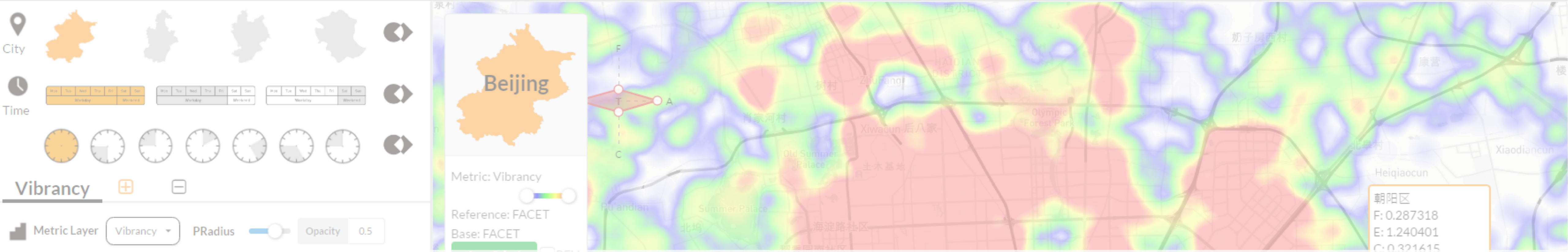


UrbanFACET: Visually Profiling Cities from Mobile Device Recorded Movement Data of Millions of City Residents

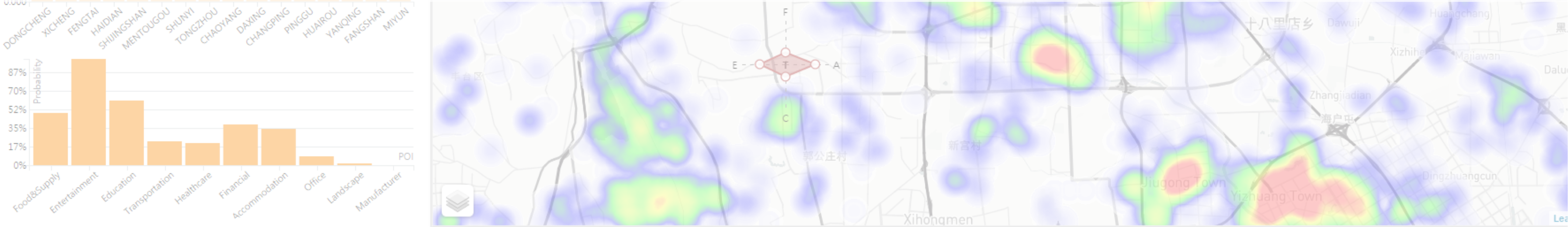
Tao Jiang[1,3,4], Lei Shi[1,4], Ye Zhao[2], Xiatian Zhang[3], Yao Lu[3]

1) State Key Laboratory of Computer Sciences, Institue of Software, Chinese Academy of Sciences

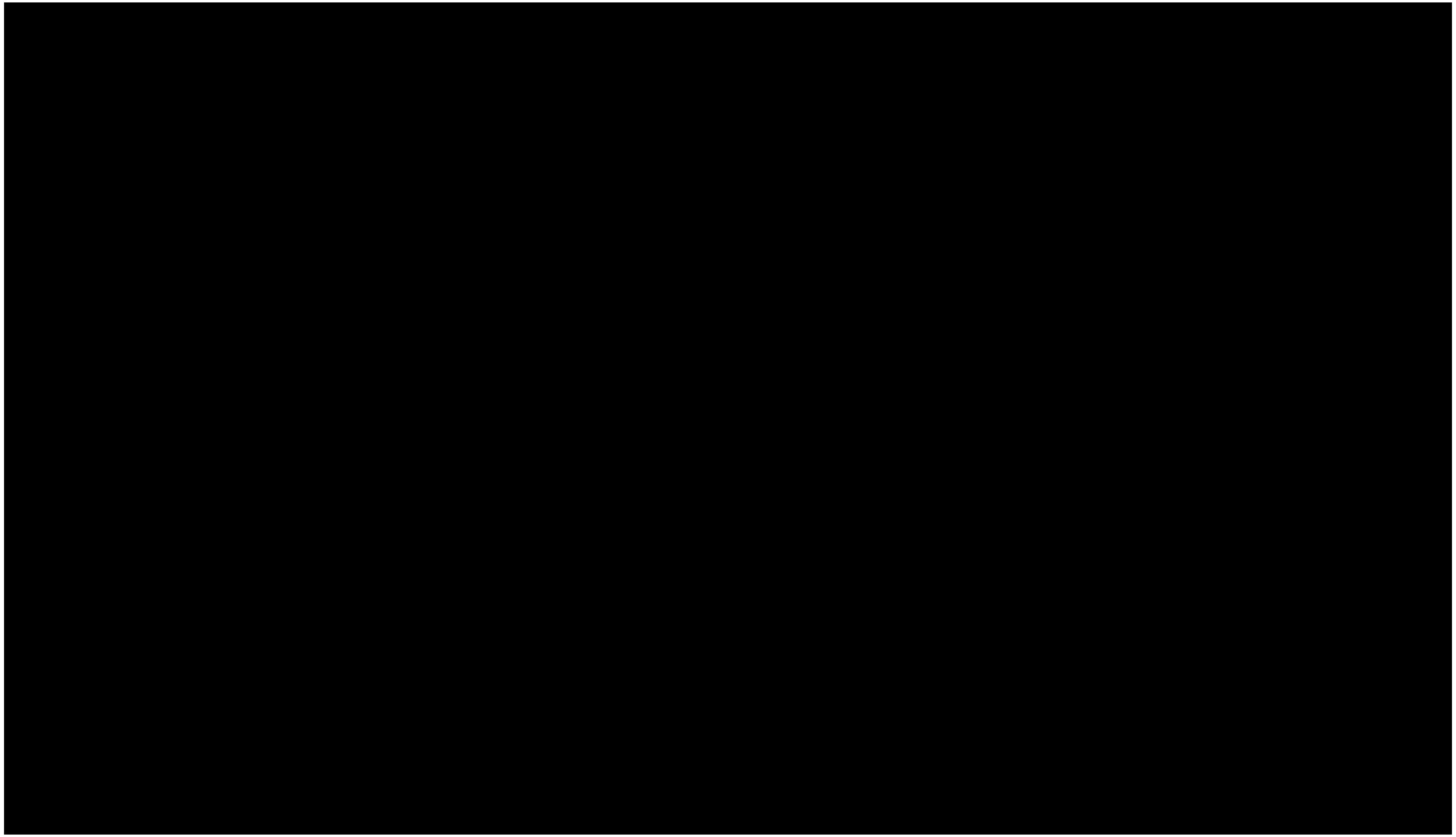
2) Department of Computer Science, Kent State University

3) Beijing Tendcloud Tianxia Technology Co., Ltd (TalkingData)

4) University of Chinese Academy of Sciences (UCAS)

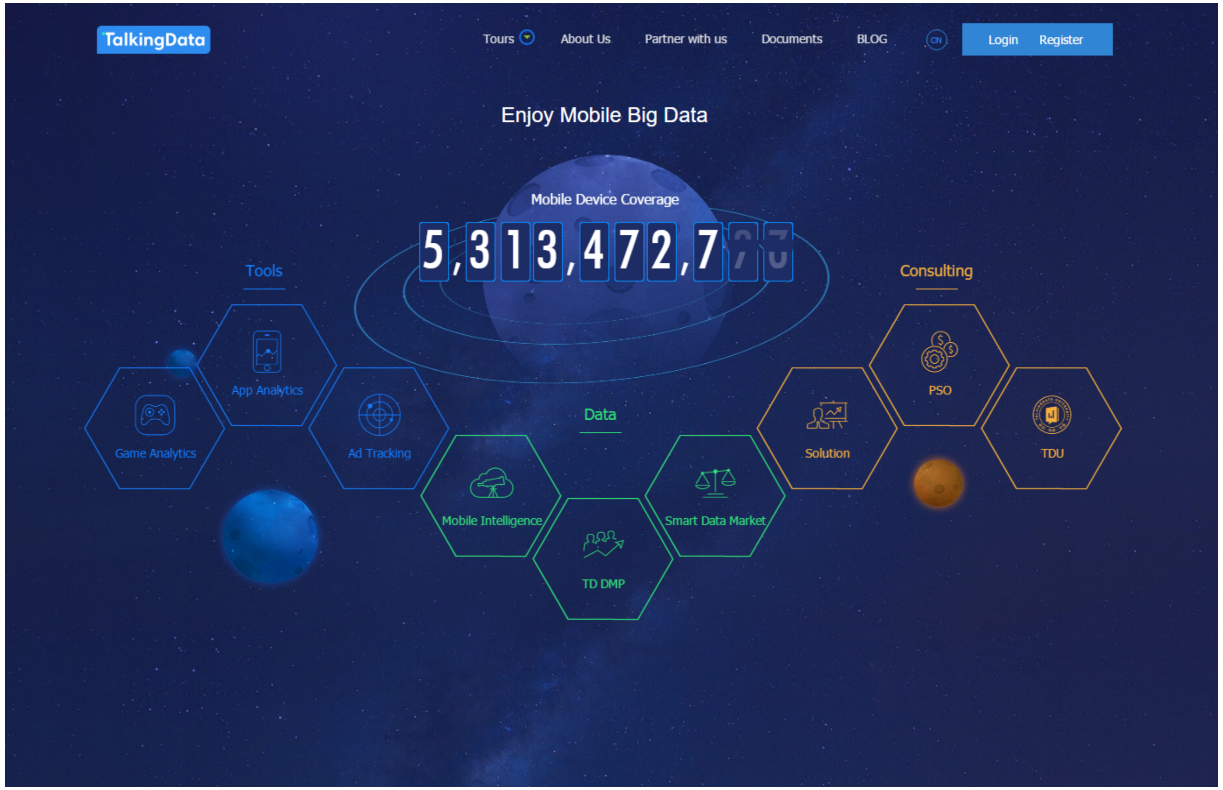


UrbanFACET Demo



Background - Big urban movement data in TalkingData

- TalkingData - **China's largest** independent Big Data service platform.
- Processes **TB$^+$** data and more than **1 billion** session requests every day.
- Covers **thousands of** mobile apps and **millions of** smart devices.
- User properties can be tracked by **unqiue** user ID.



From http://www.talkingdata.com/

ISCAS 中国科学院大学 University of Chinese Academy of Sciences KENT STATE UNIVERSITY TalkingData

Related work

- Movement data **visualization approaches**
  - Flow map, flowstrates, OD maps
- **Urban mobility** analysis
  - P. A. L., etc. Spatiotemporal analysis of bluetooth data: Application to a large urban network.
  - G. A., etc. Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data
  - J. W., etc. Visualizing the dynamics of londons bicycle hire scheme
- **Entropy** related analysis
  - C. S., etc. Limits of predictability in human mobility
  - K. C., etc. An analysis of entropy of human mobility from mobile phone data

(a) (b) (c)

Fig 1. Flow Map

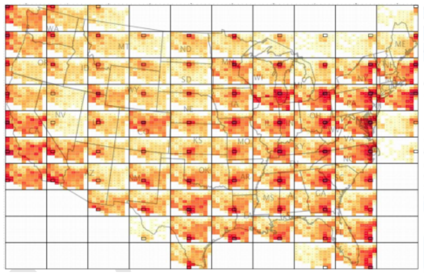


Fig 2. Flowstrates

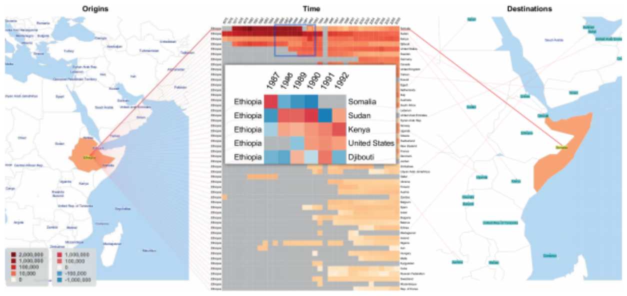


Fig 3. OD Map

Fig 4. Visualizing the Dynamics of London's Bicycle Hire Scheme

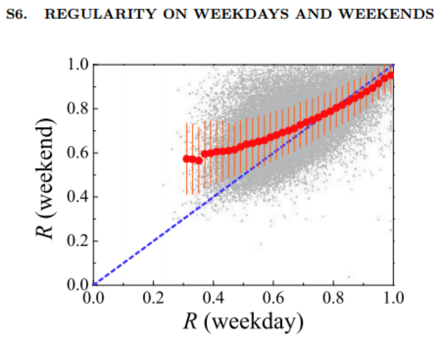


Fig 5. Limits of Predictability of Human Mobility Supplementary Material

Background - Objective

- Limits: Previous geo-tagged data are **expensive** or spatially **sparse**, it's hard to make user-level profiling and analyzing;
- Works: We use the comprehensive dataset to **profile cities**, enable experts to explore **user-level** mobilities;
- Apps: traffic scheduling, location of public facilities, situation and risk awareness;

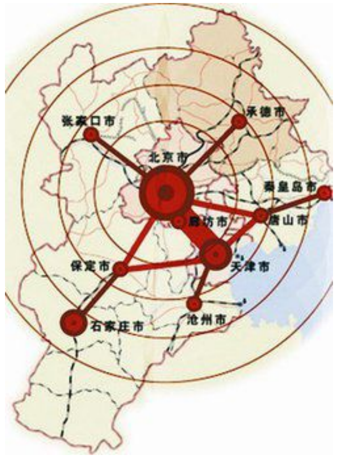


Fig 1. http://www.apptiled.com/index.php/assignment/13890/
Fig 2. http://www.aiaa.com.hk/lang-en/shuttle.html
Fig 3. http://baike.baidu.com/item/%E4%BA%AC%E6%B4%A5%E5%86%80

Dataset used in UrbanFACET

- Data Content: Continuous **90-day** user location records* (**1TB+**)
- Source Filter Condition: only **GPS/Wi-Fi**

Field	Description	Sample
Time	Timestamp of the record	20:10/07/03/2015
Lon.	Longitude of location	116.3336266
Lat.	Latitude of location	39.890955
Mid	Unique ID of the device	1470076020481
Src	Source of the location record	GPS

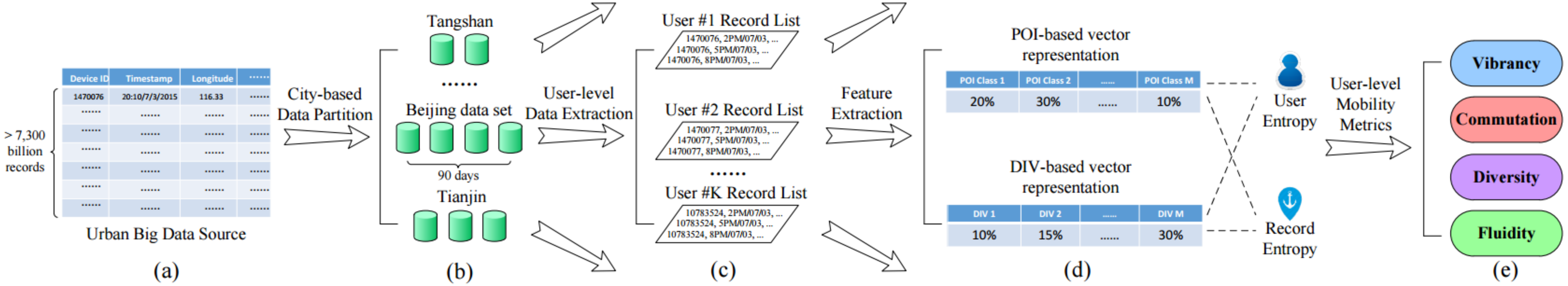
Fig 1. Metadata of location records

City	#Device	#Record	Size	Time
Beijing	31849742	8407648917	738.1G	90 days
Tianjin	8011128	2858575880	206.8G	90 days
Tangshan	2786668	920364499	64.8G	90 days
Zhangjiakou	1392236	317252149	23.1G	90 days

Fig 2. Statistics on four data sets used in this research

** User privacy is preserved by anonymizing the device ID and discretizing time.*

UrbanFACET Data Processing Framework



Feature Extraction　　*User-level Mobility Computation*

ISCAS 中国科学院大学 University of Chinese Academy of Sciences KENT STATE UNIVERSITY TalkingData

User-based Feature Extraction

- Point of Interest (POI)
- Administrative Division (DIV)

1

	POI Class									
	1	2	3	4	5	6	7	8	9	10
Probability										

2

	DIV Class								
	1	2	3	4	5	6	7	...	N
Probability								...	

3

POI Class	POI Class Type
1	Food & Supply
2	Entertainment & Leisure
3	Education
4	Transportation
5	Healthcare & Emergency
6	Financial & Bank
7	Accommodation
8	Office & Commercial
9	Natural Landscape
10	Factory & Manufacturer

4

DIV Class	DIV Class Name
1	Dongcheng
2	Xicheng
3	Chaoyang
4	Fengtai
5	Shijingshan
6	Haidian
7	Mentougou
8	Fangshan
9	Tongzhou
10	Shunyi
11	Changping
12	Daxing
13	Huairou
14	Pinggu
15	Miyun
16	Yanqing

Fig 1. User POI Feature Vector *Fig 2. User DIV Feature Vector*
Fig 3. POI Class Definition *Fig 4. DIV Class Definition (Beijing)*

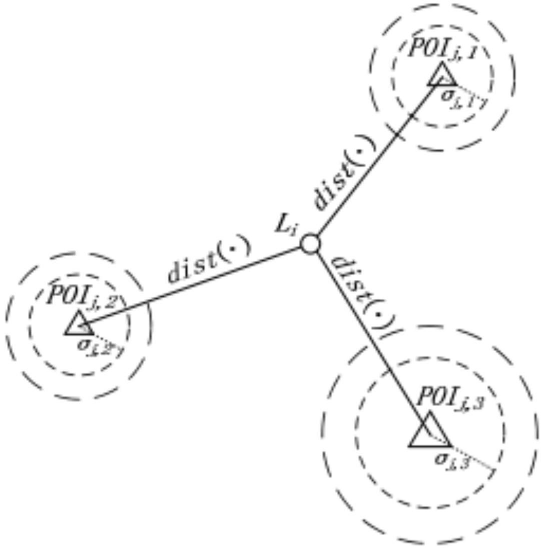
User-based Feature Extraction

Probability Definition

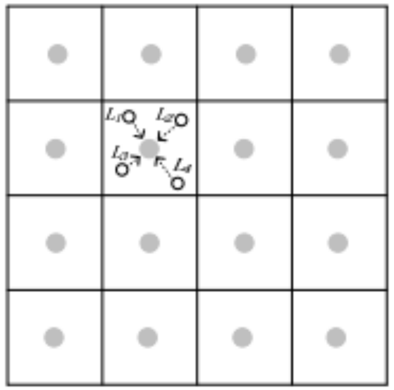
- Calculation on probability of R_i belonging to the j^{th} POI class (q_{ij})

Big Data Issues

- File-based Indexed
  - Storage and query optimization
- Grid-based Computation
  - Scalability on calculating probability



(a)


(b)

$$q_{ij} = \frac{\sum_k \varphi_{0,\sigma^2_{j,k}}(dist(L_i, POI_{j,k}))}{\sum_j \sum_k \varphi_{0,\sigma^2_{j,k}}(dist(L_i, POI_{j,k}))}$$

Mobility Definition – Classification Attempts

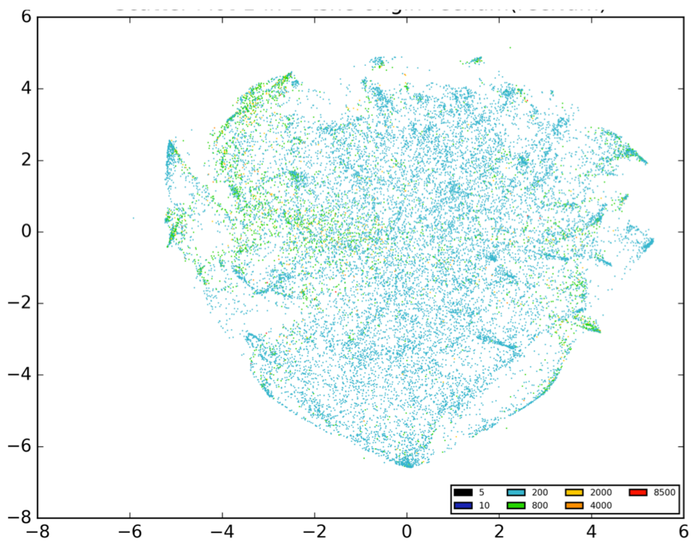


Fig 1. t-SNE of 0.1% random sample of Beijing Data

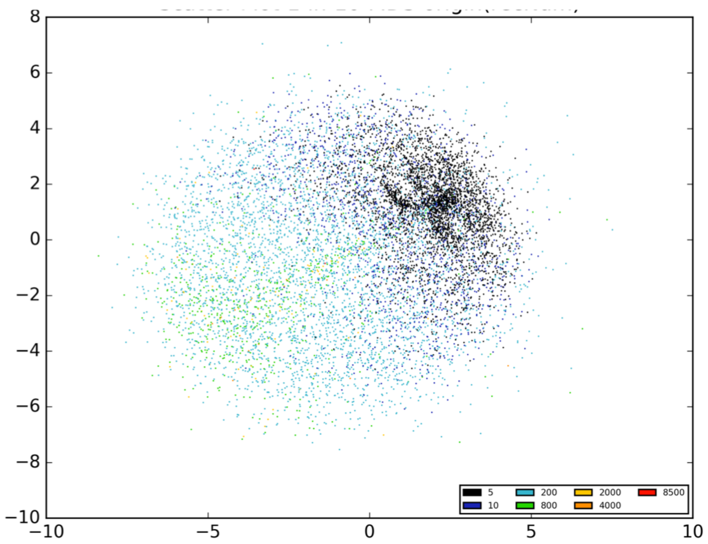


Fig 2. MDS of 0.1% random sample of Beijing Data

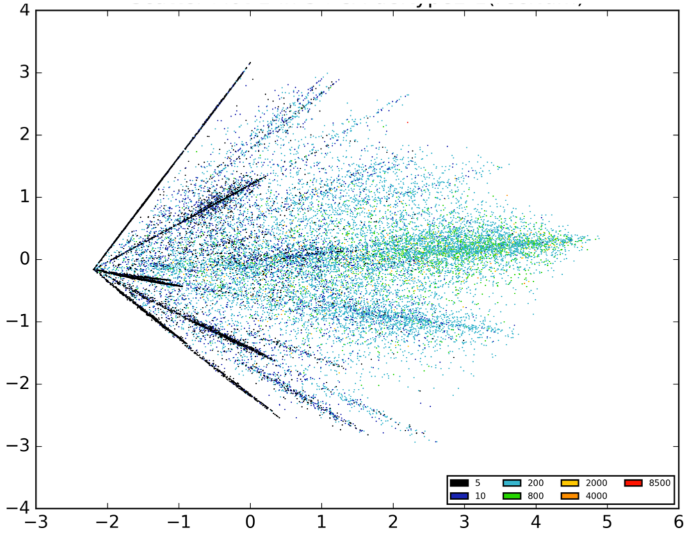


Fig 3. PCA of 0.1% random sample of Beijing Data

It is hard to determine both the **number of clusters** and their clear clustering **boundaries**

Too many **invalid** data occupied in clusters

ISCAS 中国科学院大学 University of Chinese Academy of Sciences KENT STATE UNIVERSITY TalkingData

User-level Mobility Computation

- Continuous user mobility metrics - Shannon entropy over people records

$$H_p = -\sum_{j=1}^{M} p_j \cdot \log p_j$$

- Computation and spatial mapping

User-level Mobility Computation

- Shannon entropy over **each record**

Fig. User entropy

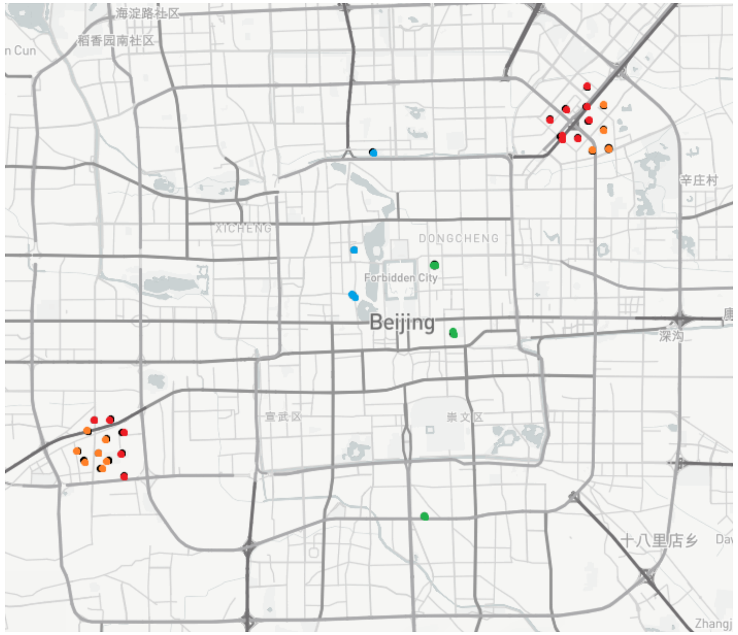
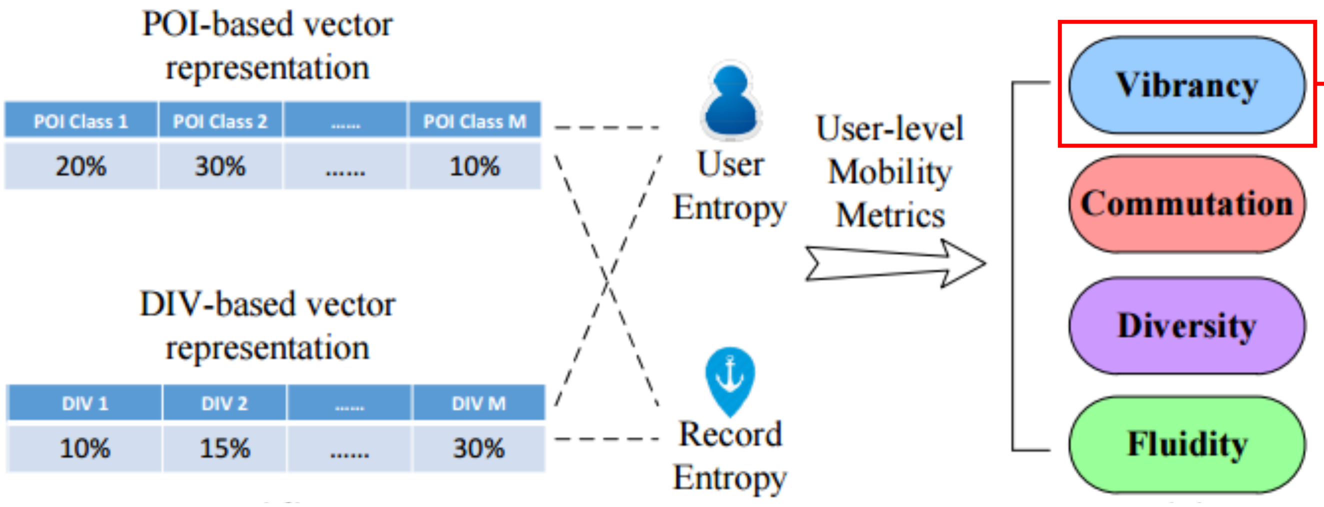


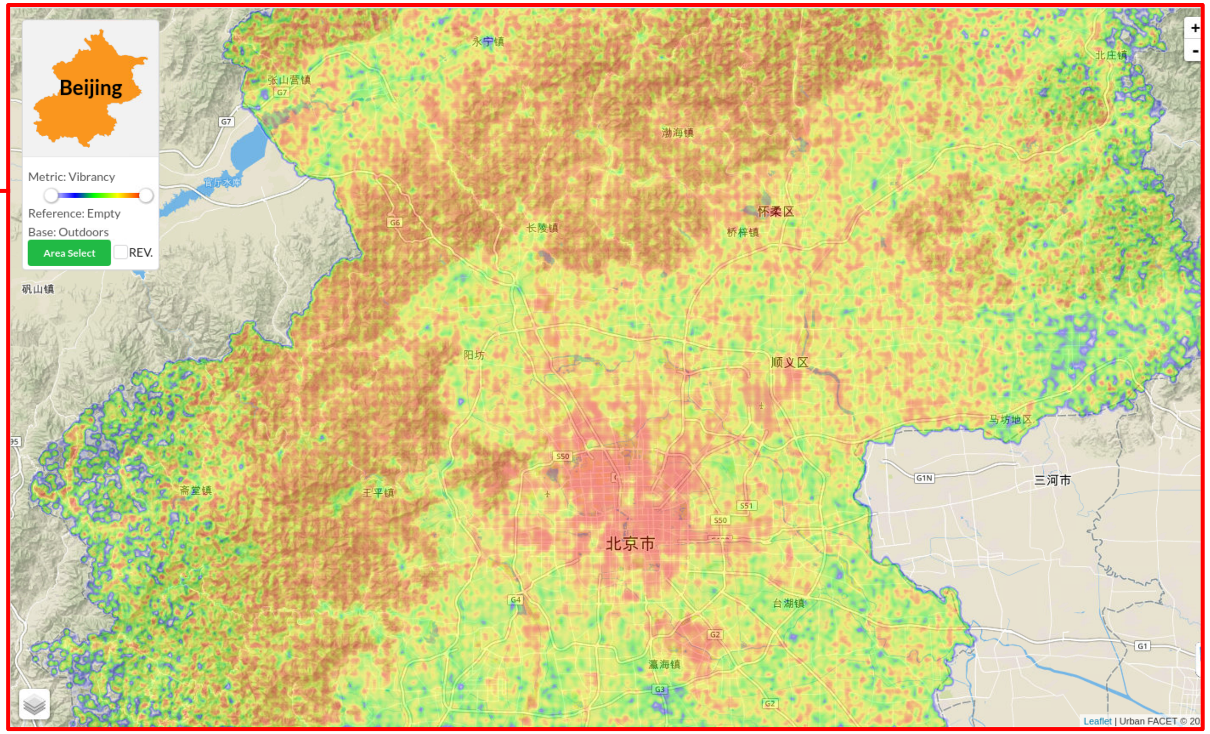
Fig. Record entropy

$$\begin{aligned} N \cdot H_p &= -\sum_{j=1}^{M} N \cdot p_j \cdot \log p_j \\ &= -\frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{M} q_{ij}} \cdot \sum_{j=1}^{M}\sum_{i=1}^{N} q_{ij} \cdot \log p_j \\ &\approx -\sum_{j=1}^{M}\sum_{i=1}^{N} q_{ij} \cdot \log p_j \\ &= \sum_{i=1}^{N}\left(-\sum_{j=1}^{M} q_{ij} \cdot \log p_j\right) \end{aligned}$$

$$H_r = -\sum_{j=1}^{M} q_{ij} \cdot \log p_j$$

User-level Mobility Computation

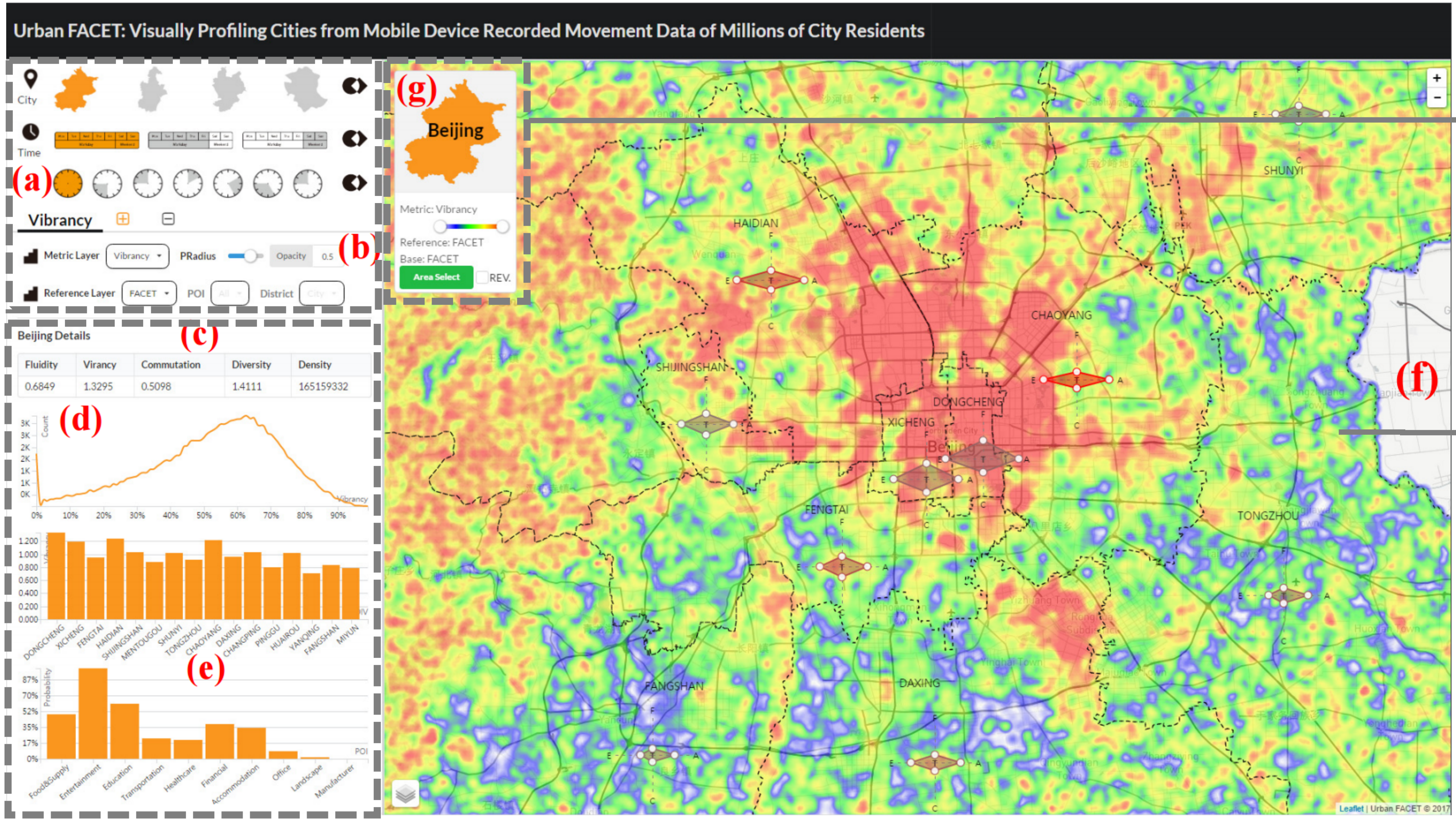




Visualization Design

Visualization configuration panel

Statistics panel



Legend and Controllers

Mobility map

ISCAS University of Chinese Academy of Sciences KENT STATE UNIVERSITY TalkingData

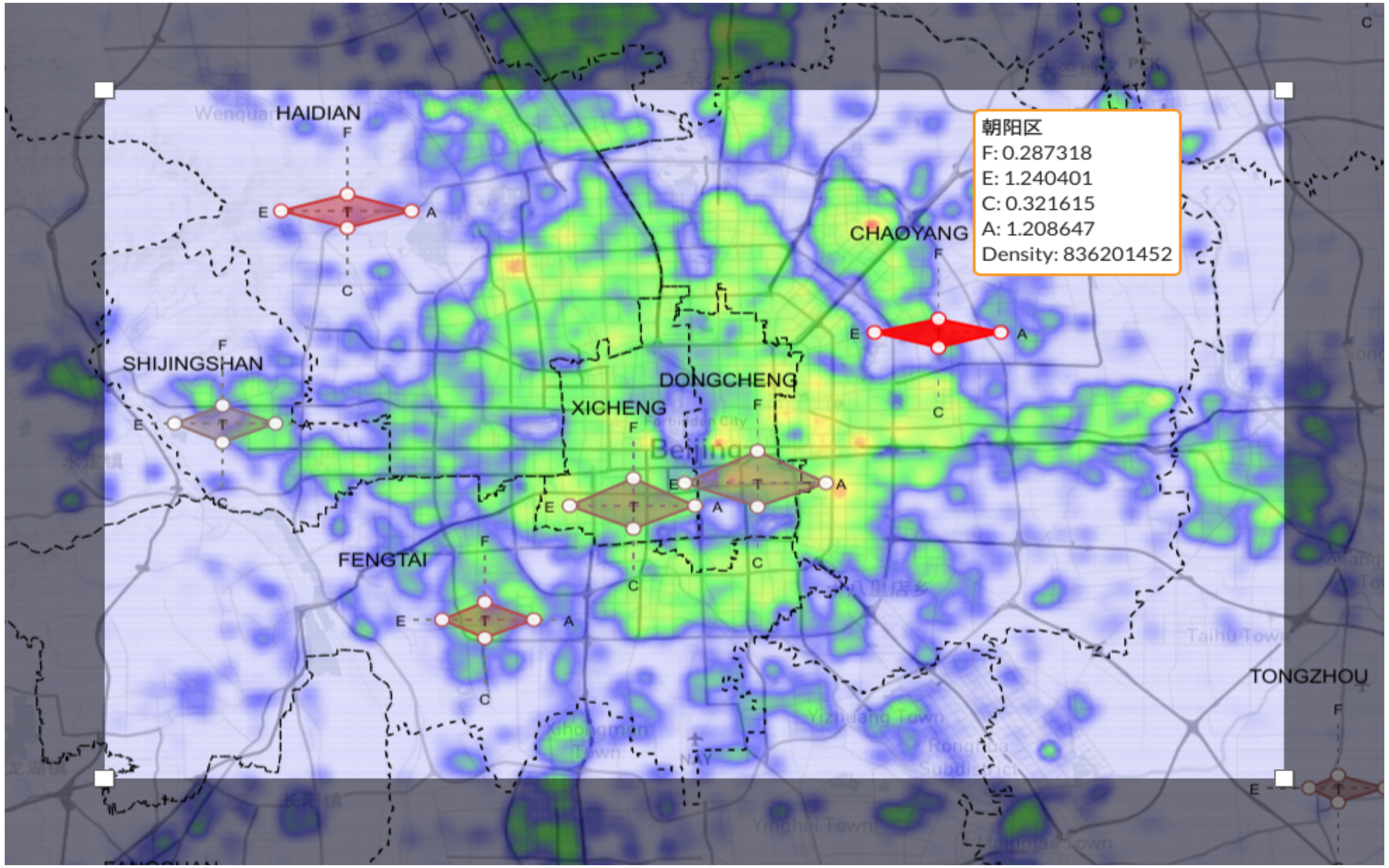
Visualization Design

- **Point diffusion** based contour map
- Double color map **filter**
- **Star plot** design to FACET simultaneously

(a) (b) (c) (d) (e)

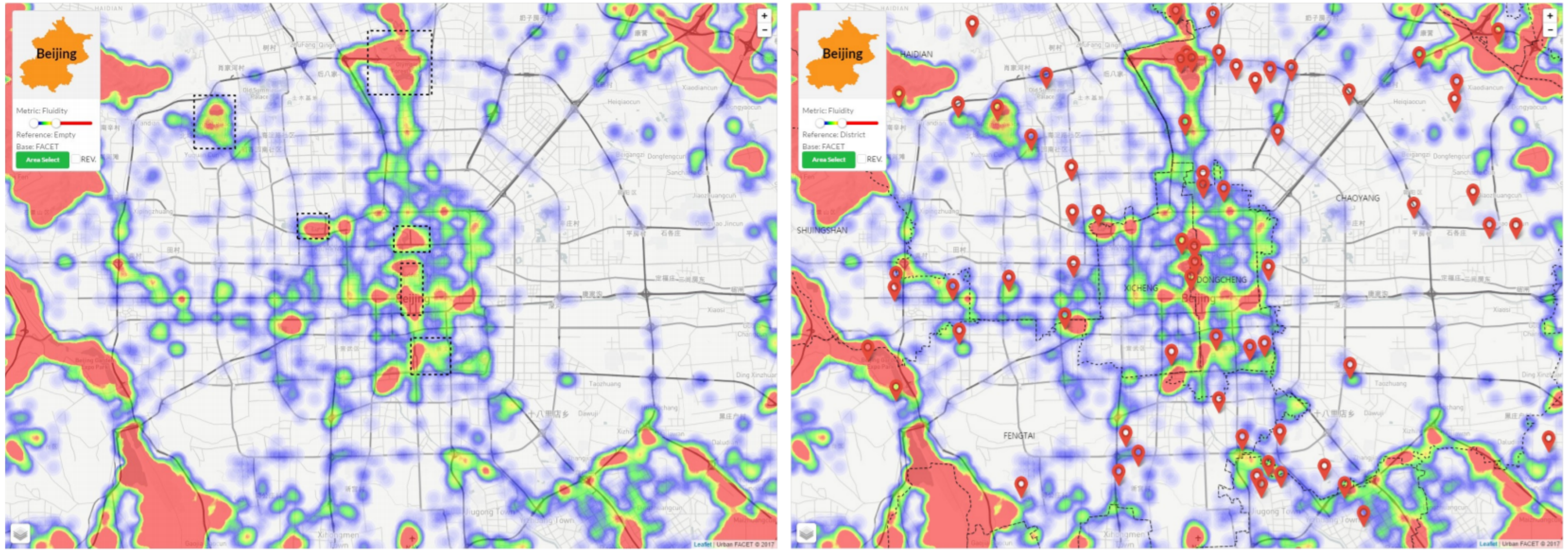
Fig. Alternative mobility metric visualization: (a) grid-based; (b) minimal contour map; (c) contour map with optimal diffusion radius and double color map filters; (d) zoom-in view with fixed diffusion radius; (e) zoom-in view with adaptive diffusion radius

Visualization Design

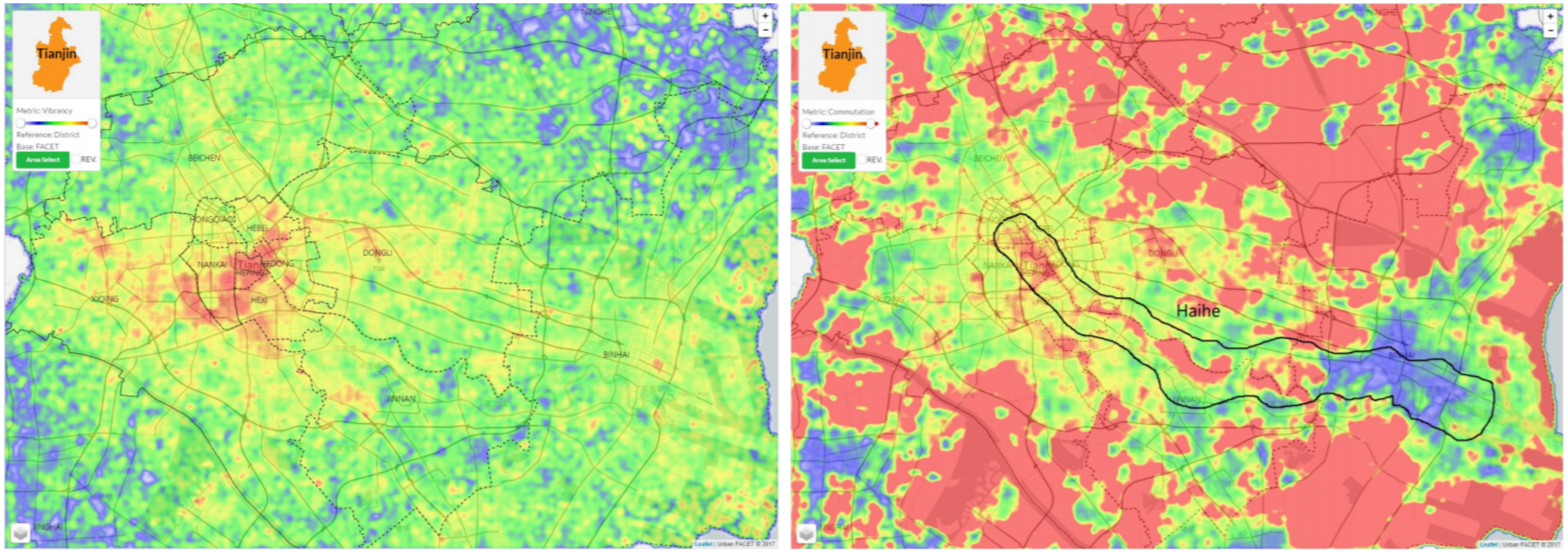


Case Study

Fig. The fluidity metric distribution in Beijing (top)



(a) Distribution (b) Correlation with tourism POIs



(a) Vibrancy (b) Commutation

Fig. The mobility metric distribution in Tianjin (left)

Case Study

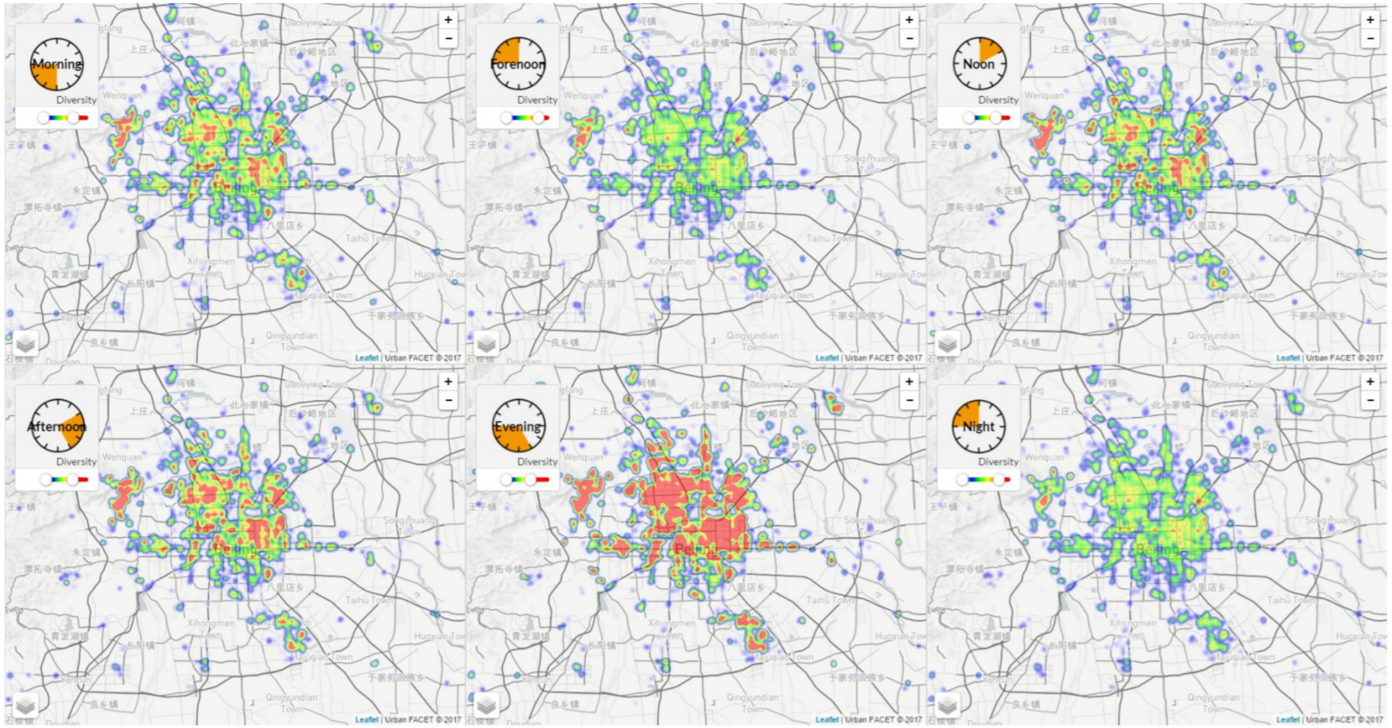


Fig. Temporal Comparison Analysis over different time periods (right)

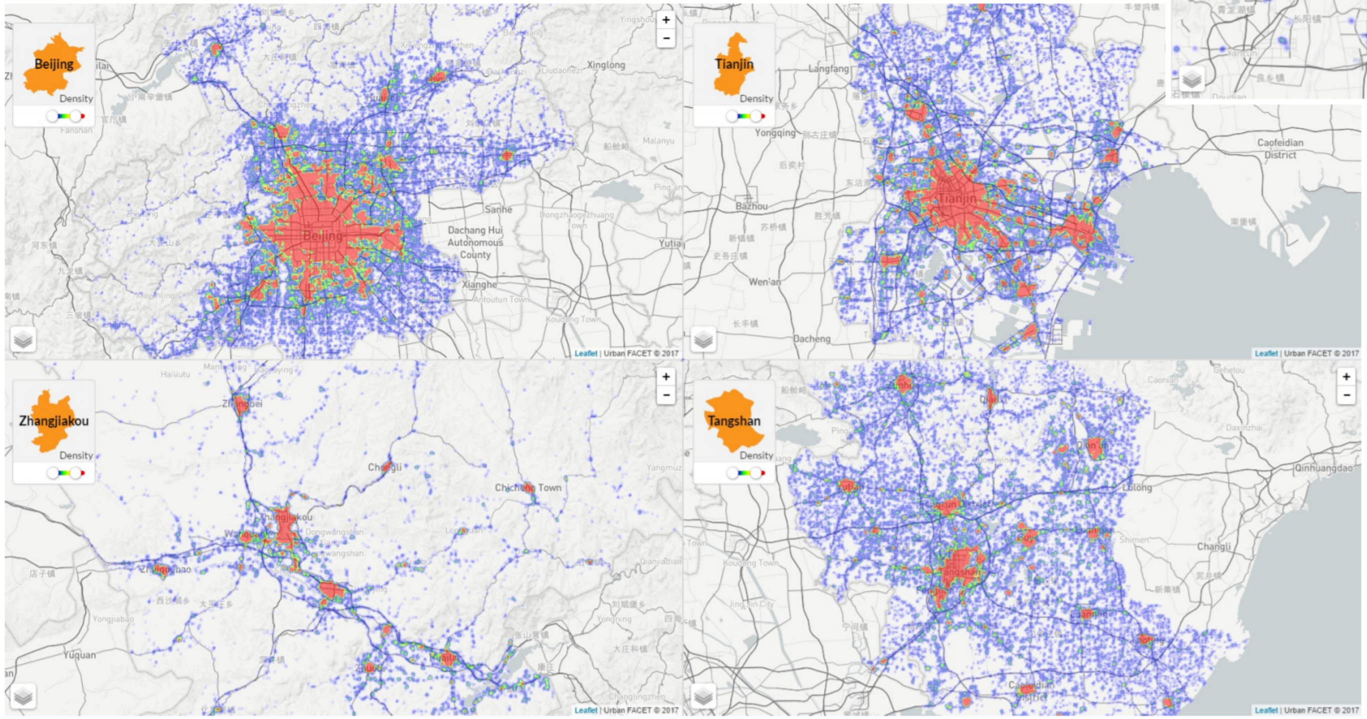


Fig. Spatial Comparison Analysis over different cities (left)

ISCAS 中国科学院大学 University of Chinese Academy of Sciences KENT STATE UNIVERSITY TalkingData

Evaluation - Controlled user experiment

- Task Design: **16 subjects** with 2 groups (mobility metrics & density metric)
- Traning Session: applied in advance
- Task: **8 tasks in 4 groups**
- Data Source: FACE/T randomly
- Result: 4 user-level mobility metrics in most cases **introduce new information** for users

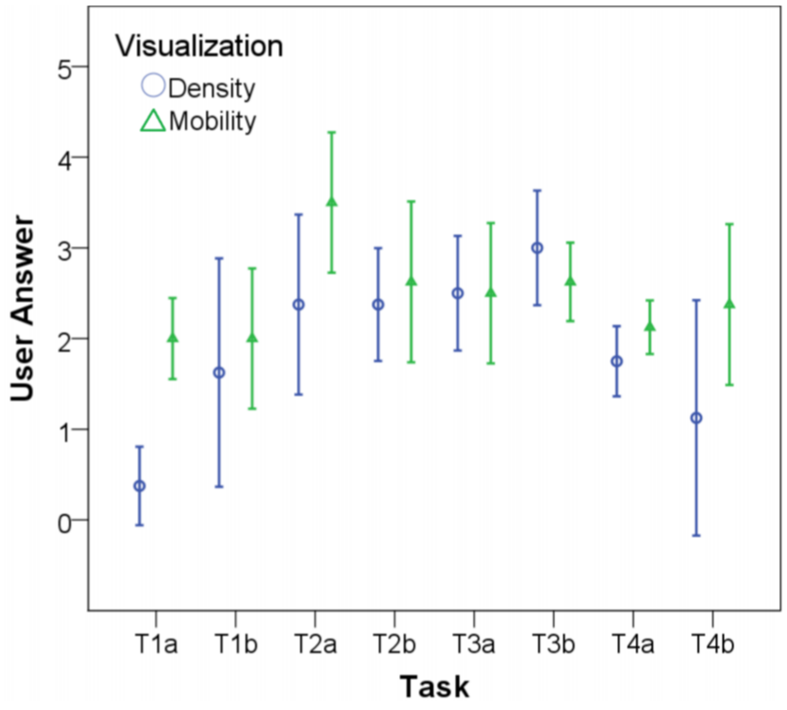


Fig. 13. Distribution of user answers in eight tasks.

Conclusion

Contribution

- Proposed a scalable, grid-based data analytics **pipeline**;
- Introduced a suite of information-theory based **metrics**;
- Developed an integrated visual analytics **system**, namely UrbanFACET.

Future Work

- **Model** fine-grained regions and conduct **analysis** on residents.
- Figure out meaningful **time-based** user mobility **metrics**.
- **Extend** the **user base** of UrbanFACET beyond domain experts.

Thanks!

UrbanFACET: Visually Profiling Cities from Mobile Device Recorded Movement Data of Millions of City Residents

hijiangtao@gmail.com

2017.05