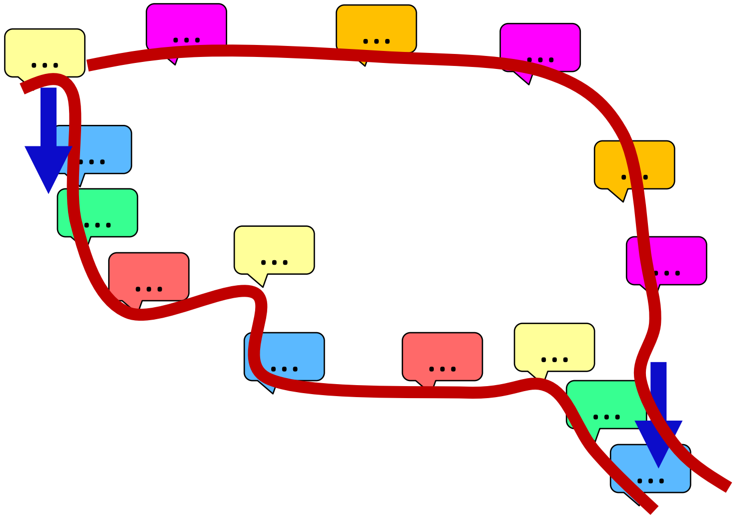
- Problem
  - Given a corpus of uncertain trajectories and
  - a user query: some point locations and a time constraint
  - Suggest the top k most popular routes



Constructing Popular Routes from Uncertain Trajectories

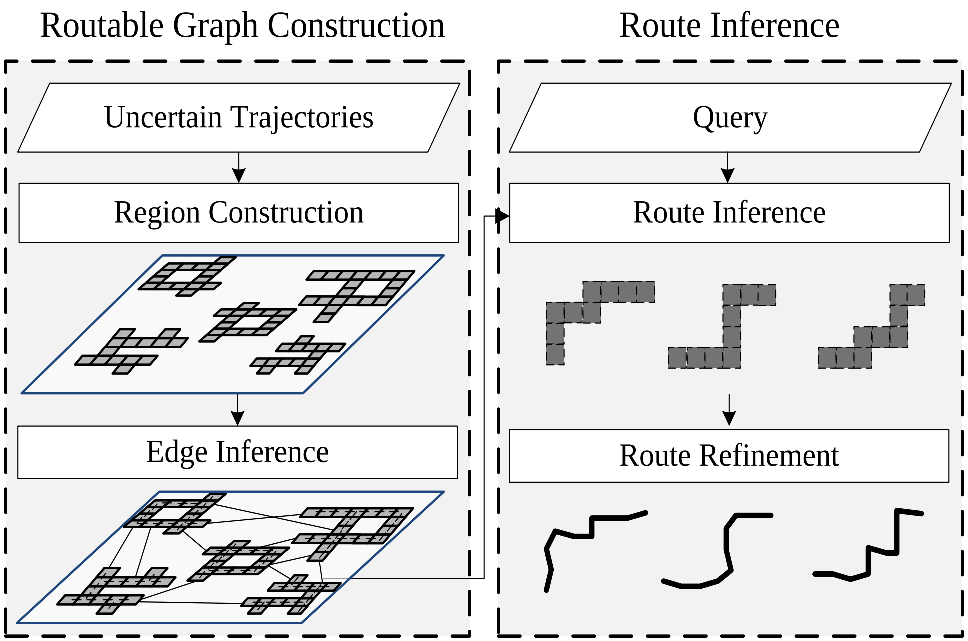
- In a road network
  - Accurate and relatively easy
  - Limited applications
  - Kai Zheng, Yu Zheng, et al. *Reducing Uncertainty of Low-Sampling-Rate Trajectories*. ICDE 2012.

- In a free space
  - Coarse and difficult
  - Wide range of applications
  - Ling-Yin Wei, Yu Zheng, et al. *Constructing Popular Routes from Uncertain Trajectories*. KDD 2012.



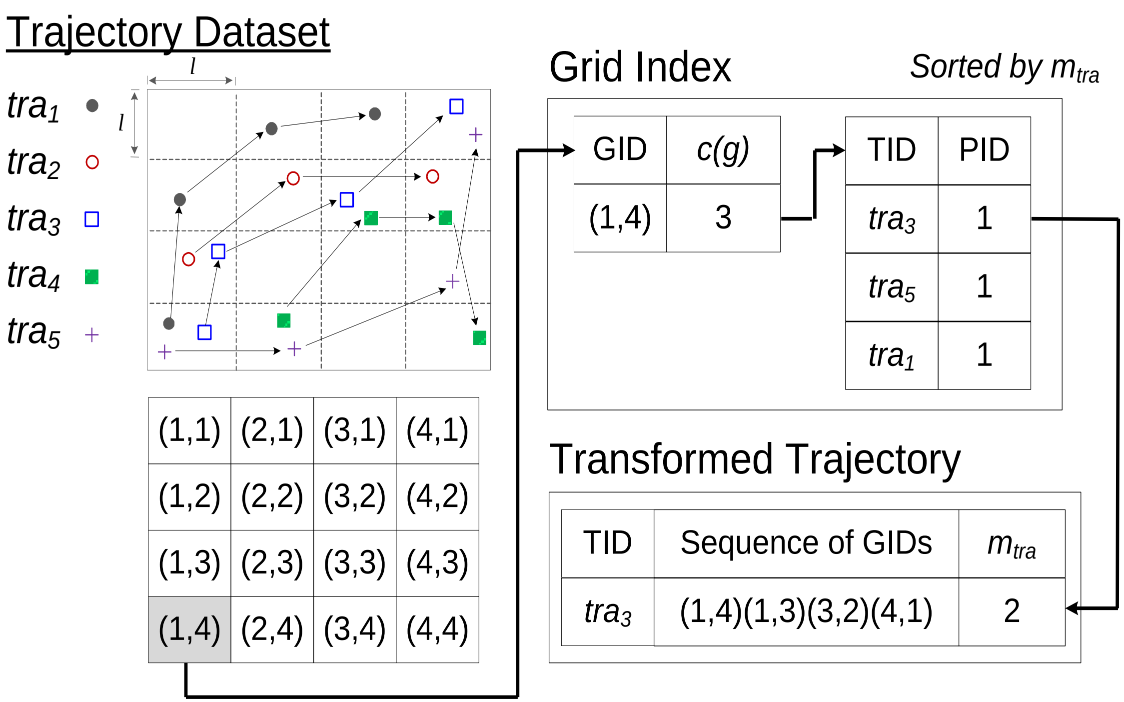
Constructing Popular Routes from Uncertain Trajectories

- Framework (for free spaces)
  - Routable graph construction
    - Space partition and spatial indexing
    - Region construction
    - Edge inference
  - Route inference
    - Routing on the graph
    - Refinement



Constructing Popular Routes from Uncertain Trajectories

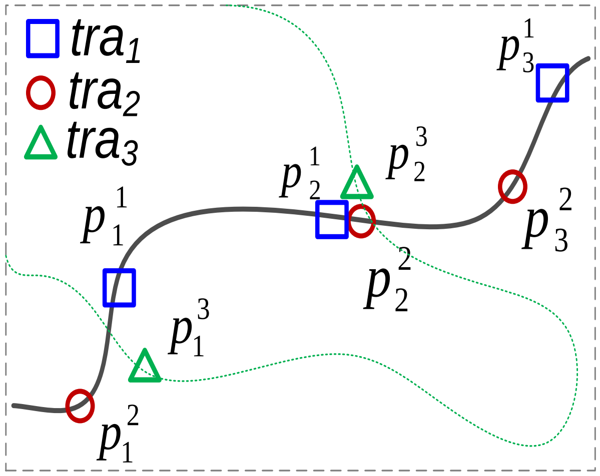
- Routable graph construction
  - Step 1: Space partition and spatial indexing
    - Approximation of an inferred route
    - Speed up the searching process

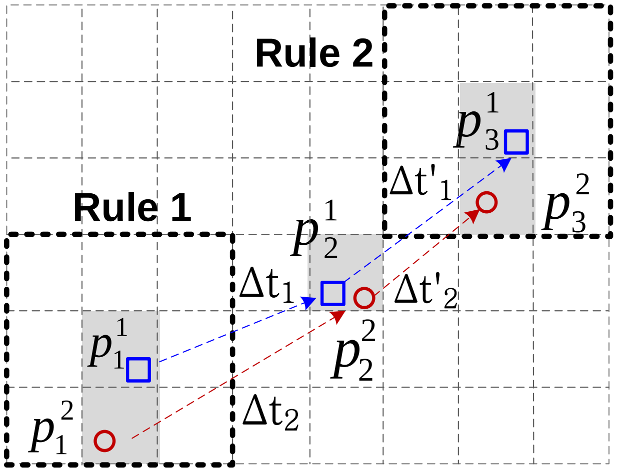


Constructing Popular Routes from Uncertain Trajectories

- Routable graph construction
  - Determine the correlated sub-trajectories
  - Using spatio-temporal correlations

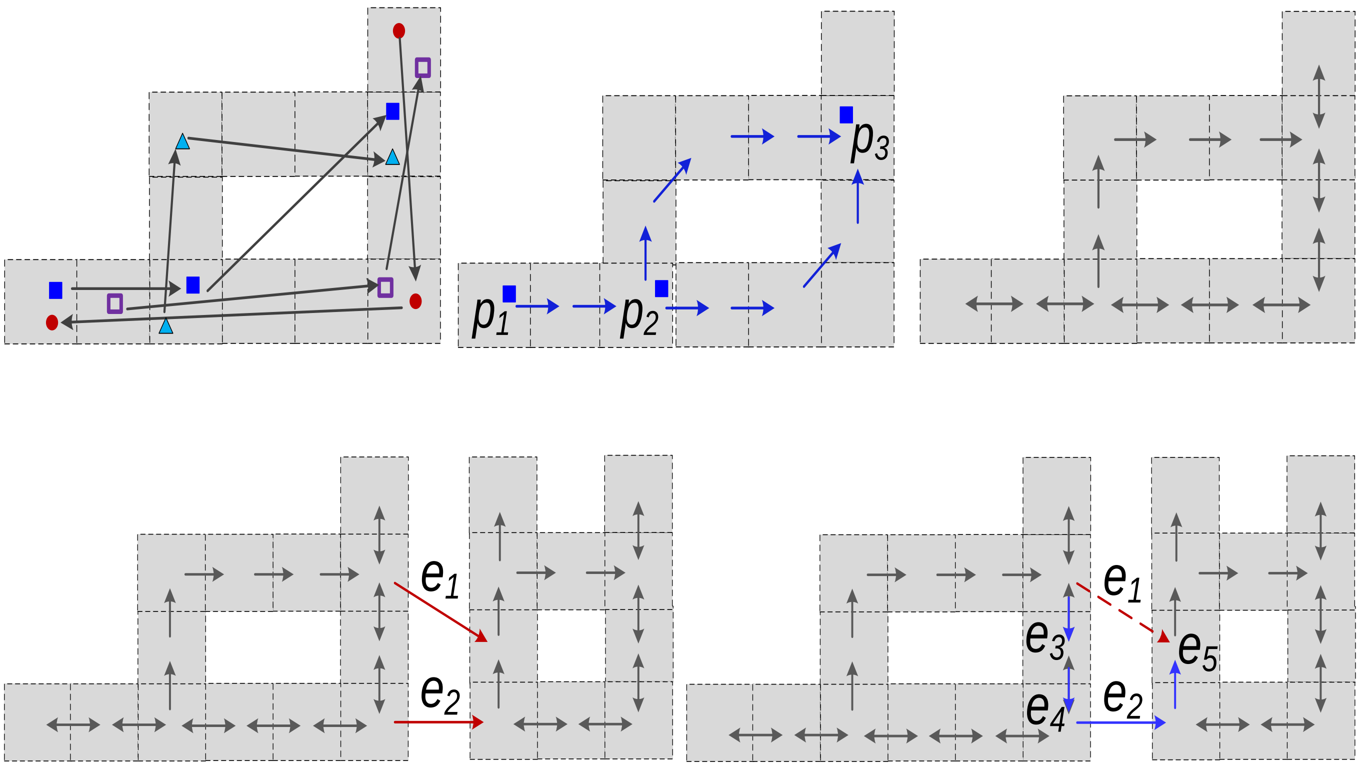
$$\frac{|\Delta t_1 - \Delta t_2|}{\max\{\Delta t_1, \Delta t_2\}} \leq \theta$$





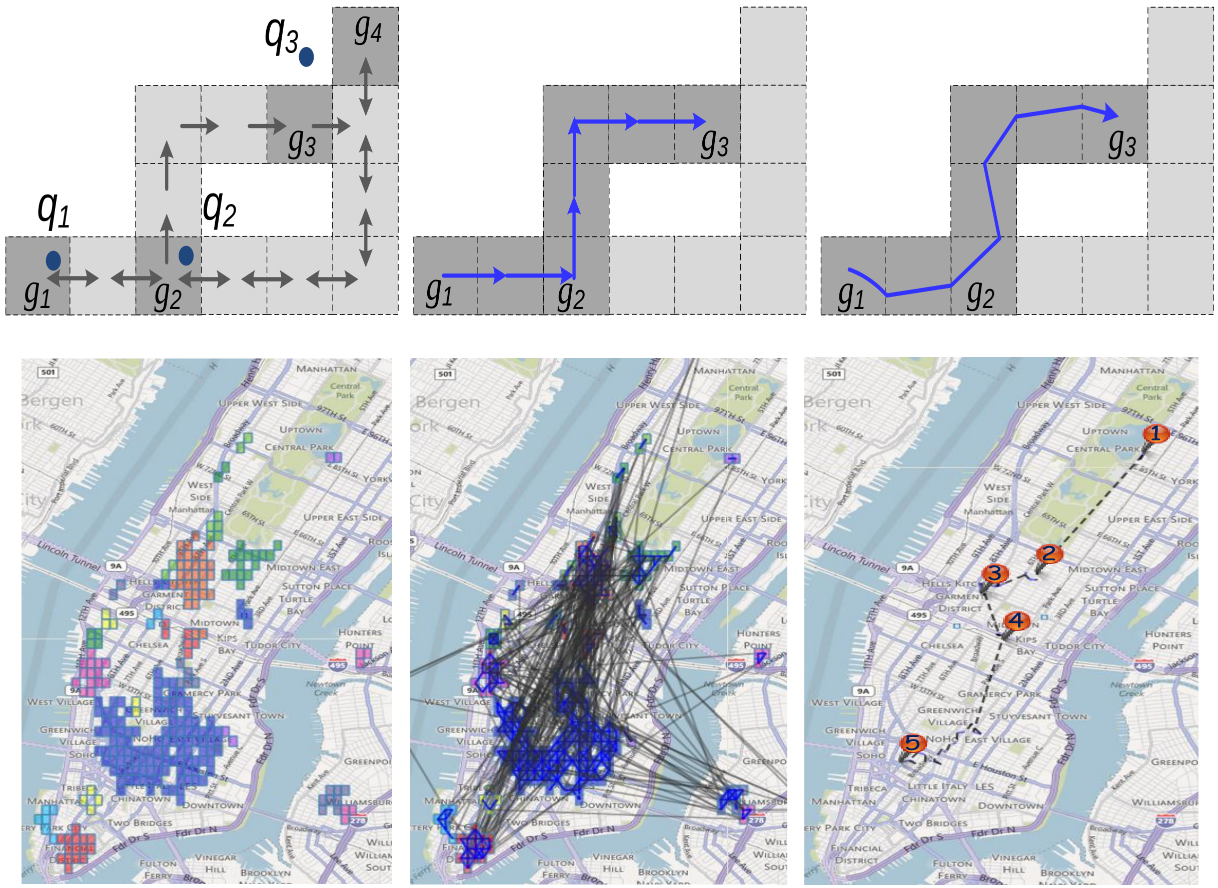
Constructing Popular Routes from Uncertain Trajectories

- Routable graph construction
  - Edge inference between grids in a region
  - Edge inference between grids from disconnected regions



Constructing Popular Routes from Uncertain Trajectories

- Route inference



References

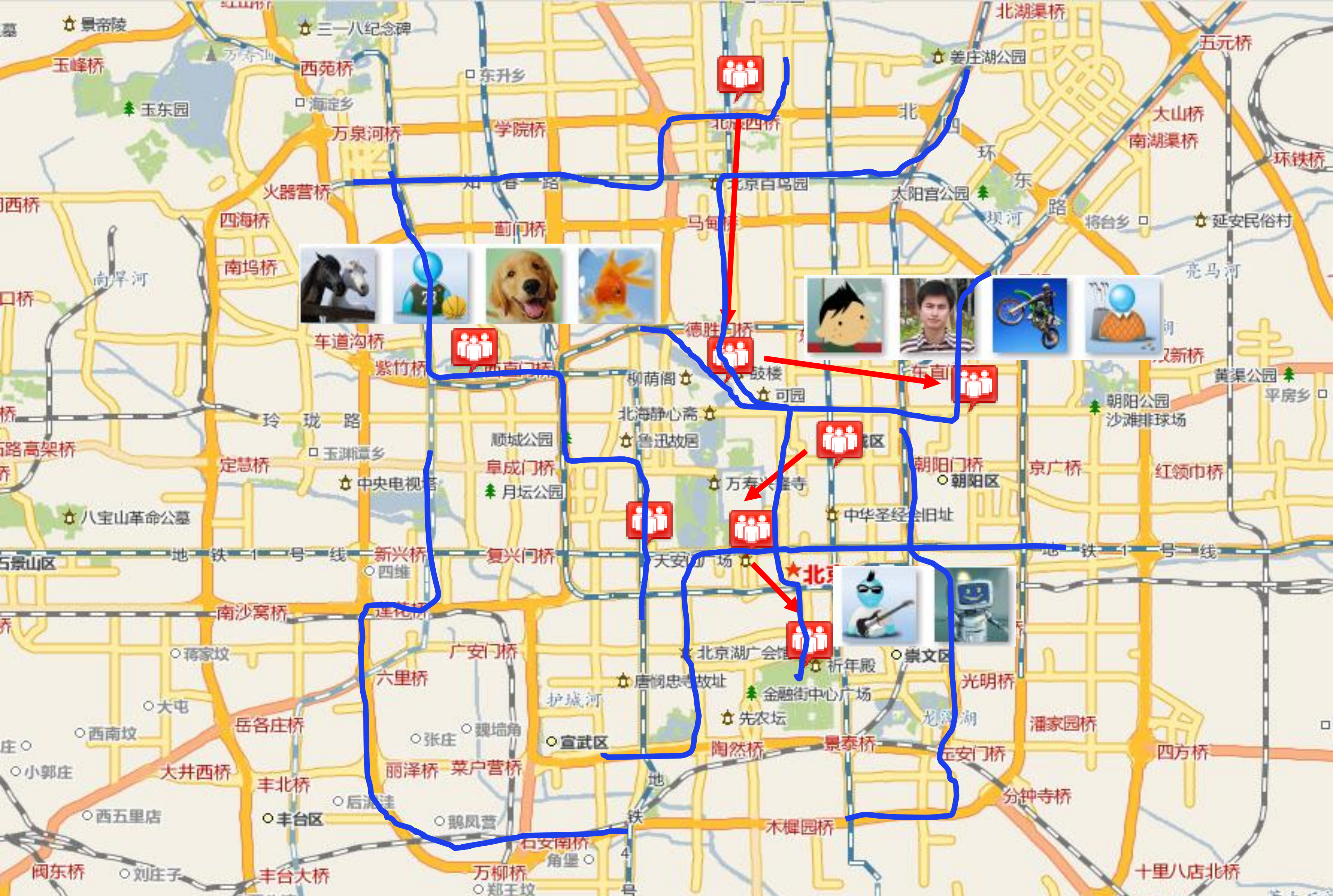
- Jing Yuan, Yu Zheng, et al. T-Drive: Driving Directions Based on Taxi Trajectories. ACM SIGSPATIAL GIS 2010.
- Jing Yuan, Yu Zheng, Xing Xie, Guangzhong Sun. Driving with Knowledge from the Physical World. KDD 2011.
- Jing Yuan, Yu Zheng, Xing Xie, Guangzhong Sun, T-Drive: Enhancing Driving Directions with Taxi Drivers' Intelligence. Transactions on Knowledge and Data Engineering.
- Jing Yuan, Yu Zheng, Liuhang Zhang, Xing Xie, Guangzhong Sun, Where to Find My Next Passenger? , UbiComp 2011.
- Nicholas Jing Yuan, Yu Zheng, Liuhang Zhang, Xing Xie. T-Finder: A Recommender System for Finding Passengers and Vacant Taxis. accepted by IEEE Transactions on Knowledge and Data Engineering .
- Ling-Yin Wei, Yu Zheng, Wen-Chih Peng, Constructing Popular Routes from Uncertain Trajectories. KDD 2012.
- Kai Zheng, Yu Zheng, et al. Reducing Uncertainty of Low-Sampling-Rate Trajectories. ICDE 2012.

Other Social Applications Using City Dynamics

- Mining interesting locations and travel sequences
- Mining user similarity based on location history
- Location-activity recommendations

Mining interesting locations and travel sequences from social media

[1] Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In WWW 2009.



Mining interesting locations, travel sequences, and travel experts from user-generated travel routes

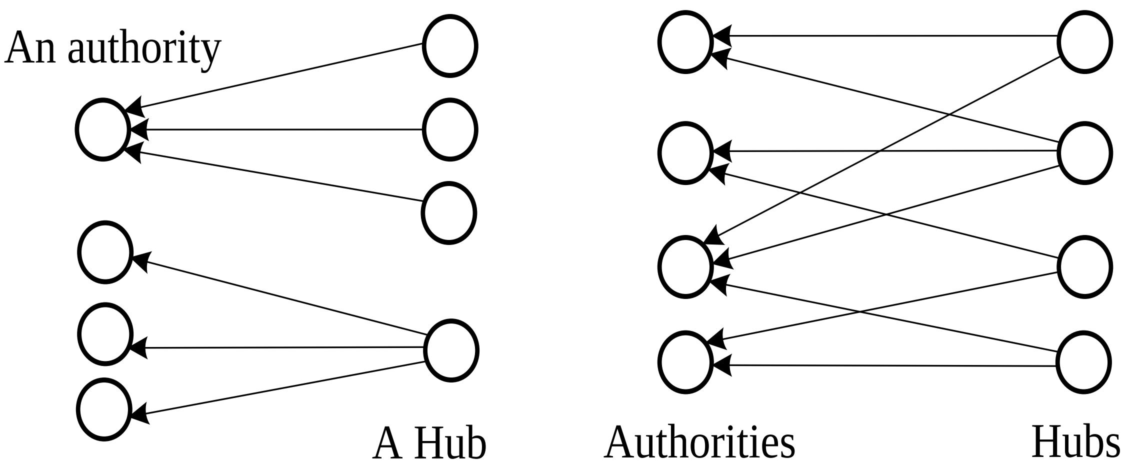

德胜门桥
北滨河公园
安定门桥
香河园
积水潭桥
北二环城市公园
雍和宫桥
鼓楼桥
北京自来水
国际友谊博物馆
南馆公园
东直门桥
后海公园
东环娱乐广
广桥
官园桥
皇城根遗址公园
港澳中心瑞
东四十条桥
东四奥林匹克
阜成门桥
景山公园
朝阳门桥
国宾酒店
中国地质博物馆
华侨大厦
北海公园
王府井大饭店
月坛北桥
北京国际展览中心
月坛公园
故宫博物院
月坛南桥
中山公园
皇城根遗址公园
北京国际饭店
首都博物馆
复兴门桥
新闻大厦酒店
长富宫饭
东单公园
首都大酒店
甘雨桥
西便门桥
翠芳园
全聚德展览馆
北京市规划展览馆

Challenges

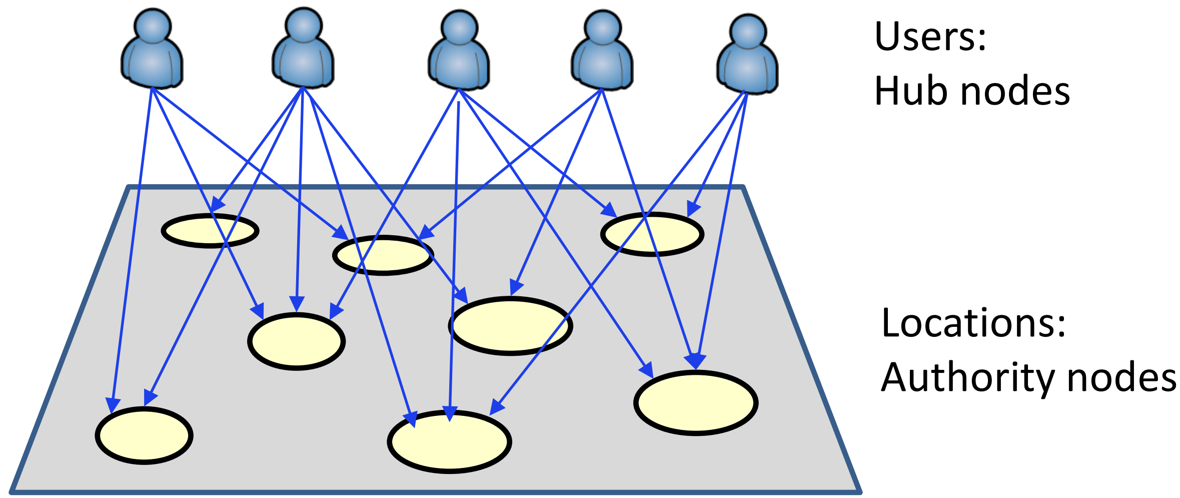
- What is a location? (geographical scales)
- The interest level of a location
  - does not only depend on the number of users who have visited this location
  - but also lie in these users' travel experiences
- How to determine a user's travel experience?
- The location interest and user travel
  - are region-related
  - are relative value (Ranking problem)

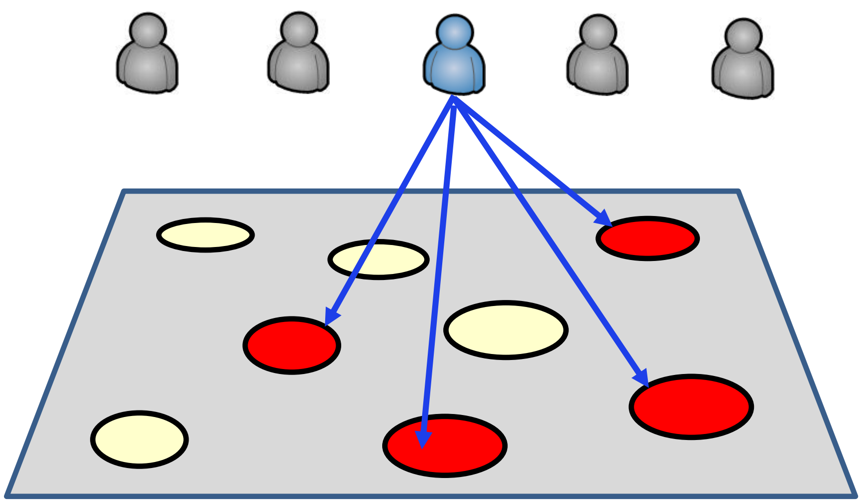
Methodology

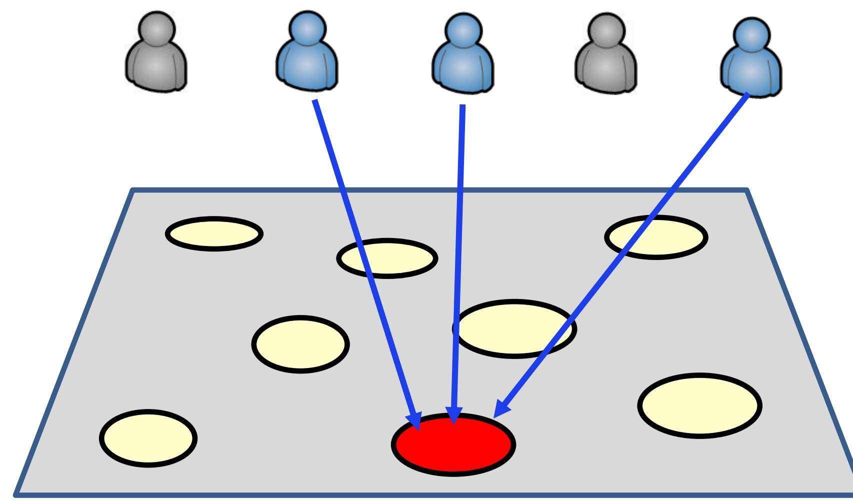
- HITS (hypertext induced topic search) model
  - Authority: a Web page with many in-links
  - Hub: is a page with many out-links
  - Mutual reinforcement relationship



The HITS-based inference model





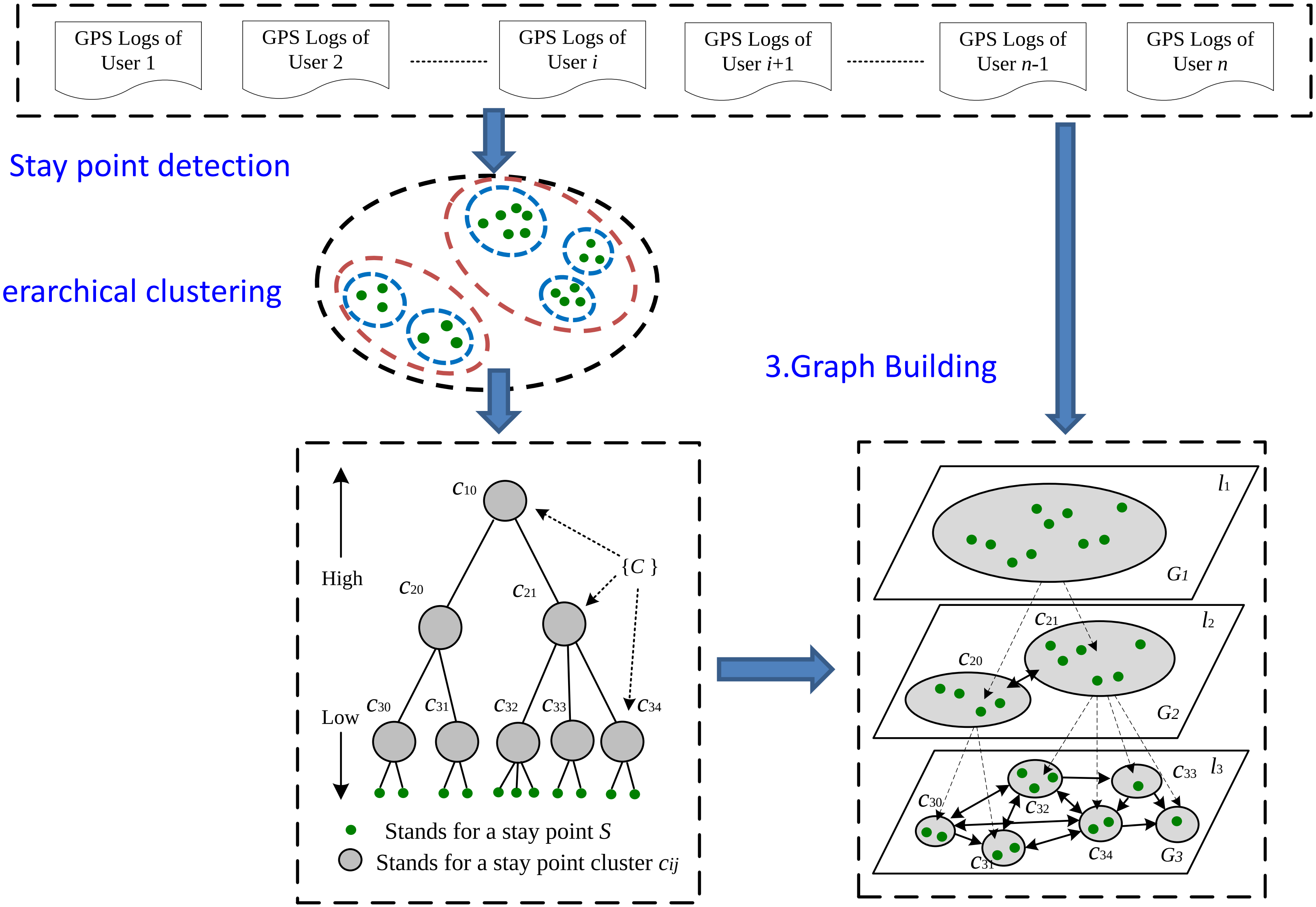


Methodology

1. Stay point detection

2. Hierarchical clustering

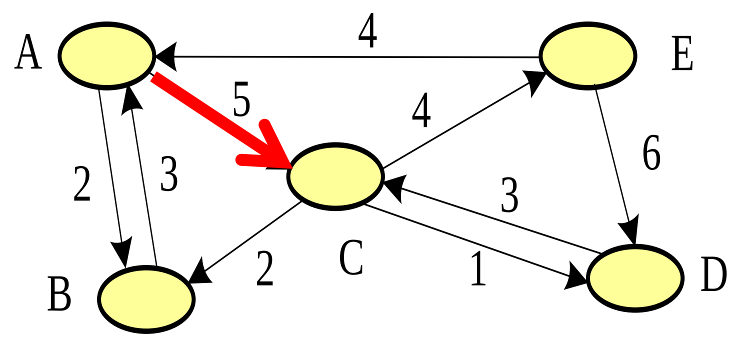
3.Graph Building



Shared Hierarchical Framework

Detecting Interesting Travel Sequences

- Three factors determining the classical score of a sequence:
  - Travel experiences (hub scores) of the users taking the sequence
  - The location interests (authority scores) weighted by
  - The probability that people would take a specific sequence



The classical score of sequence **A→C**:

$$S_{AC} = 5 \times \left(\frac{5}{7} \times a_A + \frac{5}{8} a_C\right) + \sum_{u_k \in U_{AC}} h^k .$$

a_A : Authority score of location A

h^k : User k's hub score

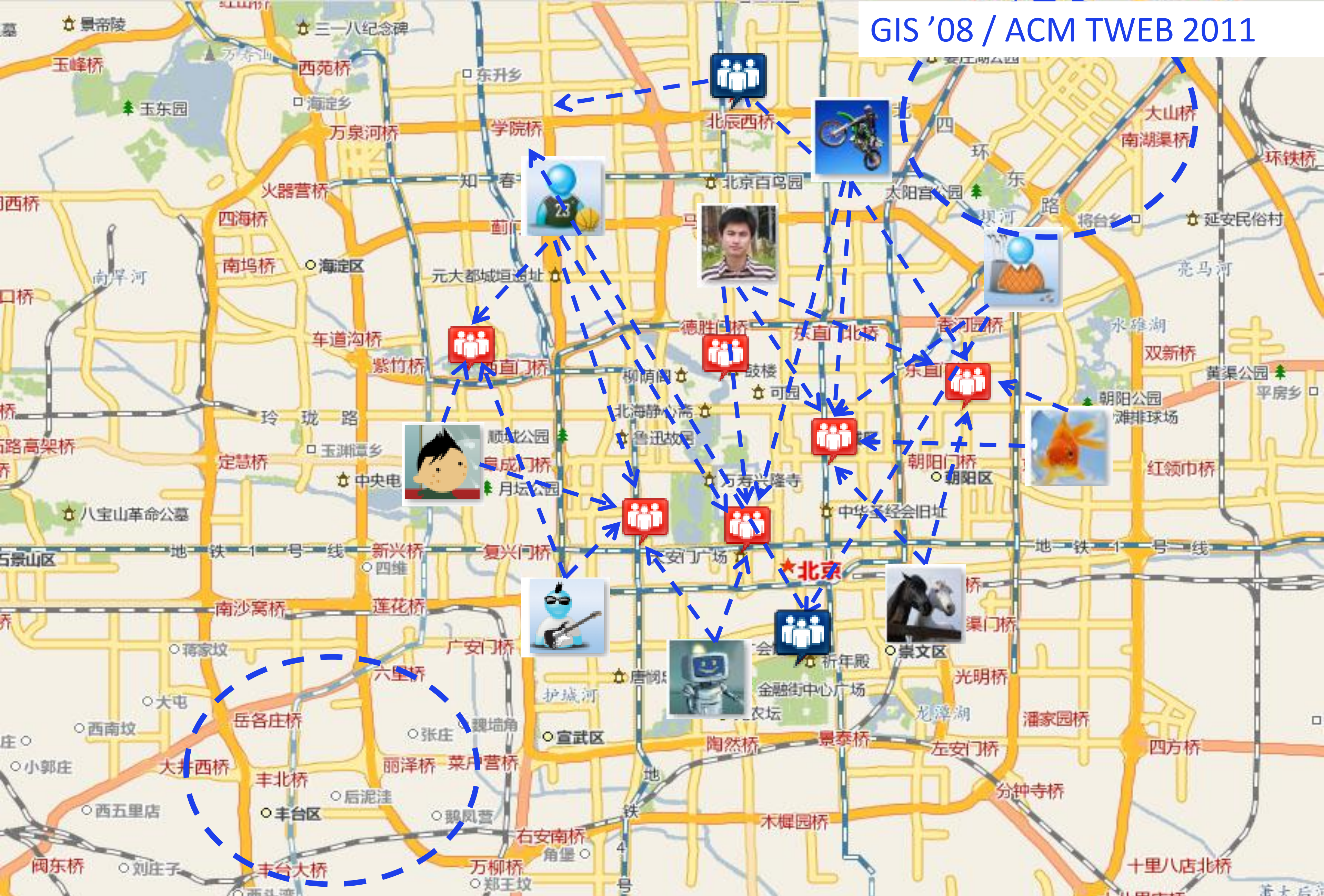
a_C : Authority score of location C

Mining User Similarity Based on Location History

References:

[1] Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, Wei-Ying Ma. Mining user similarity based on location history. In ACM SIGSPATIAL GIS 2008. ACM Press: 1-10.

GIS '08 / ACM TWEB 2011



Grouping users in terms of the similarity between their location histories, and conduct personalized location recommendations.

Mining User Similarity Based on Location History

- Some naïve methods
  - Calculated the overlapped locations
  - The Cosine similarity
  - The Pearson similarity

$$U_1 =< m_1, m_2, ... m_j, \dots , m_N >$$

$$U_2 =< m'_1, m'_2, \dots, m'_j, \dots , m'_N >$$

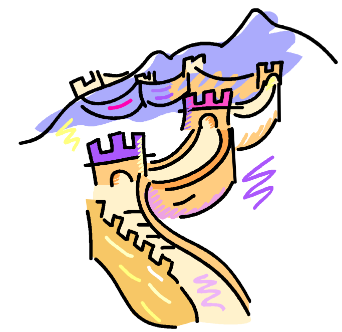
$$M = \begin{array}{c c} & \begin{array}{ccccc} l_0 & l_1 & l_2 & l_3 & l_4 \end{array} \\ \begin{array}{c} u_0 \\ u_1 \\ u_2 \\ u_3 \end{array} & \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

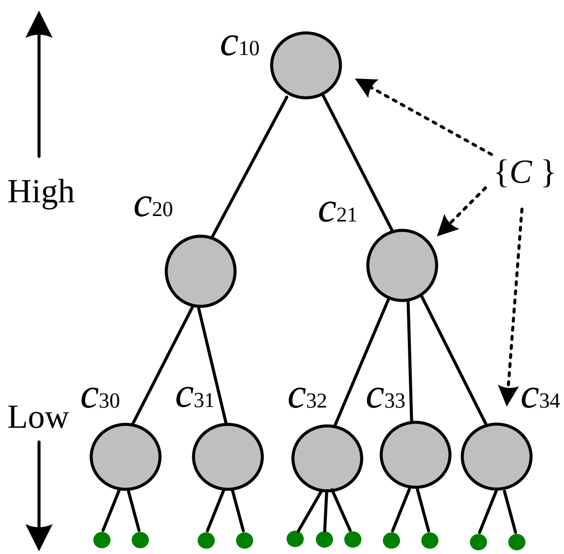
$$sim_{cosine}(u_1, u_2) = \frac{\sum_j m_j m_j'}{\sqrt{\sum_j m_j^2}\sqrt{\sum_j (m'_j)^2}}$$

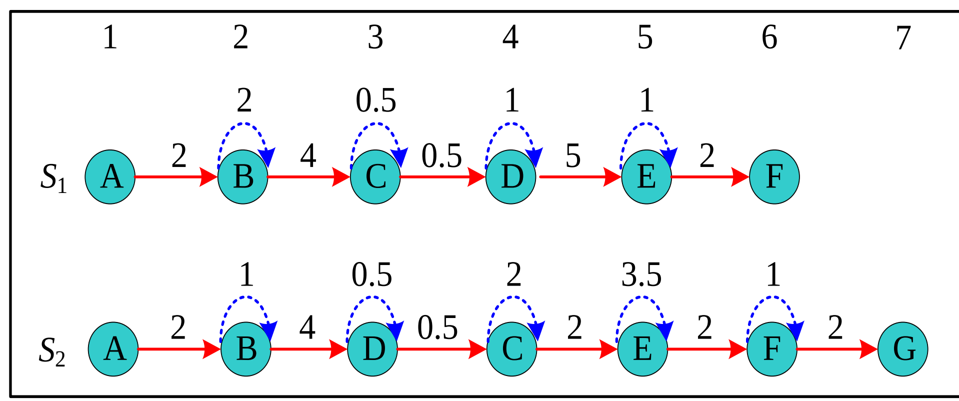
$$sim_{pearson}(u_1, u_2) = \frac{\sum_j (m_j - \overline{U}_1)(m'_j - \overline{U}_2)}{\sqrt{\sum_j (m_j - \overline{U}_1)^2 \sum_j (m'_j - \overline{U}_2)^2}}$$

Mining User Similarity Based on Location History

- Computing user similarity
  - Hierarchical properties
  - Sequential properties
  - Popularity of a location







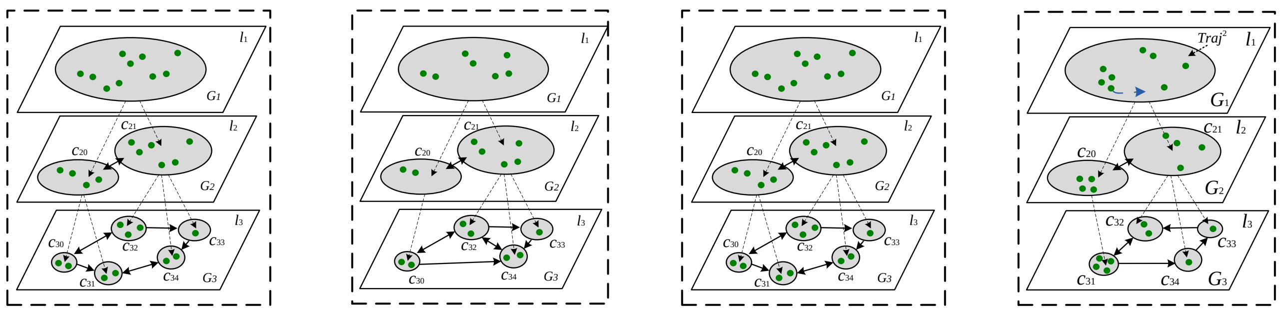
$A \rightarrow B \rightarrow C, \quad A \rightarrow B \rightarrow D \rightarrow E \rightarrow F$

Mining User Similarity Based on Location History

GPS Logs of User 1 | GPS Logs of User 2 | .... | GPS Logs of User *i* | GPS Logs of User *i*+1 | .... | GPS Logs of User *n*-1 | GPS Logs of User *n*

Modeling Location History

A Hierarchical Graph for each individual

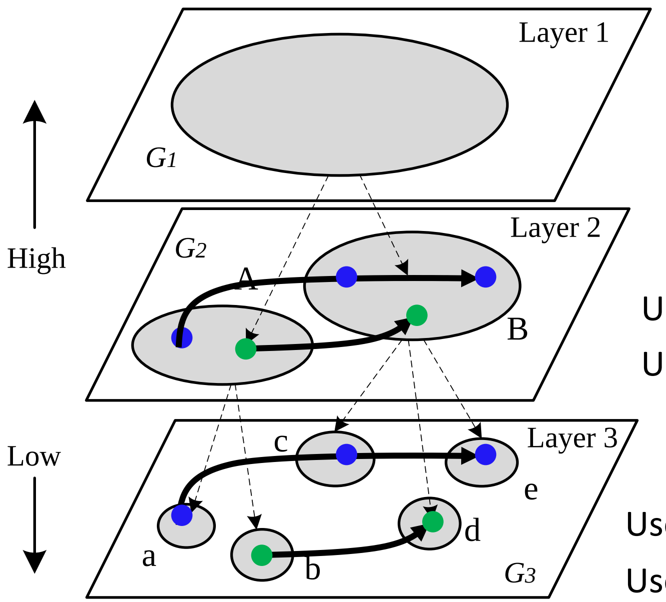


Measuring Similarity

A similarity score *Sij* for each pair of users

Mining User Similarity Based on Location History

User 1 (●): User3 (●) > User 2 (●)



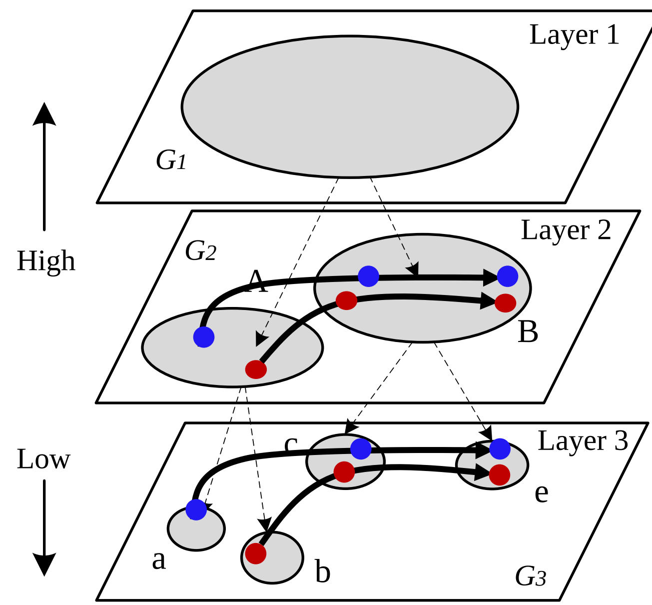
A→ B

User 1: A→ B

User 2: A→ B

User 1: a→ c → e

User 2: b→d



A→ B

User 1: A→ B

User 3: A→ B

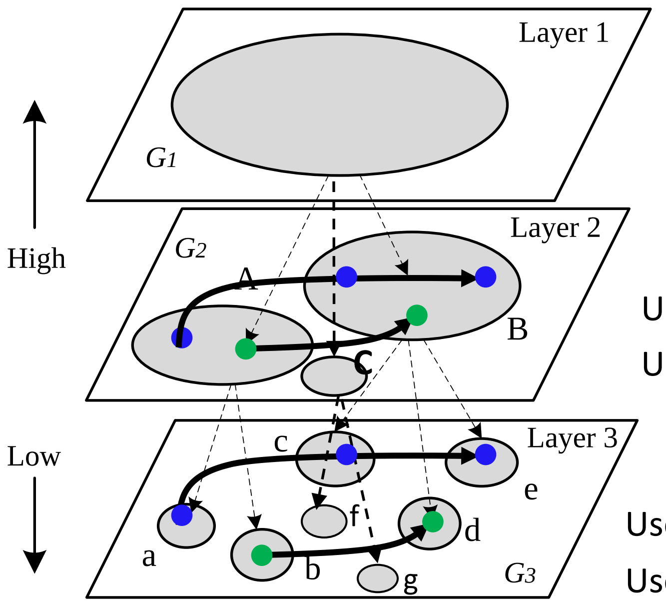
c→ e

User 1: a→ c → e

User 3: b→c → e

Mining User Similarity Based on Location History

User 1 (●): User3 (●) < User 2 (●)



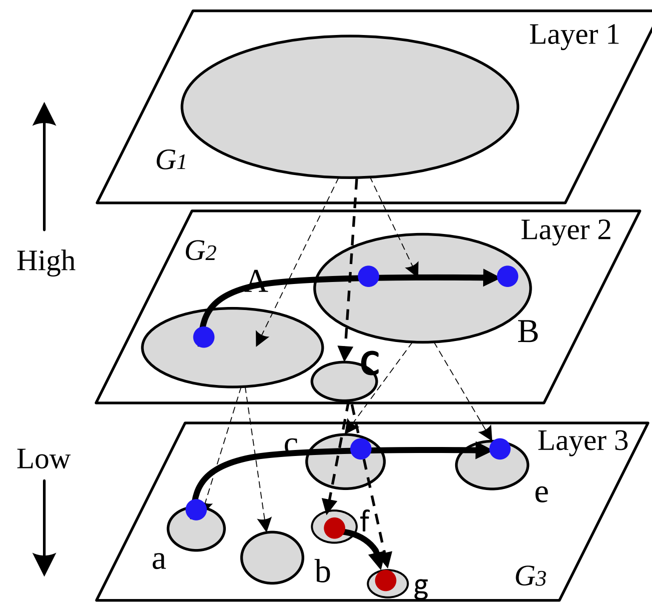
A→ B

User 1: A→ B

User 2: A→ B

User 1: a→ c → e

User 2: b→d



User 1: a→ c → e

User 3: f→g

Location-Activity Recommendation

[1] Vincent Wenchen Zheng, Yu Zheng, Xing Xie, Qiang Yang. Collaborative Location and Activity Recommendations With GPS History Data. In WWW 2010), ACM Press: 1029-1038.

[2] Vincent W. Zheng, Yu Zheng, Xing Xie, Qiang Yang. Learning from GPS Data for Mobile Recommendation. Artificial Intelligence Journal.

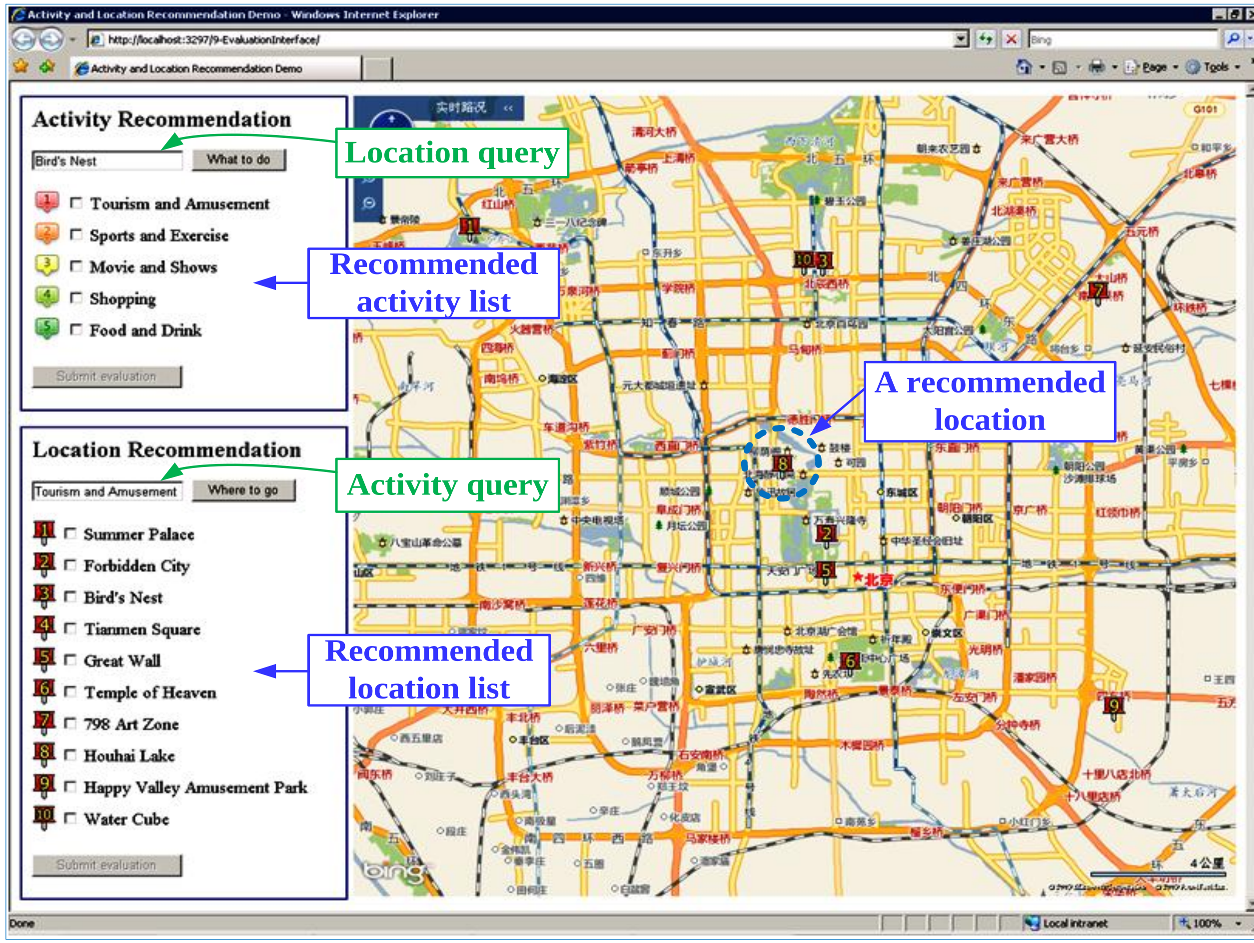
Location-Activity Recommendation

Q1: what can I do there if I visit some place?

(Activity recommendation given location query)

Q2: where should I go if I want to do something?

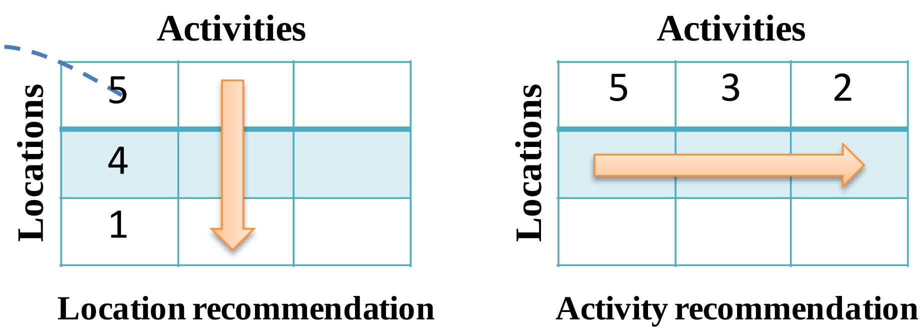
(Location recommendation given activity query)



Problem Definition

- How to well model the location-activity relation
  - Encode it into a matrix

An entry denotes how popular an activity is performed at a location



Ranking along the Columns or rows

- Example

	Tourism	Exhibition	Shopping
Forbidden City	5	4	2
Bird's Nest	4	3	1
Zhongguancun	1	2	6

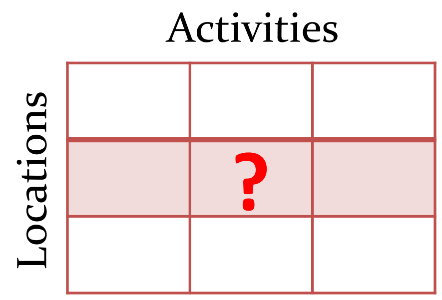
Location recommendation
Tourism:
Forbidden City > Bird's Nest > Zhongguancun

Activity recommendation
Forbidden City:
Tourism > Exhibition > Shopping

Location-Activity Recommendation

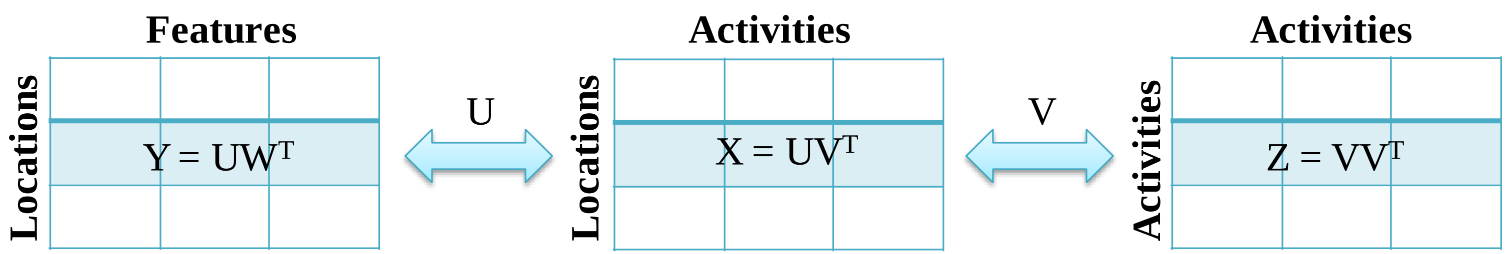
Data sparseness (<0.6% entries are filled)

	Tourism	Exhibition	Shopping
Forbidden City	5	?	?
Bird’s Nest	?	1	?
Zhongguancun	1	?	6



Solution: Collaborative Location and Activity Recommendation (CLAR)

- Collaborative filtering, with collective matrix factorization



  – Low rank approximation, by minimizing

$$L(U,V,W) = \frac{1}{2} \| I \circ (X - UV^T) \|_F^2 + \frac{\lambda_1}{2} \| Y - UW^T \|_F^2 +$$
$$\frac{\lambda_2}{2} \| Z - VV^T \|_F^2 + \frac{\lambda_3}{2} (\| U \|_F^2 + \| V \|_F^2 + \| W \|_F^2)$$

where U, V and W are the low-dimensional representations for the locations, activities and location features, respectively. I is an indicatory matrix.

References

- Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In WWW 2009.
- Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, Wei-Ying Ma. Mining user similarity based on location history. ACM SIGSPATIAL GIS 2008.
- Vincent Wenchen Zheng, Yu Zheng, Xing Xie, Qiang Yang. Collaborative Location and Activity Recommendations With GPS History Data. In WWW 2010), ACM Press: 1029-1038.
- Vincent W. Zheng, Yu Zheng, Xing Xie, Qiang Yang. Learning from GPS Data for Mobile Recommendation. Artificial Intelligence Journal.

Outline

- Background
- Fundamental algorithms
- Application scenarios for end users
  - Driving direction service
  - Taxi recommendations
  - Travel itinerary suggestion
- Application scenarios for government
  - Anomaly detection
  - Glean the problematic urban planning
  - Discover regions of different functions

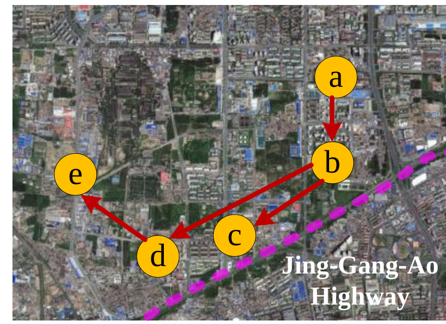
Anomaly Detection in City Dynamics

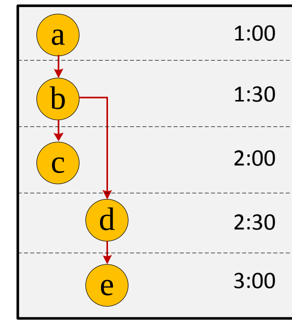
Anomaly Detection in City Dynamics

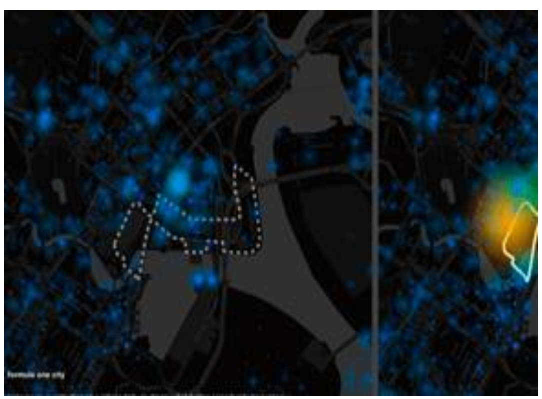
- What is an anomaly in city dynamics
  - Traffic accidents, controls, under construction
  - Disasters: downpour, surface collapse, snow storms, fires
  - Celebrations, games, and big events
  - ……
- Data sources
  - Transportation sensor data: GPS data, loop sensors
  - Social media data: tweeters, weibo, foursquare
  - Mobile phone data
  - Web log data: query log

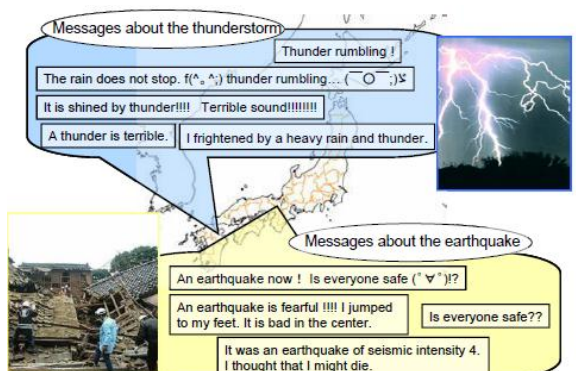
Anomaly Detection in City Dynamics

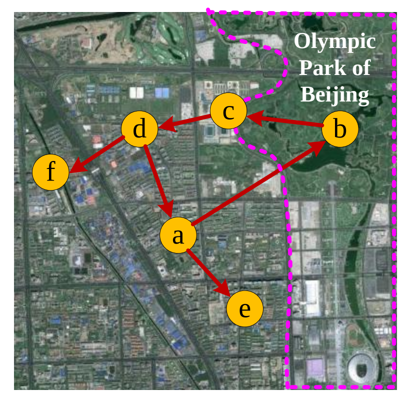
- Examples
  - Jing-Gang-Ao highway
  - Olympic park of Beijing
  - Earthquake in Japan
  - Singapore F-1 race

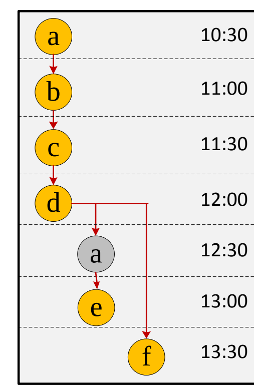










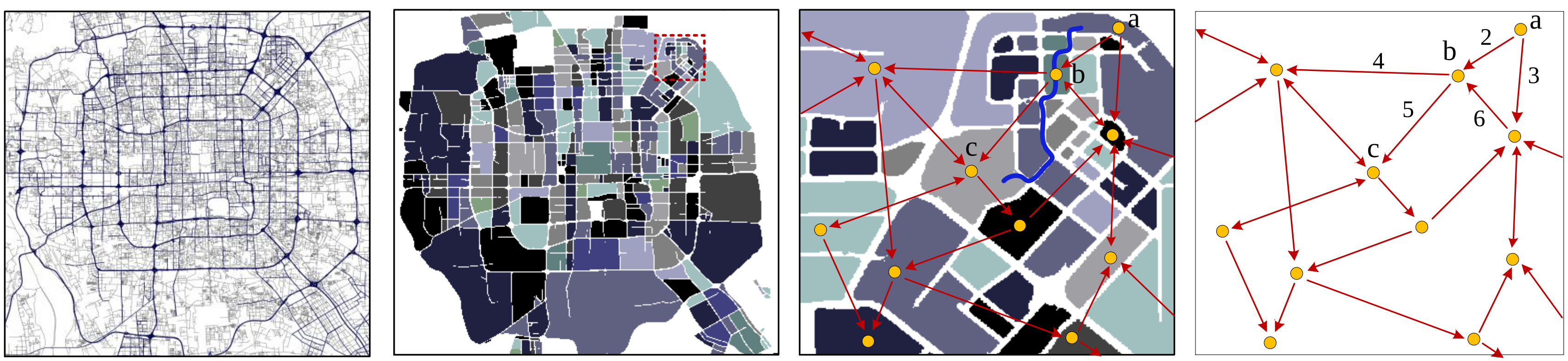


Anomaly Detection in City Dynamics

- Methods (depending on applications)
  - Spatio-temporal outlier detection methods
  - PCA, DP algorithms
- Publications
  - Wei Liu, Yu Zheng, et al. *Discovering Spatio-Temporal Causal Interactions in Traffic Data Streams*. KDD 2011

Anomaly Detection in City Dynamics

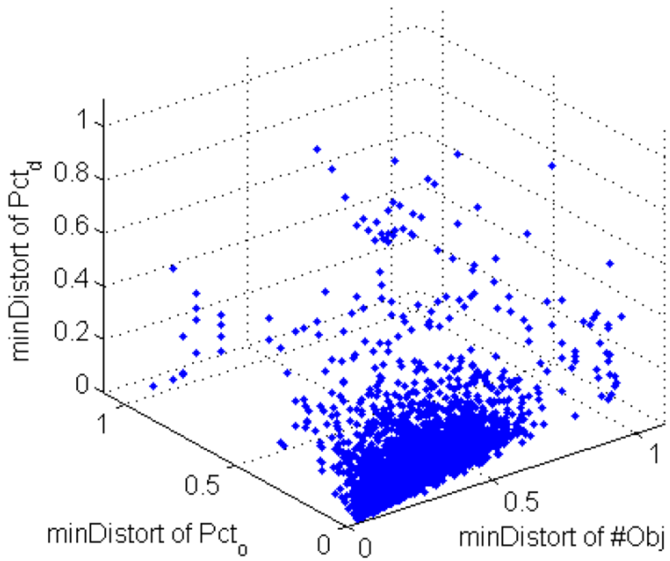
- Traffic modeling
  - Map segmentation
  - Building a region graph
  - Identify three features for each $<\#\mathrm{Obj},\ \mathrm{Pct}_o,\ \mathrm{Pct}_d>$

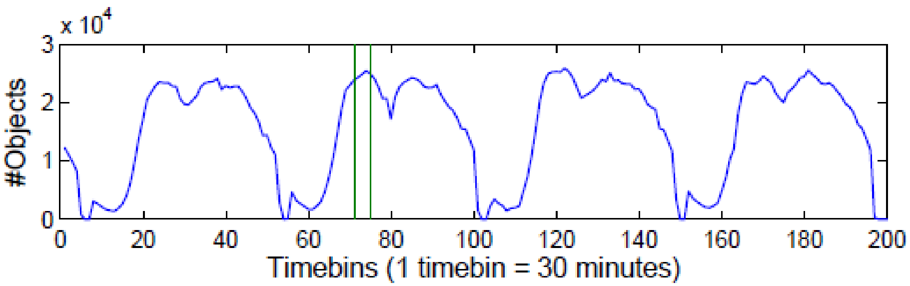


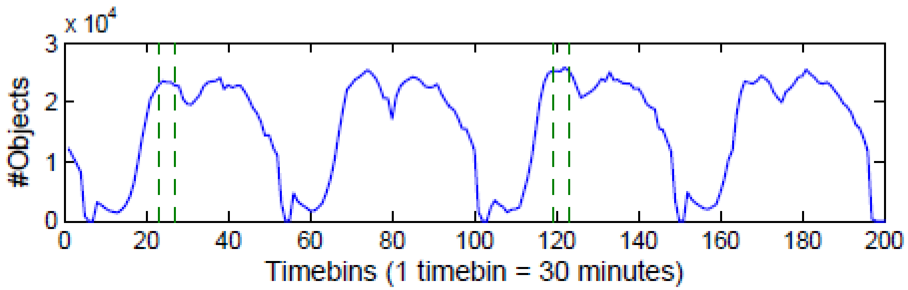
(a) Road network of Beijing (b) Partitioned regions (c)Example of traffic among regions (d) A graph of regions

Anomaly Detection in City Dynamics

- Anomaly detection
  - Calculate the distance between the corresponding item in the feature vectors of two time bins
    $$<\#\mathrm{Obj},\ \mathrm{Pct}_o,\ \mathrm{Pct}_d>$$
    $$<\#\mathrm{Obj},\ \mathrm{Pct}_o,\ \mathrm{Pct}_d>$$
  - Identify the minDistort
  - Find out the outlier points as anomalies using Mahalanobis distance

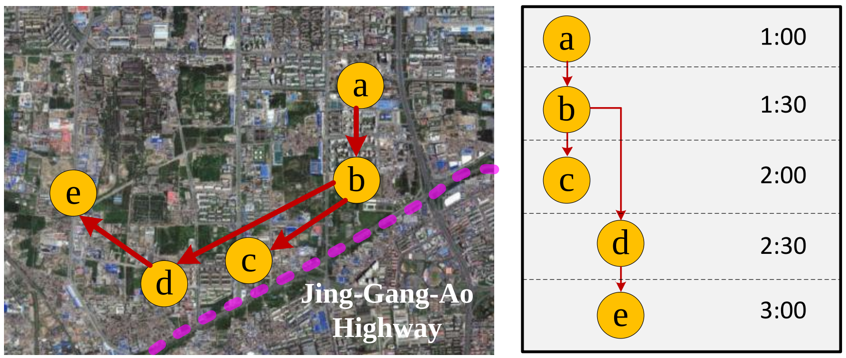


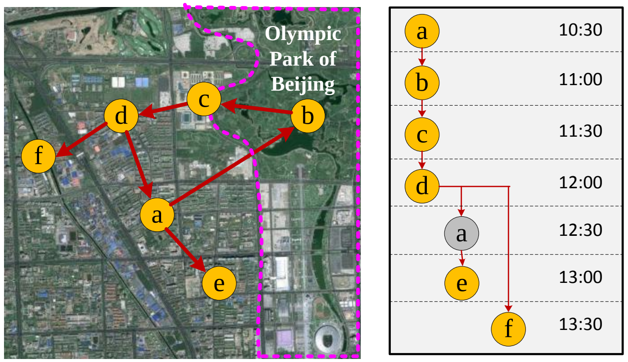




Anomaly Detection in City Dynamics

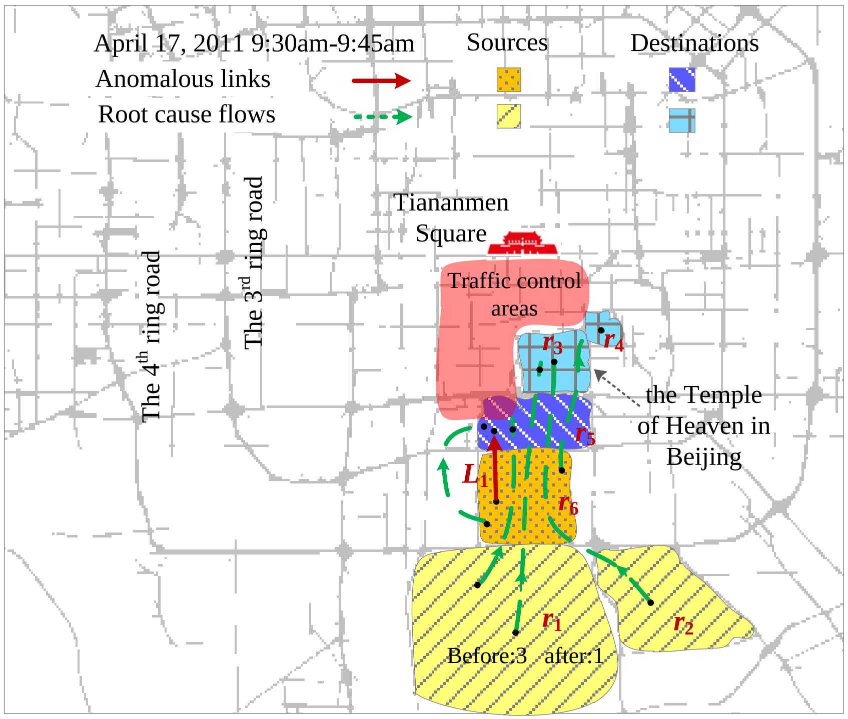
- Results



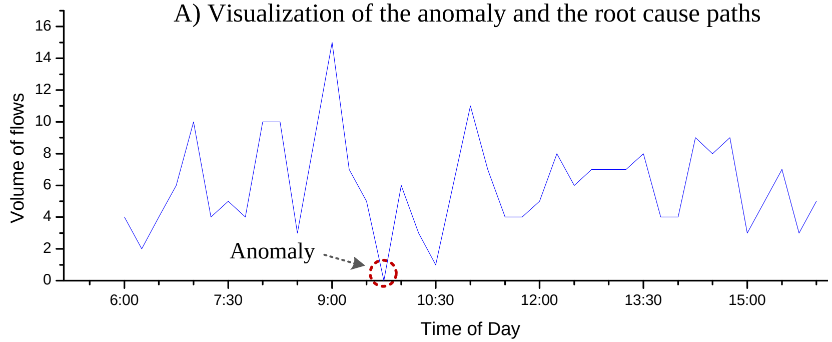


- Next step
  - Identify the root cause of the problem
  - From regions to road segments
  - Estimate the impact of an anomaly and effective visualization

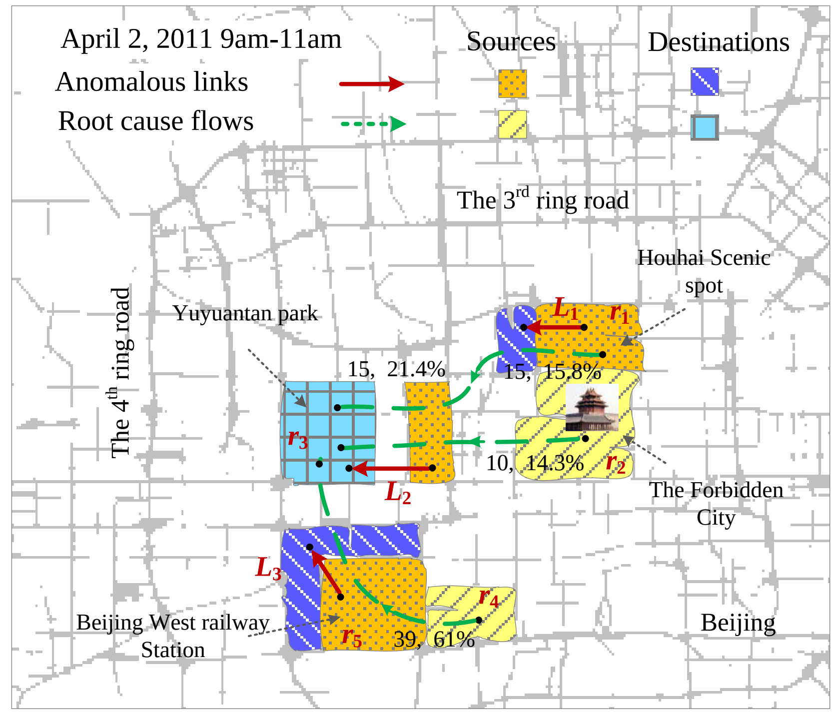
Anomaly Detection in City Dynamics



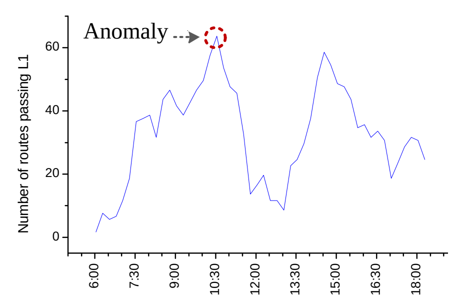
A) Visualization of the anomaly and the root cause paths



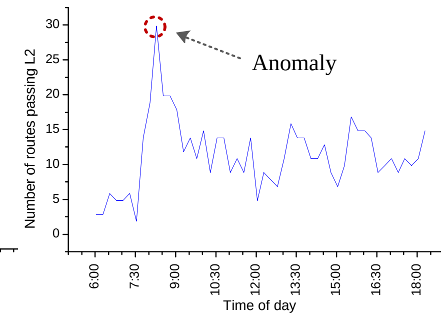
B) Traffic flows on L1



A) Visualization of the anomaly and the root cause paths



B) Traffic flows on L1



C) Traffic flows on L2

Urban Computing for Urban Planning

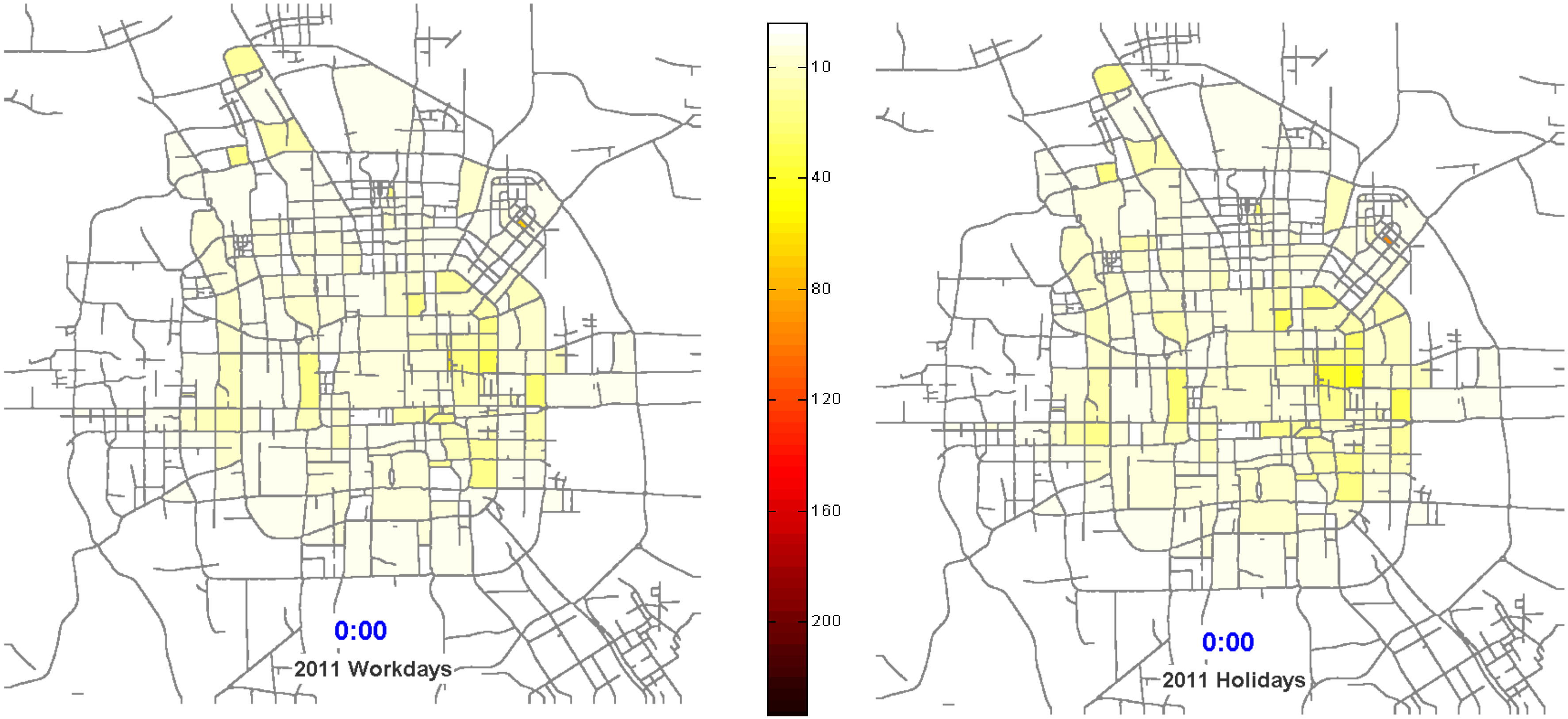


Urban Computing for Urban Planning

- Goals
  - City-wide traffic modeling
  - Evaluate city configurations
  - Suggest potential improvement to city planners
  - Identify root causes of the problem
- Datasets
  - Taxi trajectories: March to May, 2009, 2010, 2011
  - Beijing maps 2009, 2010, 2011

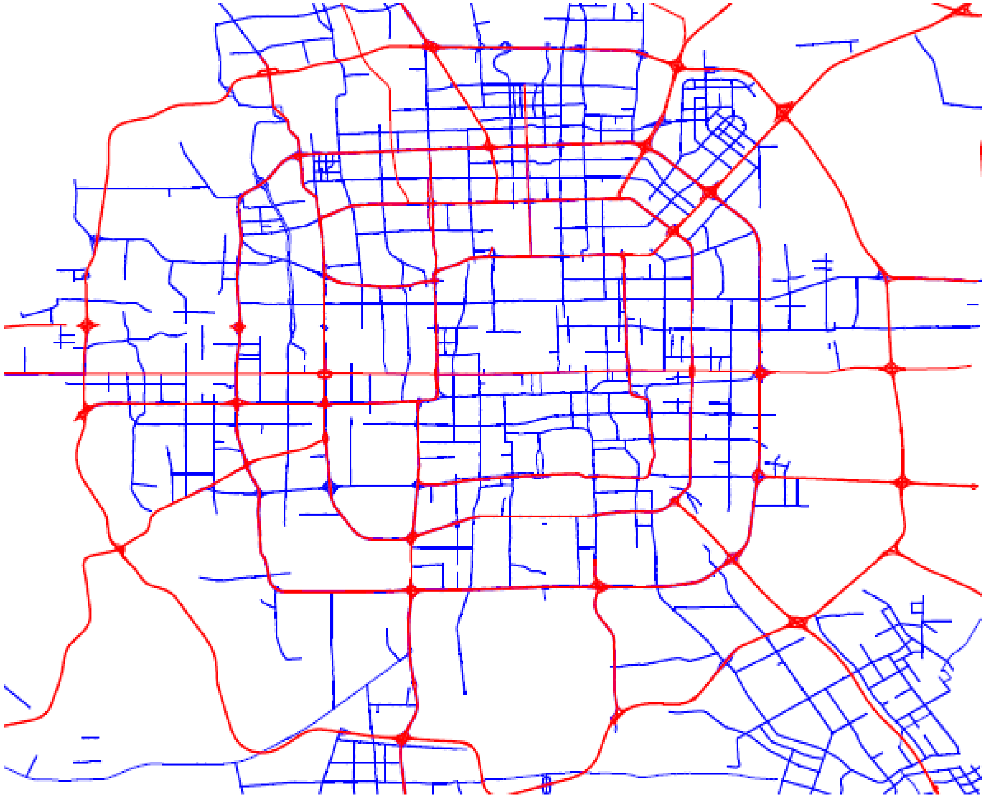
Yu Zheng, Yanchi Liu, Jing Yuan, Xing Xie, Urban Computing with Taxicabs, Best paper nominee in UbiComp 2011

Heat Maps of Beijing (2011)



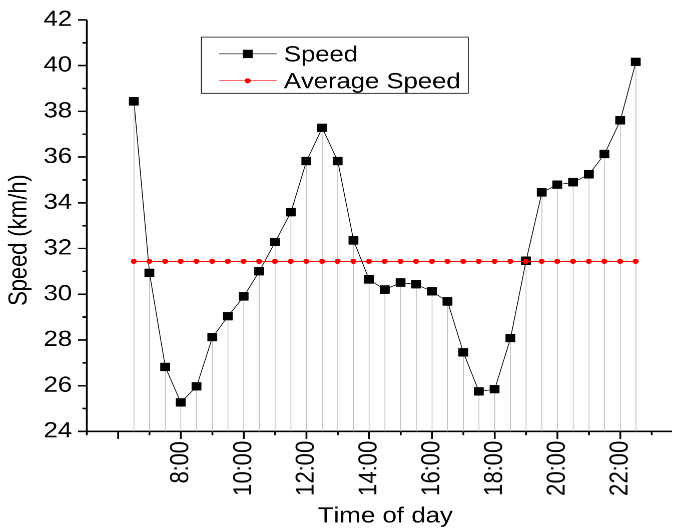
City-Wide Traffic Modeling

- Partition a city into regions with major roads
- Regions are root causes of the problem

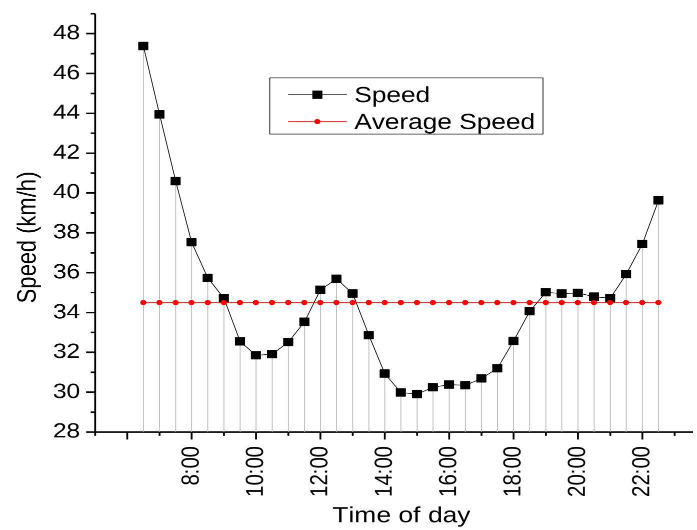


City-Wide Traffic Modeling

- Partition the dataset by time slots (a data-driven method)



Workday



Non-workday

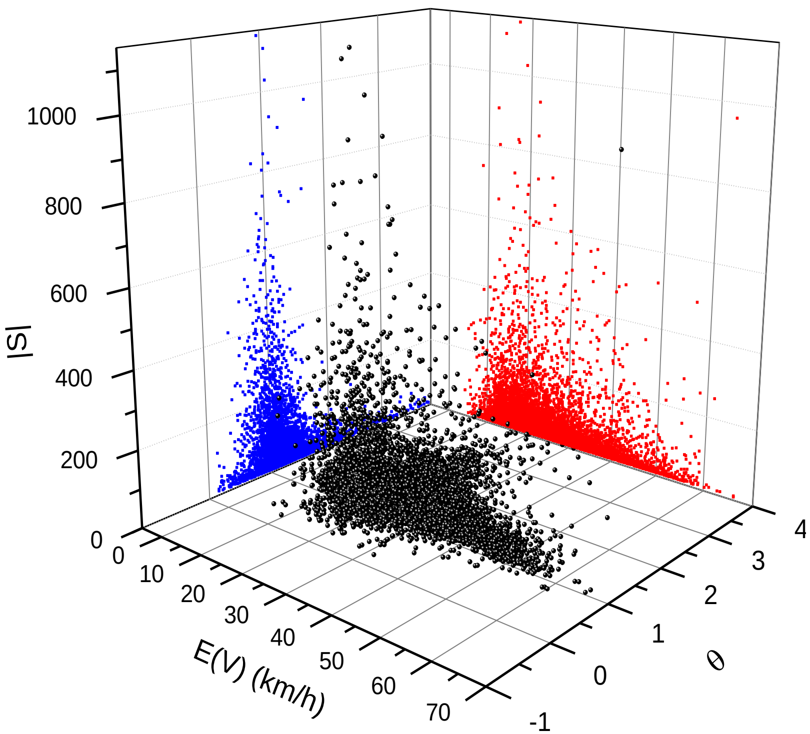
Time	Work day	Non-Workday
Slot 1	7:00am-10:30am	9:00am-12:30pm
Slot 2	10:30am-4:00pm	12:30pm-7:30pm
Slot 3	4:00pm-7:30pm	7:30pm-9:00am
Slot 4	7:30pm-7:00am	

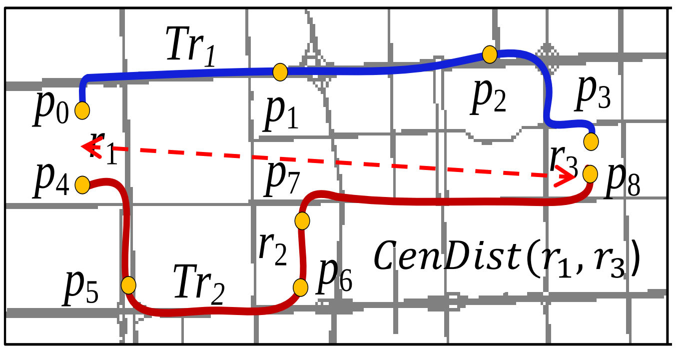
City-Wide Traffic Modeling

- Project taxi trajectories onto these regions
- Building a region graph for each time slot

Finding Problematic Edges

- Extracting features from each edge
  - |S|: Number of taxis
  - E(v): Expectation of speed
  - $\theta = E(D)/CenDist(r_1, r_3)$

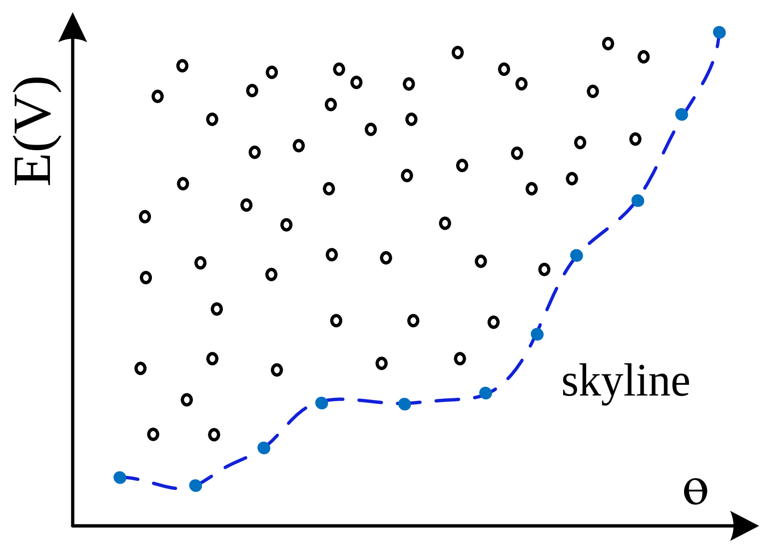




Finding Problematic Edges

- Select edges with |S| above average
- Detect Skyline edges according to $< E(V), \theta >$

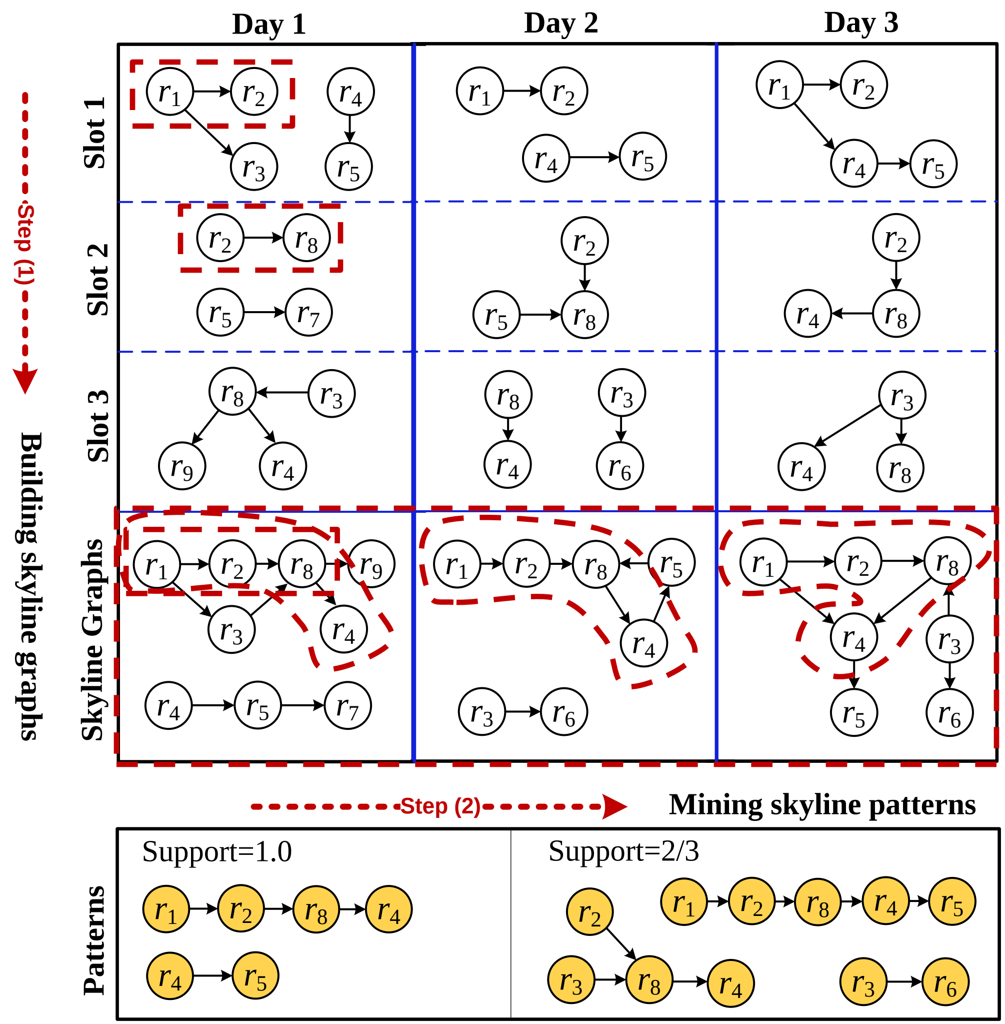
  Select edges with **big** θ and **small** $E(V)$



A) A skyline

Making Sense of Individual Problematic Edges

- Formulate skyline graphs for each day
- Mining frequent sub-graph patterns across days
  - To avoid false alert
  - Deep understanding



Top 10 most frequent problematic edges

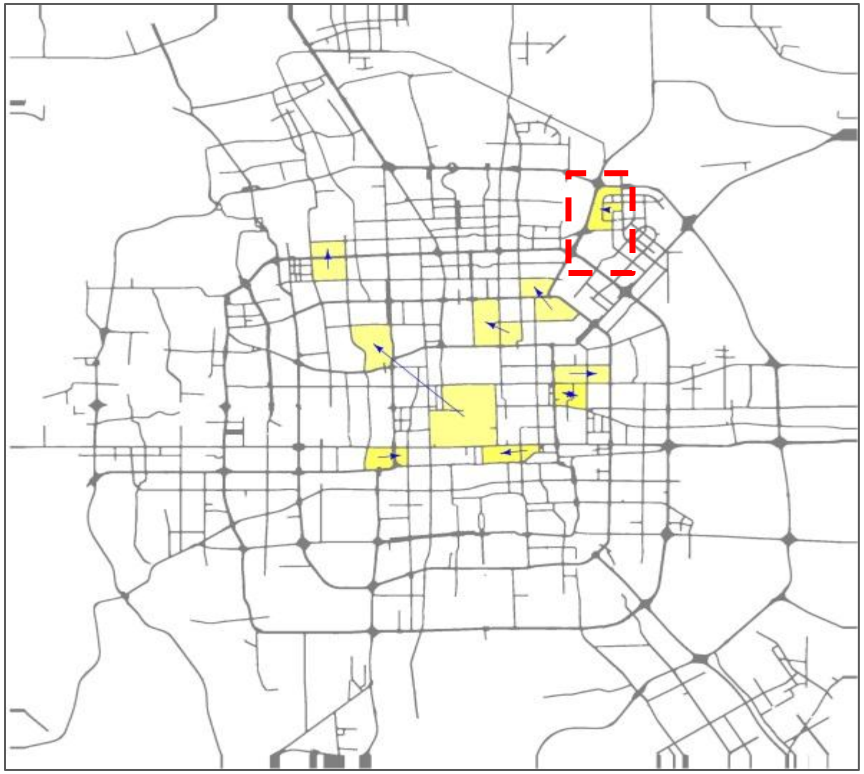
Year 2011

Top 10 most frequent problematic edges

Top 10 hottest regions

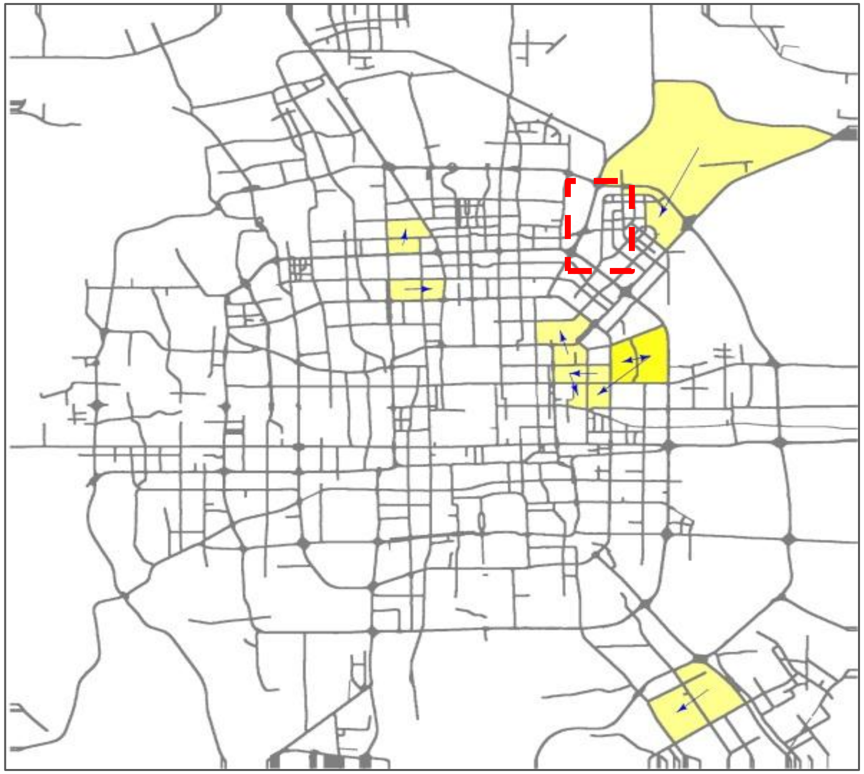
Example 1

2009

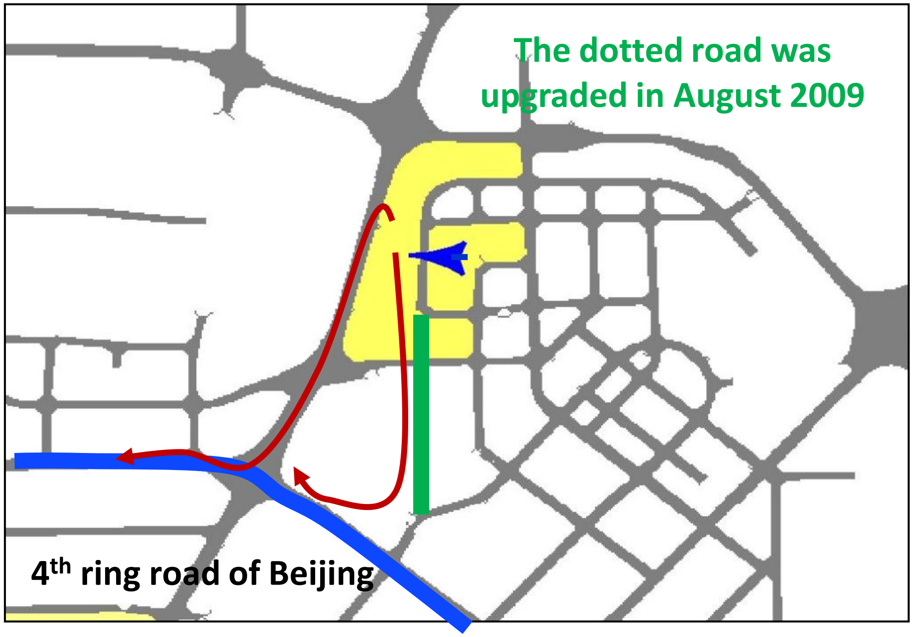


Workdays

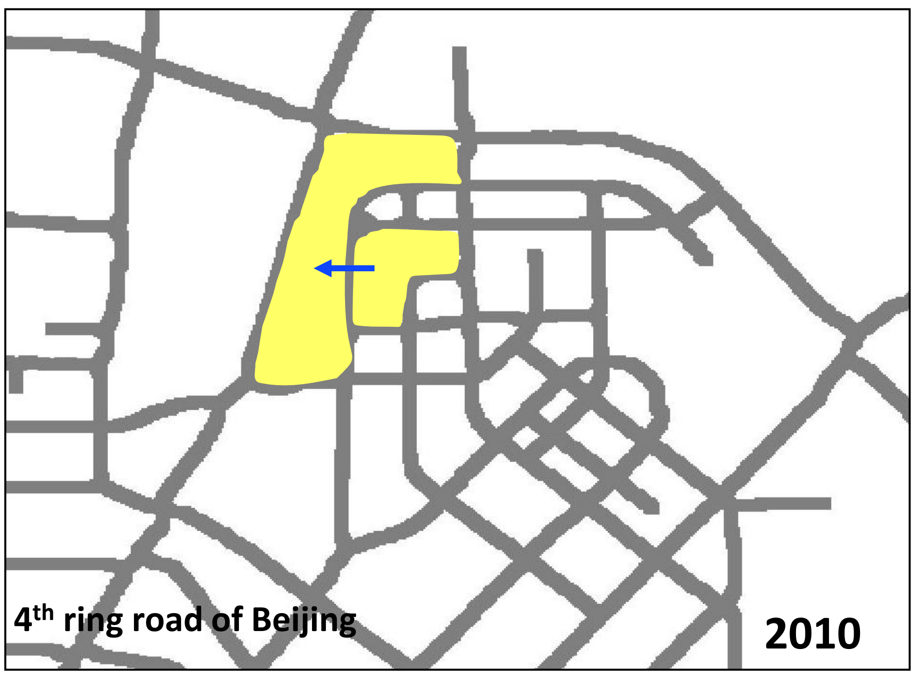
2010



Workdays

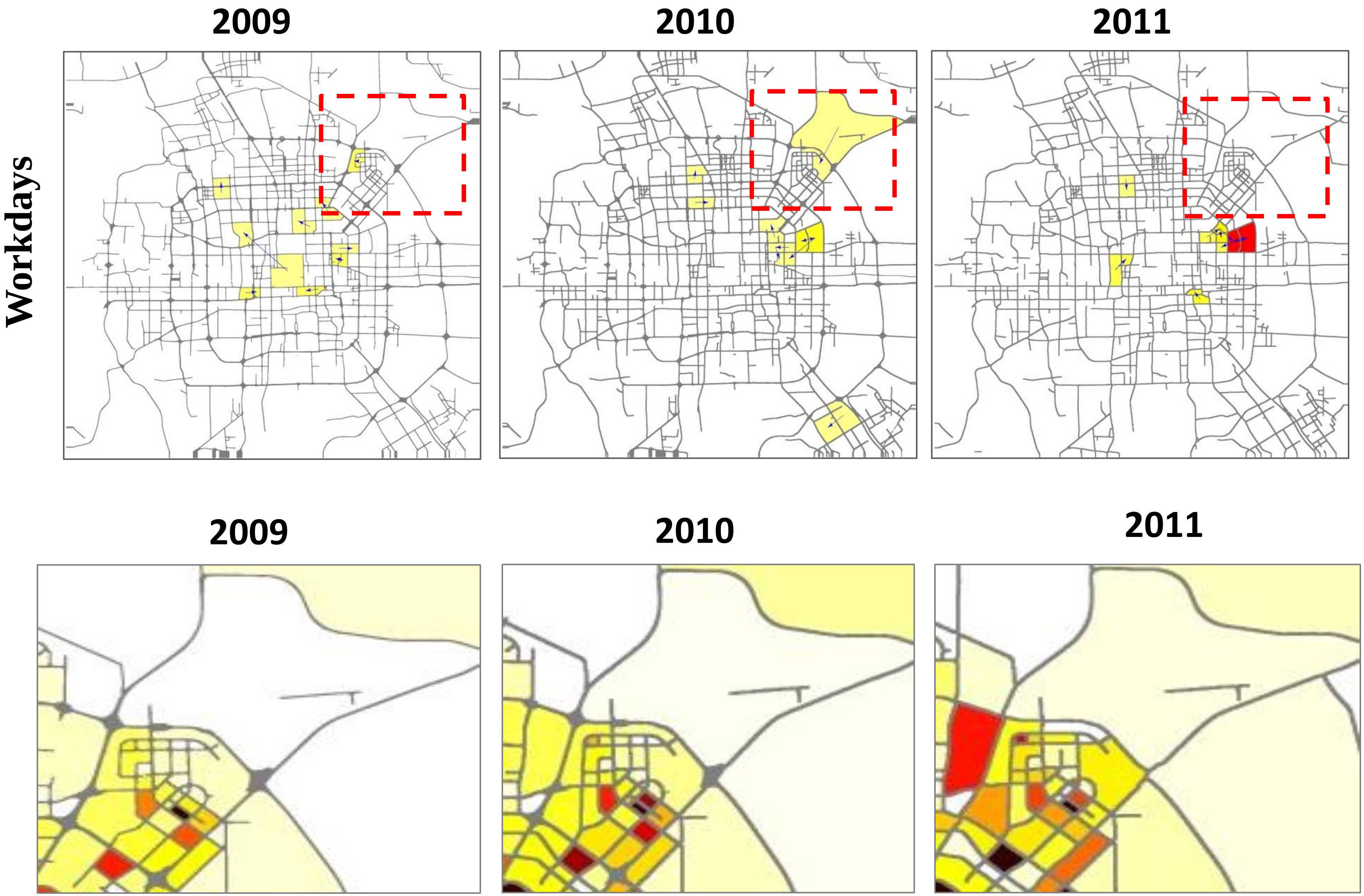

The dotted road was
upgraded in August 2009
4^{th} ring road of Beijing


2009
house.focus.cn

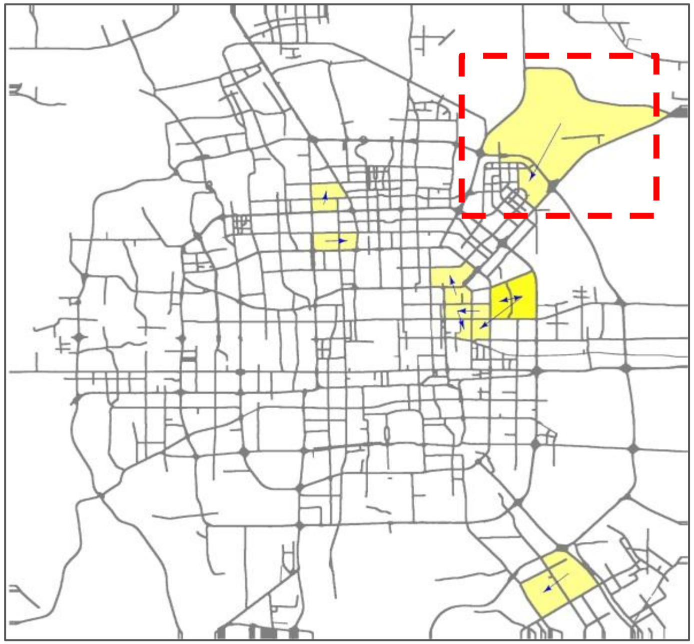

4^{th} ring road of Beijing
2010


北 N
望京西路（湖光北街）
WANGJING West Rd
湖光中街
HUGUANG Middle St
湖光中街
HUGUANG Middle St
（南湖北路 南湖南路）
（京承高速）
湖光中街中口
2010

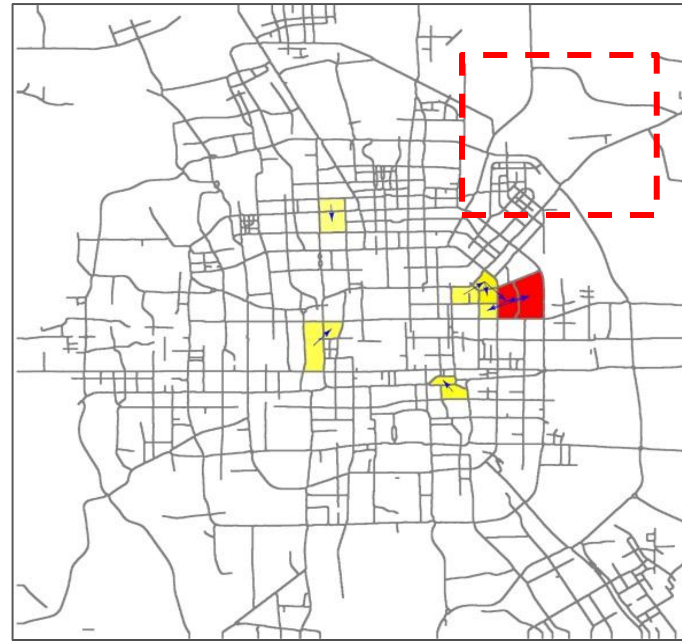
Example 2



2010



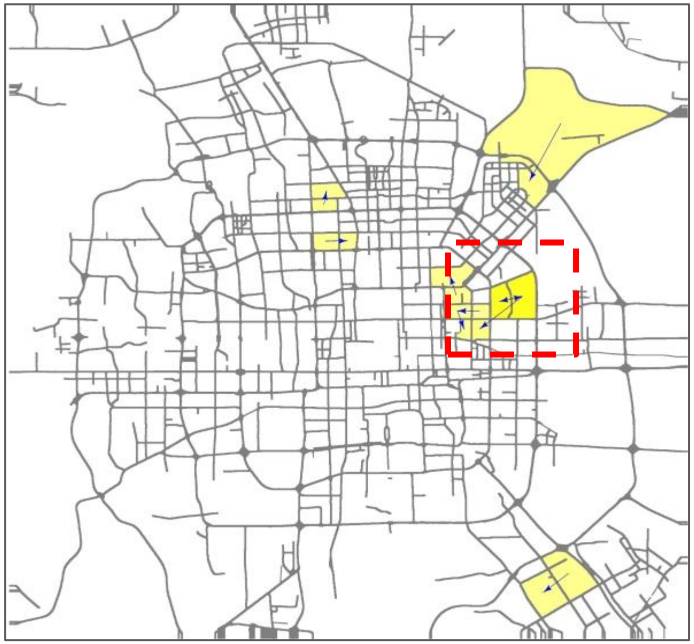
2011



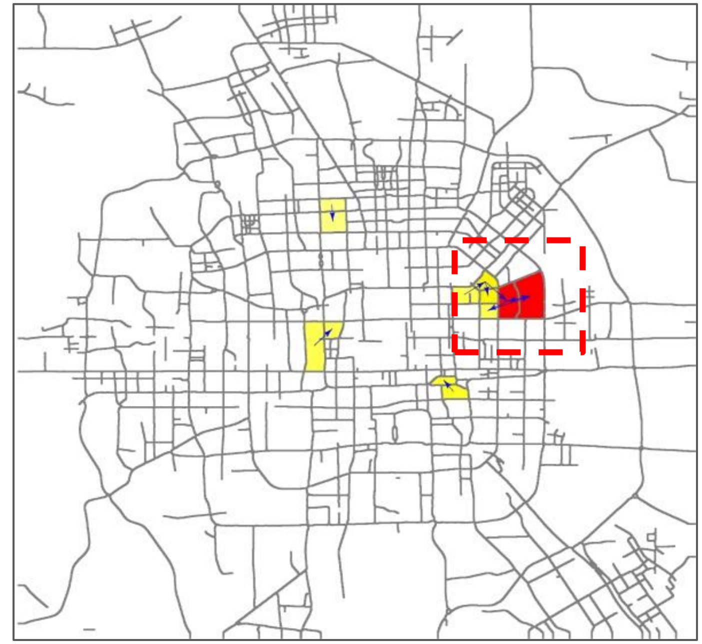
Dec. 2010 Subway Line 15 was launched

Example 3

2010



2011



Statistics on the taxi data across years

Workday

	# of trips/per taxi/per day	average distance (km)	average time (min)	average speed (km/h)	Rd	Rt
2009	16.8	6.72	16.6	26.06	0.62	0.46
2010	15.9	7.45	18.1	25.13	0.65	0.47
2011	17.5	7.97	18.6	25.74	0.68	0.62

Holiday

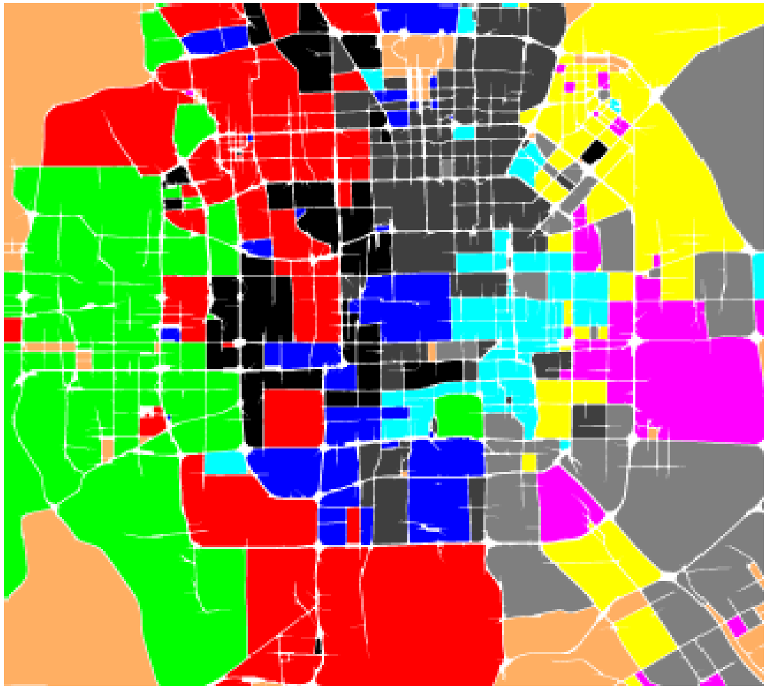
	# of trips/per taxi/per day	average distance (km)	average time (min)	average speed (km/h)	Rd	Rt
2009	13.70	6.58	15.35	27.63	0.61	0.43
2010	12.76	7.35	16.63	26.94	0.63	0.43
2011	13.20	7.87	17.23	27.33	0.67	0.60

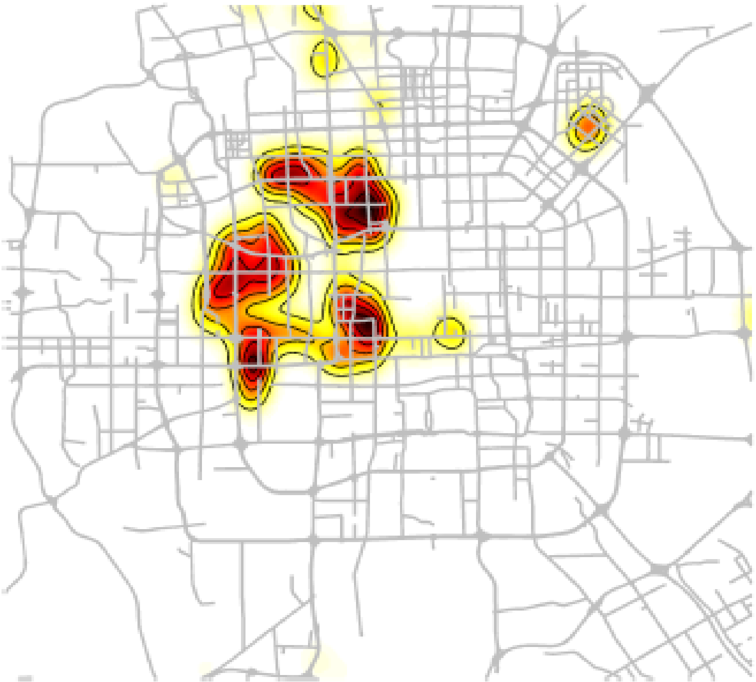
Discover Regions of Different Functions using Human Mobility and POIs

Jing Yuan, Yu Zheng, Xing Xie. Discovering regions of different functions in a city using human mobility and POIs. KDD 2012

Goals

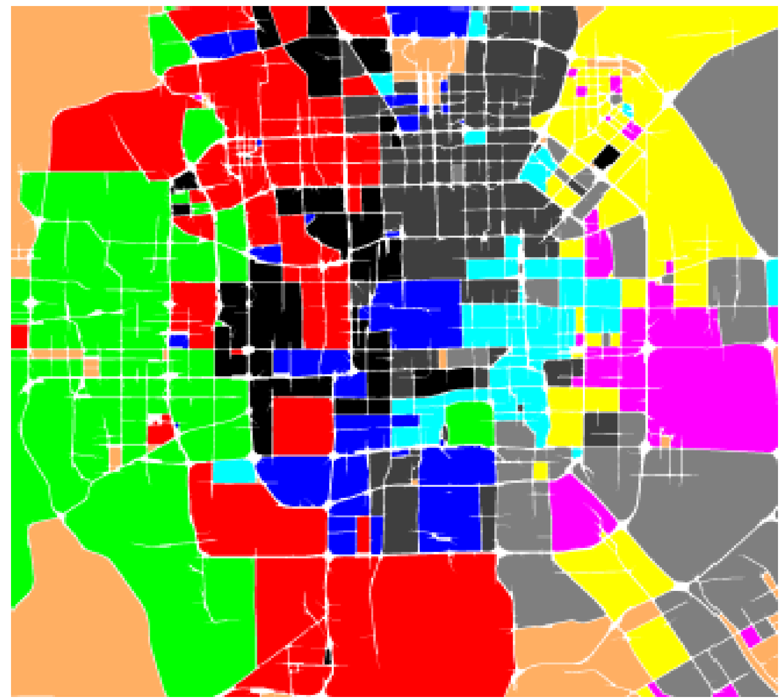
- Discovery regions of different functions
- Identify the function density in urban areas





Applications

- Calibrating urban planning
- Business allocation
- Social recommendations

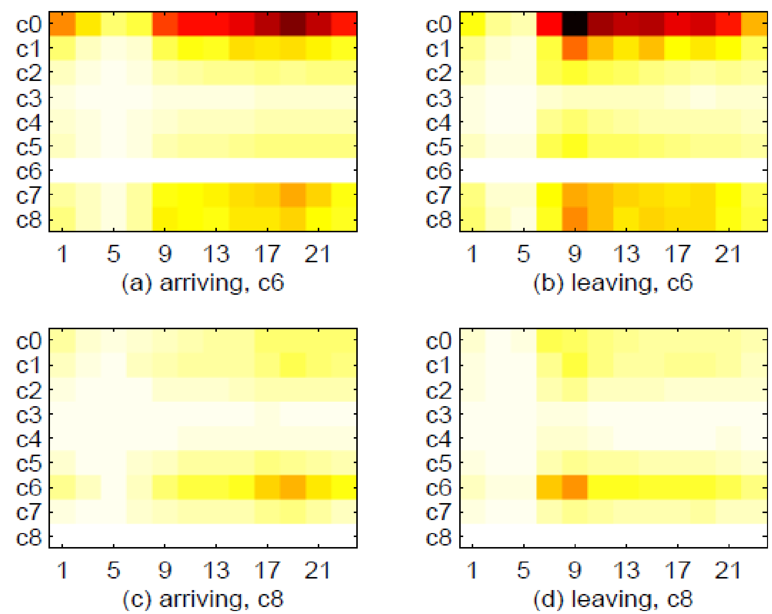


Motivation and Challenges

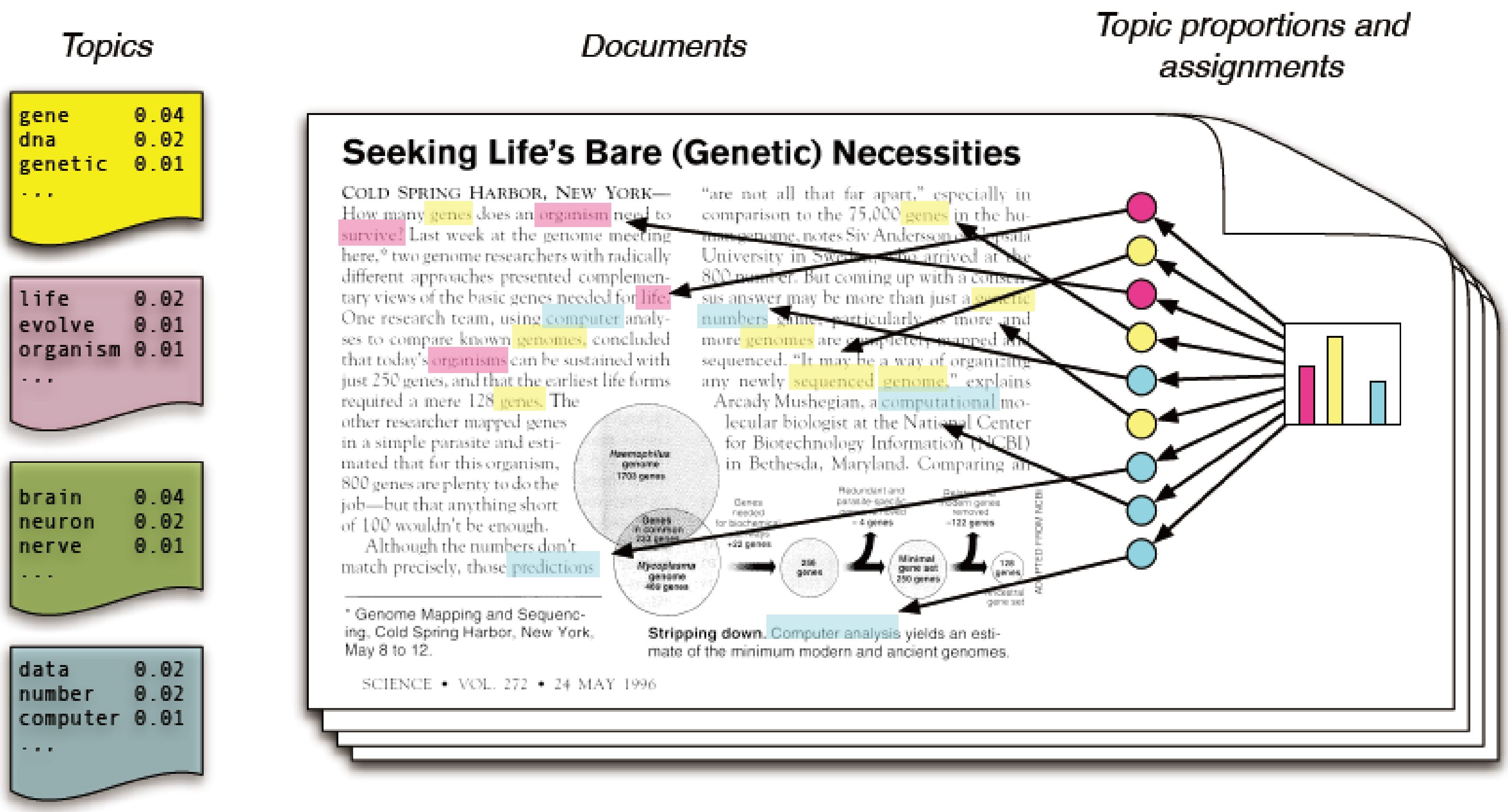
- Why POIs
  - Features the function

  - But not enough
    - Compound
    - Quality

- Why human mobility
  - Differentiate between POIs of the same category
  - Feature the function of a region

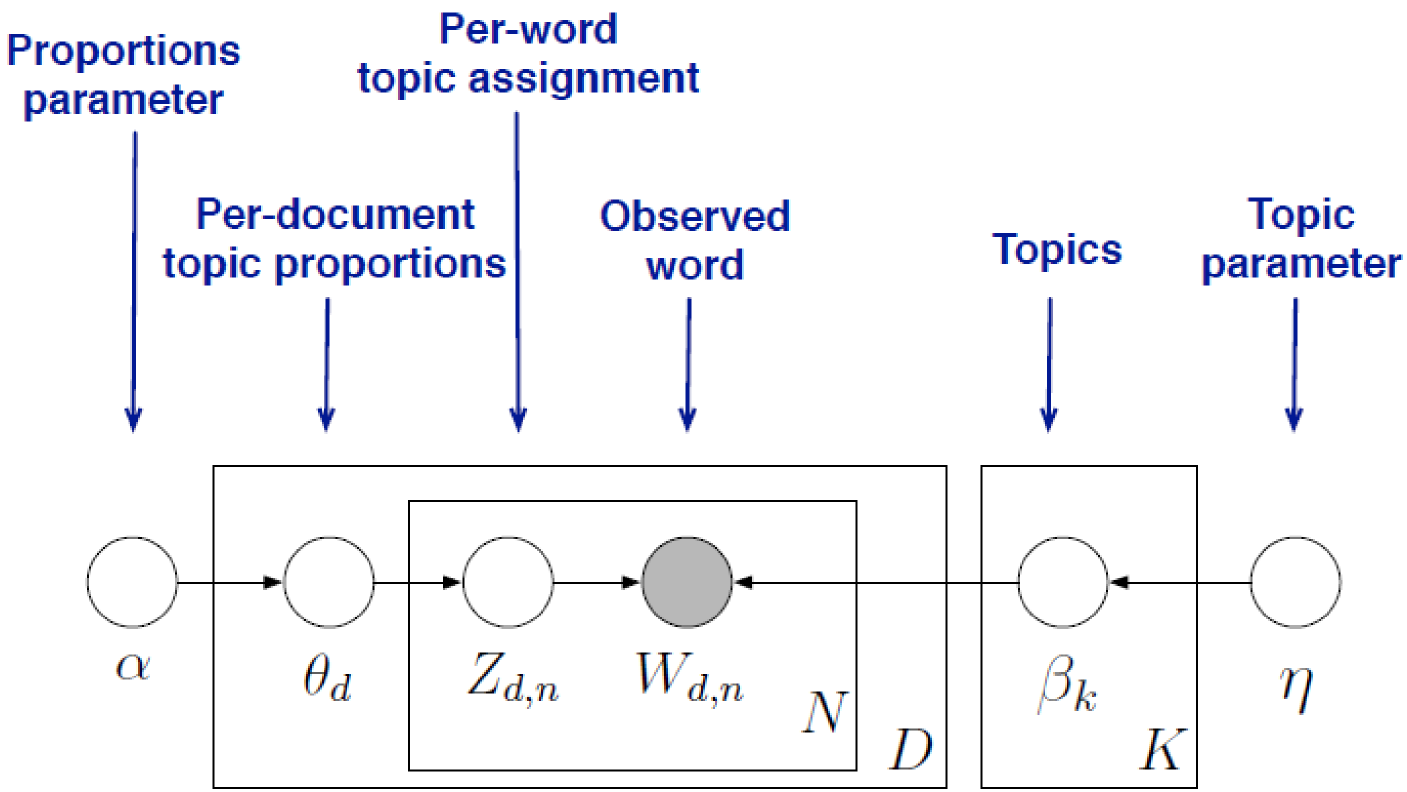


Generative model for LDA



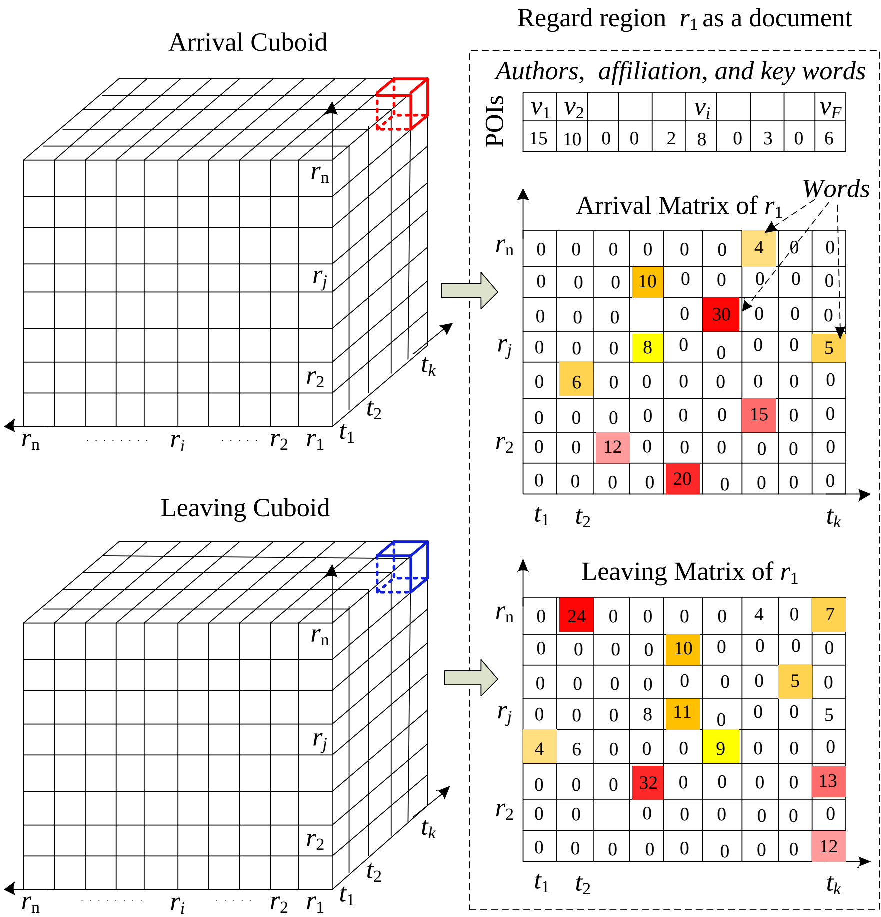
- Each **topic** is a distribution over words
- Each **document** is a mixture of corpus-wide topics
- Each **word** is drawn from one of those topics

LDA as a graphical model



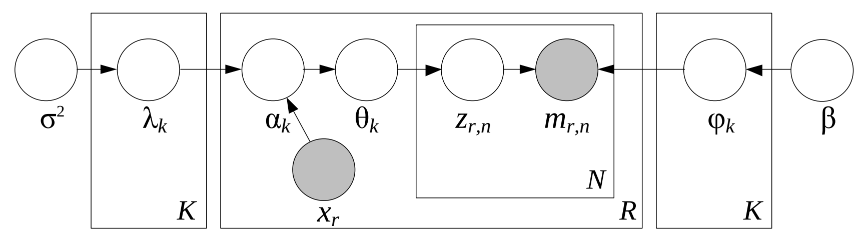
- Encodes our assumptions about the data
- Connects to algorithms for computing with data
- See *Pattern Recognition and Machine Learning* (Bishop, 2006).

Methodology overview



Mapping from regions to documents

transition cuboids	$\longrightarrow$ vocabulary
formal regions	$\longrightarrow$ documents
function of a region	$\longrightarrow$ topic of a document
mobility patterns	$\longrightarrow$ words
POI feature vector	$\longrightarrow$ metadata of a document



Infer the topic distribution using a topic model

Mobility Pattern Extraction

- Transition

  $Tr=(Tr.r_O, Tr.r_D, Tr.t_A, Tr.t_L)$

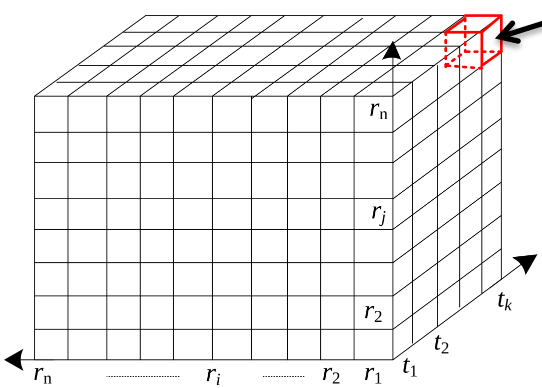
- Mobility Pattern

  $M_L=(Tr.r_O, Tr.r_D, Tr.t_L)$

  $M_A=(Tr.r_O, Tr.r_D, Tr.t_A)$

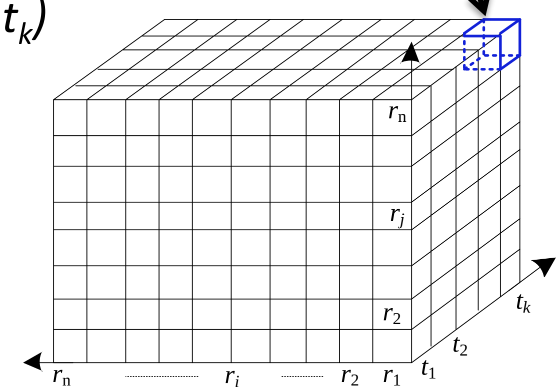
- Transition Cuboids

$C_L(1,n,k)=\#(r_1, r_n, t_k)$



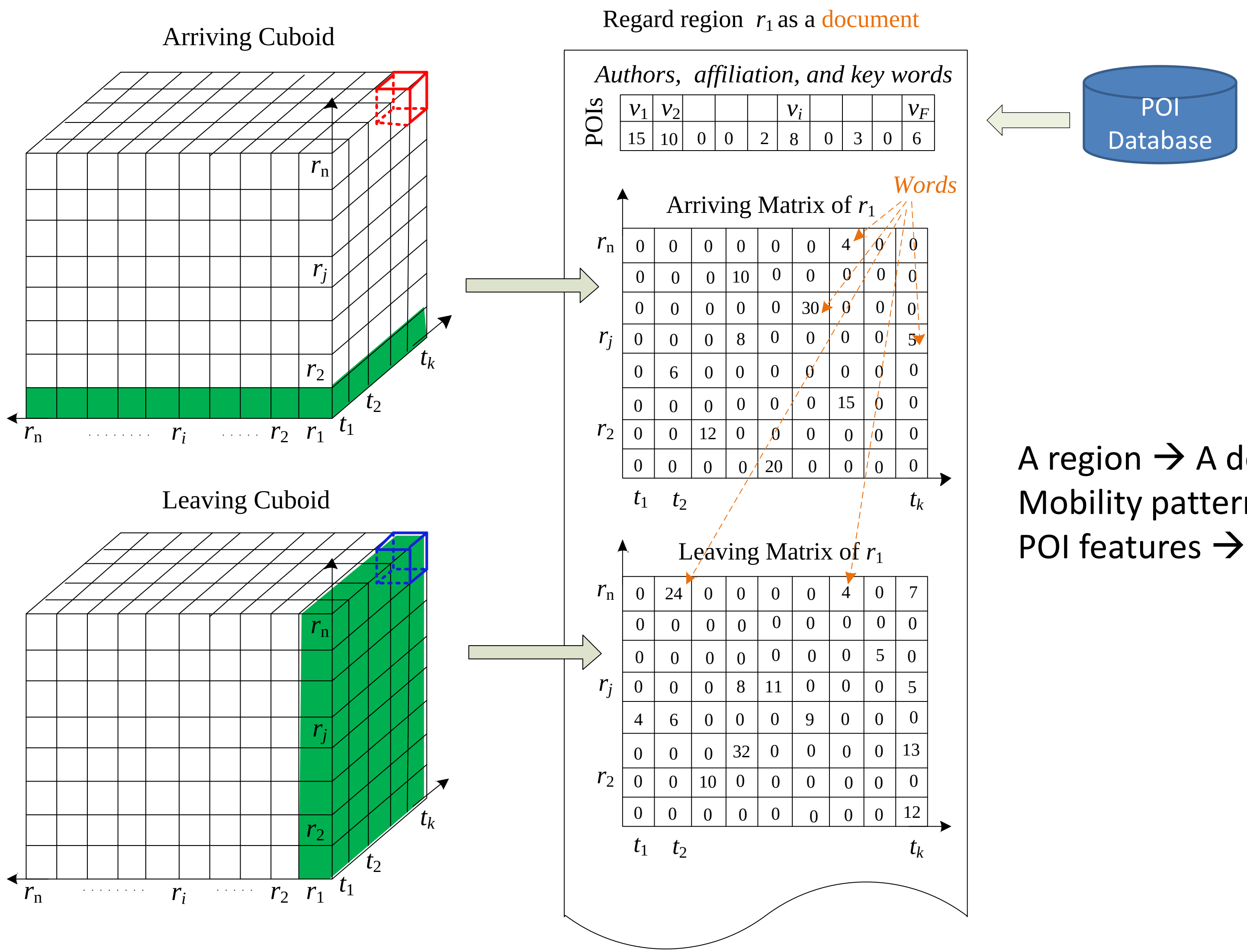
Arriving cuboid

$C_A(1,n,k)=\#(r_1, r_n, t_k)$



Leaving cuboid

Analogy from regions to documents

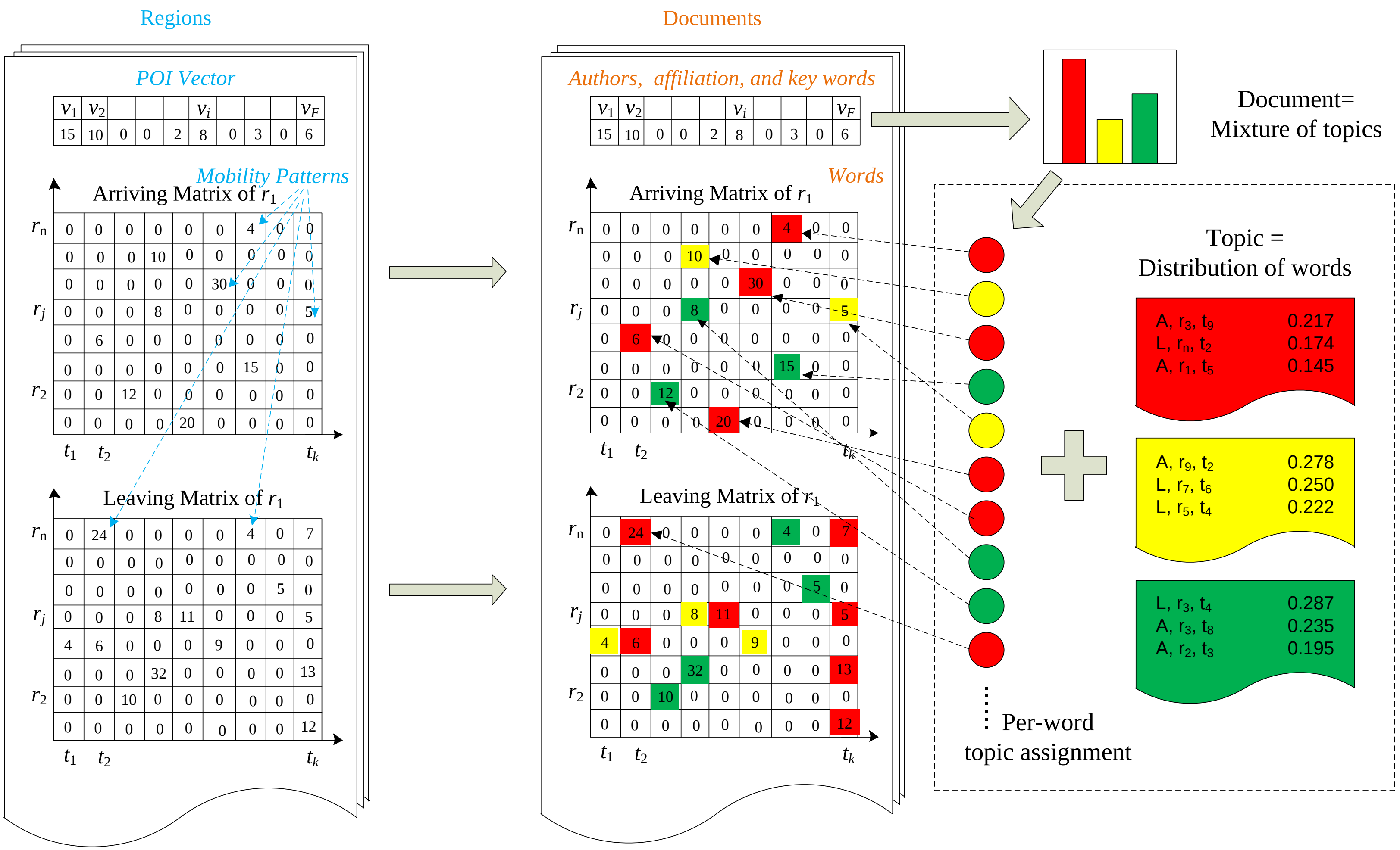


A region → A document

Mobility patterns → words

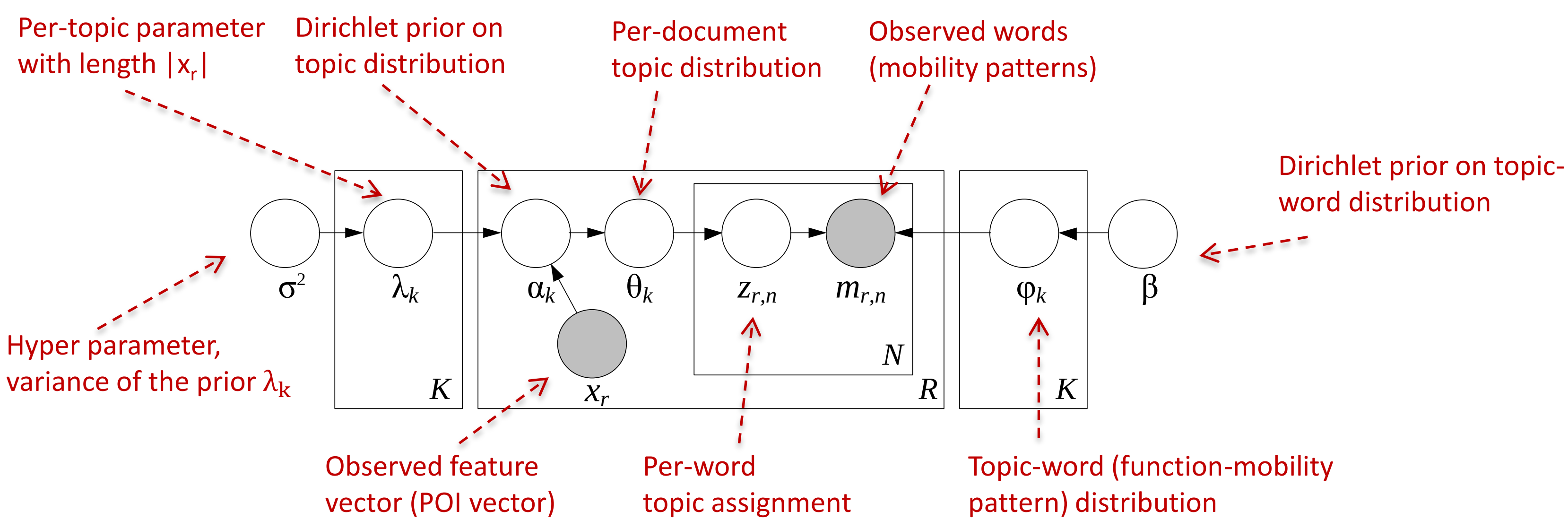
POI features → meta data

Analogy from regions to documents



Discovery of Region Topics

- Dirichlet–Multinomial-Regression(DMR)-Based Topic model (Mimno et al., 2008)
  - Variation of LDA
  - Generalized for incorporating metadata



Territory Identification

- Region aggregation
  - Regions with similar topic distributions are clustered
  - Aggregate big territories→ functional regions

Territory Identification

- Functionality intensity Estimation
  - Reason: degree of functionality vary spatially
  - Estimate the intensity for each function
  - Given x_1, x_2, …, x_n, the intensity at location s is measured by

(a) functional region c_1 (b) functional region c_4

Figure 13: Functionality intensity of functional regions

$$\lambda(s) = \sum_{i=1}^{n} \frac{1}{nr^2} K(\frac{d_{i,s}}{r}), \qquad K(\frac{d_{i,s}}{r}) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{d_{i,s}^2}{2r^2}).$$

Territory Identification

- Region annotation
  - (1) POI configuration (2) frequent mobility patterns
  - (3) Functionality density (4) human-labeled regions

Table 5: Overall POI feature vector and ranking of functional regions by DRoF. FD: frequency density, IR: internal ranking

	c0		c1		c2		c3		c4		c5		c6		c7		c8	
POI	FD	IR	FD	IR	FD	IR	FD	IR	FD	IR	FD	IR	FD	IR	FD	IR	FD	IR
CarServ	0.046	25	0.016	23	0.052	26	0.044	18	0.060	17	0.028	25	0.056	24	0.091	13	0.053	21
CarSale	0.009	28	0.005	27	0.061	24	0.006	27	0.009	27	0.005	28	0.021	27	0.015	26	0.006	27
CarRepa	0.021	26	0.011	24	0.062	23	0.042	19	0.051	20	0.023	27	0.062	23	0.057	18	0.039	25
MotServ	0.002	30	0.003	28	0.004	28	0.001	28	0.002	29	0.004	29	0.001	29	0.001	29	0.003	28
Caf/Tea	0.226	14	0.121	9	0.226	12	0.066	15	0.113	13	0.252	6	0.237	13	0.052	19	0.153	10
StaStor	0.135	17	0.037	20	0.127	17	0.037	20	0.058	18	0.080	19	0.100	19	0.073	15	0.072	17
LivServ	1.289	1	0.581	2	1.322	2	0.399	1	0.698	1	0.780	2	1.345	2	0.430	2	0.886	2
Sports	0.054	23	0.035	21	0.092	21	0.030	22	0.041	22	0.033	23	0.080	20	0.035	20	0.093	16
Hospital	0.244	13	0.088	13	0.222	13	0.069	14	0.137	12	0.144	15	0.246	12	0.070	16	0.194	8
Hotel	0.202	15	0.063	16	0.115	18	0.058	16	0.071	16	0.086	18	0.211	15	0.059	17	0.049	22
SceSpo	0.048	24	0.007	26	0.032	27	0.012	25	0.016	25	0.029	24	0.044	25	0.012	27	0.031	26
Residen	0.795	3	0.230	5	0.638	6	0.203	5	0.323	5	0.398	5	0.797	4	0.221	4	0.440	3
Gov/Pub	0.442	7	0.103	11	0.276	11	0.094	10	0.188	9	0.169	12	0.375	7	0.177	6	0.150	11
Sci/Edu	0.315	11	0.139	7	1.084	3	0.109	9	0.323	6	0.251	8	0.530	6	0.124	9	0.266	6
TrasFac	0.459	6	0.115	10	0.397	7	0.091	11	0.150	11	0.191	11	0.364	8	0.113	10	0.257	7
Bank/Fina	0.376	9	0.128	8	0.383	8	0.078	13	0.107	14	0.197	10	0.320	10	0.083	14	0.135	12
CopBusi	1.128	2	0.593	1	1.947	1	0.334	2	0.348	4	0.548	4	1.738	1	0.475	1	0.977	1
StrFur	0.002	29	0.000	30	0.001	30	0.001	30	0.000	30	0.001	30	0.000	30	0.001	30	0.000	30
Entr/Bri	0.296	12	0.065	14	0.210	14	0.081	12	0.160	10	0.160	14	0.228	14	0.133	7	0.097	15
PubUti	0.405	8	0.101	12	0.285	9	0.112	8	0.238	7	0.209	9	0.314	11	0.132	8	0.132	13
ChiRes	0.692	5	0.252	4	0.926	5	0.294	3	0.399	3	0.813	1	0.829	3	0.235	3	0.370	4
ForRes	0.098	18	0.050	17	0.054	25	0.010	26	0.009	26	0.163	13	0.063	21	0.018	25	0.101	14
FasRes	0.095	19	0.046	18	0.141	16	0.034	21	0.050	21	0.126	16	0.132	17	0.026	22	0.057	20
ShopMal	0.724	4	0.268	3	0.929	4	0.242	4	0.476	2	0.559	3	0.734	5	0.203	5	0.306	5
ConvStor	0.370	10	0.157	6	0.281	10	0.128	7	0.234	8	0.251	7	0.362	9	0.108	11	0.160	9
E-Stor	0.056	21	0.017	22	0.107	20	0.029	23	0.037	23	0.037	22	0.063	22	0.018	24	0.040	23
SupMar	0.055	22	0.008	25	0.065	22	0.020	24	0.025	24	0.042	21	0.040	26	0.021	23	0.040	24
FurBuil	0.086	20	0.065	15	0.151	15	0.192	6	0.093	15	0.088	17	0.142	16	0.099	12	0.064	19
Pub/Bar	0.179	16	0.043	19	0.114	19	0.044	17	0.053	19	0.060	20	0.120	18	0.031	21	0.071	18
Theater	0.011	27	0.001	29	0.002	29	0.001	29	0.006	28	0.025	26	0.007	28	0.002	28	0.002	29

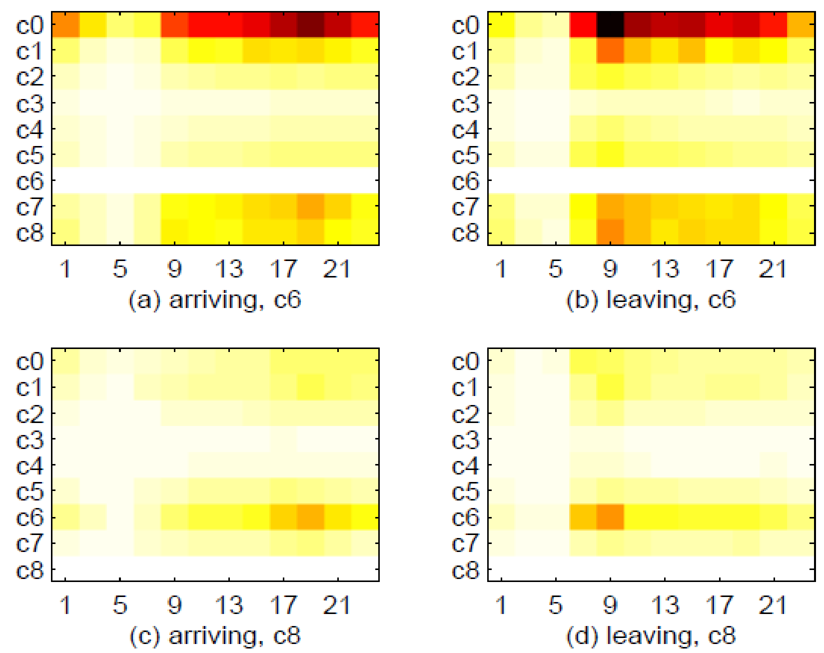
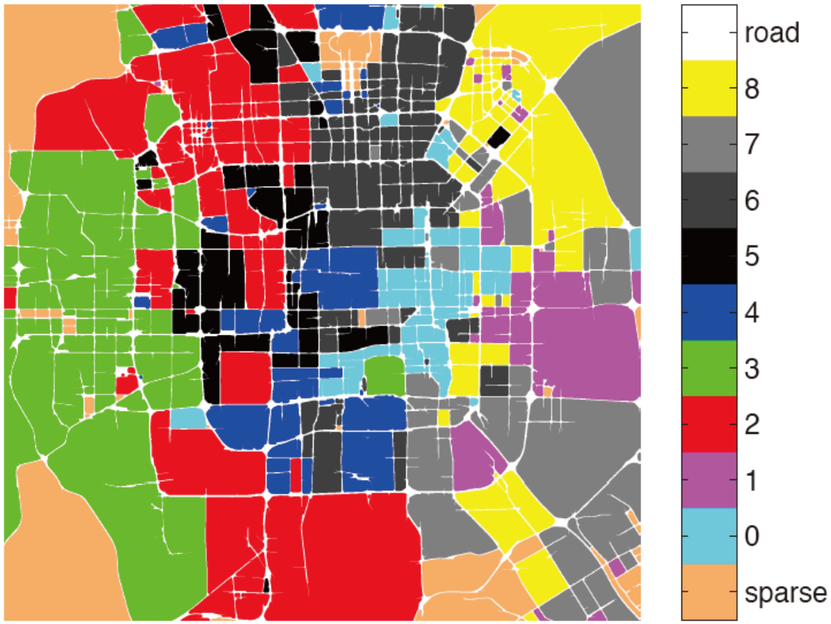


Figure 10: Transitions of c_6 and c_8

Annotation of Territories

- Diplomatic and embassy areas[c0]
- Education and science areas[c2]
- Developed residential areas[c6]
- Emerging residential areas[c8]
- Developed commercial/entertainment areas[c5]
- Developing commercial/business/entertainment areas [c1]
- Regions under construction[c7]
- Areas of historic interests[c4]
- Nature and parks[c3]

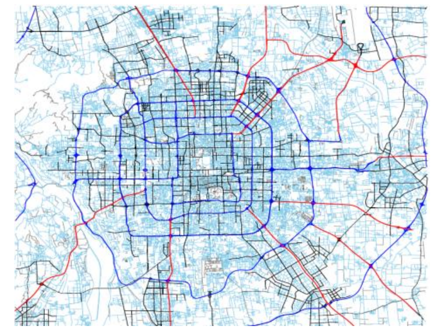


Evaluation

- Datasets (2010 and 2011, Beijing)

POI Data

Taxi trajectories



Road Networks

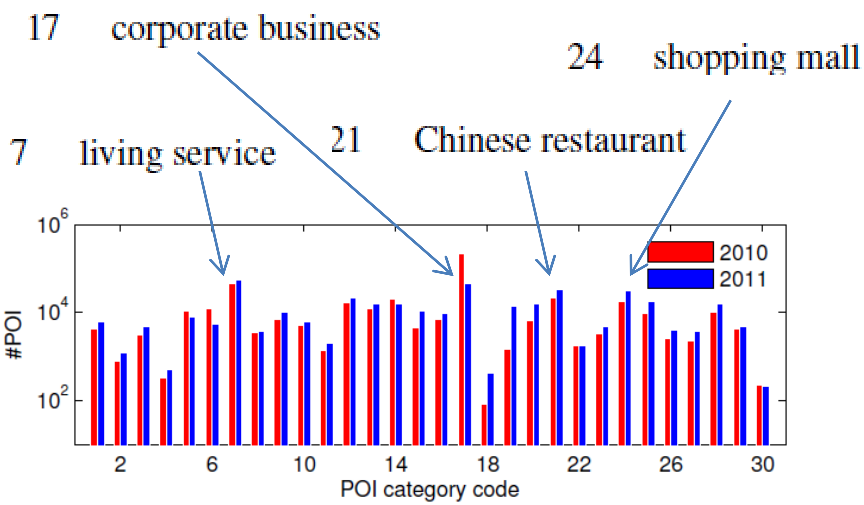


Table 3: Statistics of taxi trajectories and road networks

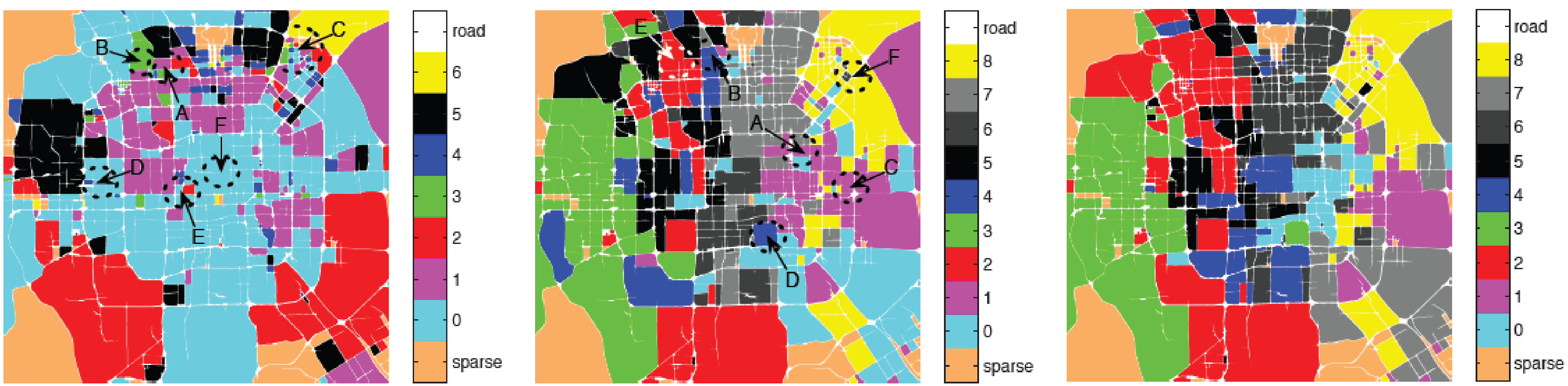
	year	2010	2011
Trajectories	#taxis	12,726	13,597
	#occupied trips	21,678,203	8,202,012
	#effective days	112	92
	average trip distance(km)	7.22	7.47
	average trip duration(min)	15.98	16.1
	average sampling interval(sec)	74.46	70.45
Roads	#road segments	150,357	162,246
	percentage of major roads	18.9%	17.1%
	#segmented formal regions	565	554
	size of "vocabulary" (non-0 items)	3,318,331	3,244,901

Results

- Baselines

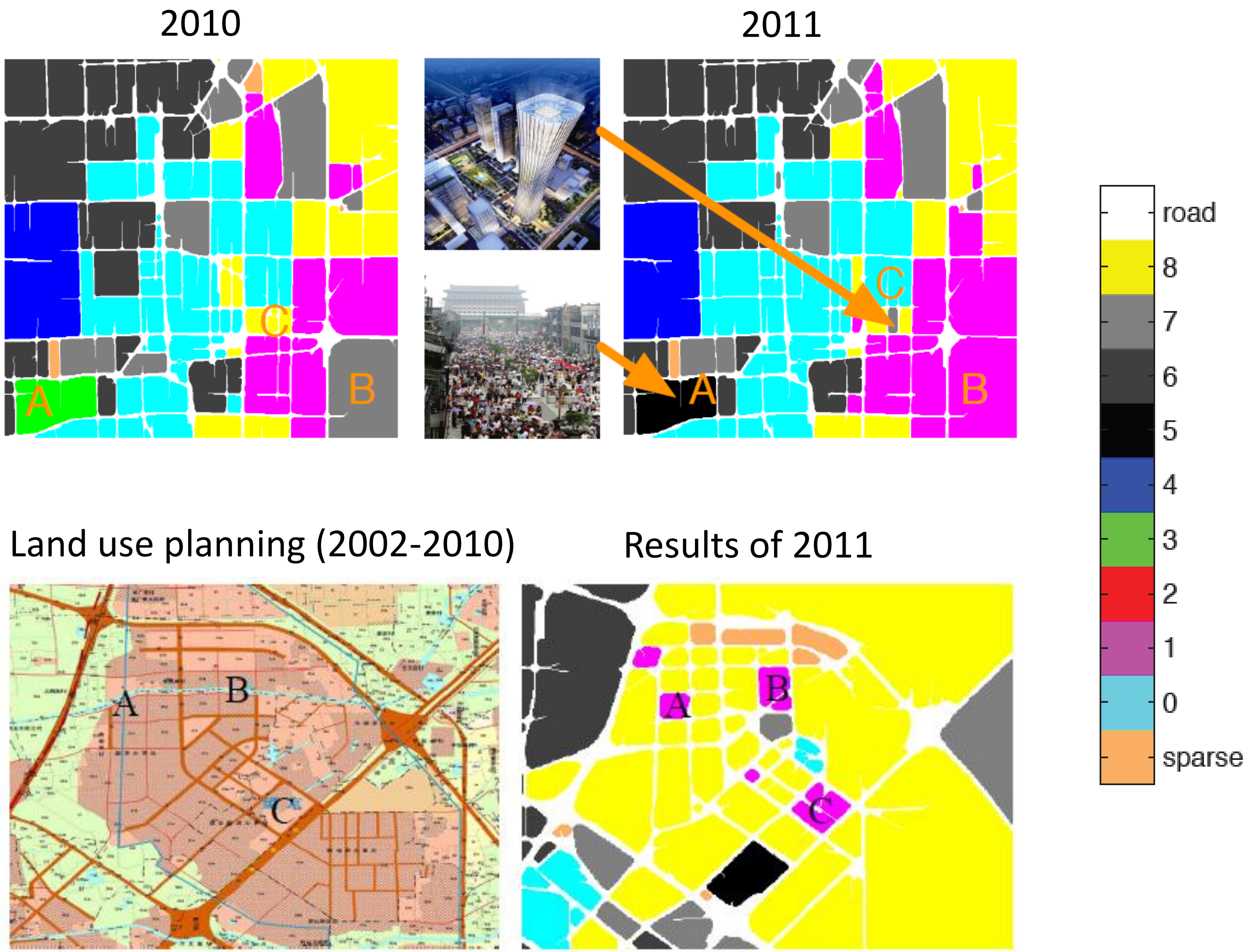
  Only using POI data (TF-IDF)

  Only using mobility data (LDA-based method)



(a) TF-IDF (b) LDA (c) DRoF

Evaluation

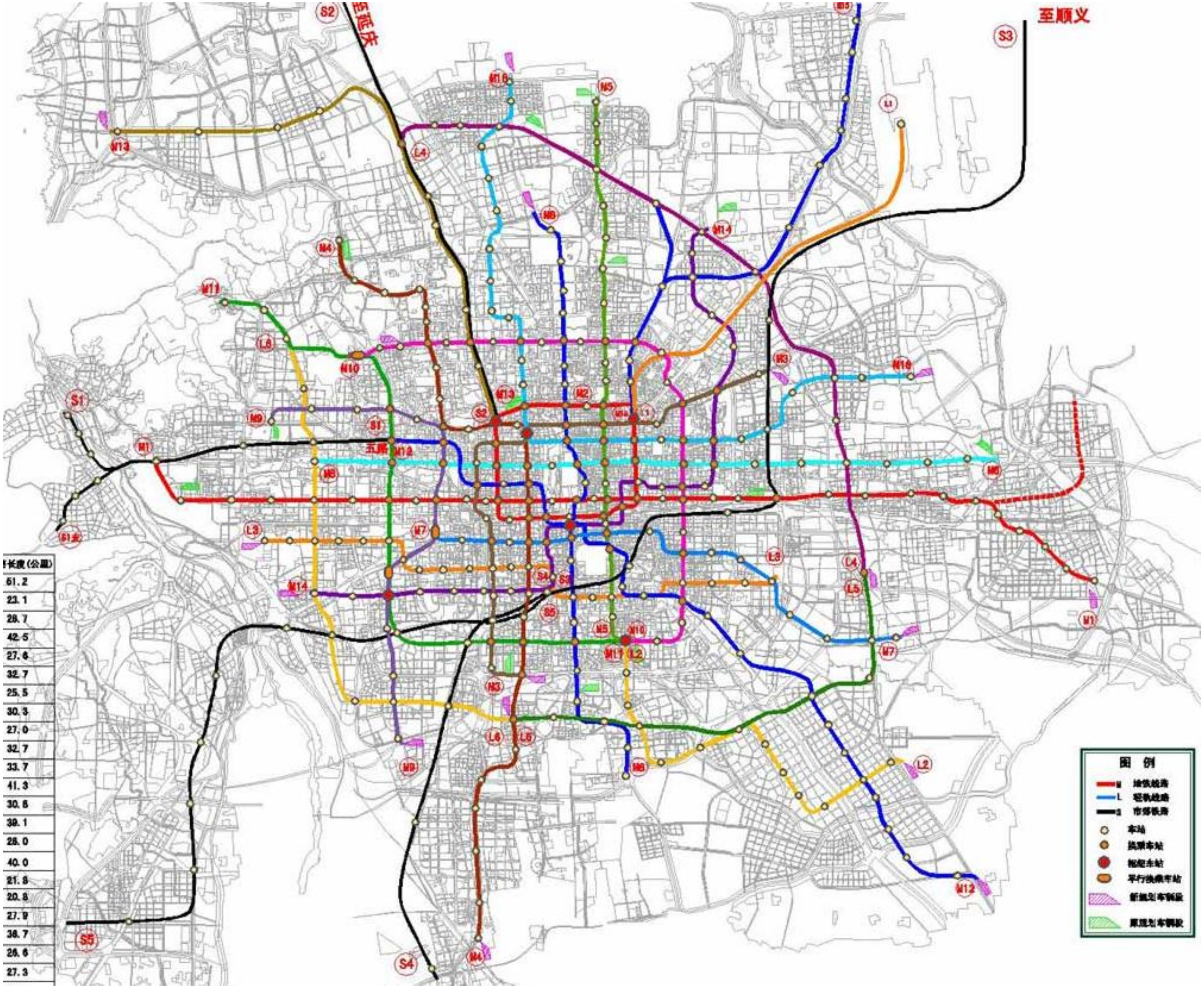

2010
2011
A
B
C
road
8
7
6
5
4
3
2
1
0
sparse
Land use planning (2002-2010)
Results of 2011

References

- Wei Liu, Yu Zheng, et al. Discovering Spatio-Temporal Causal Interactions in Traffic Data Streams. KDD 2011
- Yu Zheng, Yanchi Liu, Jing Yuan, Xing Xie, Urban Computing with Taxicabs, UbiComp 2011(Best paper nominee)
- Jing Yuan, Yu Zheng, Xing Xie. Discovering regions of different functions in a city using human mobility and POIs. KDD 2012

What's the Next

- Energy consumption
- Pollution monitoring and management
- Economy analysis

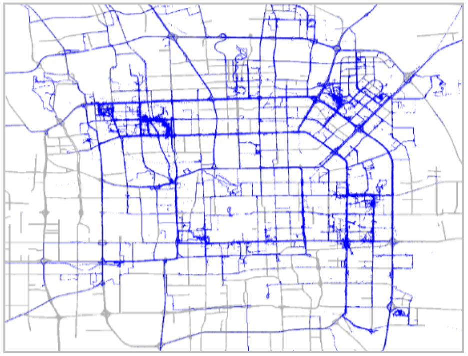


Beijing Subway by 2015: The city with the longest distance of subway (561km)

Two times longer than that of Paris(221.6KM)

Datasets Released

- GeoLife GPS trajectories
  - Generated by 178 users over 3 years
  - With transportation mode labels: driving, walking, biking, bus…
  - Annual release



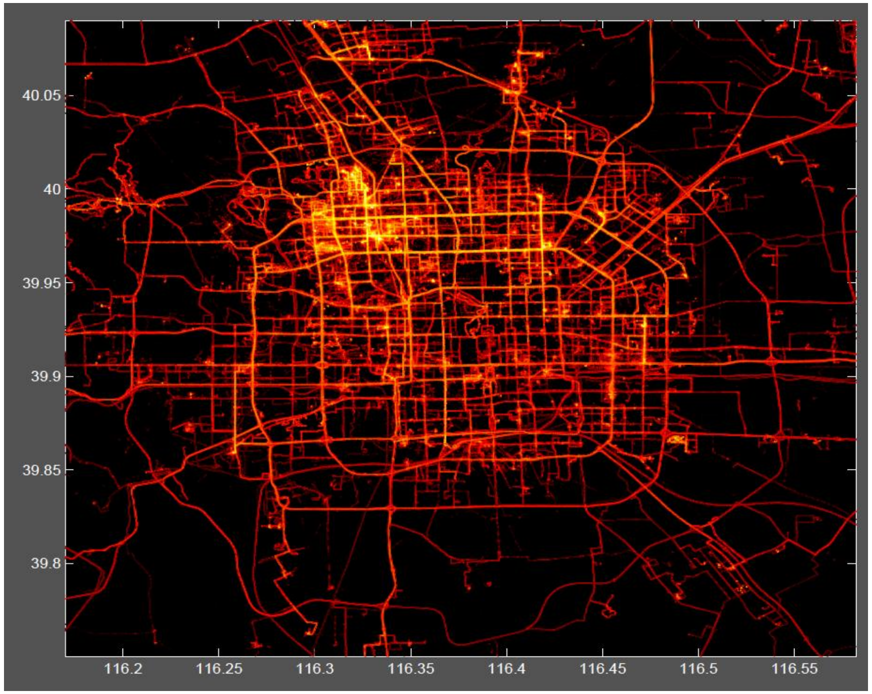
Link to the data

- T-Drive Taxi trajectories
  - Generated by Over 10,000 taxis in one week in Beijing
  - 15 million points
  - Distance > 9 million km

Link to the data

GeoLife Trajectory Dataset (1.2)

[Link to the data]

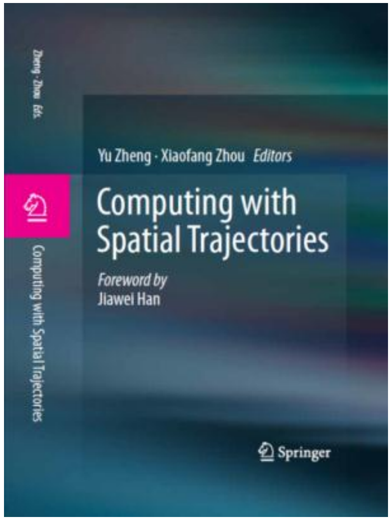


Transportation mode	Distance (km)	Duration (hour)
Walk	10,123	5,460
Bike	6,495	2,410
Bus	20,281	1,507
Car & taxi	32,866	2,384
Train	36,253	745
Airplane	24,789	40
Other	9,493	404
Total	14,0304	12,953

	Version 1.1	Version 1.2	Incremental
Time span of the collection	04/2007 – 10/2011	04/2007 – 8/2012	+10 months
Number of users	178	182	+4
Number of trajectories	17,621	18,670	+1,049
Number of points	23,667,828	24,876,978	+1,209,150
Total distance	1,251,654km	1,292,951km	+41,297 km
Total duration	48,203hours	50,176hours	+1,973 hour
Effective days	10,413	11,129	+716

Miscellaneous

- International Workshop on Urban Computing
  - In conjunction with KDD 2012 at Beijing China
  - **August 12, 2012**
  - http://www.cs.uic.edu/~urbcomp2012/
- The special issue on Urban Computing at ACM TIST
  - Top-tier international journal
  - Submission Due: **Oct. 7, 2012**
  - http://tist.acm.org/CFP.html
- The 4th international workshop on location-based social networks (LBSN 2012)
  - In conjunction with UbiComp 2012 in CMU, USA
  - **Sept. 8, 2012**
- A related text book:
  - Computing with spatial trajectories
  - Free tutorial slides download (here)



Thanks!

Yu Zheng

yuzheng@microsoft.com

Homepage

Homepage of Urban Computing:
http://research.microsoft.com/en-us/projects/urbancomputing/default.aspx