Urban Computing

-Providing a better life in cities and a better city for lives

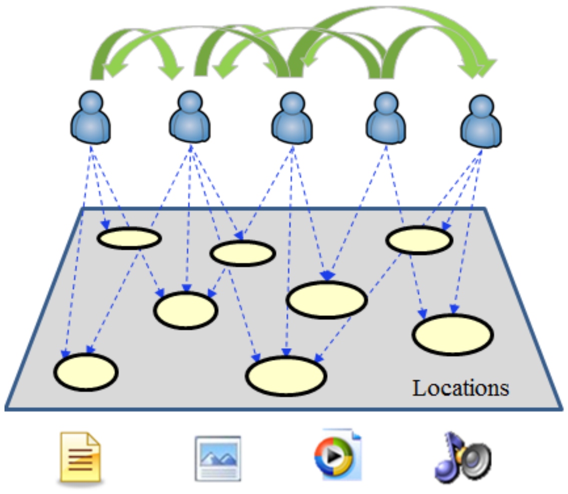
Visit its homepage

Yu Zheng

Ph.D. Researcher

Microsoft Research Asia

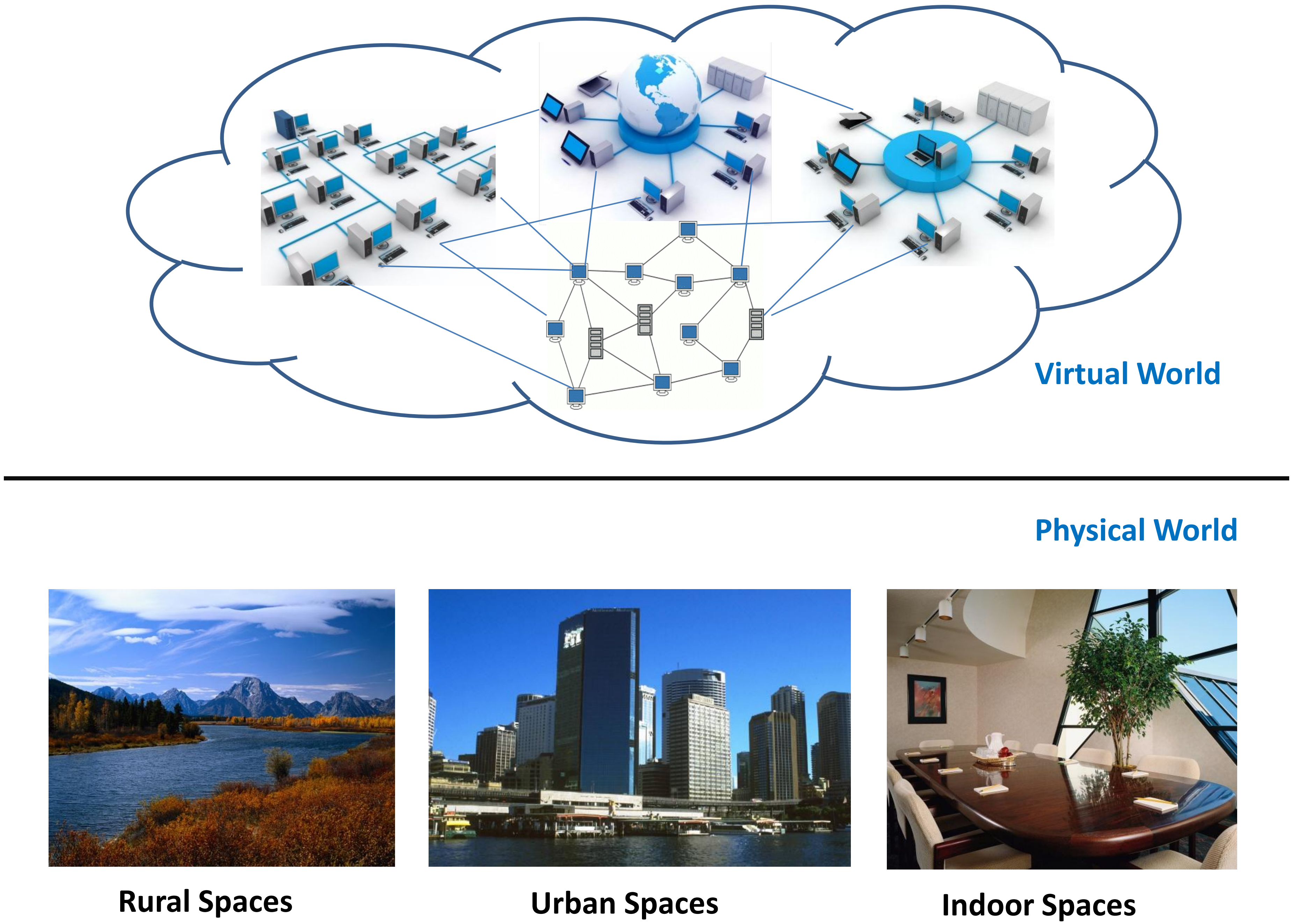
My Research Background



Location-Based Social Networks

Urban Computing

Trajectory Computing


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

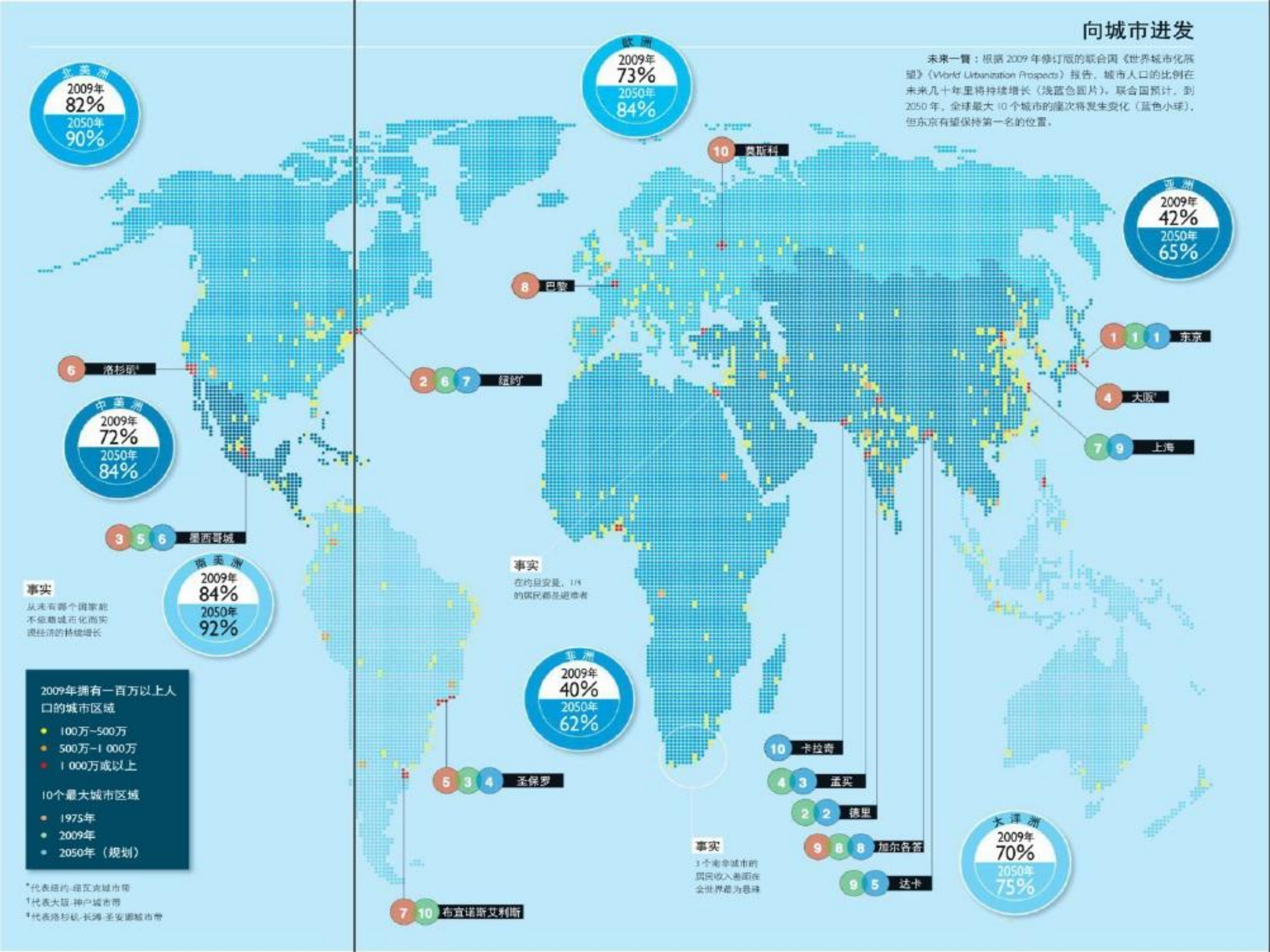
Shanghai

Why Urban Computing

- **2008**
  - The majority of people live in cities
  - Expanded 10 times in 20th century (0.25→2.8 billion)
- **2050**
  - Total population is over 9 billion
  - 6 billion urban residences

向城市进发

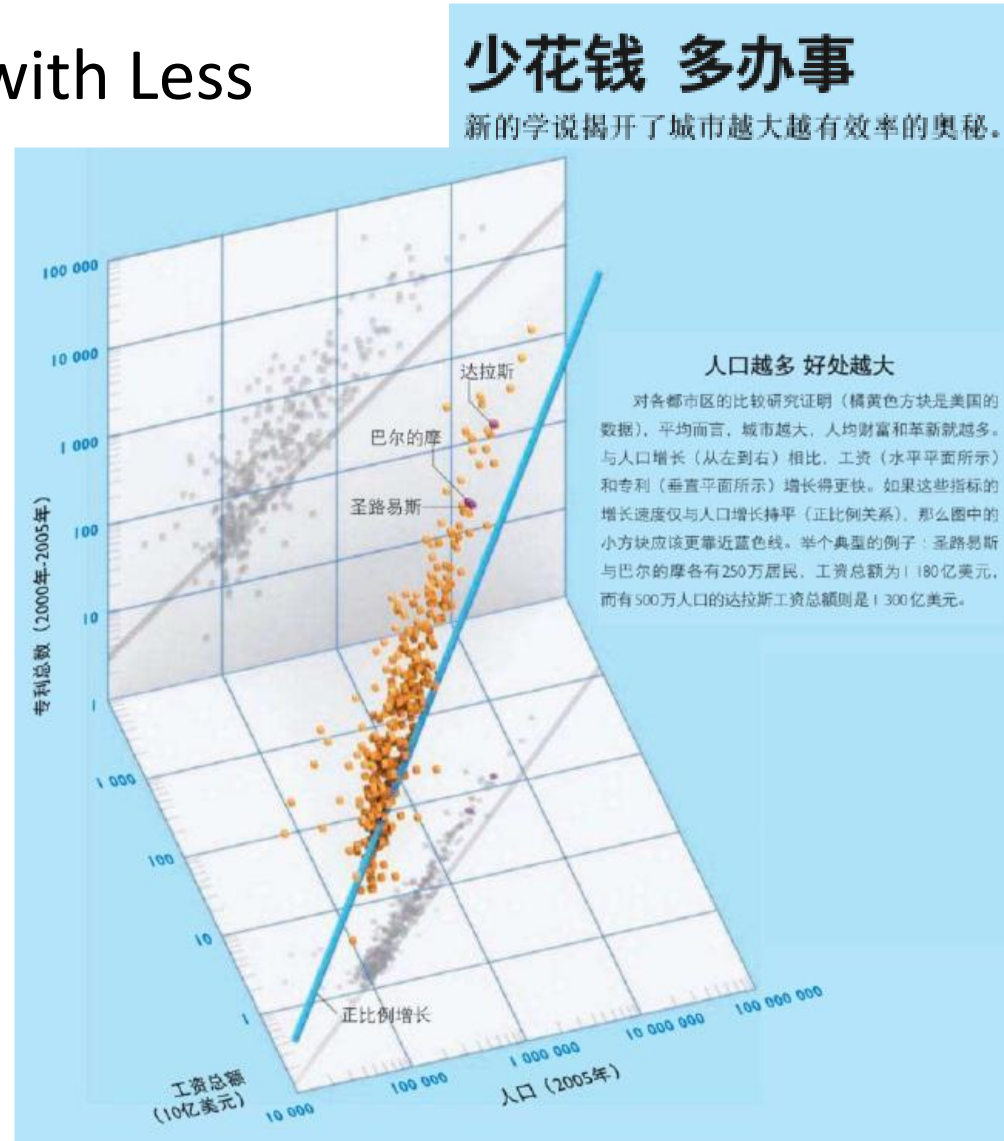
未来一瞥：根据 2009 年修订版的联合国《世界城市化展望》（*World Urbanization Prospects*）报告，城市人口的比例在未来几十年里将持续增长（浅蓝色圆片）。联合国预计，到 2050 年，全球最大 10 个城市的座次将发生变化（蓝色小球），但东京有望保持第一名的位置。



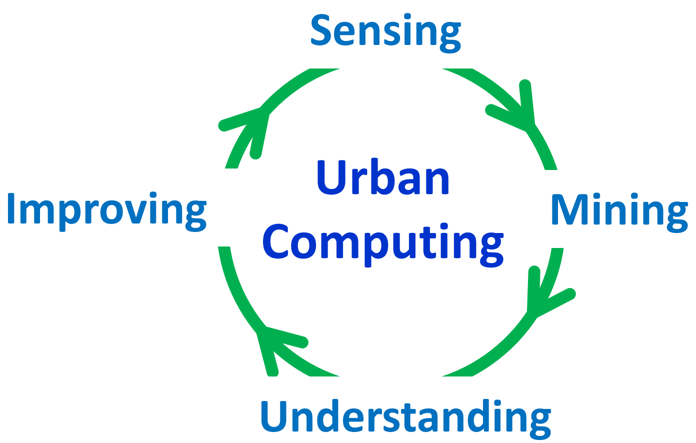
Why Urban Computing

- Bigger Cities Do More with Less
  - 15%
  - 15%
  - Energy-efficient

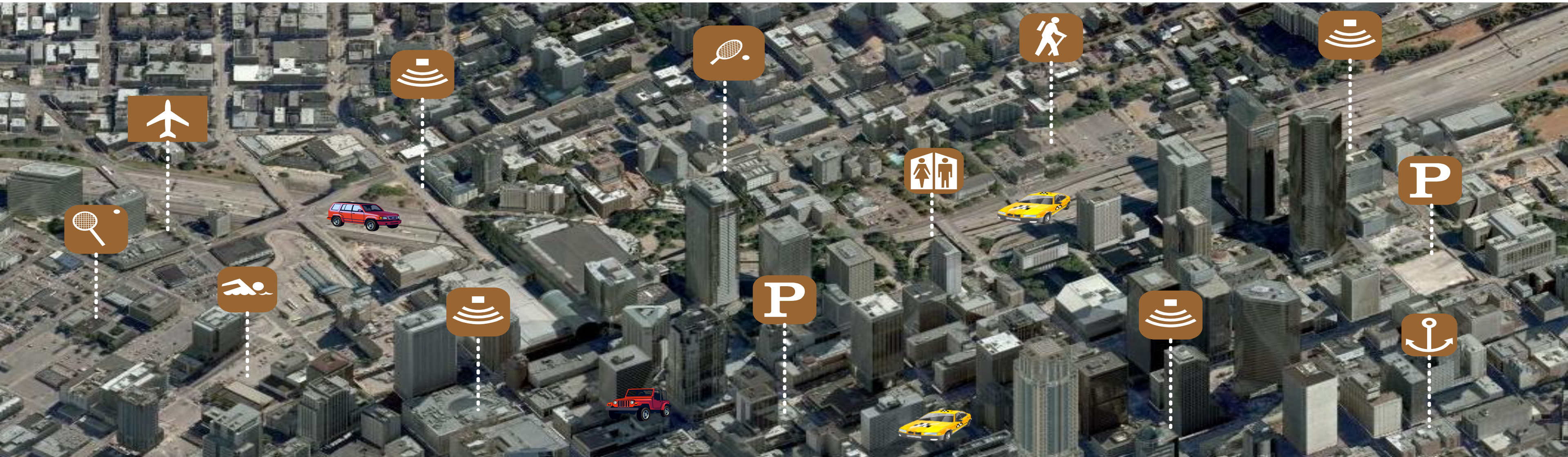
[Link]



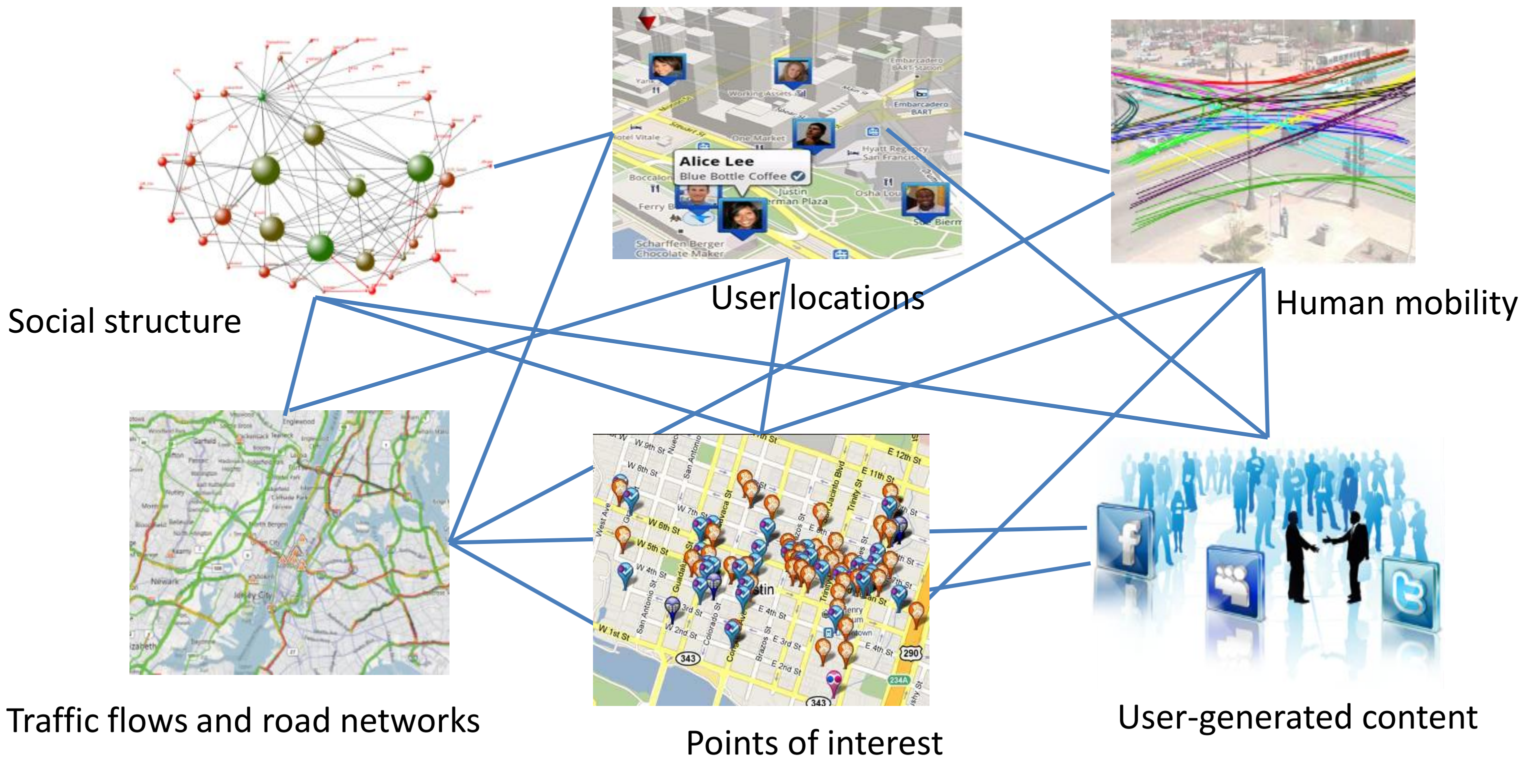
What's Urban Computing



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 enable a city-wide computing for serving people and cities.

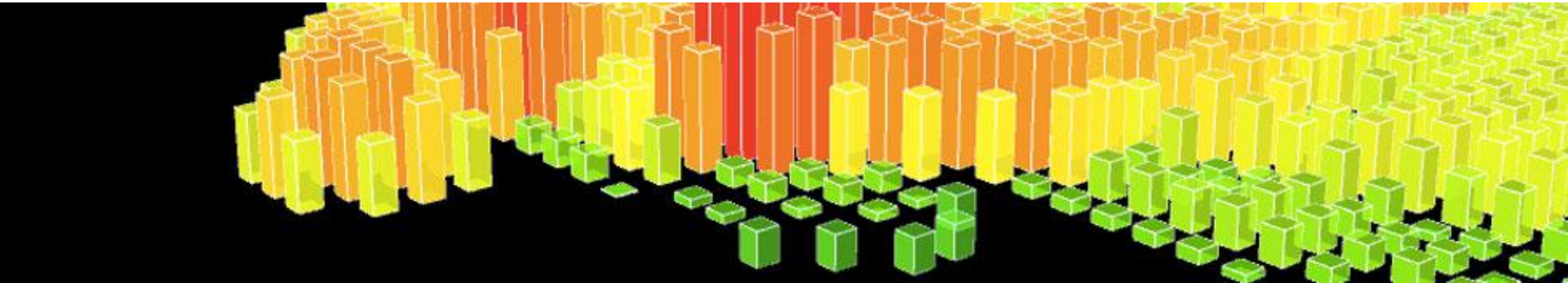


Knowledge discovery from huge and heterogeneous data

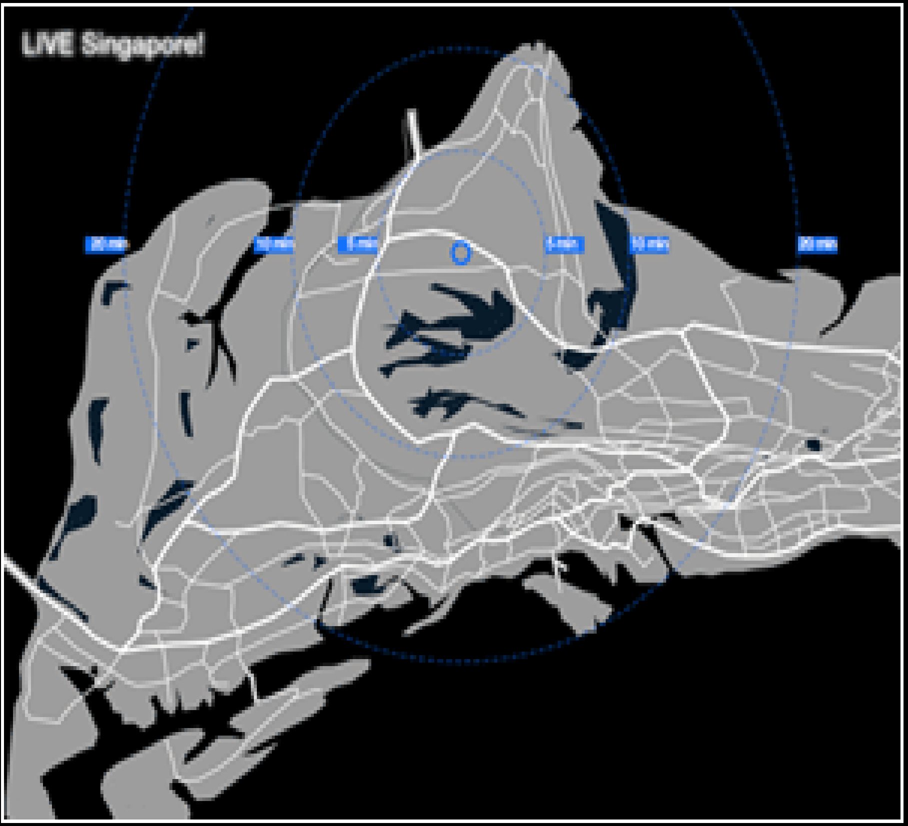


Live Singapore

Video

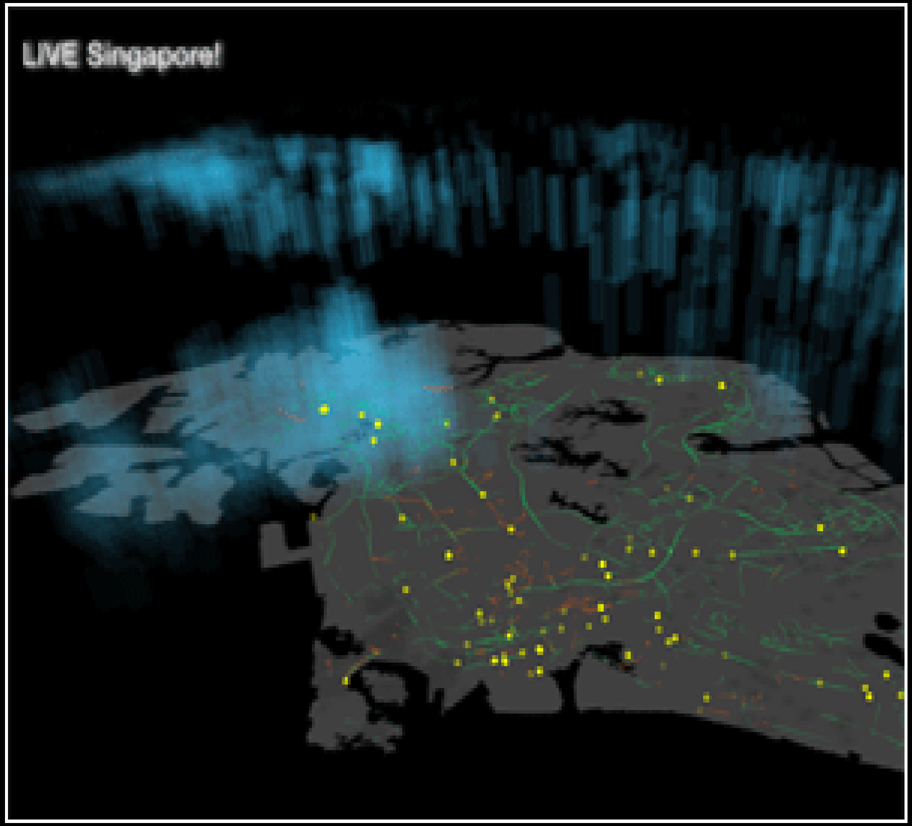


ISOCHRONIC SINGAPORE



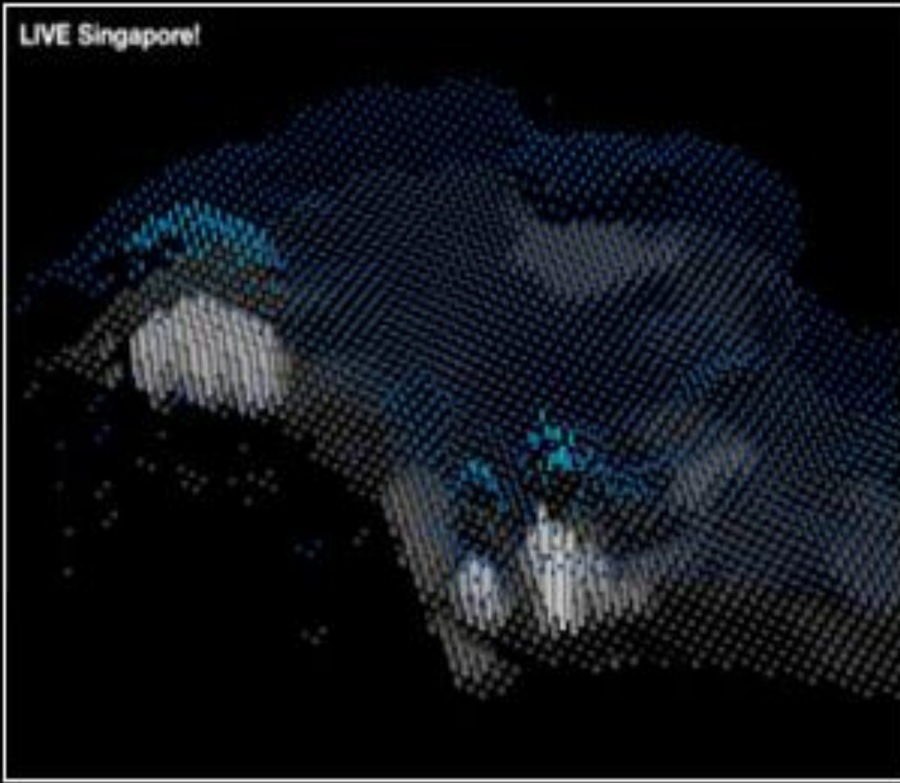
As vehicular traffic opens up and jams in the course of the day, the time we need to move in Singapore shrinks and expands. How long will it take you to go from home to any other destination? Find out with this isochronic map, where the deformations are proportional to travel time - and reveals the changes in the course of a weekend/week day.

RAINING TAXIS



Singapore's mobility is heavily reliant on taxis, but what happens when it rains? Getting hold of a cab is not the easiest thing in the world. We are exploring how our transportation system behaves by combining taxi and rainfall data, and investigating how in the future the system can streamline in order to better match taxi supply and demand.

URBAN HEAT ISLANDS

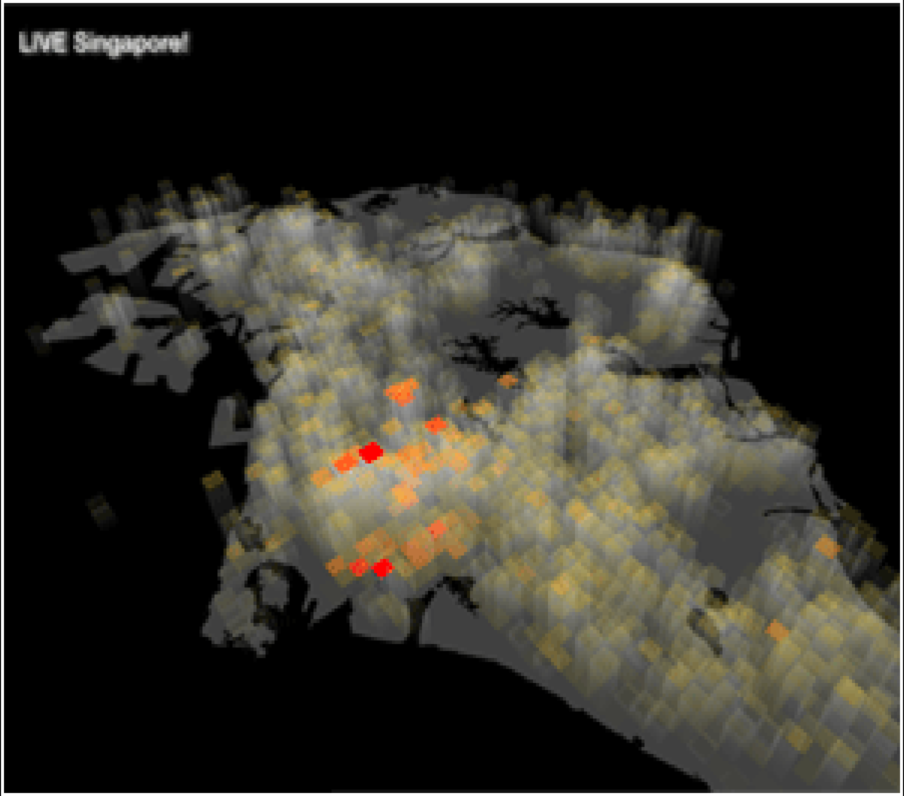


It is documented that temperatures in cities are several degrees higher than in the surrounding countryside, but as temperature rises we use more air-conditioning, which in turn results in even higher temperatures. Check out this effect on the map, which shows estimated temperature rise (top) and energy consumption (bottom) in different parts of Singapore.

FORMULA ONE CITY

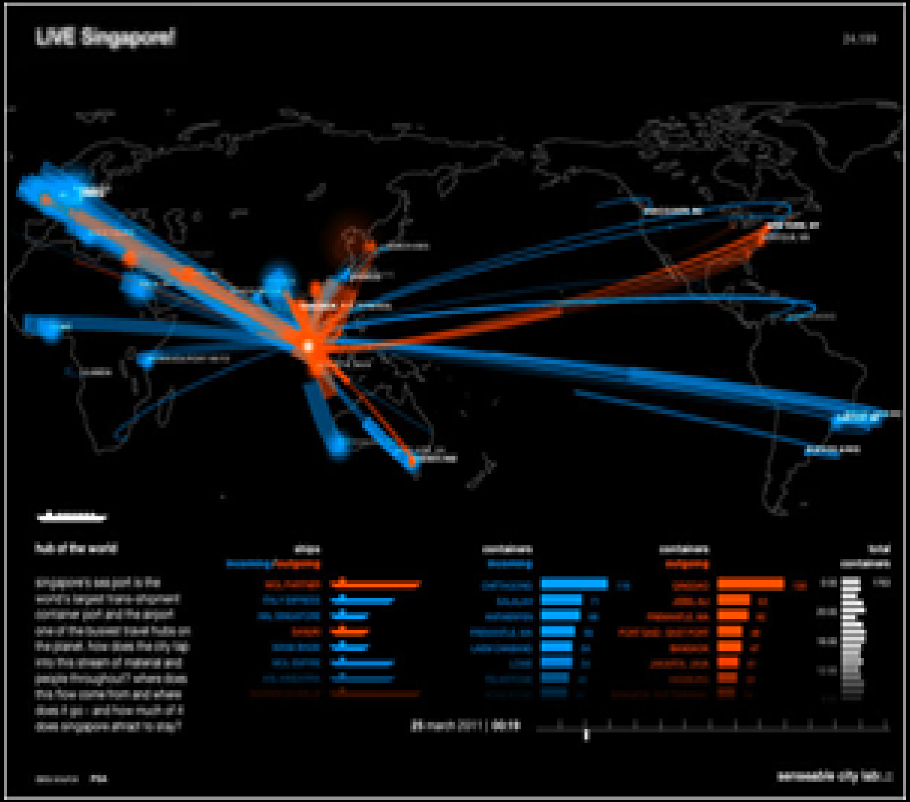
Large scale events disrupt a cities' daily routines. What better opportunity to explore this effect than Singapore's Formula One Grand Prix? How does this event impact our daily life? How do we respond to it? How do we share our excitement via cellphone? Find out by looking at this map, where the color and size of the glow are proportional to the amount of text messaging during the Formula One race.

REAL-TIME TALK



Singapore's mobile phone penetration is above 140%, many own more than one device. How do we make use of the island's cellphone network via voice calls and text messages? How can this inform us about the usage of urban space in real-time? Find out by looking at this map, where height and color intensity indicate the level of cellphone network usage.

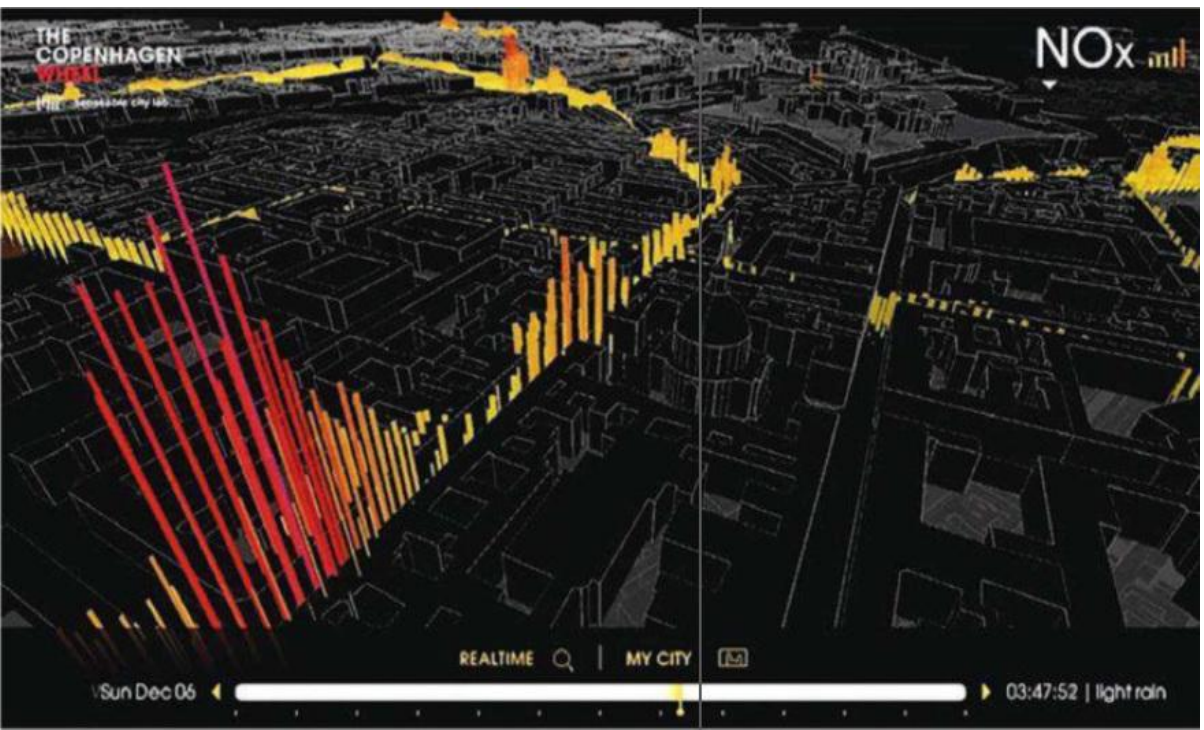
HUB OF THE WORLD



Singapore is the world's largest trans-shipment container port and one of the busiest airport hubs in the world. How is our island affected by this constant stream of people and goods passing through? Where do these flows come from and go to and how many of them are here to stay? Find out by looking at this map which shows the port and airport's global reach.

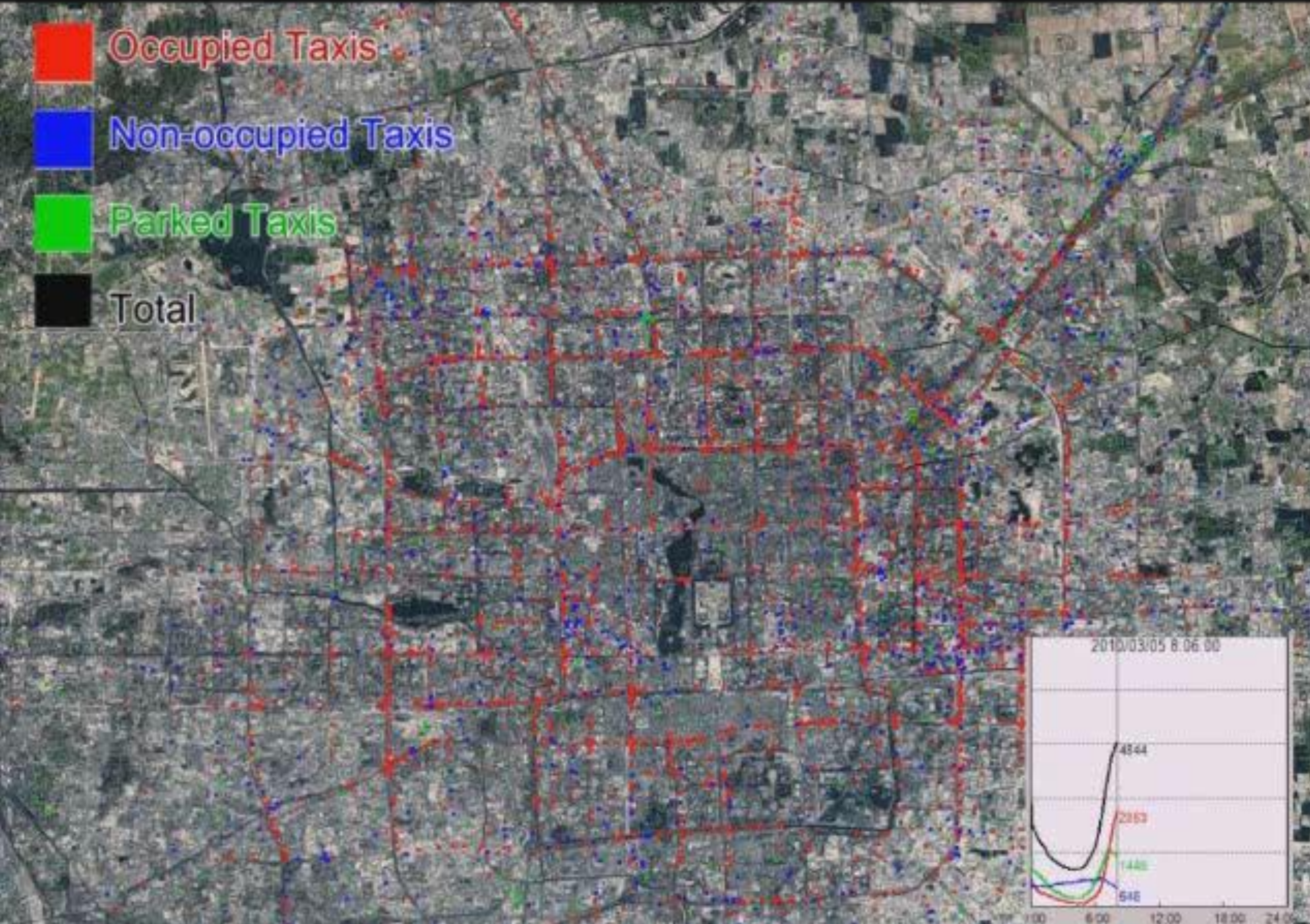
The Copenhagen Wheel

- Sense your fitness
- Urban environments
- Social connections



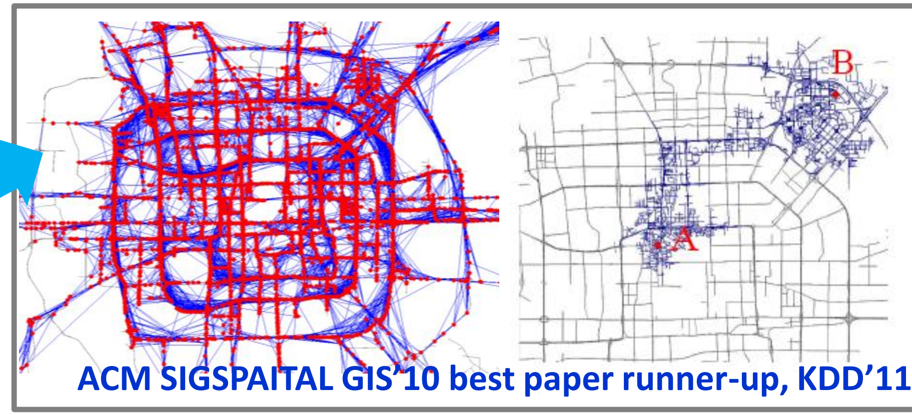
[Video]


GPS-equipped taxis are **mobile sensors**

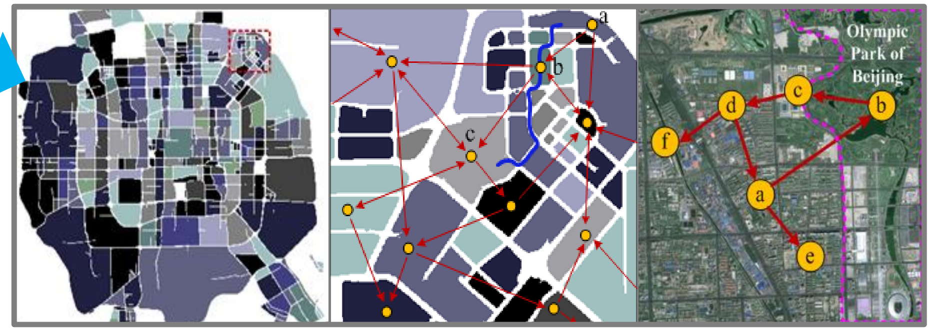


Occupied Taxis
Non-occupied Taxis
Parked Taxis
Total
2010/03/05 8:06:00
4844
646
6:00
12:00
18:00

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

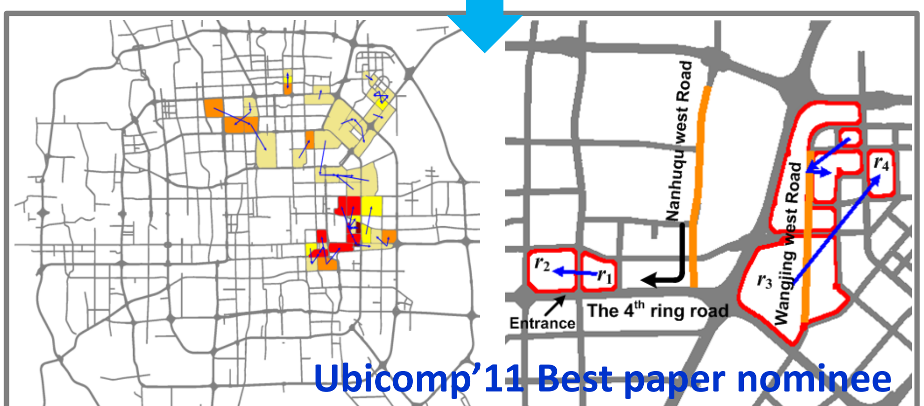
Urban Computing with Taxicabs



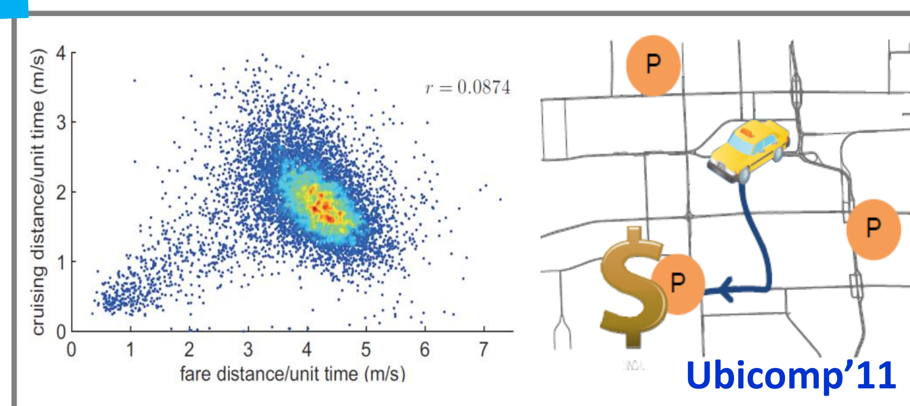
Finding Smart Driving Directions



Anomalous Events Detection KDD'11



Urban Computing for Urban Planning



Passengers-Cabbie Recommender system

T-Drive: Driving Directions Based on Taxi Trajectories

- Finding the *customized* and *practically fastest* driving route for a particular user using
  - (Historical and real-time) Traffic conditions
  - Driver behavior (of taxi drivers and end users)

Physical Routes

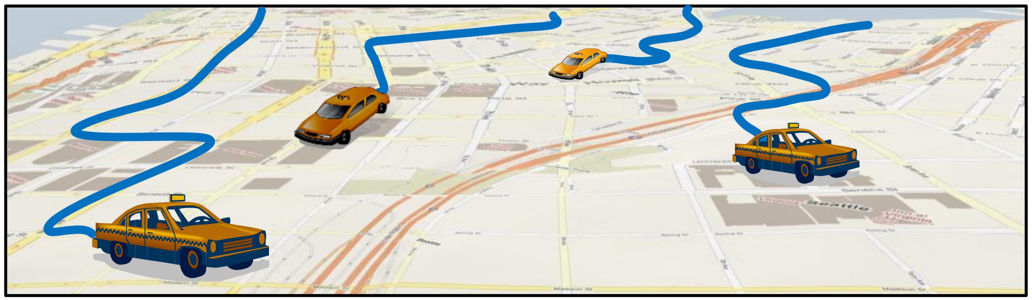
Traffic flows

Drivers

[Read the Related papers]

Motivation

- Taxi drivers are **experienced** drivers
- GPS-equipped taxis are **mobile sensors**

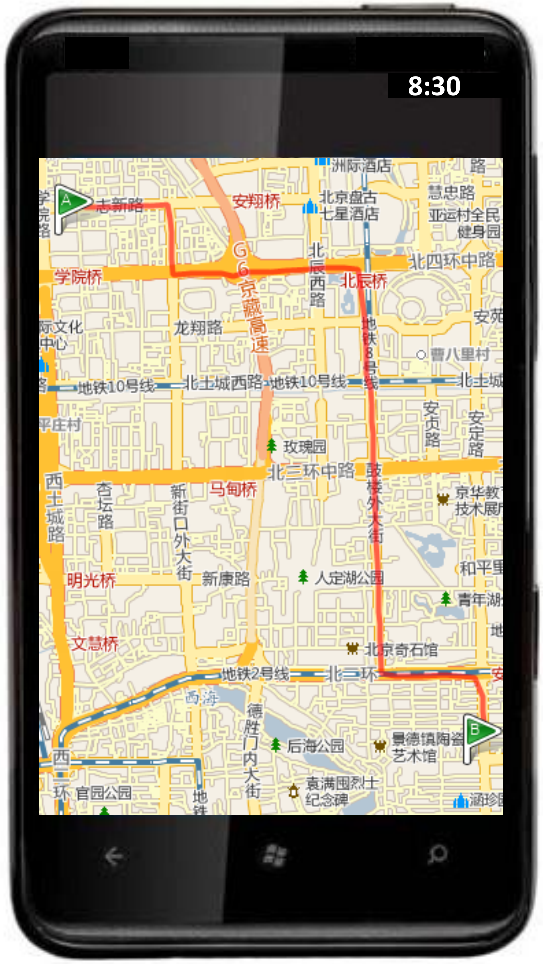


Human Intelligence

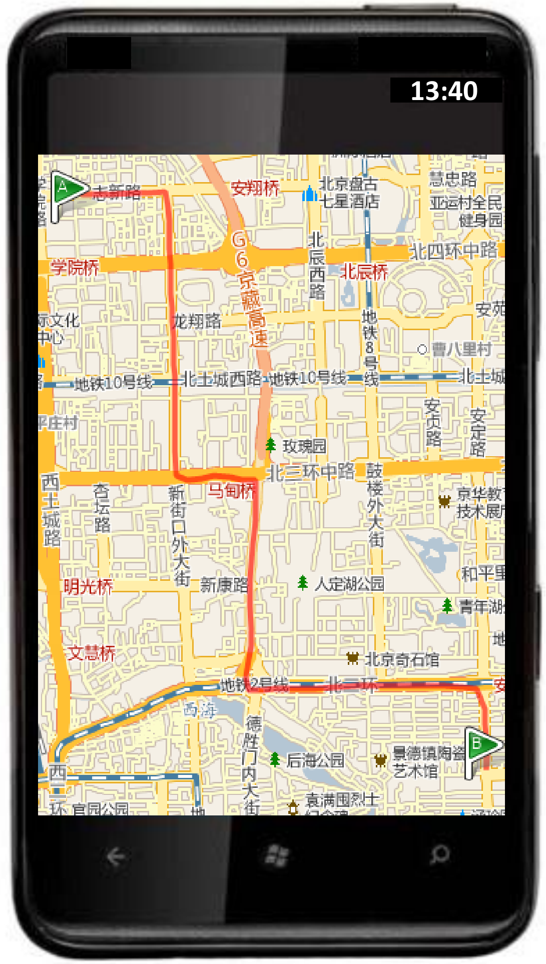
Traffic patterns

Application Scenarios

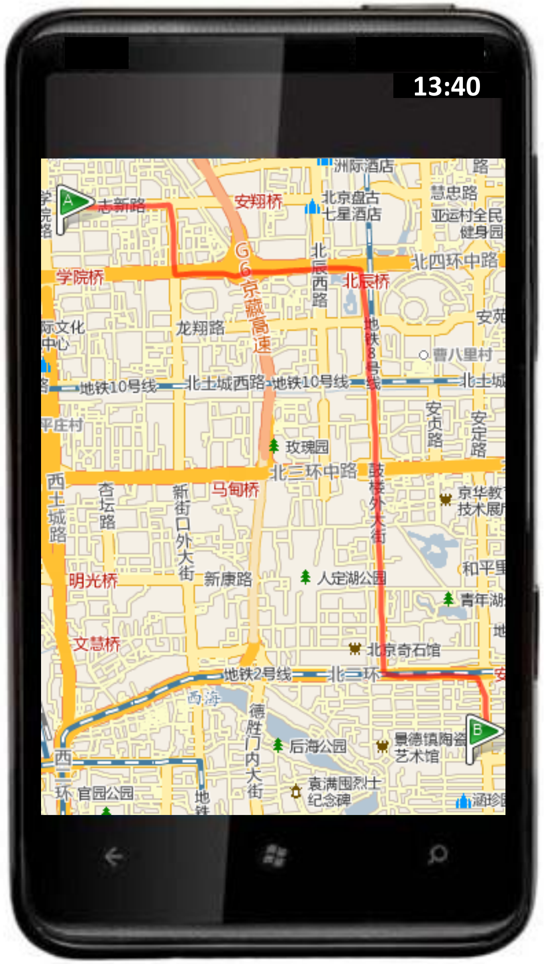
Driver A



Driver A

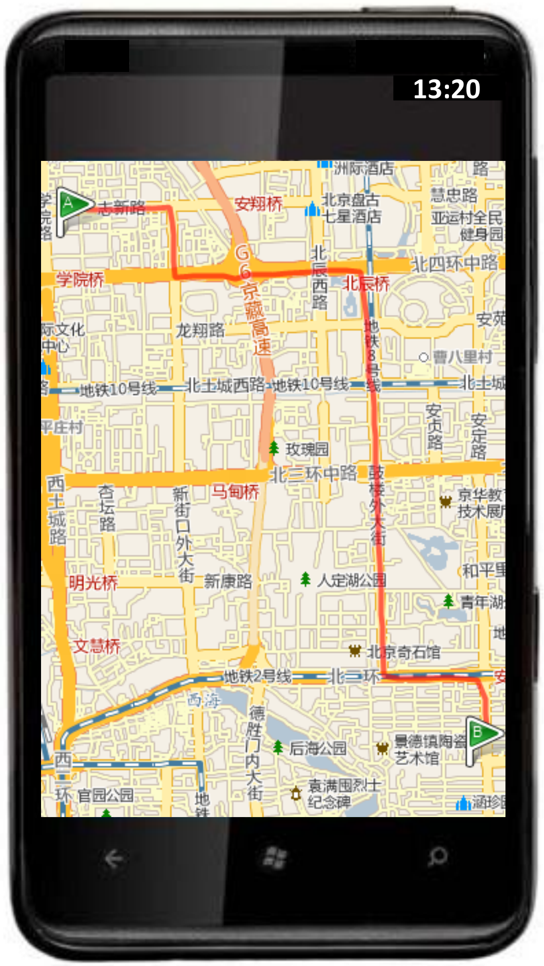


Driver B



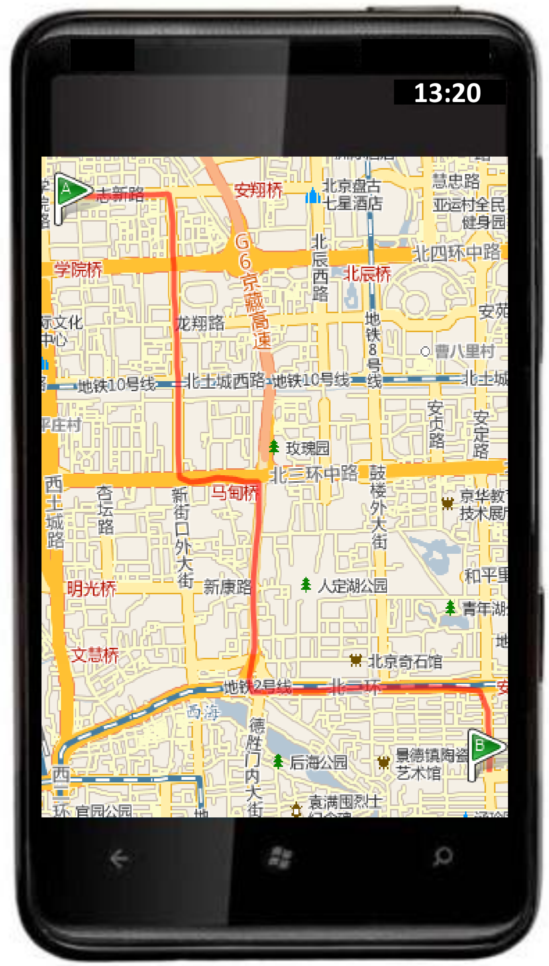
Application Scenarios

Driver B

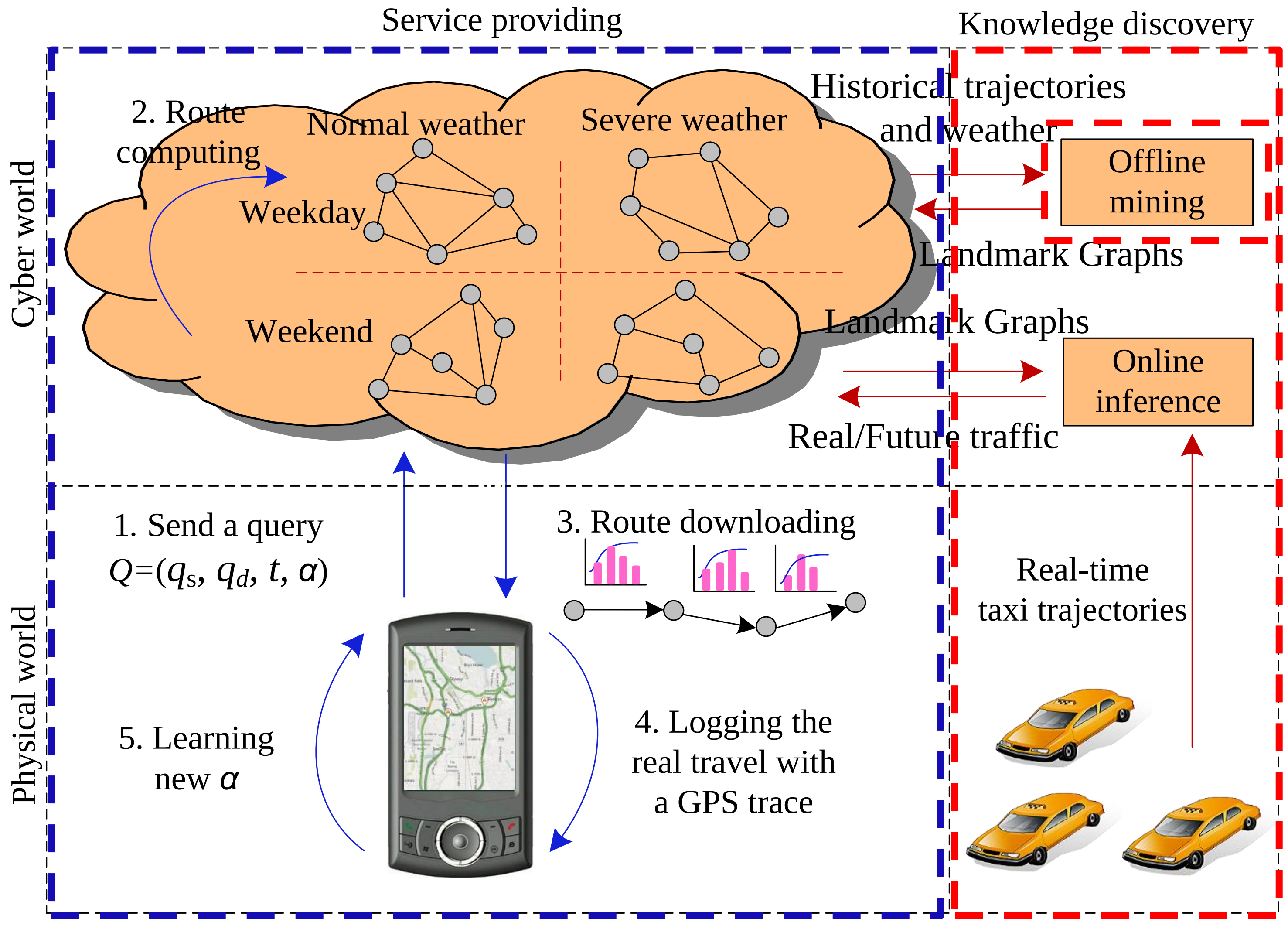


Log user B’s driving routes for 1 month

Driver B



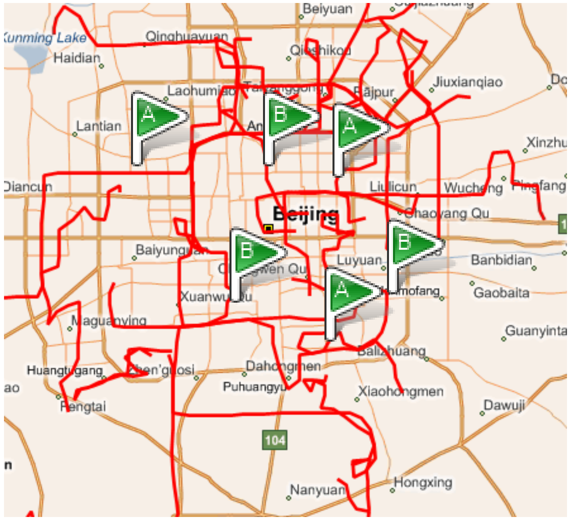
System Overview

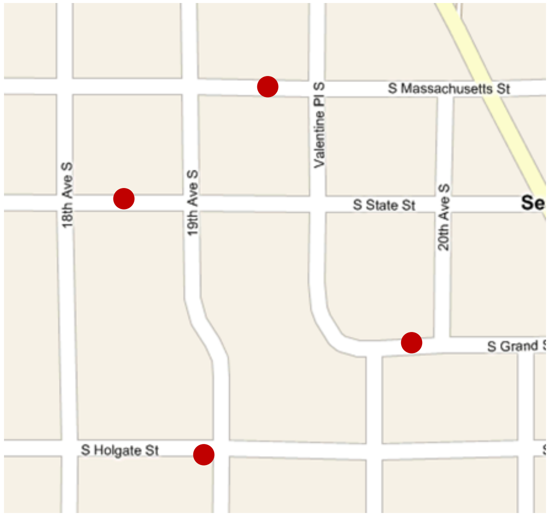


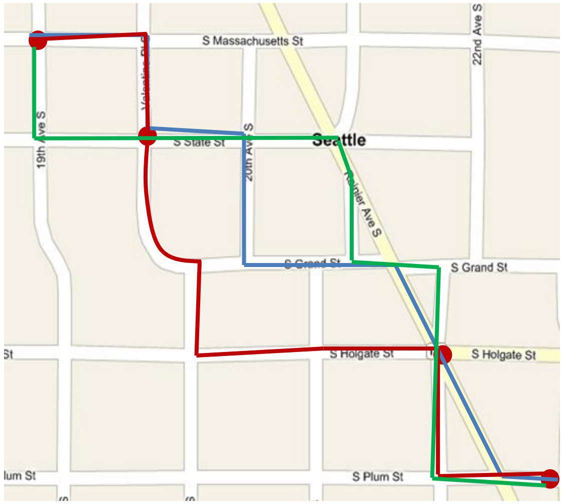
Offline Mining

- Building landmark graphs
- Mining taxi drivers' knowledge
- Challenges

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

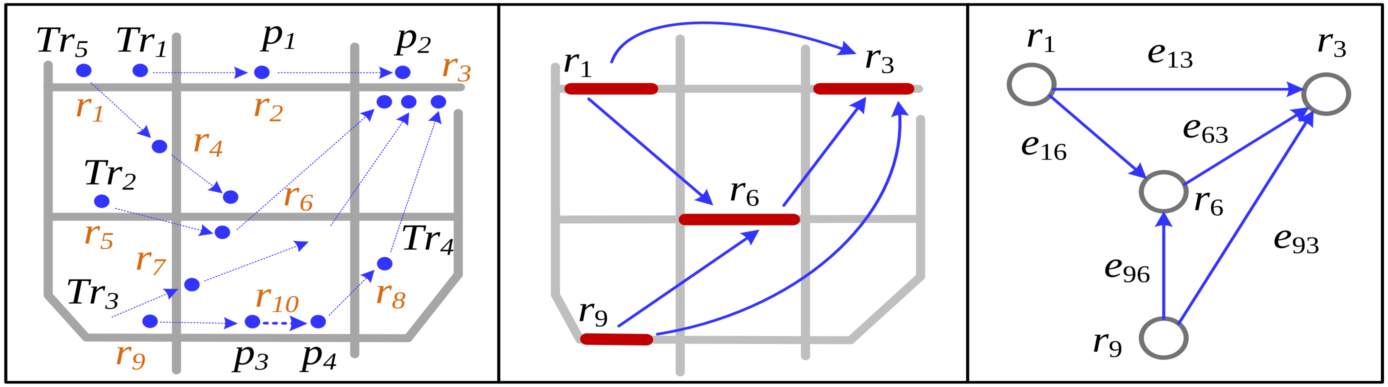






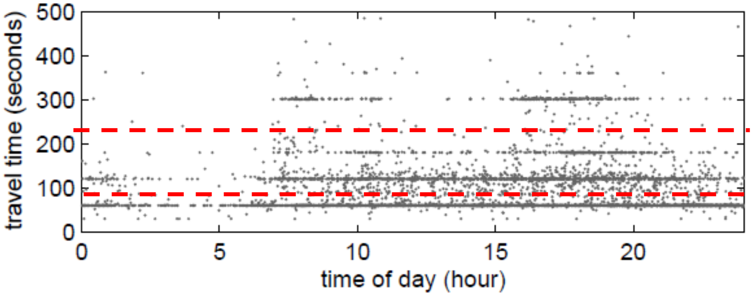
Offline Mining

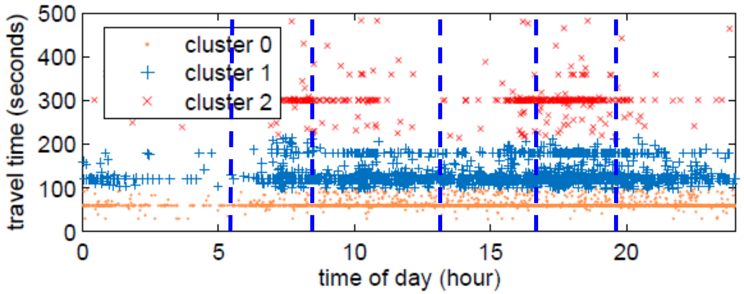
- Building landmark graphs

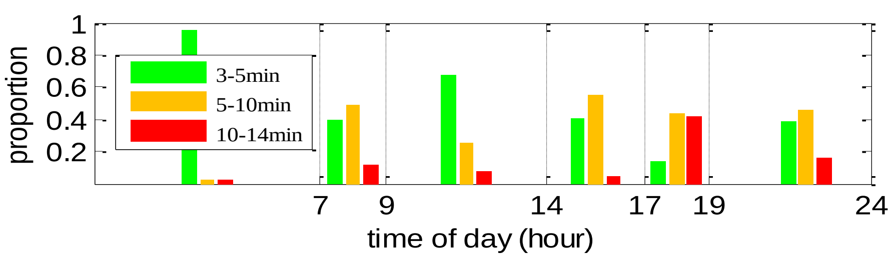


A) Matched taxi trajectories B) Detected landmarks C) A landmark Graph

Mining Taxi Drivers' Knowledge





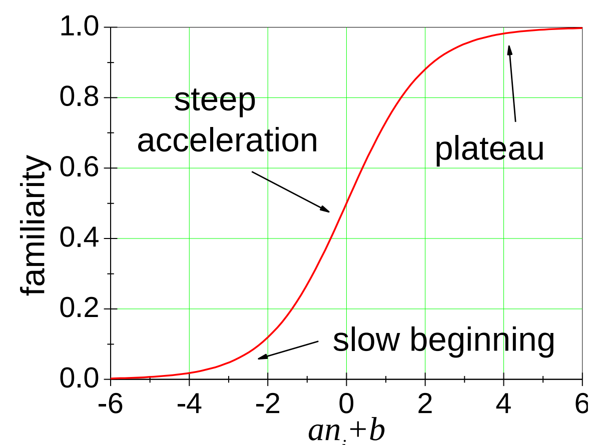


C) Distributions of travel time

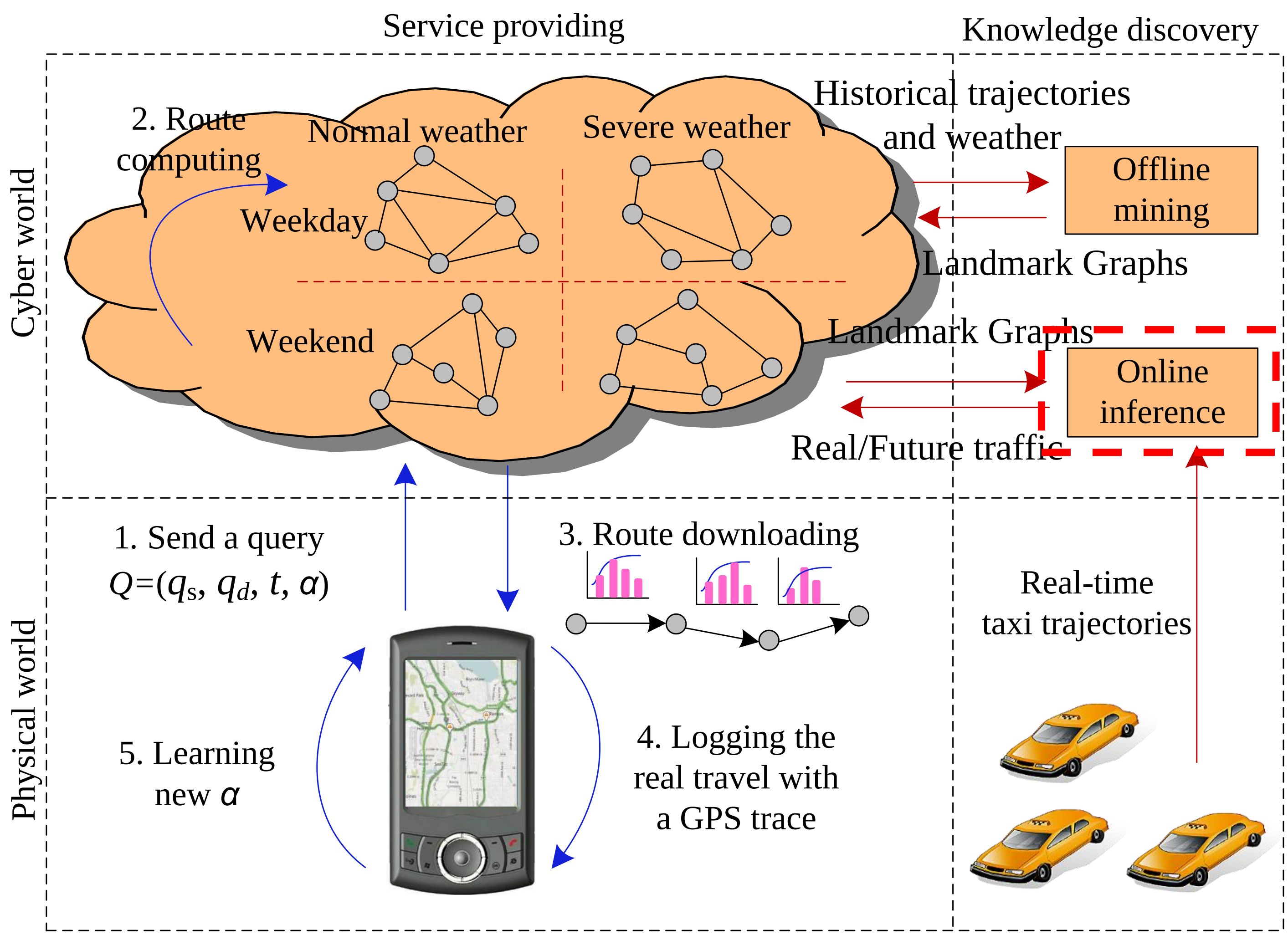
- Learning travel time distributions for each landmark edge
  - Traffic patterns vary in time on an edge
  - Different land edges have different distributions
- Differentiate taxi drivers' experiences in different regions

Sigmoid learning curve

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

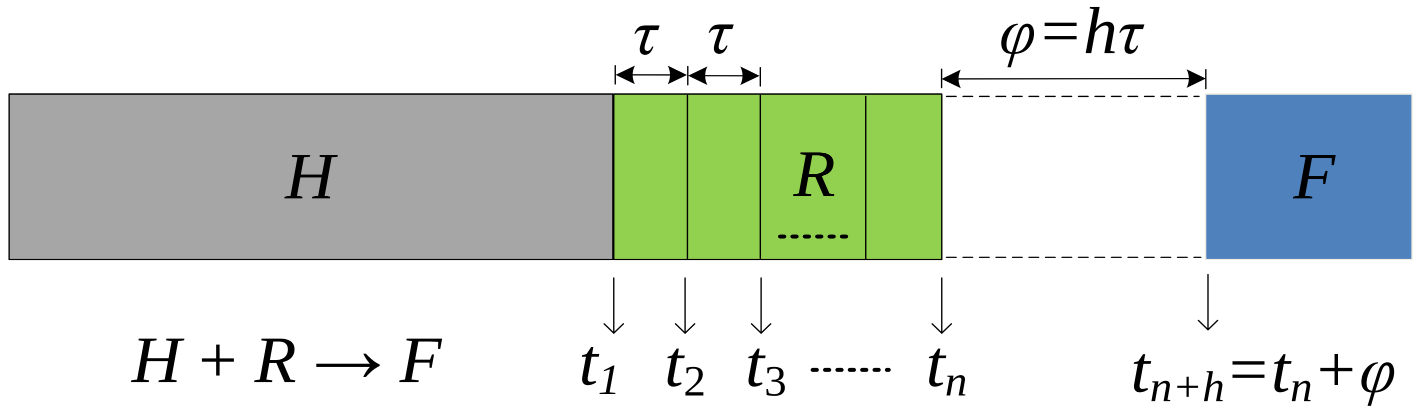


System Overview

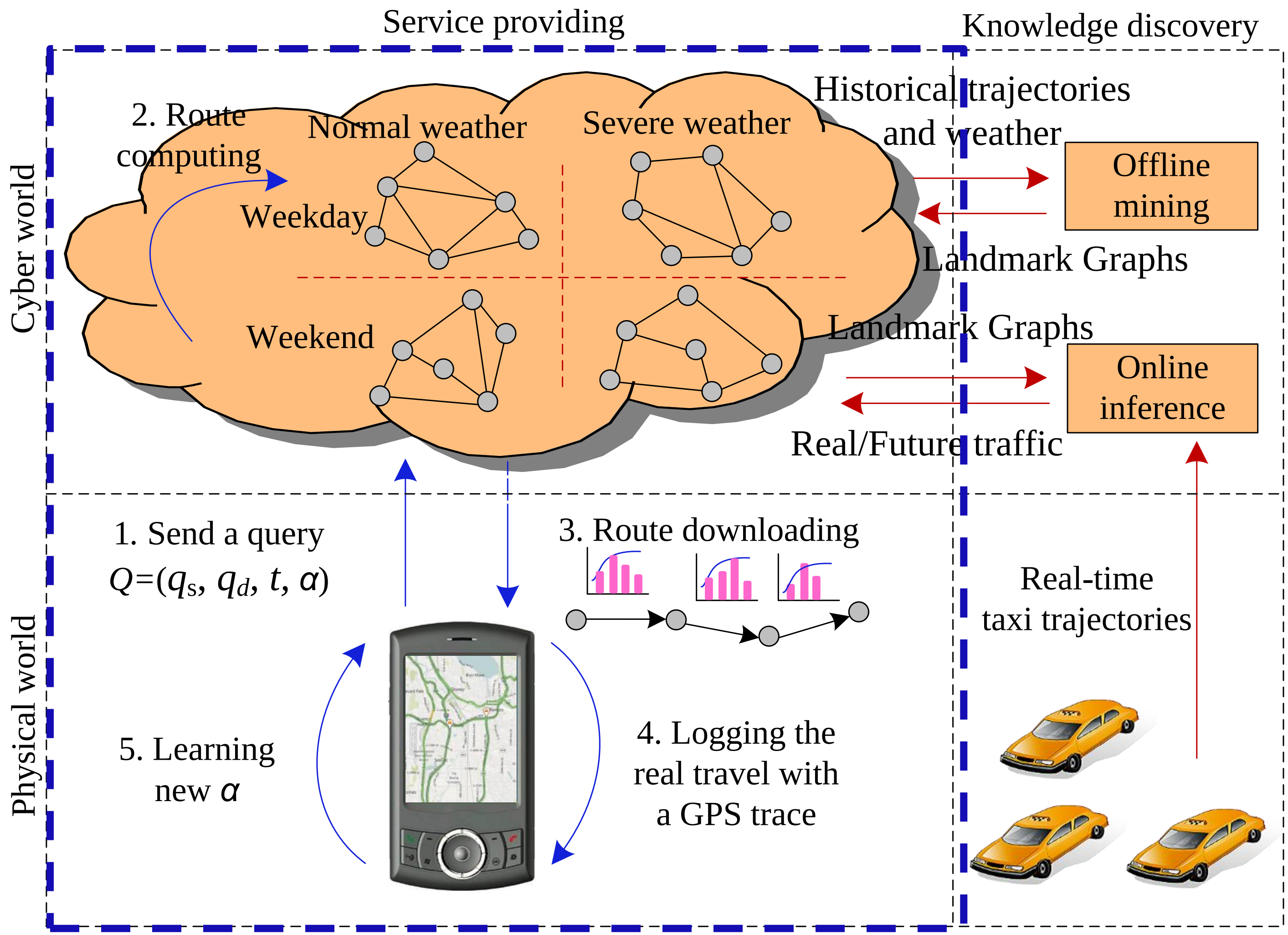


Online Inference

- 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

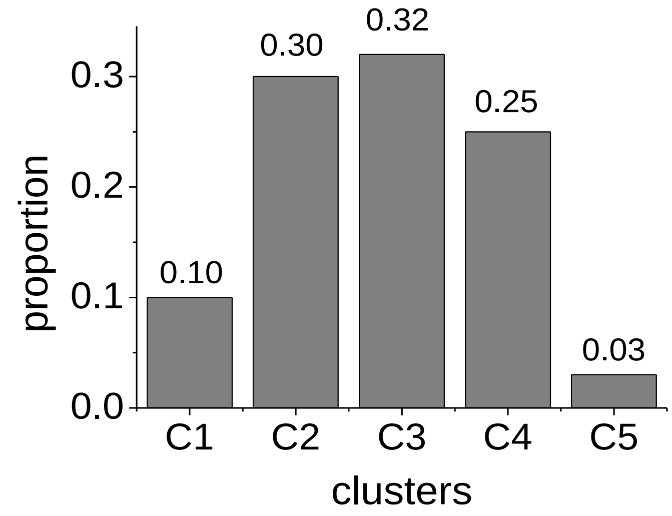


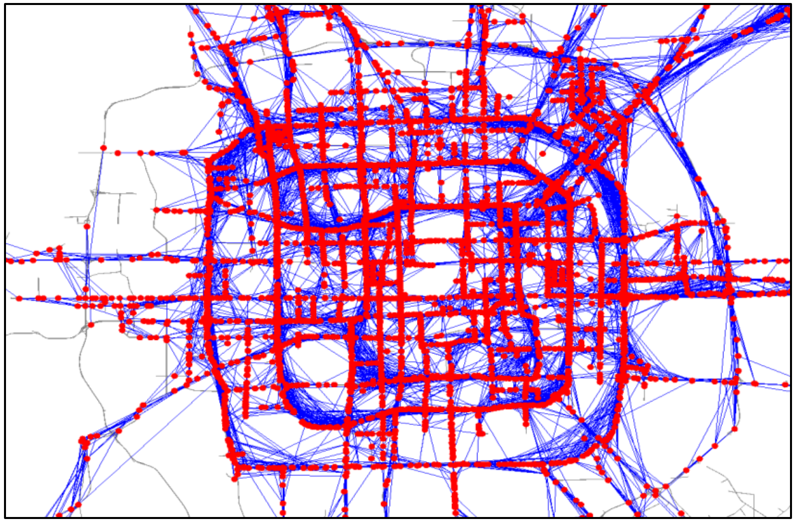
System Overview



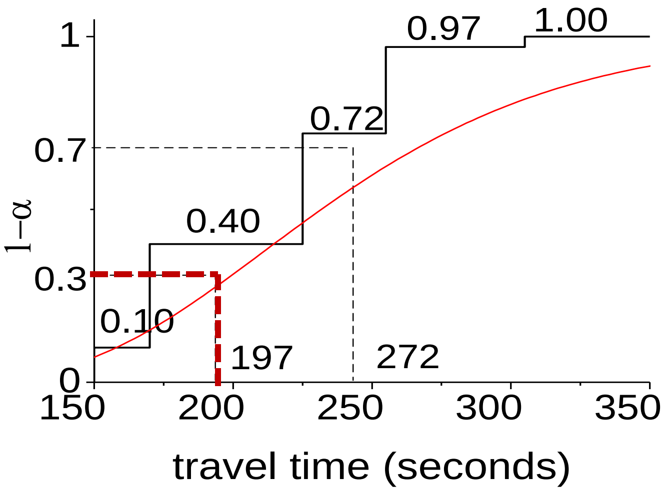
Route Computing

- 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



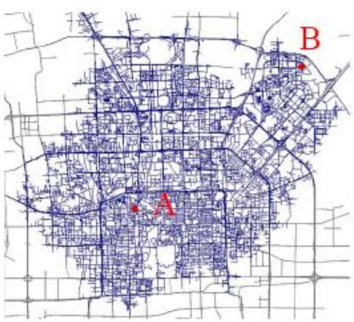


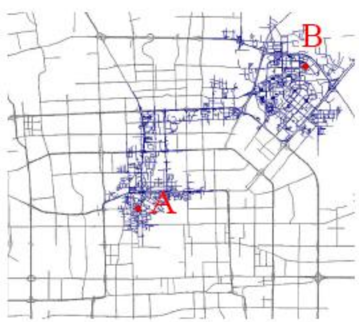
A landmark graph

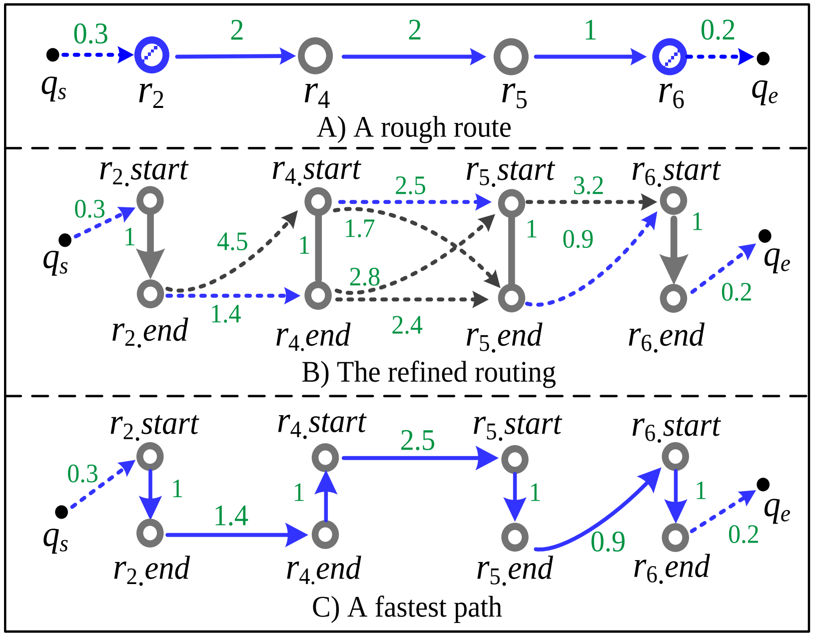


Route Computing

- 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

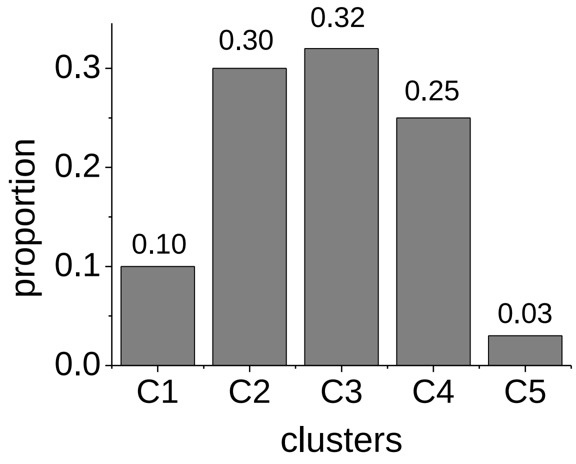






Learning an end user's drive behavior

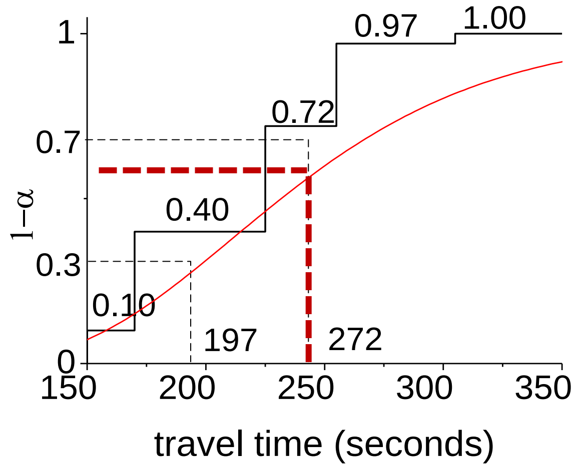
- Drive behavior
  - Vary in persons and places
  - Vary in progressing driving experiences
  - Custom factor: $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$



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

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 – Beijing Datasets

- Beijing Taxi Trajectories
  - 33,000 taxis in 3 months
  - Total distance: 400 million km
  - Total number of points: 790M
  - Average sampling interval:
  - 3.1 minutes, 600 meters
- Beijing Road Network
  - 106,579 road nodes
  - 141,380 road segments
- Driving history of users
  - GPS trajectories from GeoLife project (Data released)

Evaluations – Singapore Dataset

- For evaluating traffic prediction on road segments

Summary

- Model traffic patterns and taxi drivers' intelligence with landmark graphs
- Historical + Real time → Future (m-th order Markov model)
- Two stage routing algorithm
- Self-adaptive to a user's drive behavior

- The practically fastest path is
  - Time-dependent
  - User-specific (for a particular user)
  - Self-adaptive

Urban Computing for Urban Planning

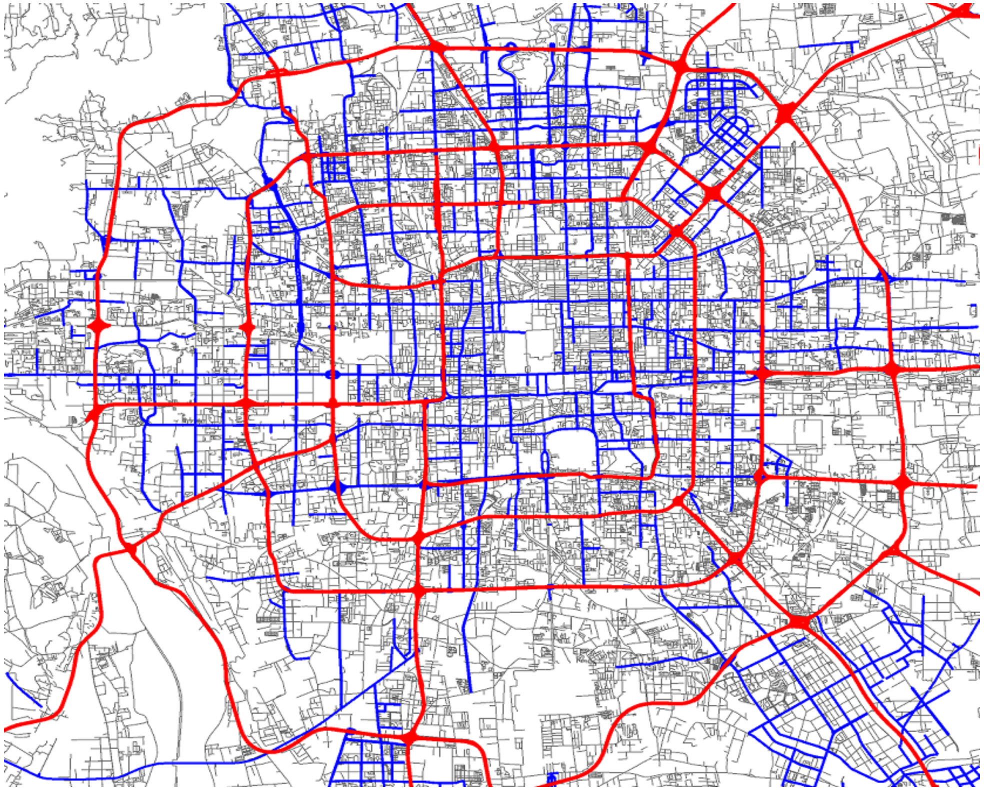
- Goals
  - City-wide traffic modeling
  - Evaluate city configurations
  - Suggest potential improvement to city planners
  - Identify root causes of the problem
- Datasets
  - Taxi trajectories: March to May, 2009 and 2010
  - Beijing maps and POIs

Read the Related papers



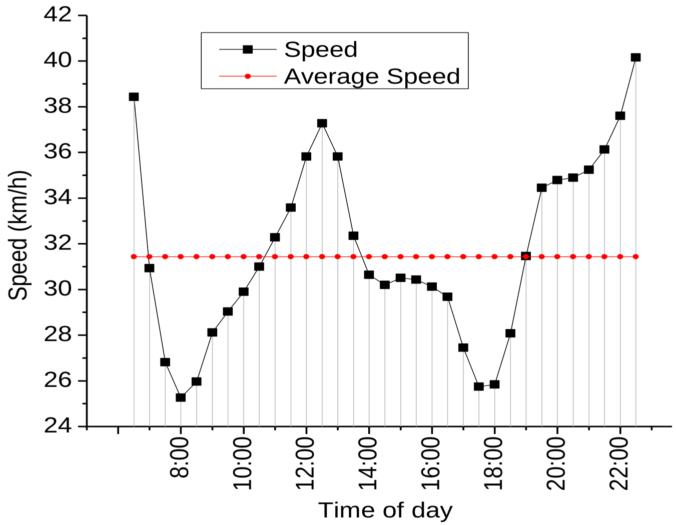
Methodology

- Partition a city into regions with major roads

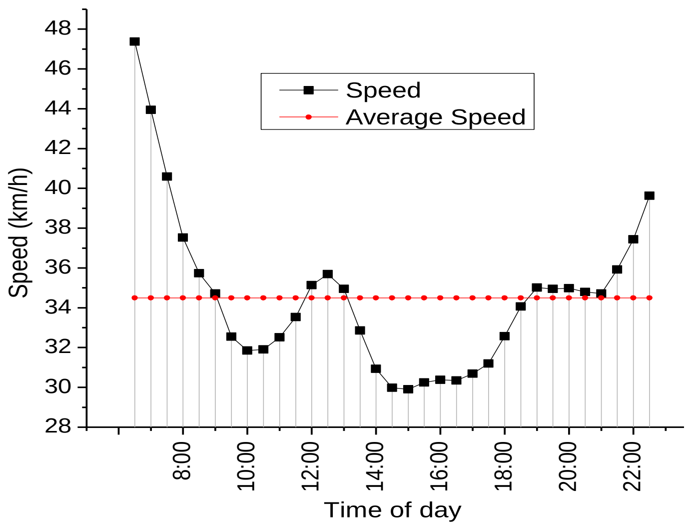


Methodology

- Partition the trajectory dataset into some portions



Workday



Rest day

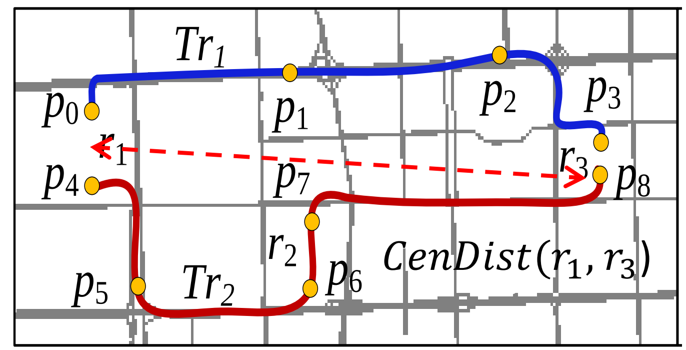
Time	Work day	Rest day
Slot 1	7:00am-10:30am	9:00am-12:30pm
Slot 2	10:30am-4:00pm	12:30pm-7:30pm
Slot 3	4:00pm-7:30pm	7:30pm-9:00am
Slot 4	7:30pm-7:00am	

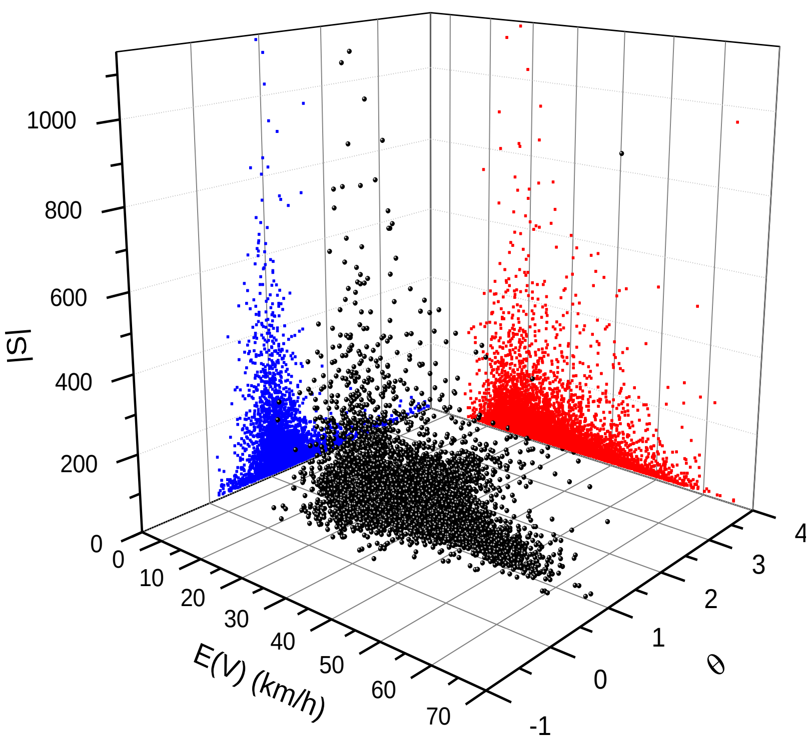
Methodology

- 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
  - $\theta = E(D)/CenDist(r_1, r_3)$





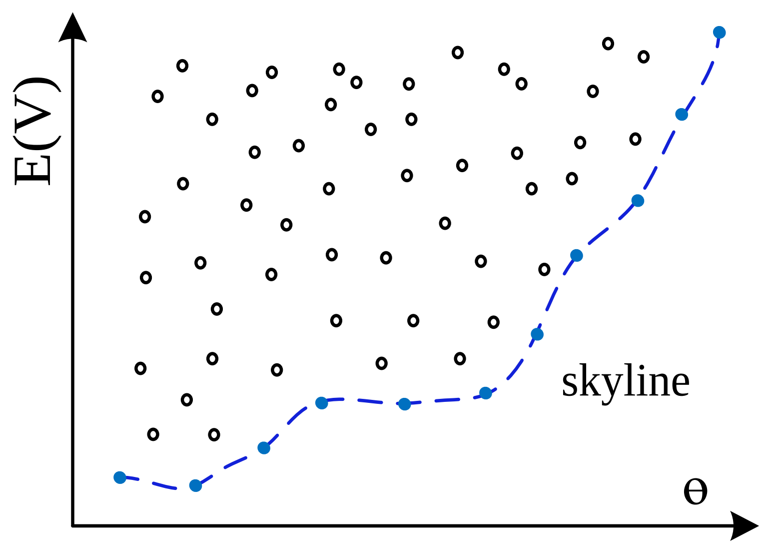
$$
\boldsymbol{M} = \begin{array}{c} \\ r_0 \\ r_1 \\ \vdots \\ r_i \\ \vdots \\ r_{n\text{-}1} \\ r_n \end{array}
\begin{array}{c} \begin{array}{cccccc} r_0 & r_1 \cdots & r_j & \cdots & r_{n\text{-}1} & r_n \end{array} \\
\left[\begin{array}{cccccc}
\phi & a_{0,1} & \cdots & & & a_{0,n} \\
a_{1,0} & \phi & \cdots & & & a_{1,n} \\
\cdots & & & & & \\
a_{i,0} & a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,n} \\
\cdots & & & & & \\
a_{n\text{-}1,0} & \cdots & & & \phi & a_{n\text{-}1,n} \\
a_{n,0} & \cdots & & & & \phi
\end{array} \right] \end{array}
$$

$$a_{ij} = < |\boldsymbol{S}|, E(V), \theta >$$

Finding Problematic Edges

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

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

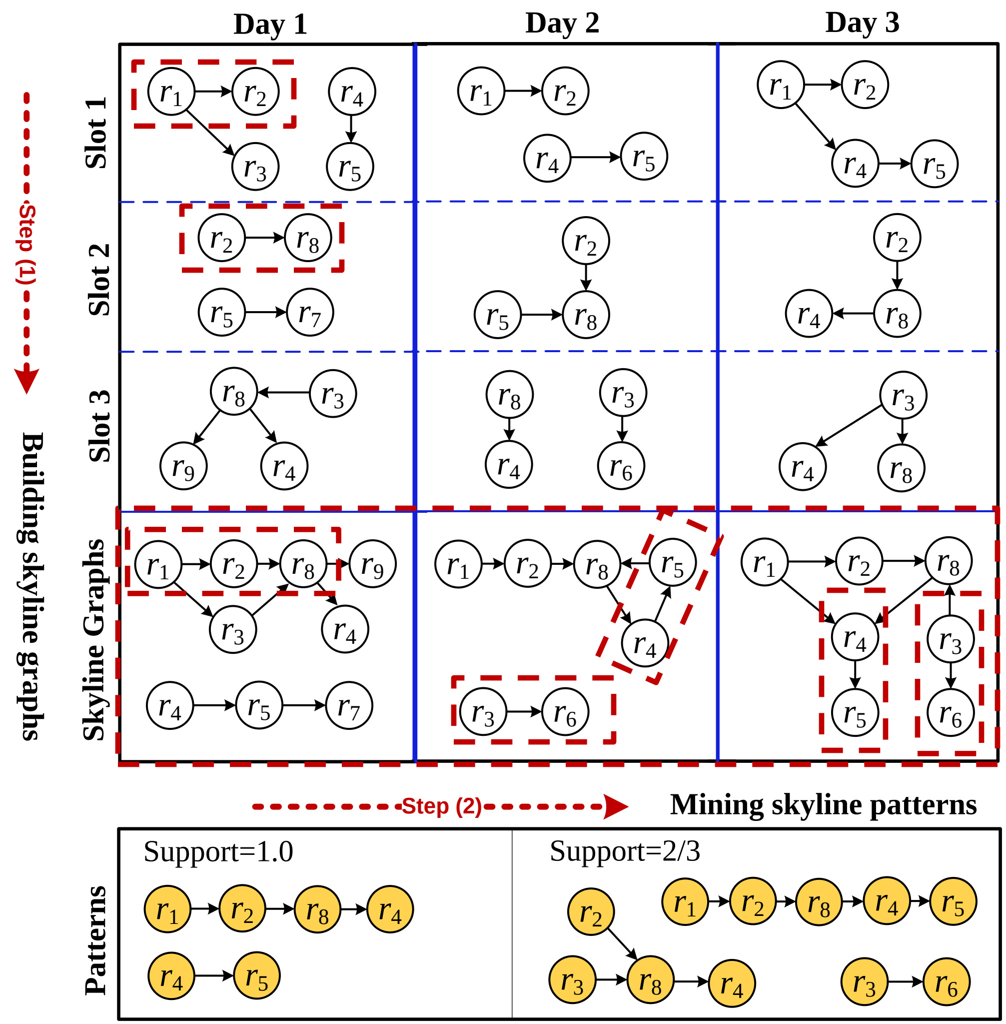


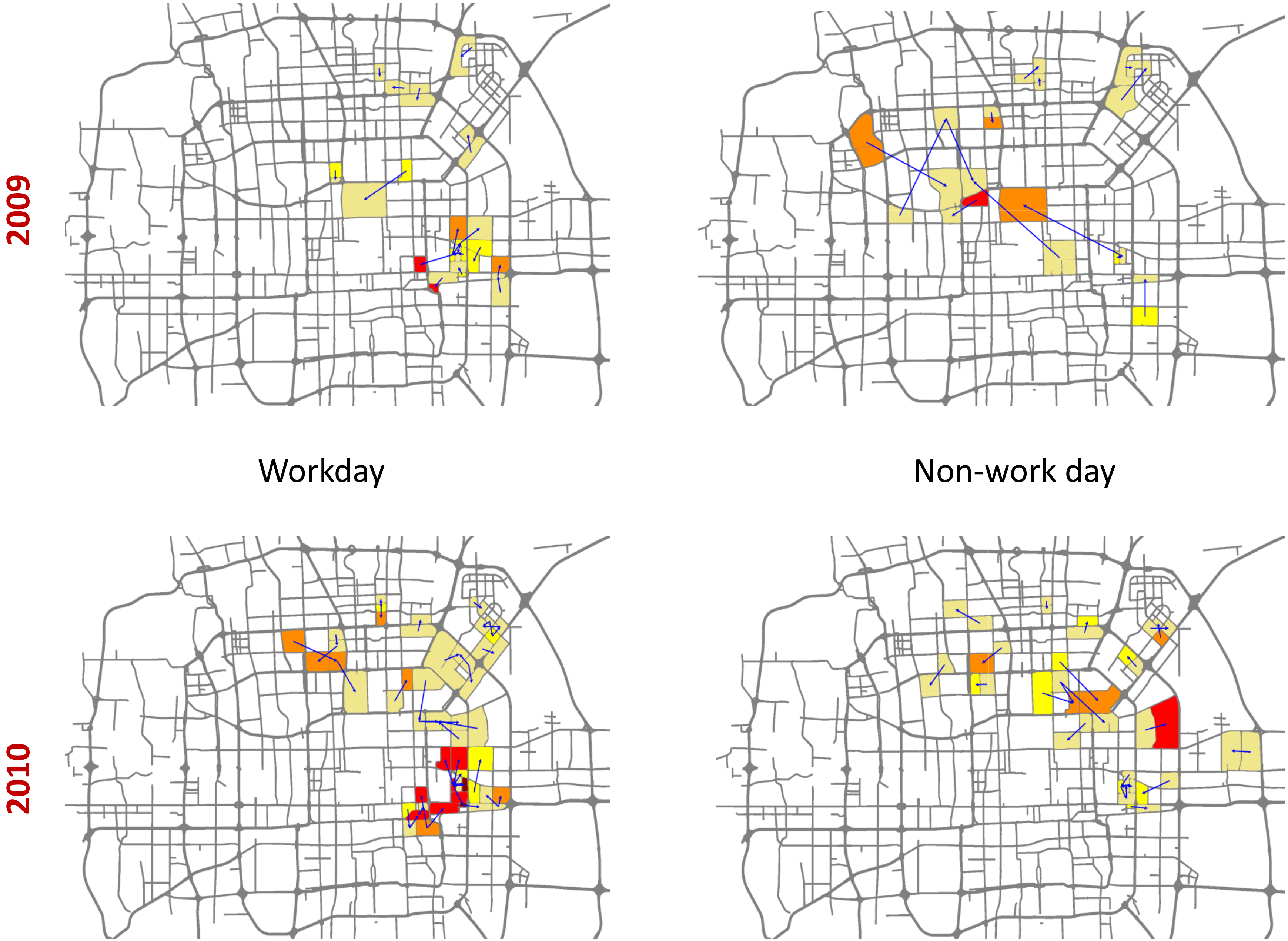
A) A skyline

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
- Find correlation among patterns

$$(r_3 \rightarrow r_6) \Rightarrow (r_4 \rightarrow r_5)$$




2009
Workday
Non-work day
2010

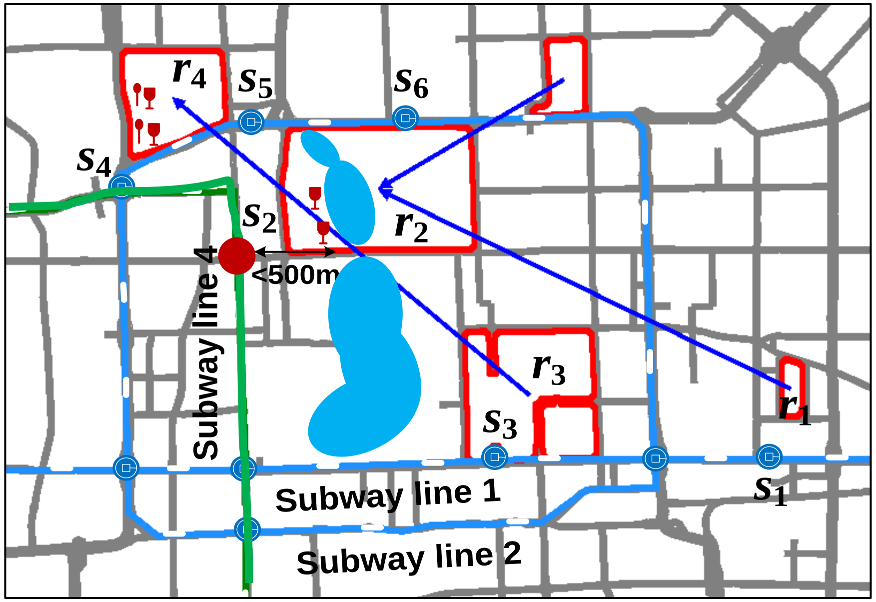

A new road launched in August 2009
An additional entrance
r_2
r_1
Entrance
The 4th ring road
2009
house.focus.cn
北 N
北湖渠路
BEIHUQU Rd
（辛店路）
关 庄 路
GUANZHUANG Rd
（北苑路）
关庄路东口
2010

Example 2

r_1 : Guomao (In CBD of Beijing)

r_2: Houhai (A bar street)

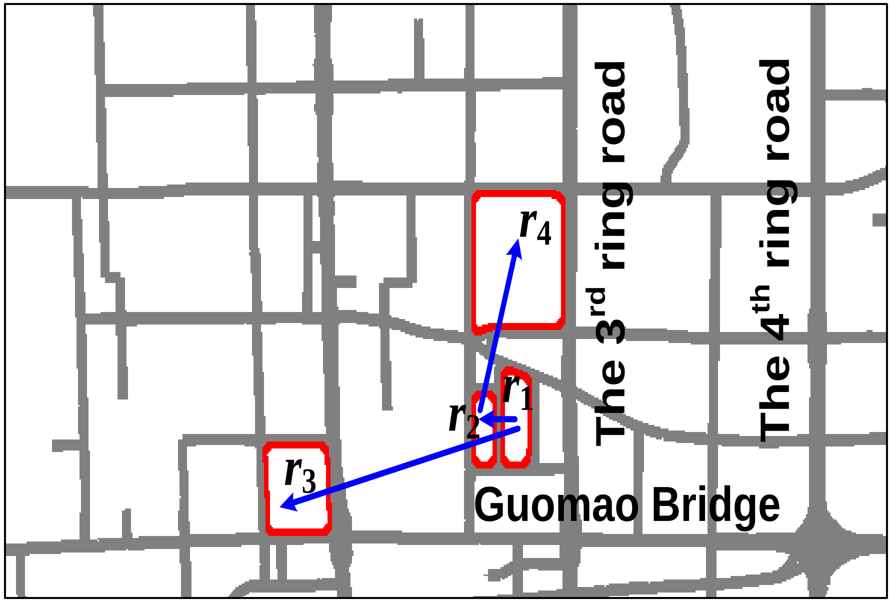
Line 4 launched in 2010



Example 3

Finding correlation

$$(r_1 \rightarrow r_2) \Rightarrow [(r_2 \rightarrow r_4),(r_1 \rightarrow r_3)]$$



Computing with Spatial Trajectories

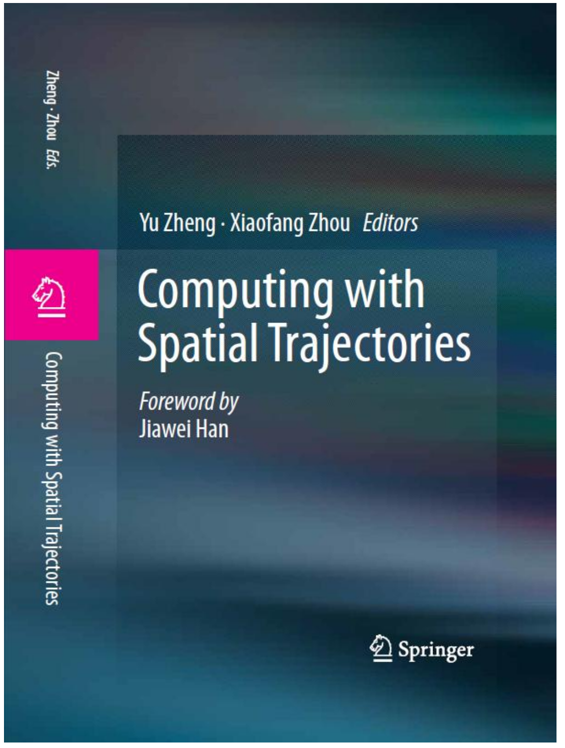
- A tutorial on ACM SIGSPATIAL GIS 2011
- [Check it out online]

Tutorial Session: Computing with Spatial Trajectories

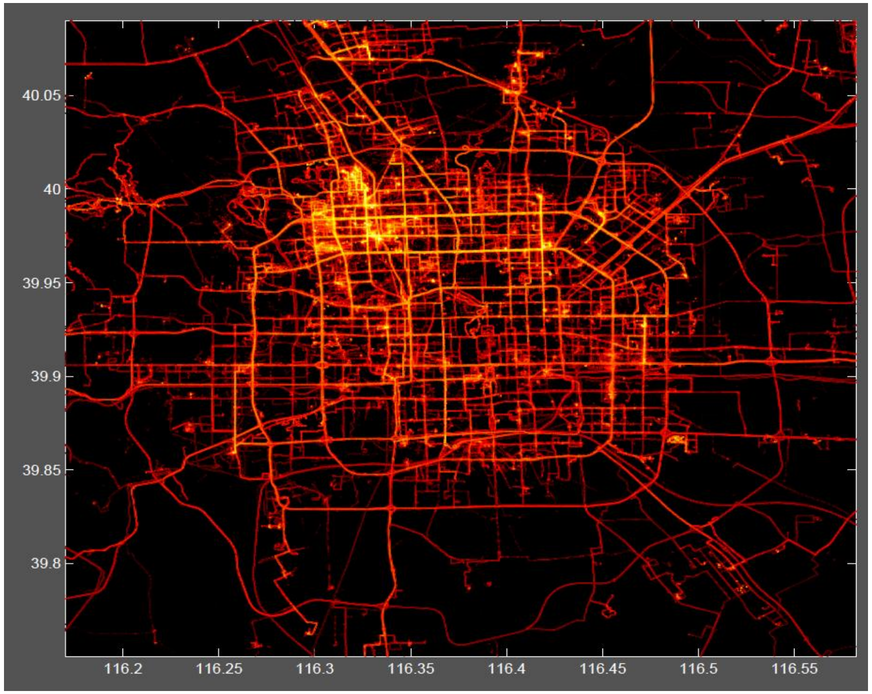
In ACM SIGSPATIAL GIS 2011, 3pm-6pm, Nov. 4, 2011

Session Chair: Ouri Wolfson

15:00-15:05	Book overview Yu Zheng
15:05-15:30	Trajectory Preprocessing (Chapter 1) John Krumm
15:30-15:55	Trajectory Indexing and Retrieval (Chapter 2) Ke Deng
15:55-16:20	Uncertainty in Spatial Trajectories (Chapter 3) Goce Trajcevski
16:20-16:45	Trajectory Pattern Mining (Chapter 5) Christian S. Jensen
16:45-17:10	Trajectory Analysis for Driving (Chapter 7) John Krumm
17:10-17:40	Location-Based Social Networks (Chapter 8,9) Yu Zheng



GeoLife Trajectory Dataset (1.1)



Transportation mode	Distance (km)	Duration (hour)
Walk	11,457	5,126
Bike	6,335	2,304
Bus	21,931	1,430
Car & taxi	34,127	2,349
Train	74,449	459
Airplane	28,493	37
Other	10,886	335
Total	187,679	12,041

	Version 1.0	Version 1.1	Incremental
Time span of the collection	04/2007 – 08/2009	04/2007 – 12/2010	+16 months
Number of users	155	167	+12
Number of trajectories	15,854	17,355	+1,501
Number of points	19,304,153	22,294,264	2,990,111
Total distance	600,917 km	1,070,406 km	+469,489 km
Total duration	44,776 hour	48,349 hour	+3,573 hour
Effective days	8,977	9,694	+717

[Link to the data]

	Version 1.1	Version 1.2	Change
Time span of the collection	04/2007 – 12/2010	04/2007 – 10/2011	+10 months
Number of users	167	178	+11
Number of trajectories	17,355	18,465	+1,110
Number of points	22,294,264	24,910,206	+2,615,942
Total distance	1,070,406 km	1,132,980km	+62,574 km
Total duration	48,349 hour	50, 465hour	+2,116 hour
Effective days	9,694	10,422	+728

Video and Demos

Visit [the homepage of Urban Computing]

Thanks!

Yu Zheng

yuzheng@microsoft.com