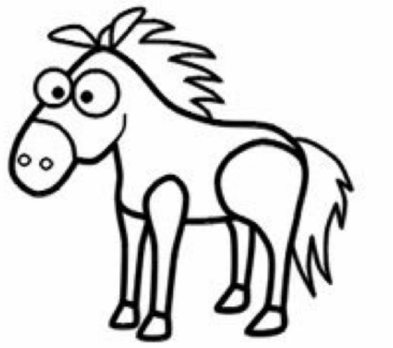
Invited Talk @ Lancaster University

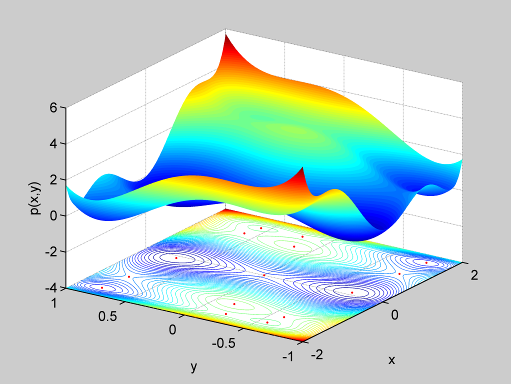
Spatiotemporal Urban Data: Sensing, Analytics, and Applications

Longbiao CHEN   PhD, Assistant Professor

Xiamen University, China

September, 2019





Short Bio

陈 龙彪

博士，助理教授

厦门大学 信息学院

福建省智慧城市感知与计算重点实验室

福建省厦门市厦门大学海韵园行政楼B601办公室

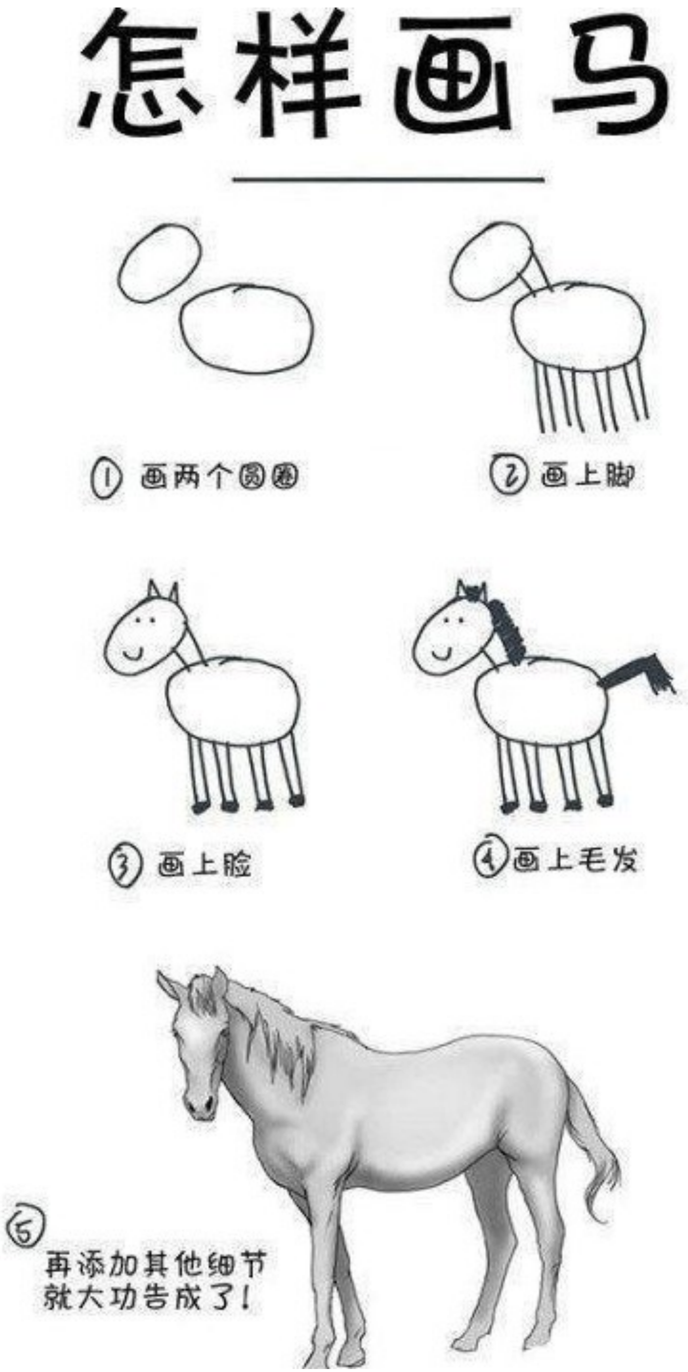
教育经历

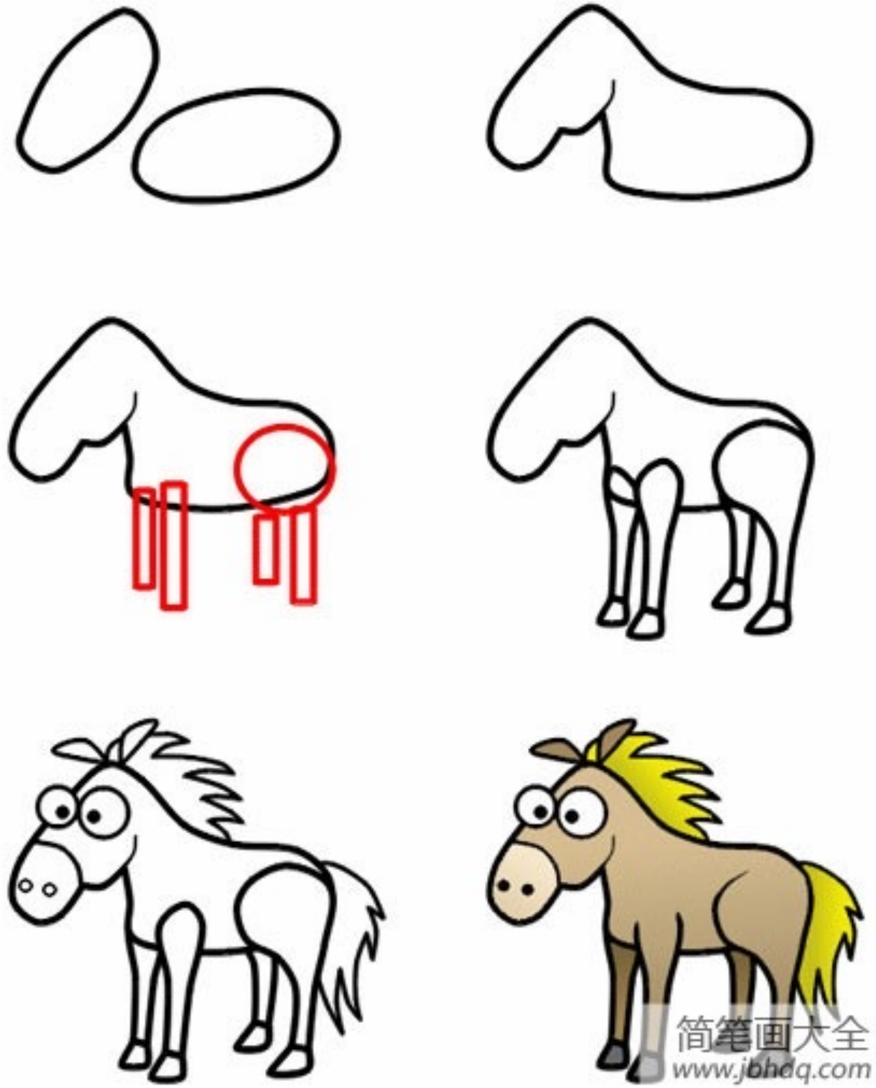
2015–2018	博士学位	法国索邦大学（Sorbonne University）（原巴黎第六大学）
		导师：张大庆（*Fellow, IEEE*）、Thi-Mai-Trang NGUYEN
2010–2016	博士学位	浙江大学 计算机学院CCNT实验室
		导师：潘 纲
2006–2010	学士学位	浙江大学 竺可桢学院（混合班） 计算机科学专业

工作经历

2013–2015	助理研究员	法国科学院国立电信研究院，合作导师：Jérémie Jakubowicz
2009–2010	研究实习生	IBM 上海研发中心

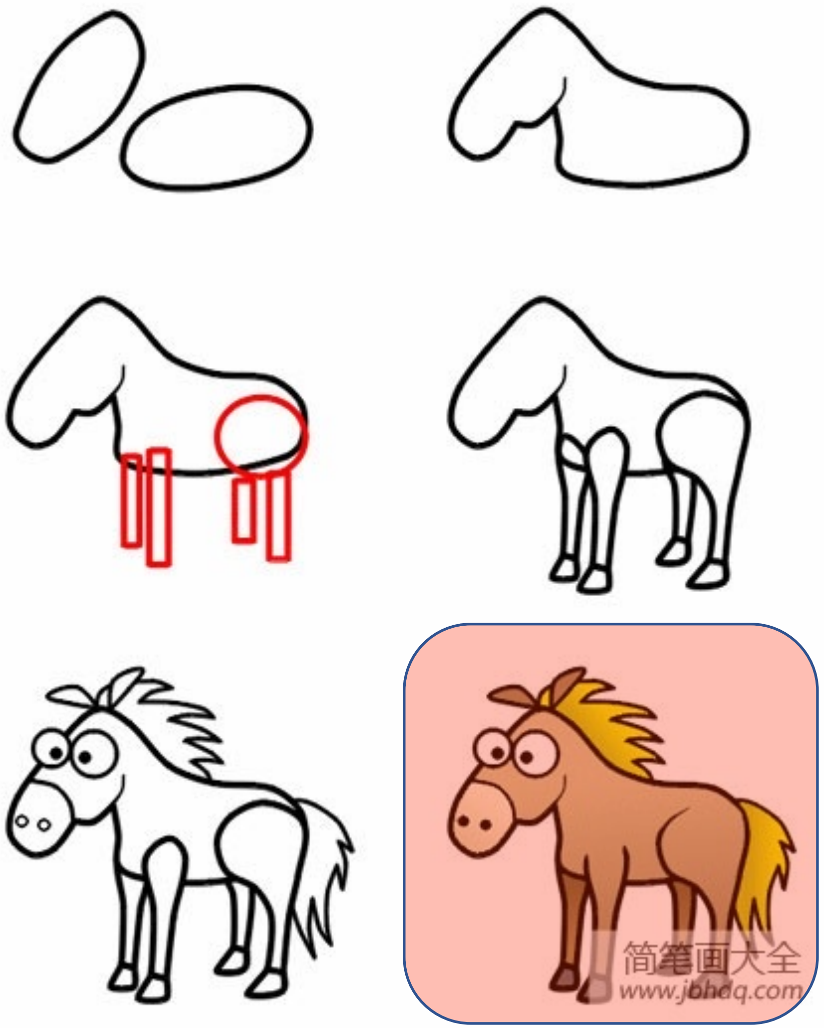
Xiamen, China


怎样画马
① 画两个圆圈
② 画上脚
③ 画上脸
④ 画上毛发
⑤ 再添加其他细节
就大功告成了!


简笔画大全
www.jbhdq.com

A Glance

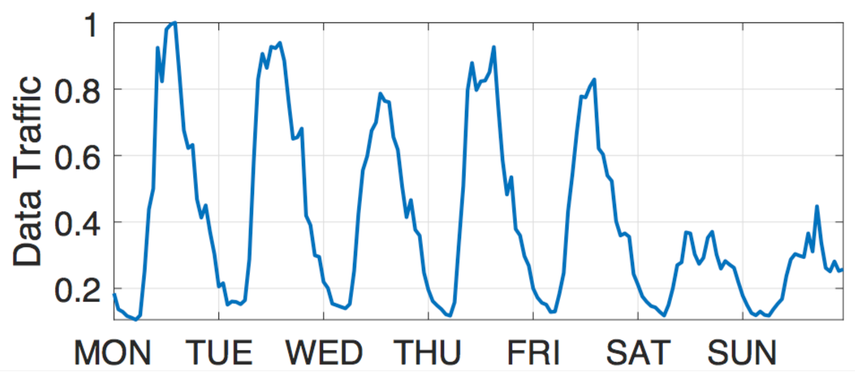
Spatiotemporal Urban Data



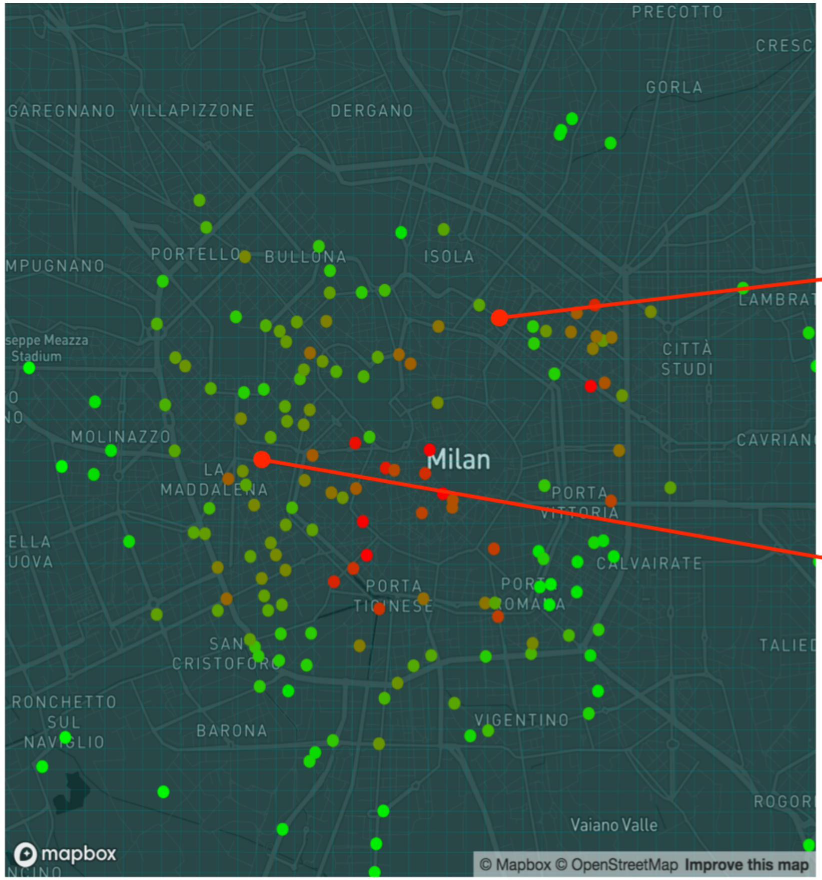
Spatiotemporal Urban Data: A Glance

https://cityos.io/

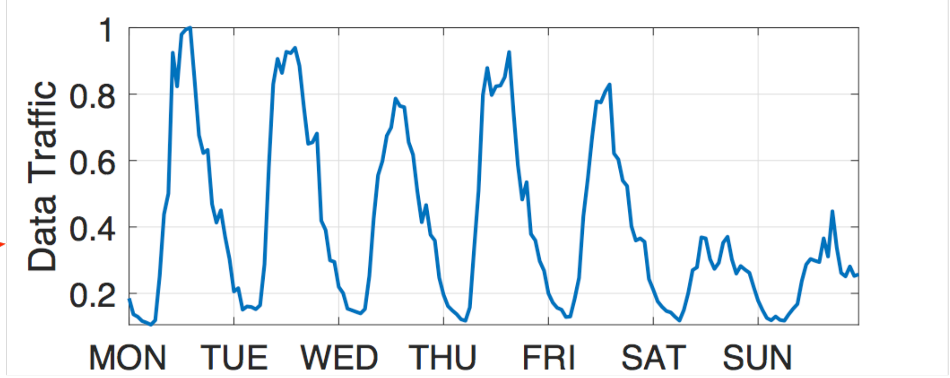
Temporal Urban Data



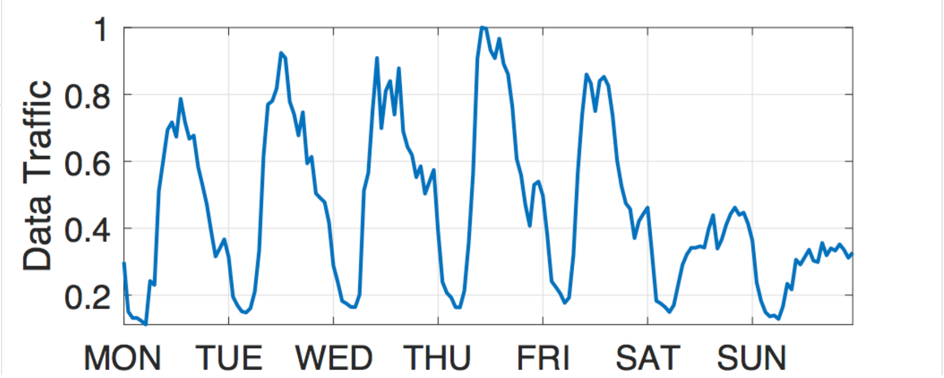
Spatiotemporal Urban Data



(a) Base stations in Milan (11/01/2013 - 12/31/2013)

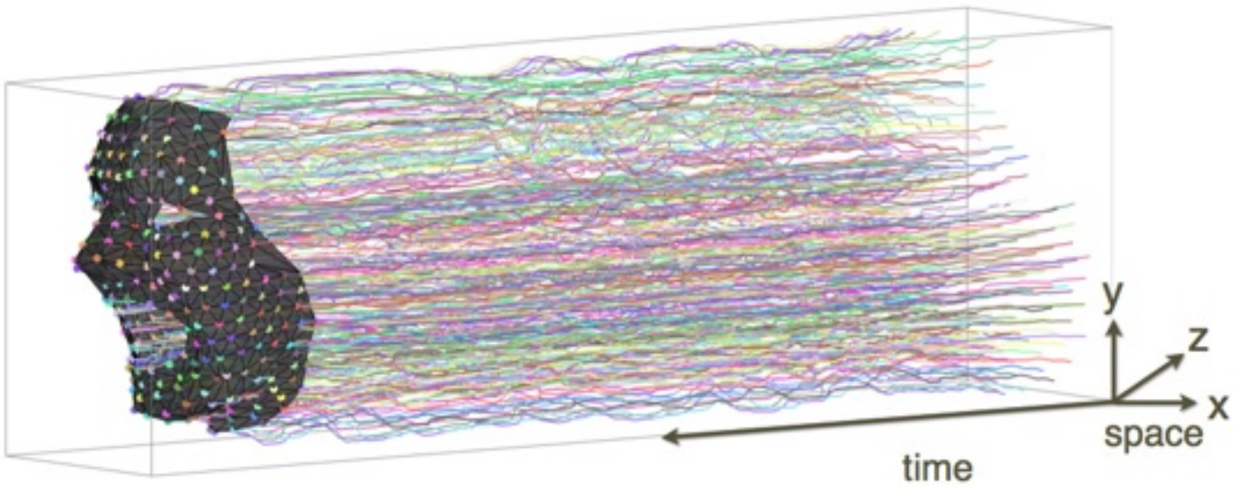


(b) One week of traffic in Centro Direzionale

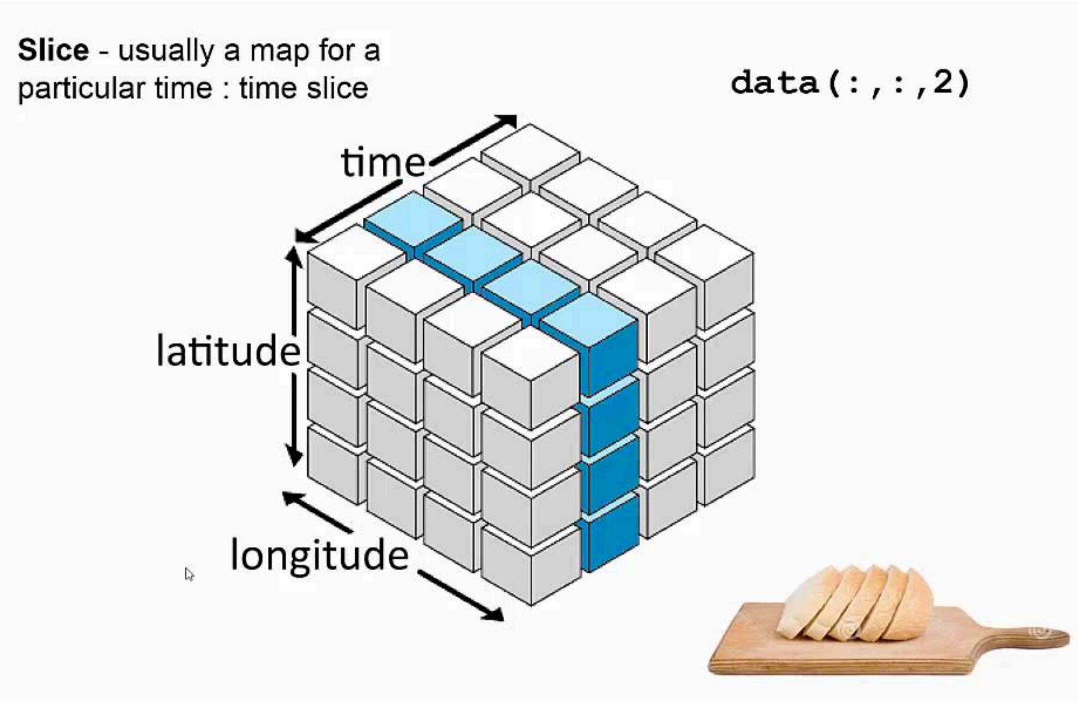


(c) One week of traffic in Piazza Wagner

Spatiotemporal Urban Data: Look and Feel



[Akhter, 2012]



[ENVS3019, ANU Course]

Spatiotemporal Data: Definitions

- Spatiotemporal data models the evolution of spatial objects in time.

[Newell, Temporal GIS, 1992]

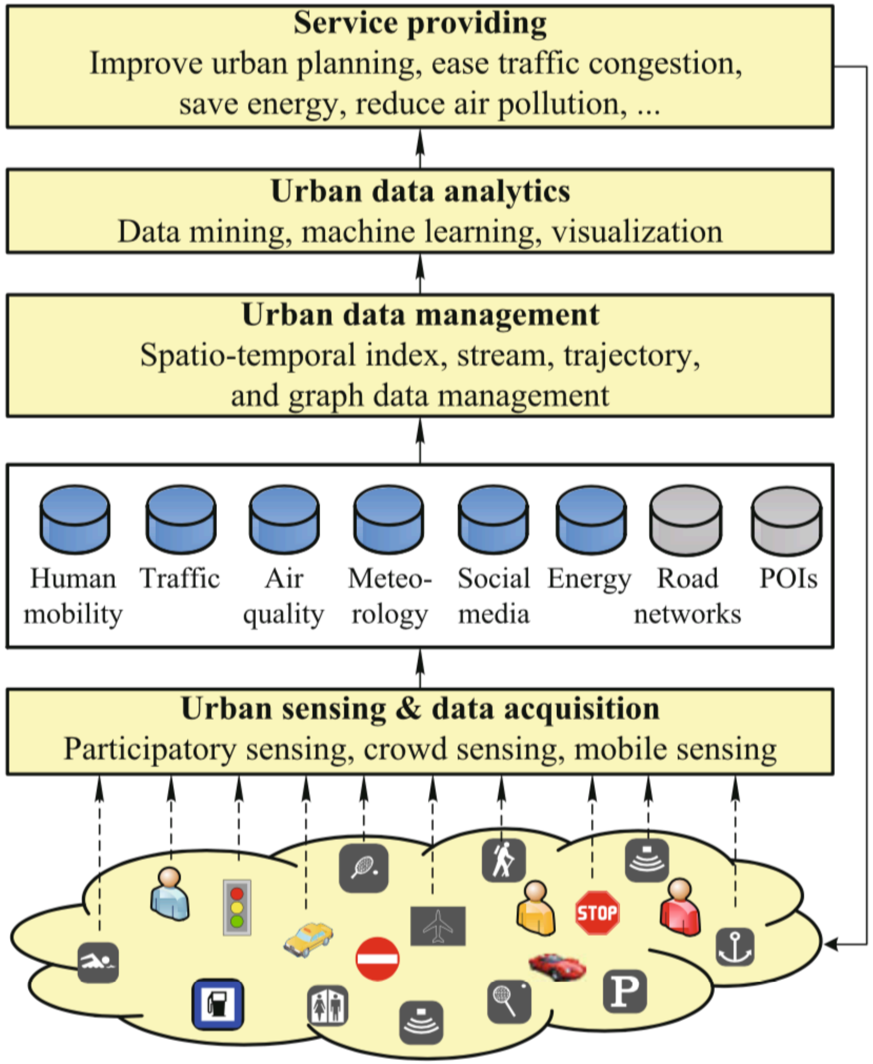
- Spatiotemporal database manages spatiotemporal objects and supports corresponding query functionalities.

[Xiong, Spatiotemporal Database, 2017]

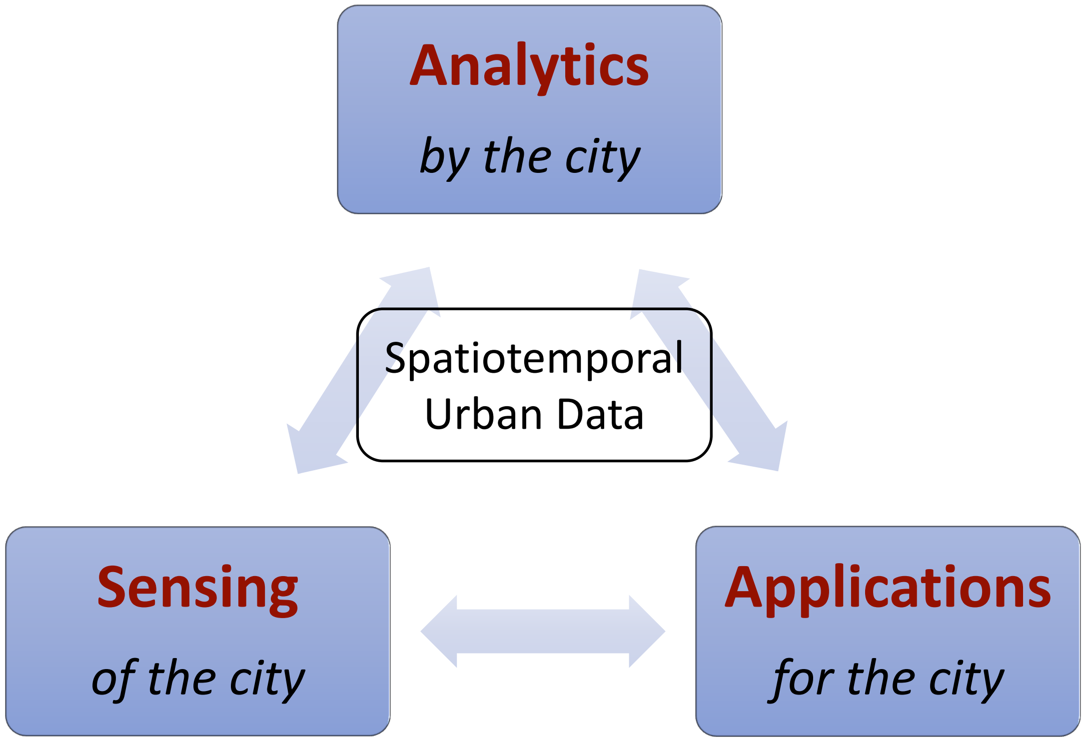
- Spatiotemporal data mining refers to the process of discovering patterns and knowledge from spatiotemporal data.

[Han, Data Mining, 2012]

Spatiotemporal Data and Urban Computing

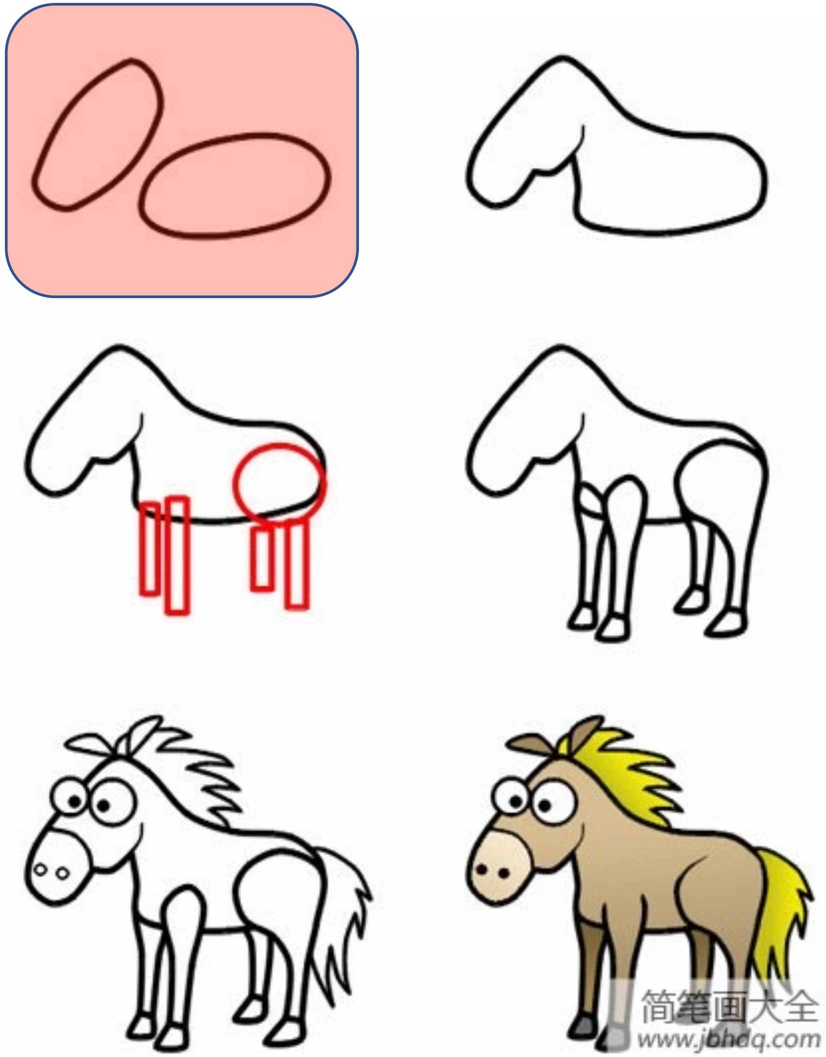


[Yu, Urban Computing, 2016]



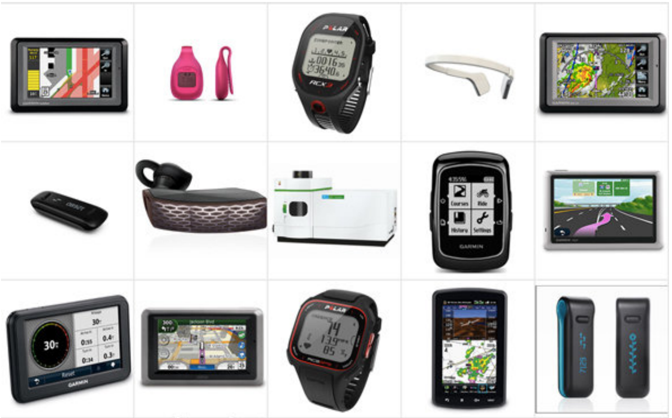
Sensing and Management

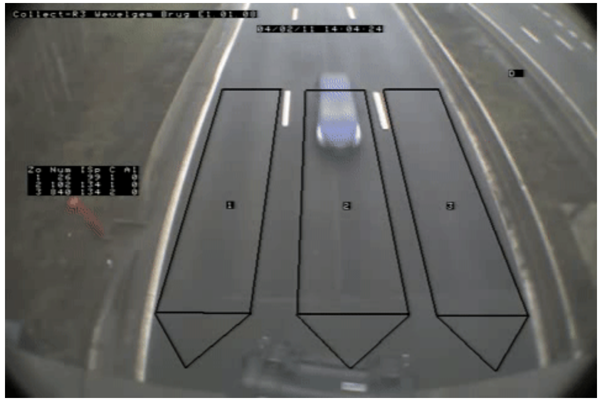
Spatiotemporal Urban Data



Spatiotemporal Sensors

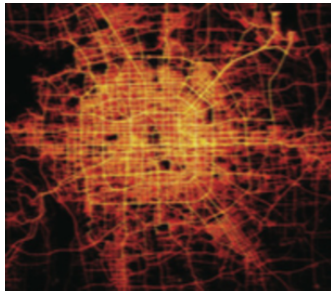
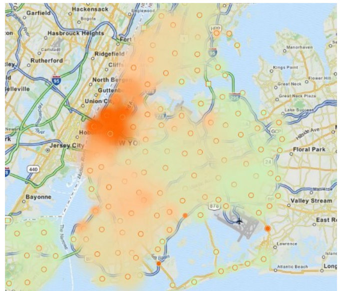
- **Smart devices**
  - Mobile phones, GPS trackers, RFIDs, …
- **Infrastructure**
  - Parking meters, Traffic cameras, License Plate Recognizers (LPR), …
- **Crowdsensing**
  - Human-as-a-Sensor (HaaS), Vehicle-as-a-Sensor (VaaS), …





Data Types

- **Trajectory data**
  - GPS traces: mobile phones, taxicabs, container ships
  - Cell ID sequences: mobile network, public Wi-Fi
- **Origin-Destination data**
  - Ridership: metro, bus, train, flight
  - Pick-up/drop-off records (PDRs): taxicabs, bike-share, car-share
- **Check-in data**
  - Geo-tagged tweets/photos/videos: Twitter, Instagram, Facebook
  - Mobile payment: bank cards, QR codes, NFC devices

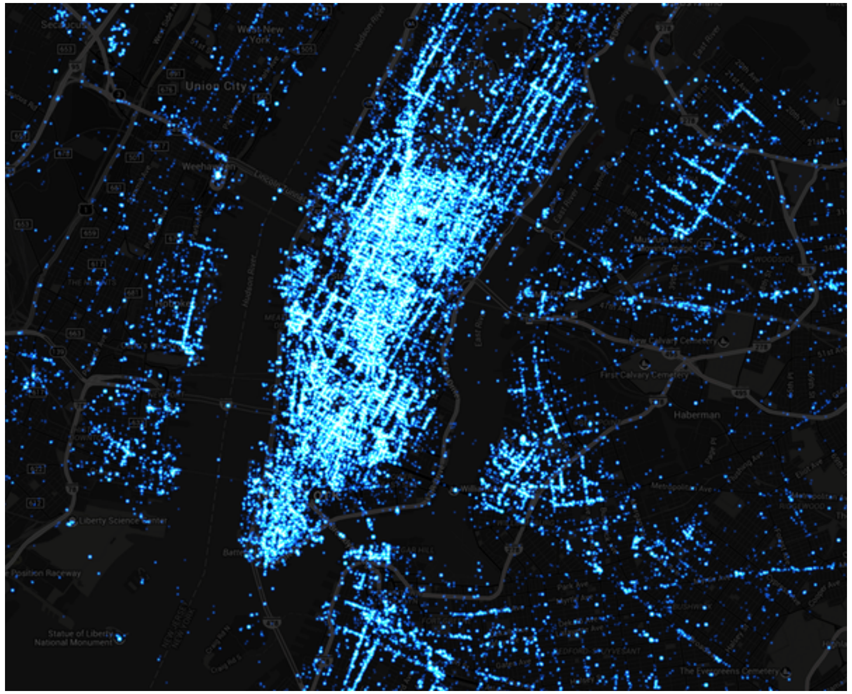



Example: Trajectories in Xiamen

- **Vehicle trajectory data from Xiamen transportation police**
  - 1.2 million vehicles, 3.8 million people
  - 5,486 taxis, 136,406 pick-ups/day

Example: Check-ins in Foursquare

- **Crawled from twitter geo-tweets since 2011**



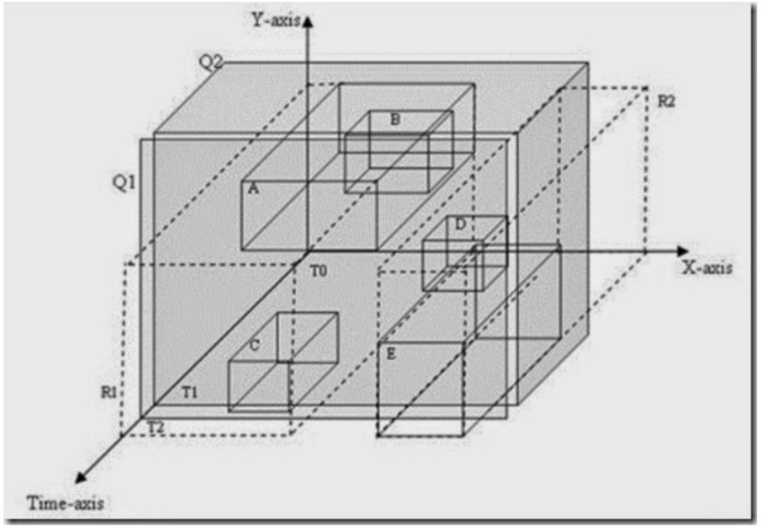
New York City, 2014

Washington D.C., 2012

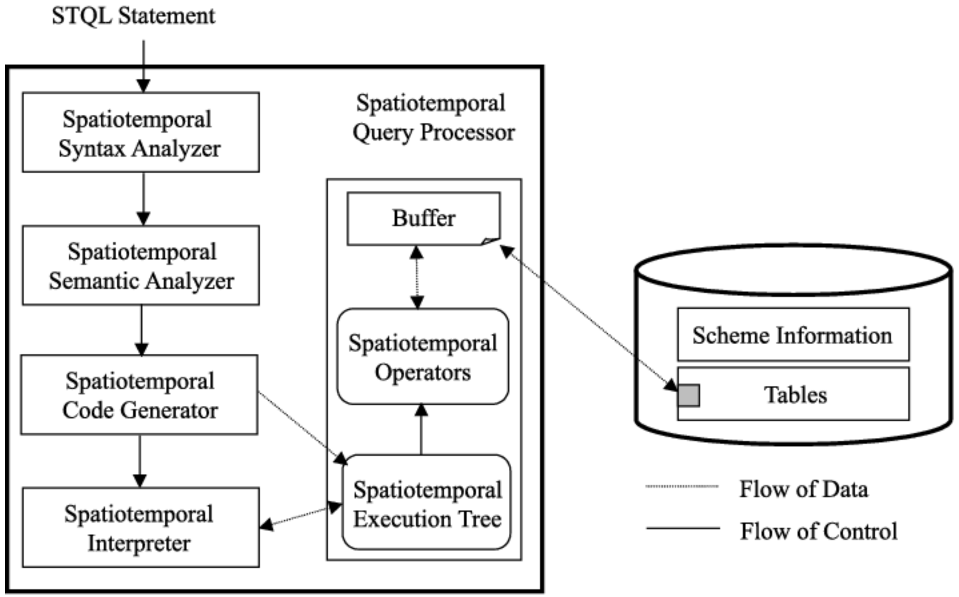
[THMS 2016] [JNCA 2015]

Data Management

- **Spatiotemporal Database**
  - Distributed storage
  - Query process
  - User interface
  - …



[Xiong, 2017]

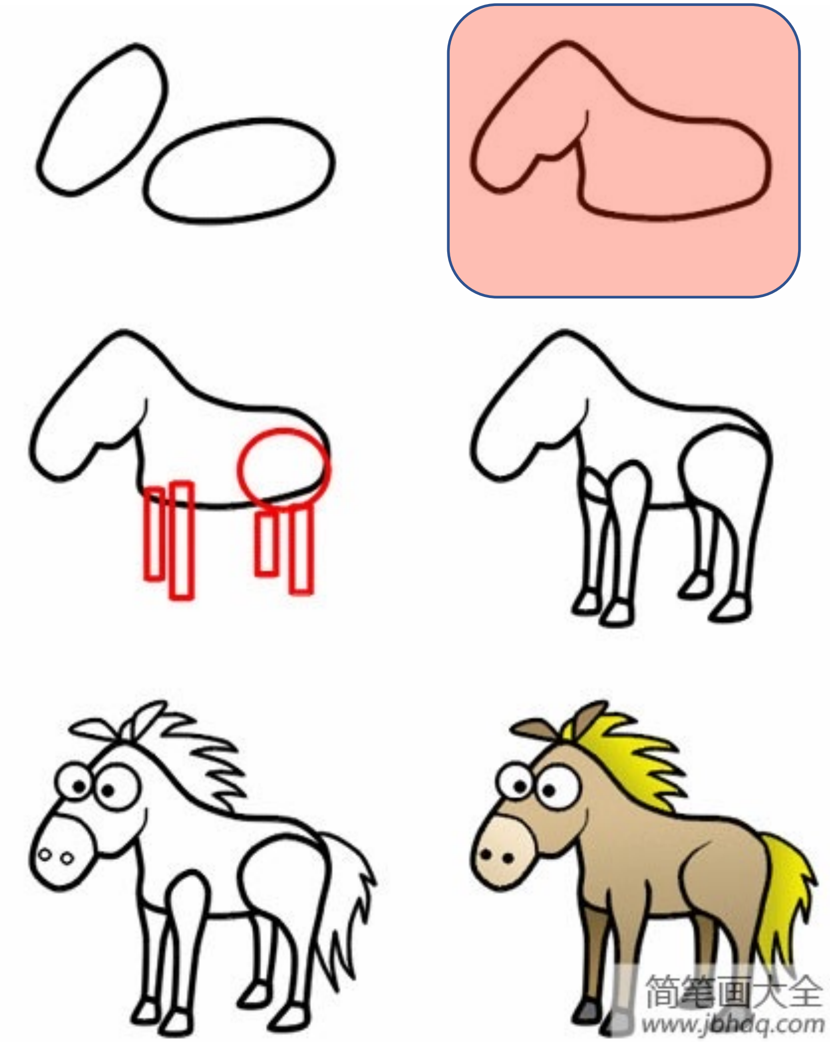


[Kim, 2002]

Read this Book: Xiong, Spatiotemporal Database, 2017

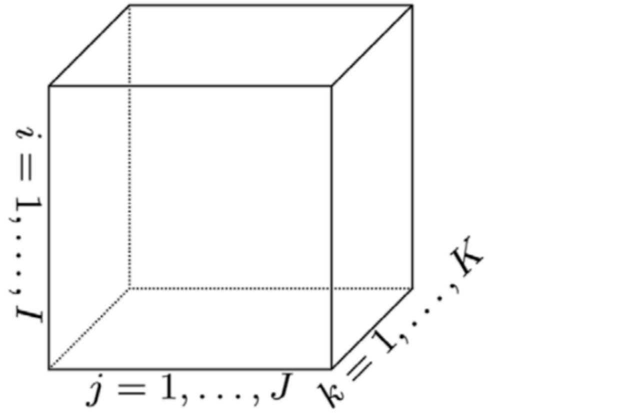
Representation and Modeling

Spatiotemporal Urban Data

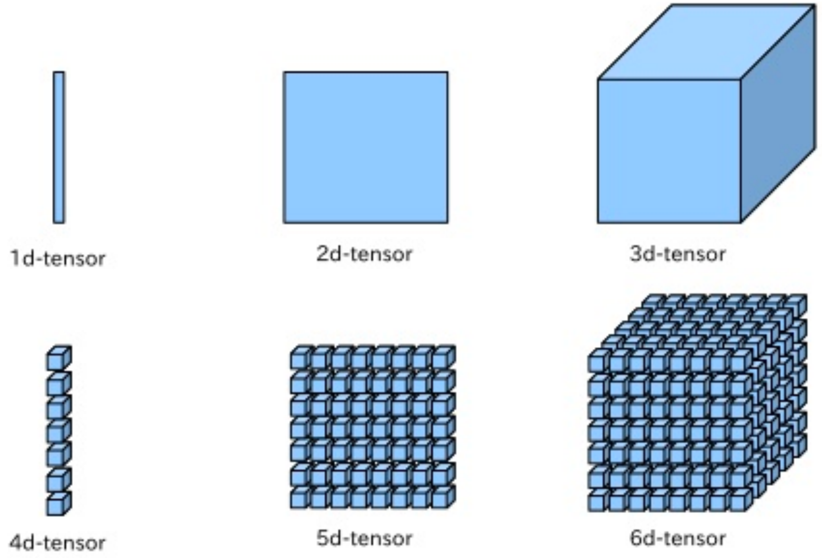


A Unified Model: Tensor

- **Spatiotemporal Tensor**
  - Spatial: latitude and longitude, regional
  - Temporal: time of day, day of week
  - Categorial: user, activity



A third-order tensor: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$.



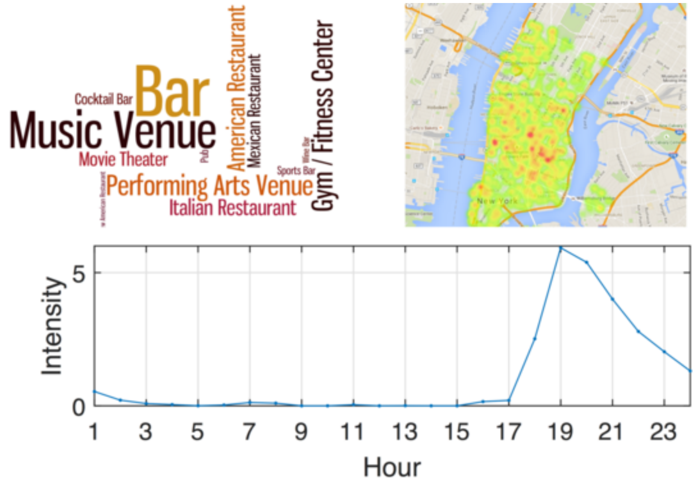
[Zhang, 2018]

Type	Scalar	Vector	Matrix	Tensor
Definition	a single number	an array of numbers	2-D array of numbers	k-D array of numbers
Notation	x	$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$	$\boldsymbol{X} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,n} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ X_{m,1} & X_{m,2} & \cdots & X_{m,n} \end{bmatrix}$	$\mathbf{X}$ $X_{i,j,k}$
Example	1.333	$\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ 9 \end{bmatrix}$	$\boldsymbol{X} = \begin{bmatrix} 1 & 2 & \cdots & 4 \\ 5 & 6 & \cdots & 8 \\ \vdots & \vdots & \vdots & \vdots \\ 13 & 14 & \cdots & 16 \end{bmatrix}$	$\boldsymbol{x} = \left[\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \\ 70 & 80 & 90 \end{bmatrix} \begin{bmatrix} 100 & 200 & 300 \\ 400 & 500 & 600 \\ 700 & 800 & 900 \end{bmatrix} \right]$

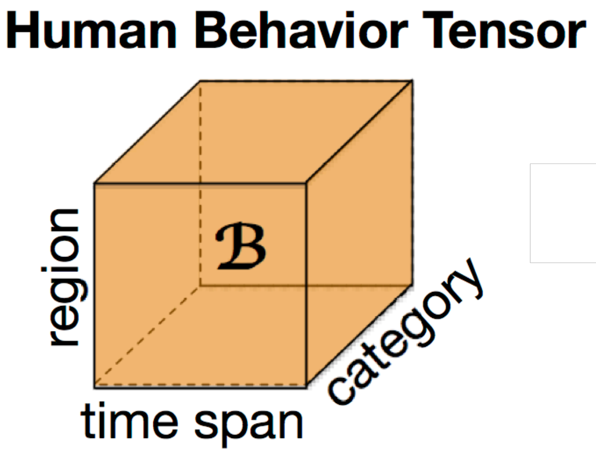
[Knoldus, 2018]

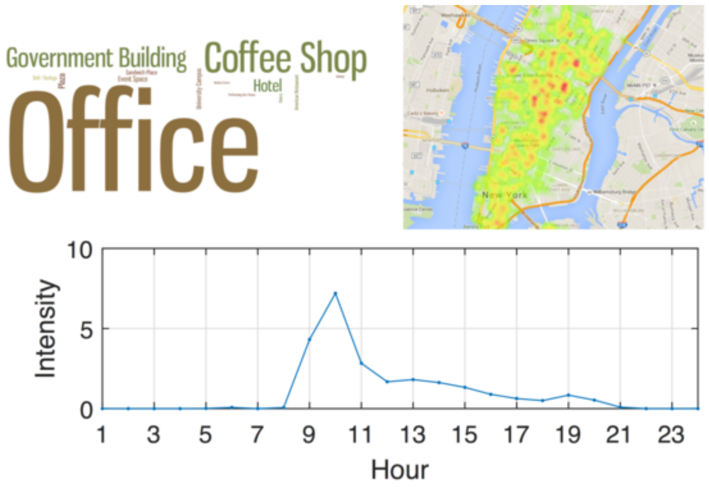
A Unified Model: Tensor

- **Example: Human Behavior Tensor**
  - Spatial: regions
  - Temporal: time spans
  - Categorial: activity categories



Category #2: Nightlife



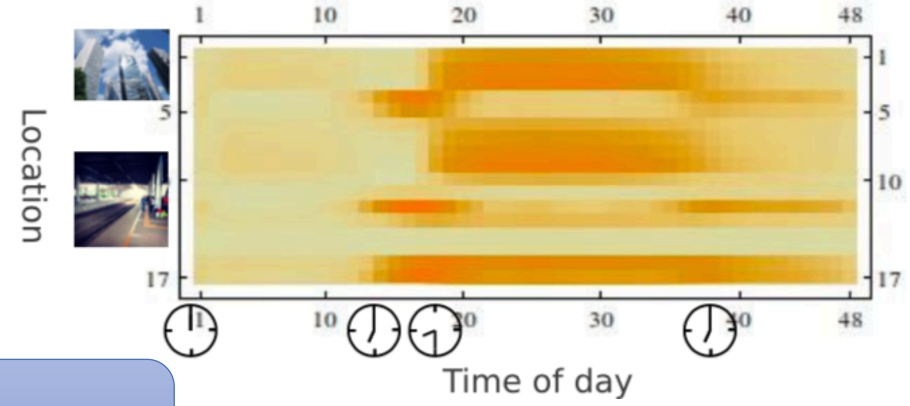


Category #1: Working

[Chen, 2016]

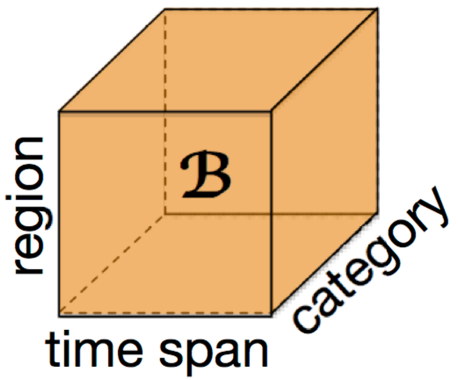
A Unified Model

- **How to determine tensor dimensions?**



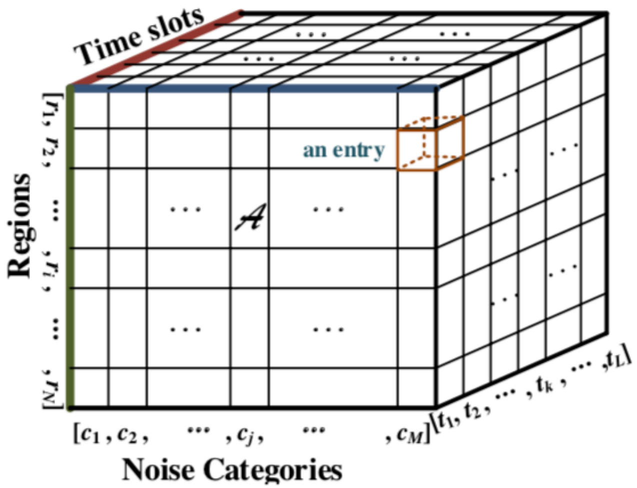
[Fan, 2014]

Human Behavior Tensor

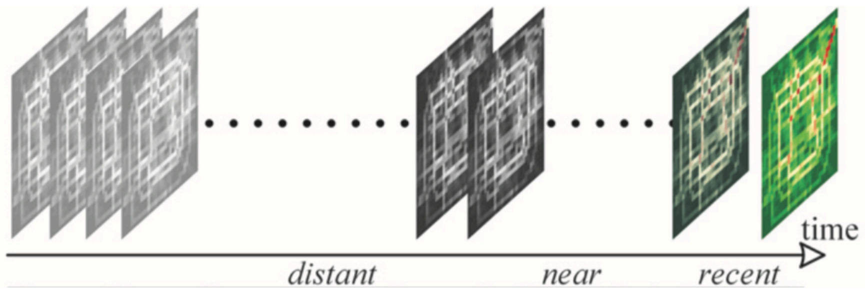


[THMS, 2016]

Just Follow Your Heart



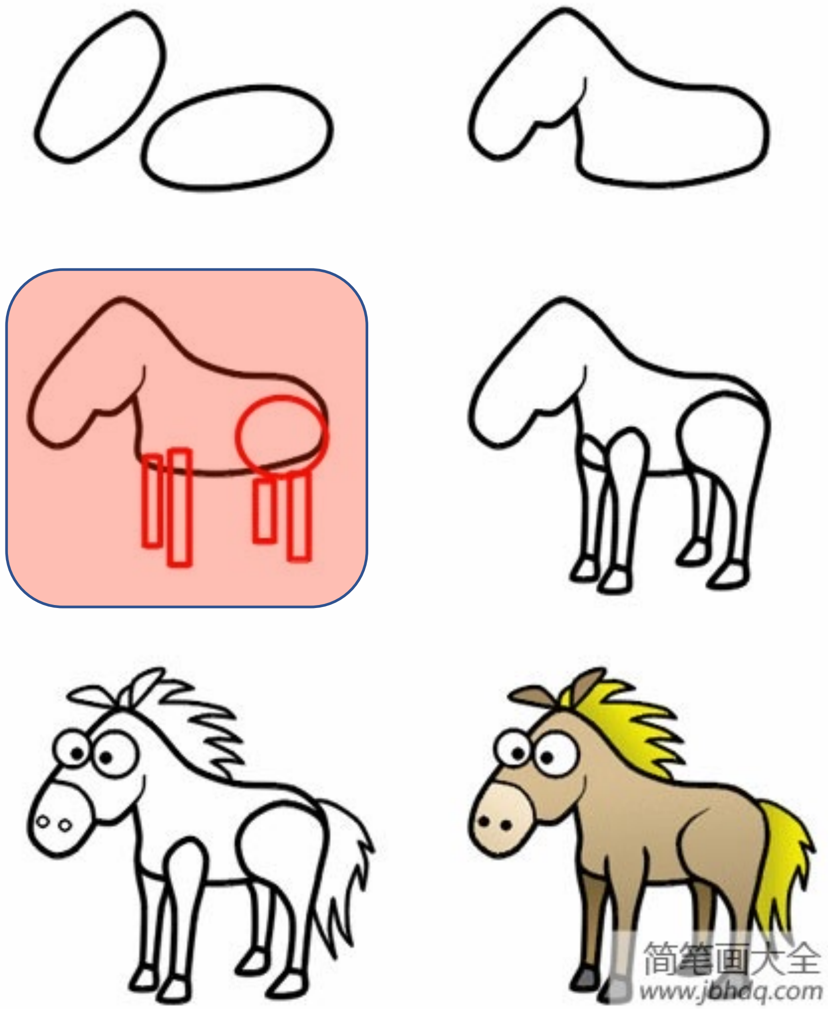
[Zheng, 2014]



[Zheng, 2017]

Aggregation and Clustering

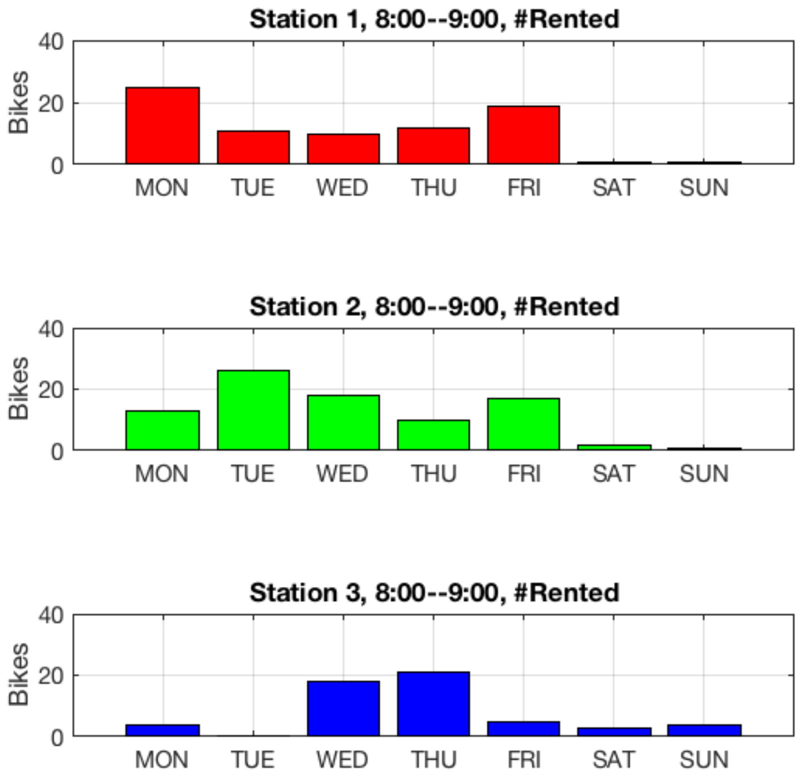
Spatiotemporal Urban Data



Aggregation and Clustering

- **Raw data**
  - Spatially sparse
  - Temporally fluctuating

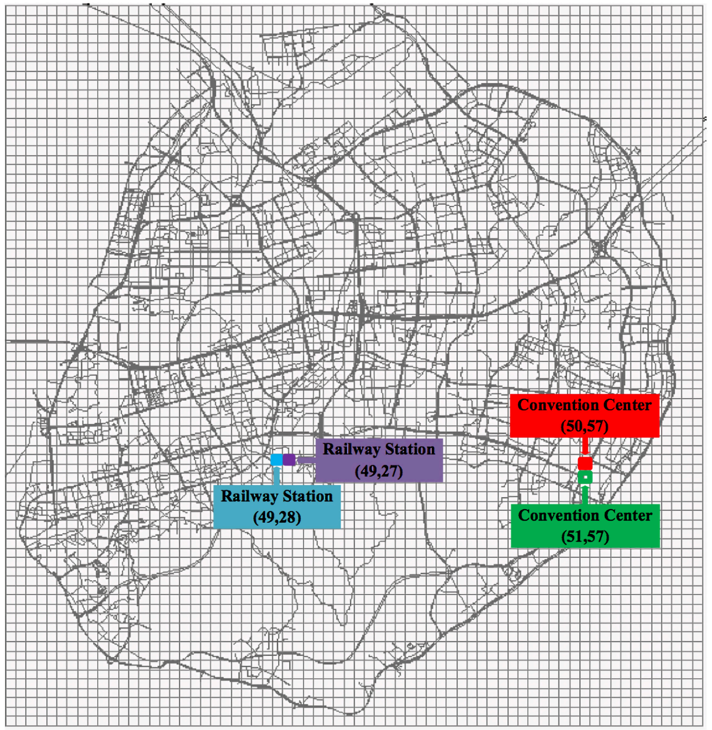
[Pan, 2013]

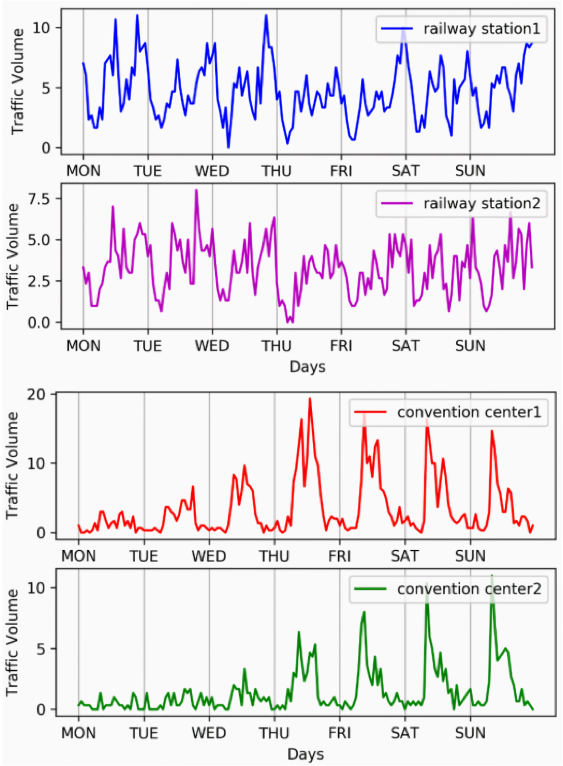


[UbiComp, 2015]

Aggregation and Clustering

- **Aggregation**
  - Spatial grids
  - Temporal profiles



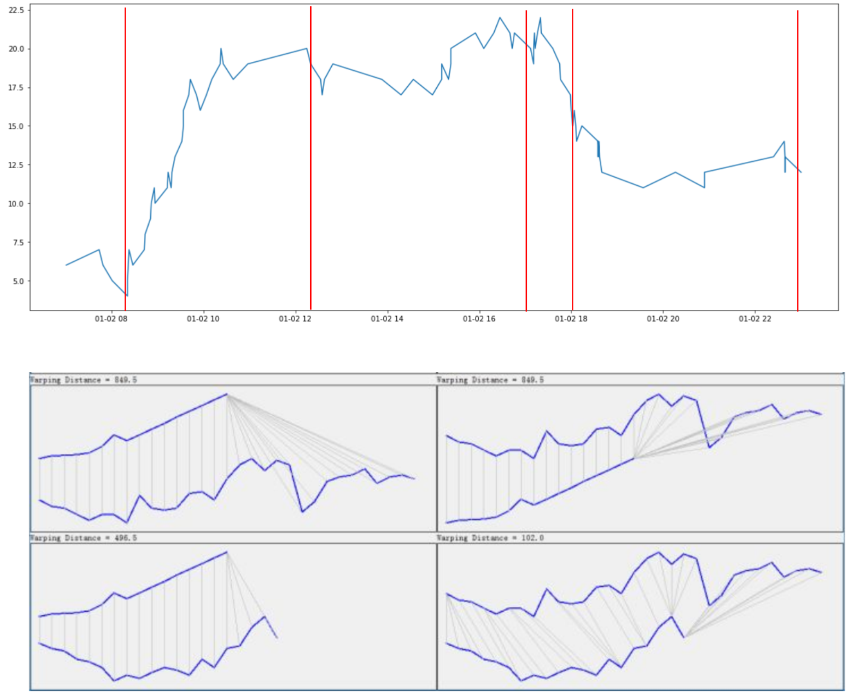


[FCS, 2018]

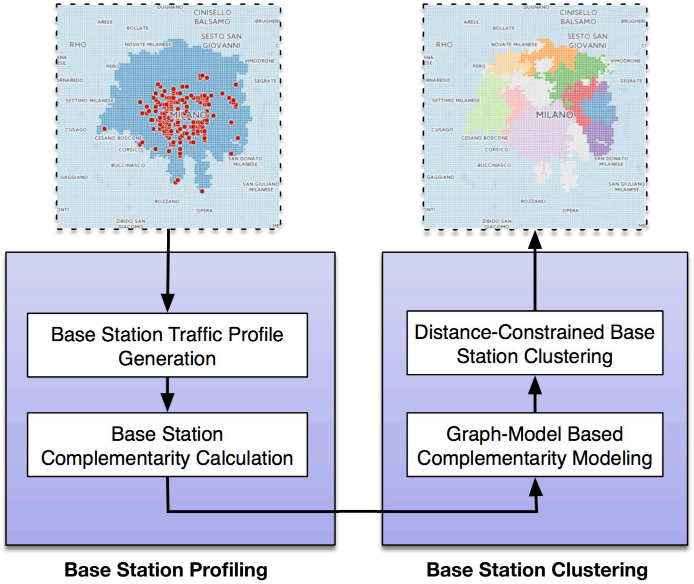
Aggregation and Clustering

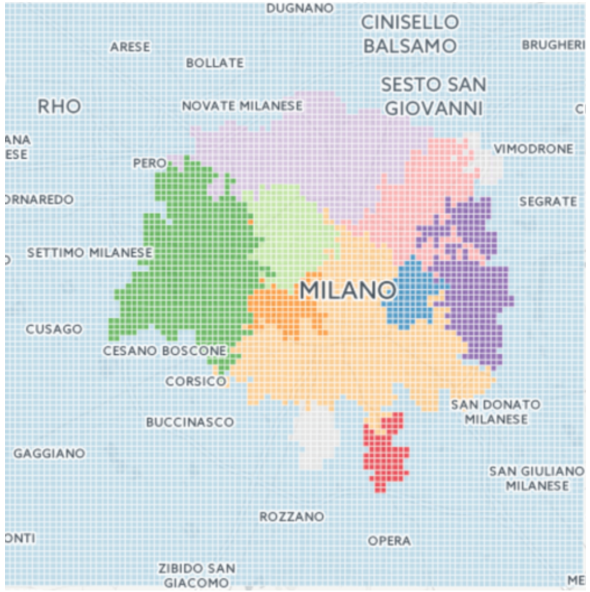
- **Clustering**
  - Spatial: density-based (DBSCAN), distance-based (K-Means)…
  - Temporal: trend extraction (TreNet), dynamic time warping (DTW)…

[Didi, 2018]



[UIC, 2019]



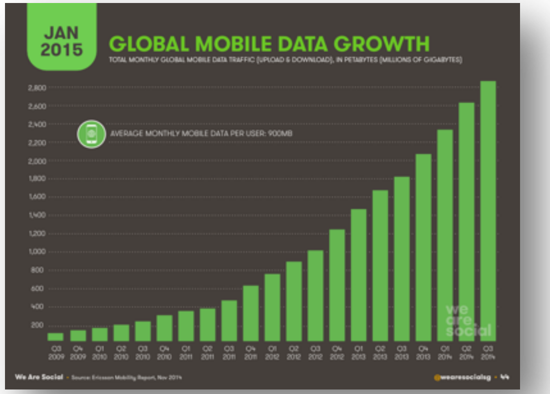


Aggregation and Clustering: Application #1

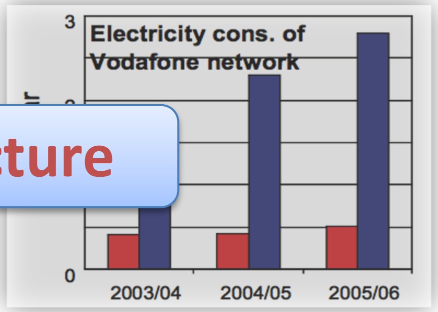
Complementary Traffic Clustering for Energy-Efficient Cloud Radio Networks (JNCA 2018)

Background

- **Surging mobile data traffic**
  - smart phones, IoT devices
  - 18 fold over 5 years [Cisco, 2016]



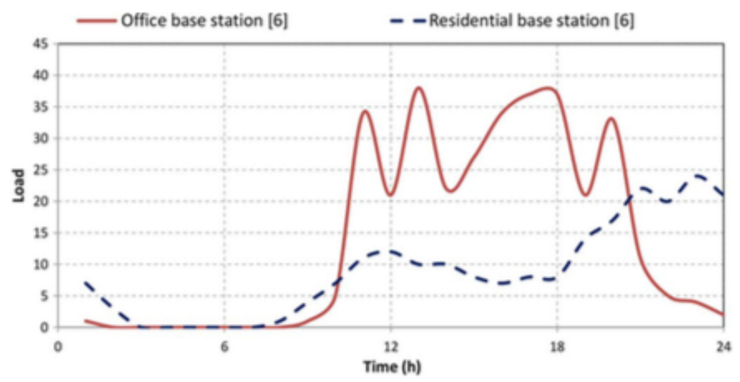
- **more base stations → higher energy consumption**
  - The total energy consumed by cellular networks … takes up more than 3 percent of the worldwide electric energy consumption nowadays. [Li, IEEE Wireless Comm., 2011]



Energy-efficient mobile network architecture

Traffic Demand Patterns

- **Non-uniform**
  - Spatial: office vs. residential areas
  - Temporal: peak hours vs. off-peak hours

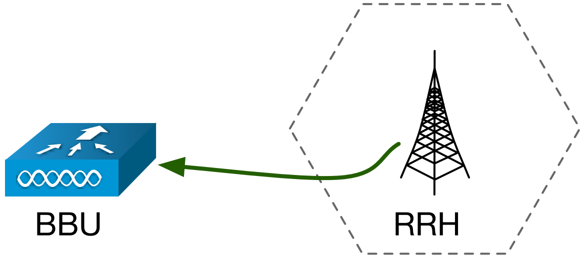


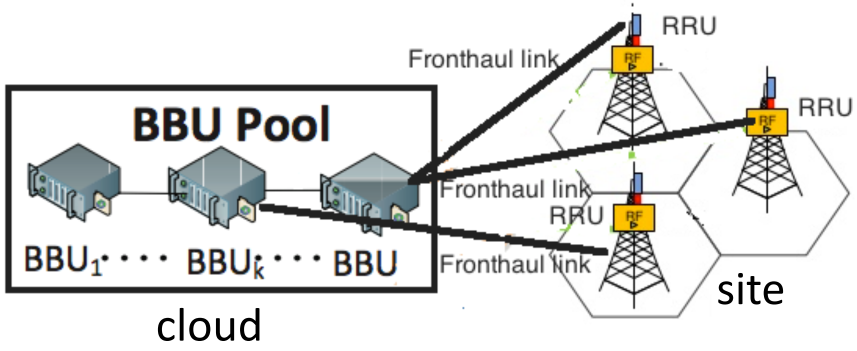
- **Problems**
  - over-provision: cover peak traffic volume
  - low-utilization: resources wasted in off-peak hours

- **Cognitive Cellular Networks (CCNs) [Thomas 2007] [Fiore 2016]**
  - Demand-responsive data processing capacity
  - Optimized deployment cost and energy consumption

C-RAN

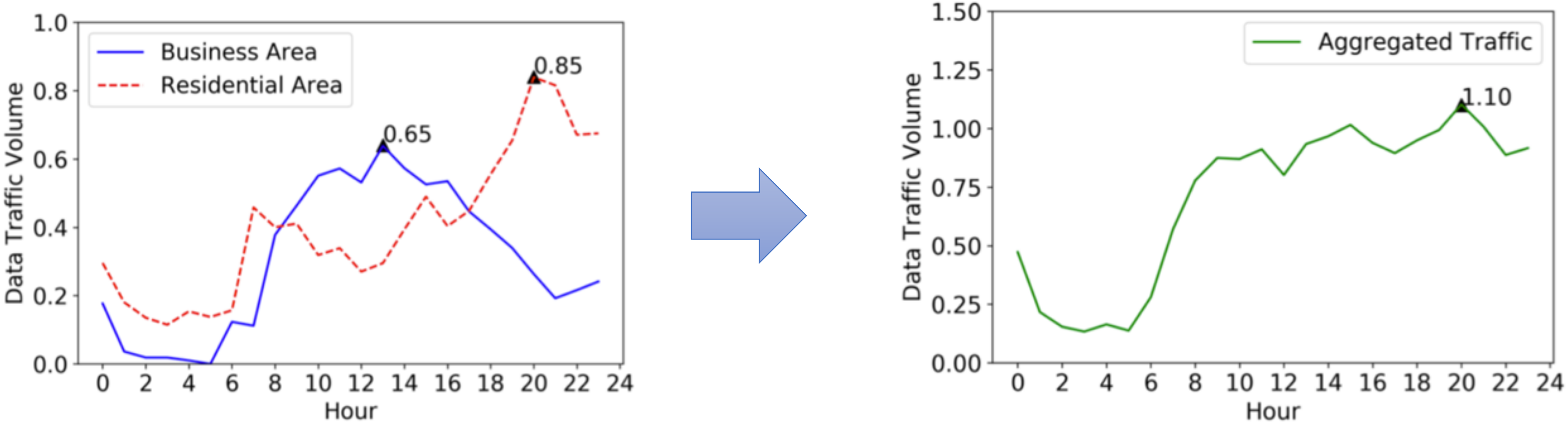
- **Cloud Radio Access Network (C-RAN)**
  - Base station → RRH + BBU
  - Remote Radio Head (RRH): light-weight frontend
  - Baseband Unit (BBU): centralized pool
  - Connection: optical fiber





Benefits of C-RAN: energy-efficient

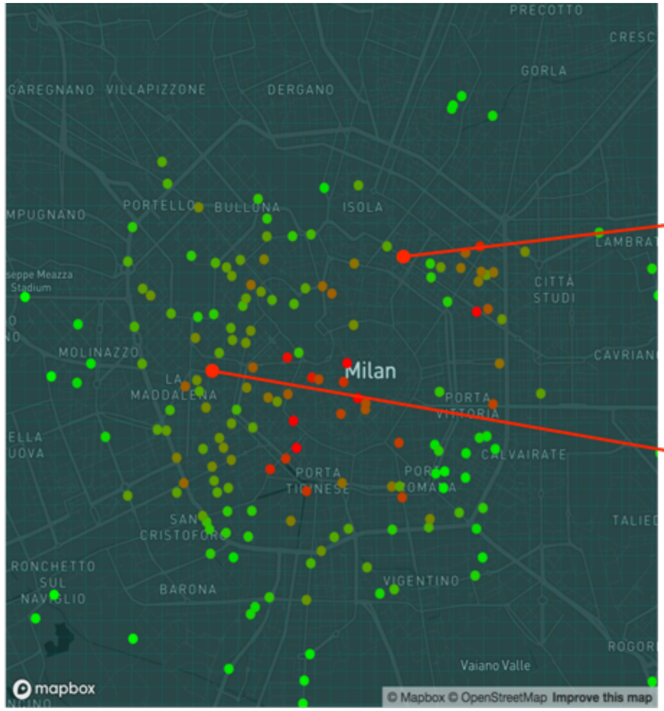
- **Cluster complementary RRHs to a BBU pool**
  - Non-overlapping peaks → shared maximum capacity
  - Stable traffic → increased utilization rate



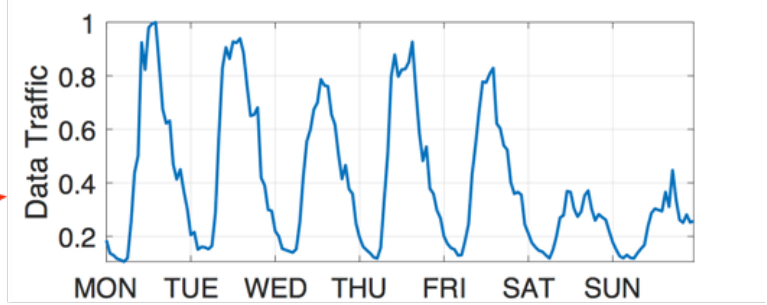
Mapping two base stations to a BBU pool

Challenge #1

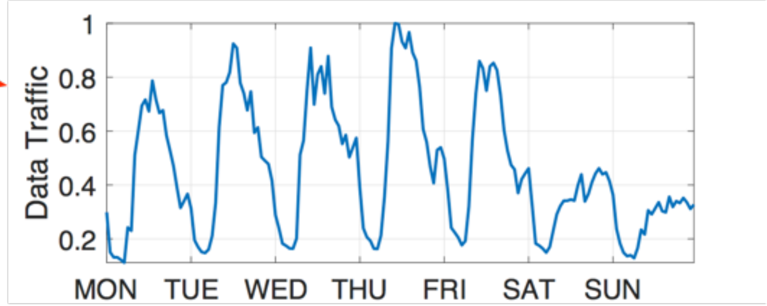
- **How to characterize RRH traffic pattern?**
  - spatial: residential/business areas, …
  - temporal: weekday/weekends, am/pm, …



(a) Base stations in Milan (11/01/2013 - 12/31/2013)



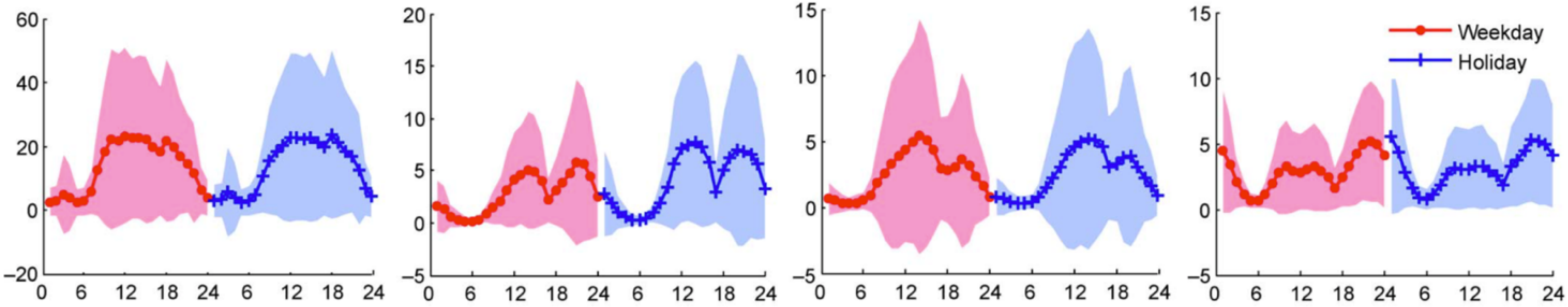
(b) One week of traffic in Centro Direzionale



(c) One week of traffic in Piazza Wagner

RRH Traffic Profiling

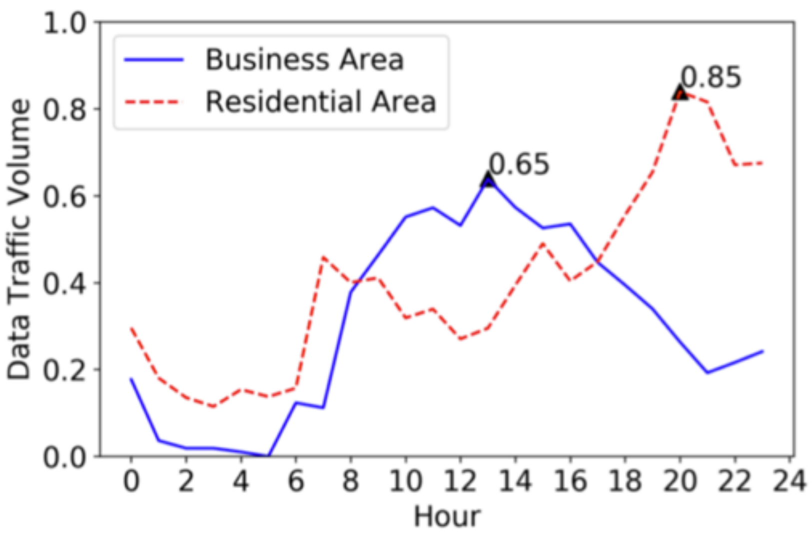
- **RRH traffic profiling**
  - A typical weekday profile: $\mathbf{f}_w(s_i) = [u_1, u_2, \ldots, u_{24}]$
  - A typical weekend profile: $\mathbf{f}_n(s_i) = [v_1, v_2, \ldots, v_{24}]$
  - Concatenated traffic profile: $\mathbf{f}(s_i) = [\mathbf{f}_w(s_i), \mathbf{f}_n(s_i)]$

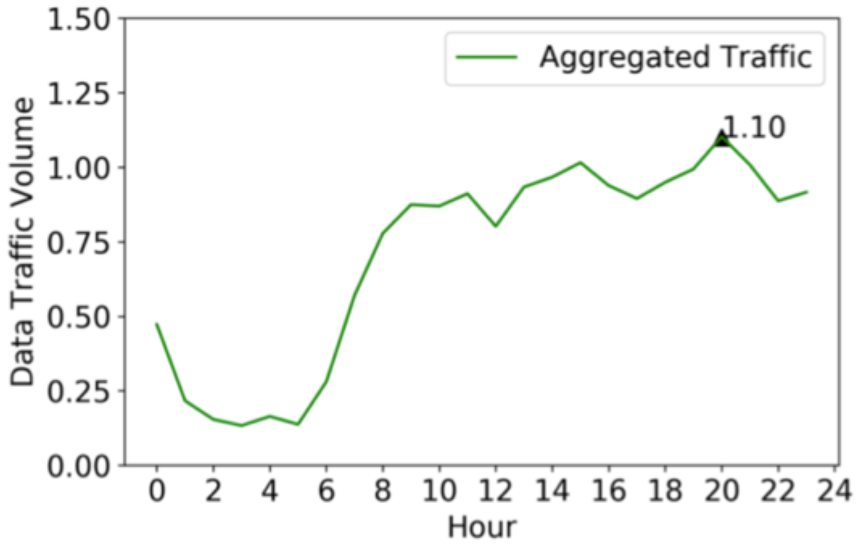


(a) (b) (c) (d)

Challenge #2

- **How to measure complementarity of RRHs?**
  - non-overlapping peaks: temporal distribution
  - stable traffic: capacity utilization





RRH complementarity measurement

- **Peak distribution:**
  - Extract peak-times of members:
  - Calculate the Shannon entropy:

$$T(s_i) = \{t_{i_1}, t_{i_2}, \dots, t_{i_m}\}, \quad 1 \le i_m \le 24$$

$$H(S) = -\sum_{k=1}^{K} p_k \log p_k$$

- **Capacity utilization:**

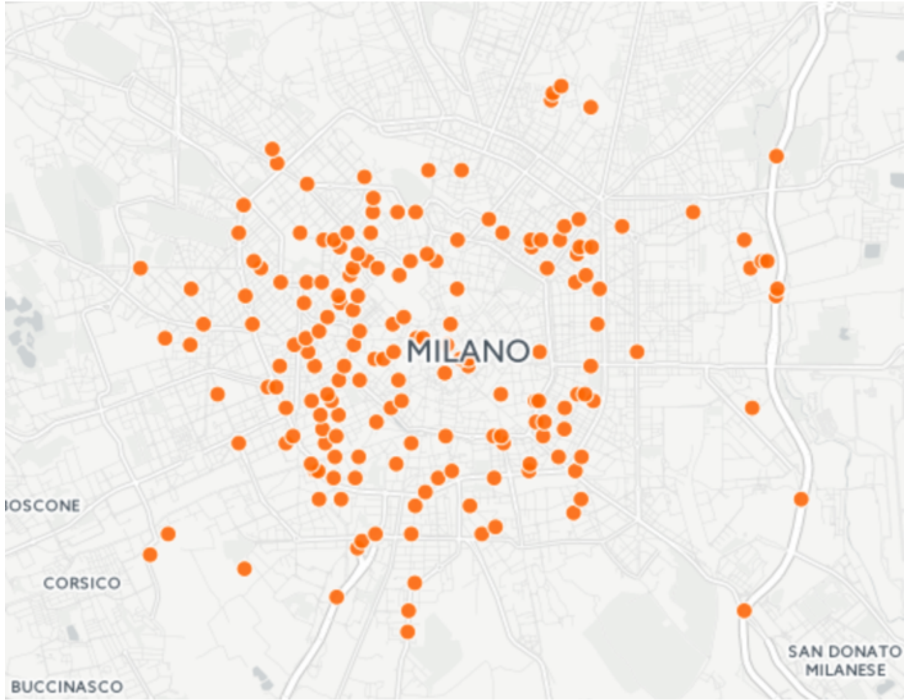
$$U(S) = \frac{\text{mean } \mathbf{f}(S)}{\max\ \mathbf{f}(S)}$$

- **Final measurement:**

$$M(S) = U(S) * H(S) = -\frac{\text{mean } \mathbf{f}(S)}{\max\ \mathbf{f}(S)} \sum_{k=1}^{K} p_k \log p_k$$

Challenge #3

- **How to cluster complementary RRHs to BBU pools?**
  - Various clustering schemes
  - Distance constraints: base station location, pool location



Base station positions in Milan, Italy

Complementary RRH Clustering

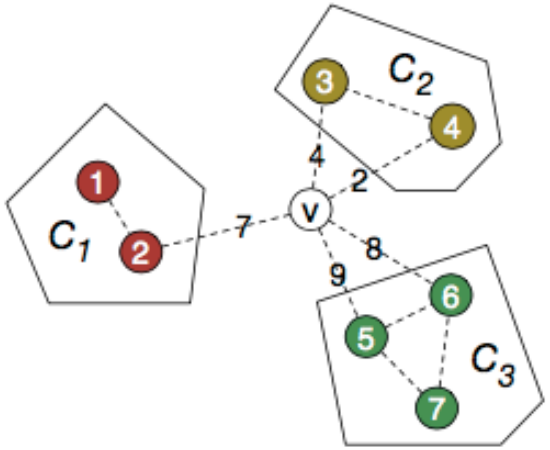
- **Graph model: node partitioning**

Problem: Given graph $G = (V, E)$, we first define a set of clusters $\mathbb{P} = \{C_1, \ldots, C_K\}$, where

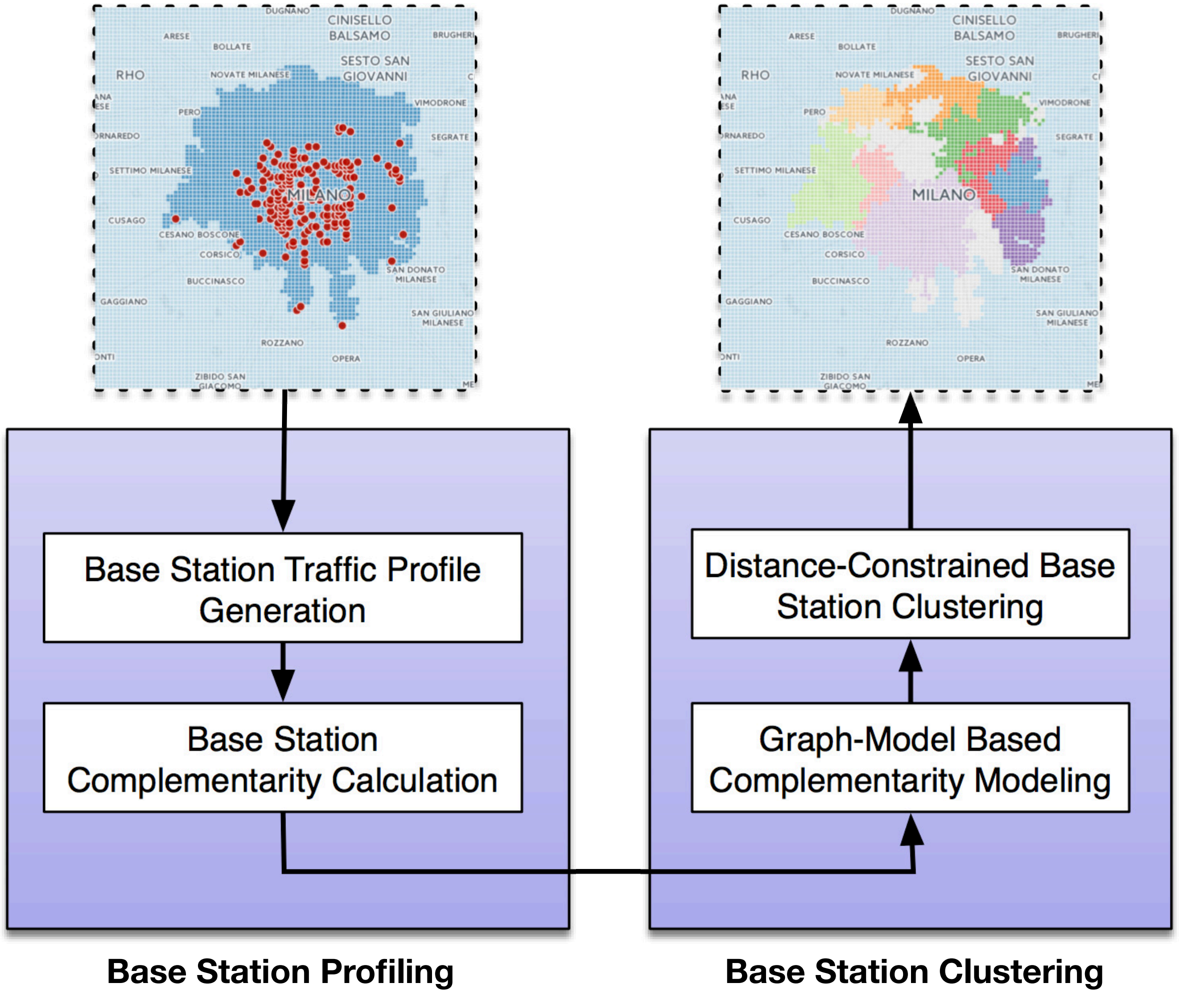
$$\cup_{\forall C_k \in \mathbb{P}} = V \quad \text{and} \quad \cap_{\forall C_k \in \mathbb{P}} = \emptyset \tag{10}$$

- **Distance-Constrained Complementarity-Aware clustering algorithm (DCCA)**

$$value(v,C) = con(v,C) \times log\left(\frac{\tau}{max\{dist(v,v')\}}\right)$$



Proposed Framework



Evaluation

- **Real-world dataset**
  - Mobile data traffic: Telecom Italia (TIM) Big Data Challenge
    - 11/01/2013 to 12/31/2013 (2 months), Italy
  - Base station Positions: CellMapper.net1

Table 1. Dataset Description

Item	Milan	Trentino
# Grids	10,000	11,466
Grid size	55,225 m^2	1000,000 m^2
# RRH	182	522
# Covered grids	2,918	2,035
Average coverage	885,420 m^2	3,932,950 m^2
Average traffic volume	0.19	0.13
Data collection period	11/01/2013–12/31/2013	

1. https://www.cellmapper.net/map

Evaluation

- **Metrics**
  - capacity utilization before and after RRH clustering

$$Util(\mathbb{P}) = \frac{\mathrm{mean}_{C_k} U(C_k)}{\mathrm{mean}_{s_i} U(s_i)}$$

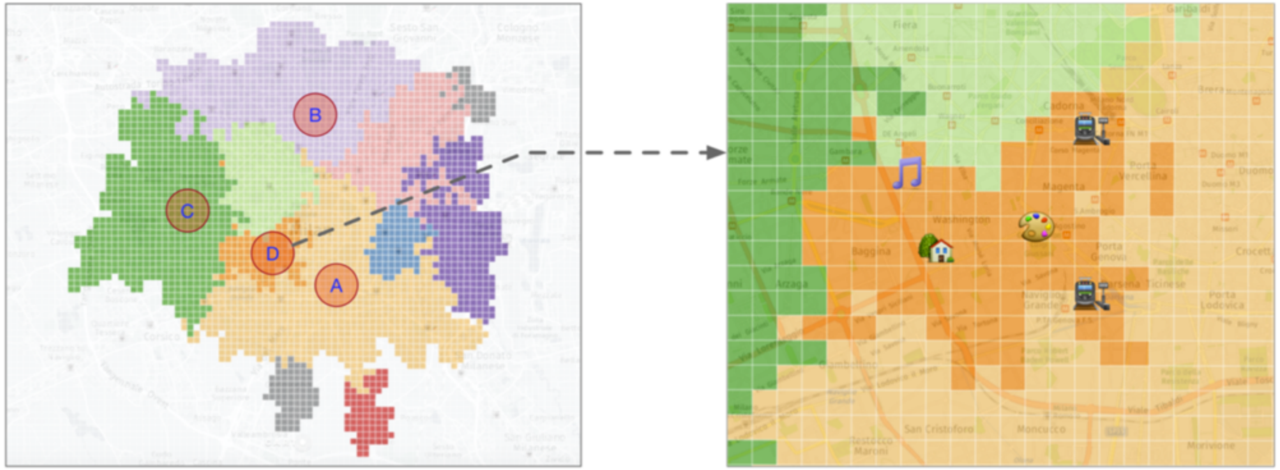
- **Baseline**
  - Distance-constrained clustering (without complementarity)

- **Results**

Utilization	Milan	Trentino
DC	58.7%	39.2%
DCCA (Proposed)	83.4%	76.7%

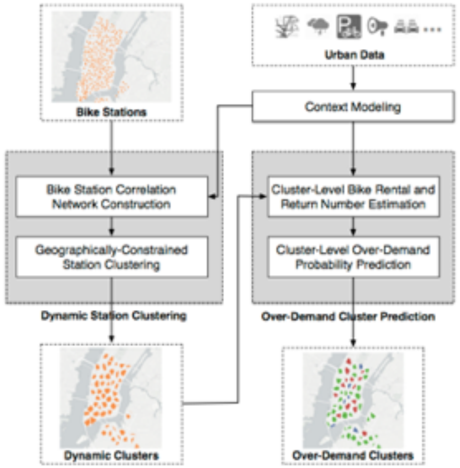
A Case Study in Milan

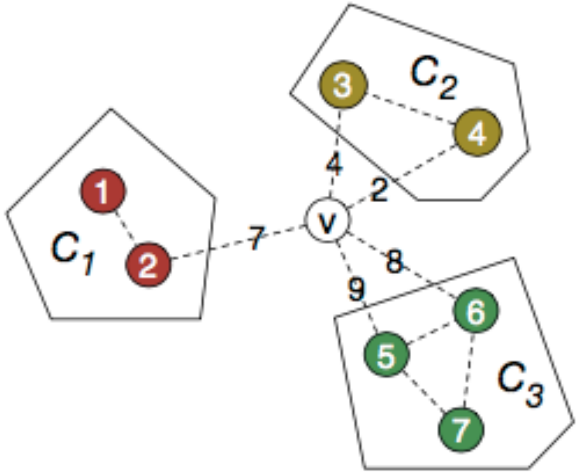
- **An example clustering scheme in Milan, Italy**
  - 12 clusters for setting BBU pools
  - Cluster A, B, and C: urban + residential
  - Cluster D: small but hybrid area
    - Residential: the Washington neighborhood
    - Theater: Teatro Nazionale CheBanca
    - Metro hub station: Wagner Station



(a) (b)

(a)





Aggregation and Clustering: Application #2

ACM International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp'16)

Honorable Mention Award

Users' Problem: Over-Demand Stations

- **No bikes or docks available in stations**

Over-demand stations (full and empty)

- In DC, 60% users are unsatisfied with **over-demand** stations[1]
- In NYC, operators get fined when **over-demand** stations occur[2]

---

[1] Capital Bikeshare 2014 Member Survey Report

[2] S. Kaufman, Citi Bike: The First Two Years, 2015

Operators' Task: Preventing Over-Demand Stations

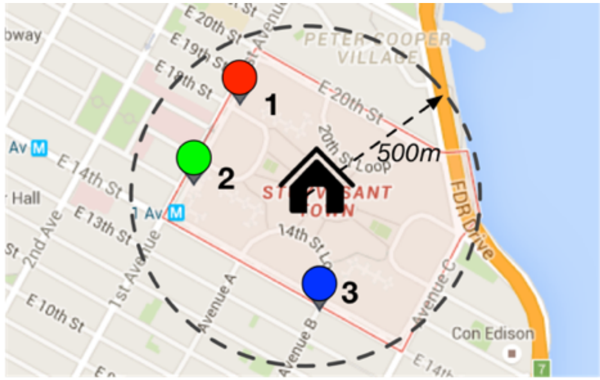
- **Know: which station will be over-demand?**
- **Act: trucks scheduling, temporary docks, etc.**

Trucks moving bikes

Temporary docks for special events

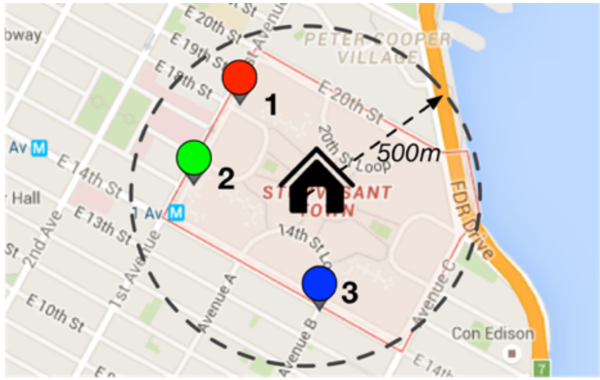
Challenge: Fluctuating Traffic

- **Densely distributed stations → fluctuating traffic**



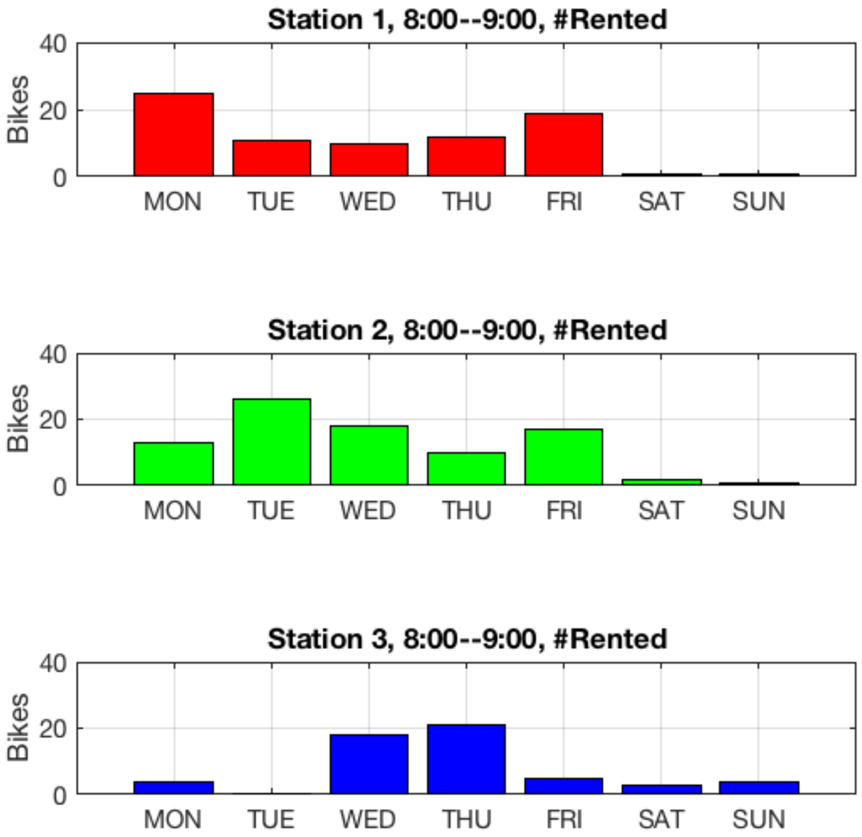
Challenge: Fluctuating Traffic

- **Densely distributed stations → fluctuating traffic**

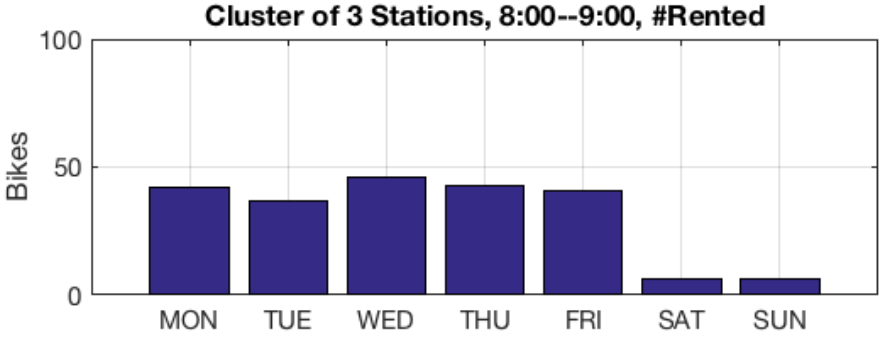


rush hours

station-level



area-level

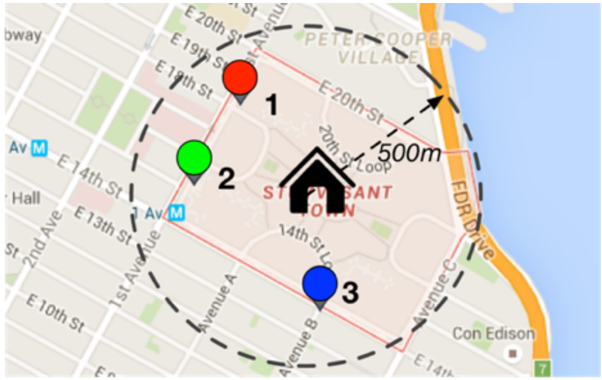


more predictable

highly variant

Challenge: Fluctuating Traffic

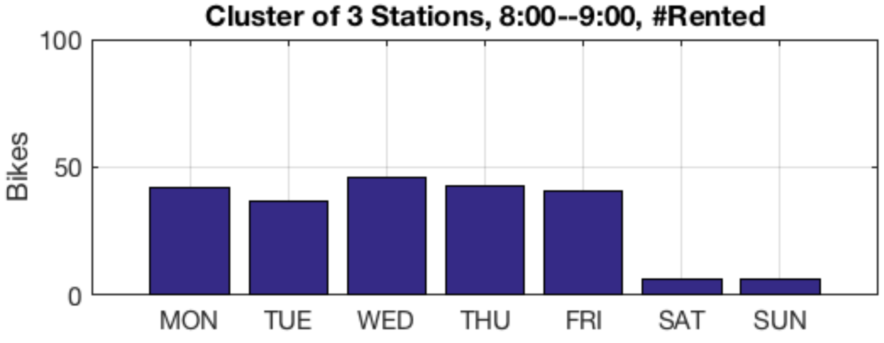
- **Densely distributed stations → fluctuating traffic**



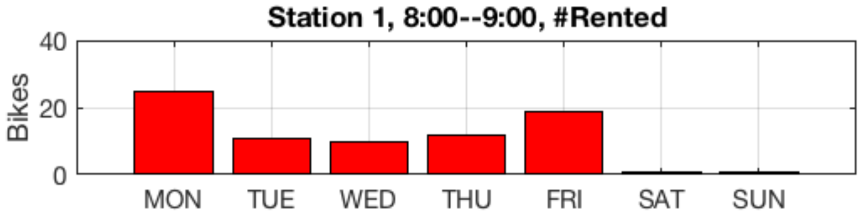
rush hours

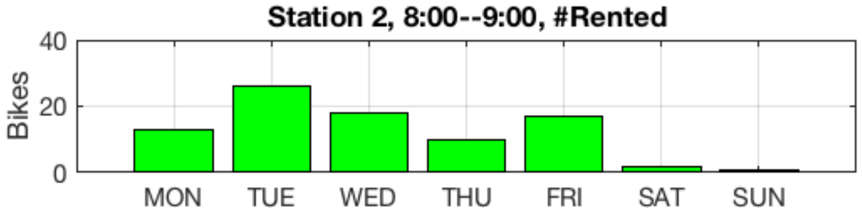
station-level

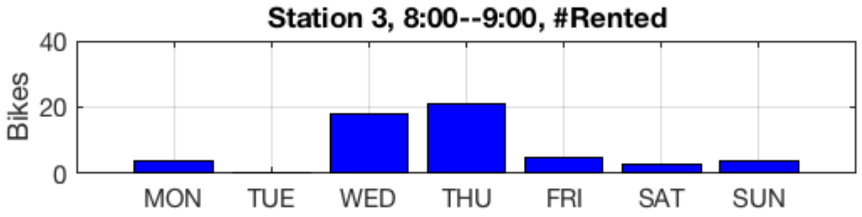
area-level



more predictable



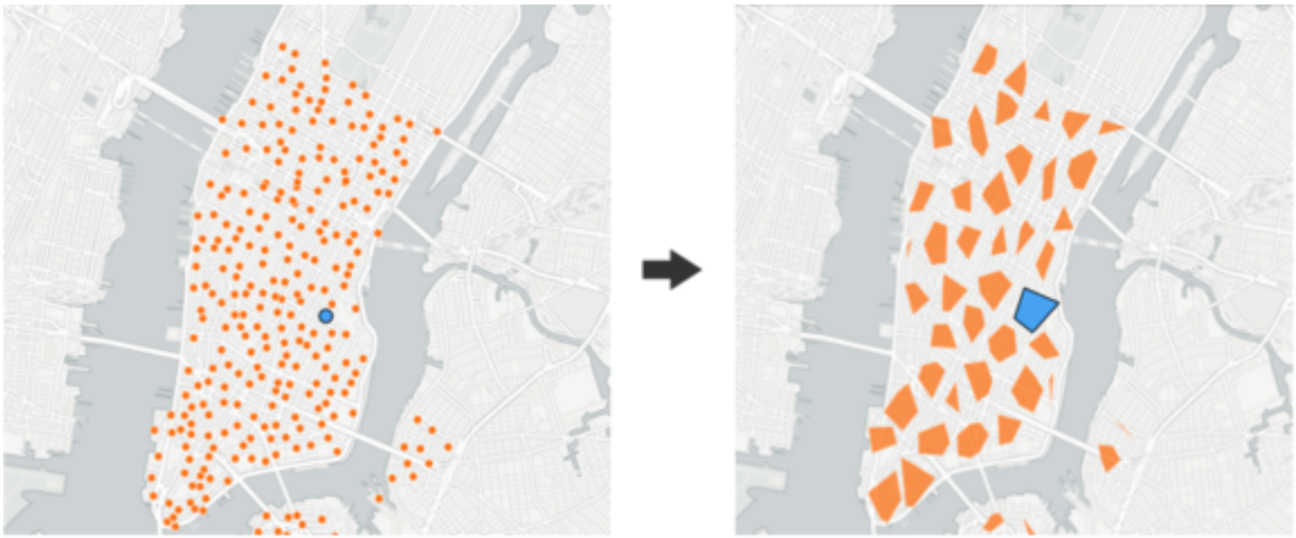




highly variant

Idea: Cluster for Stable Traffic

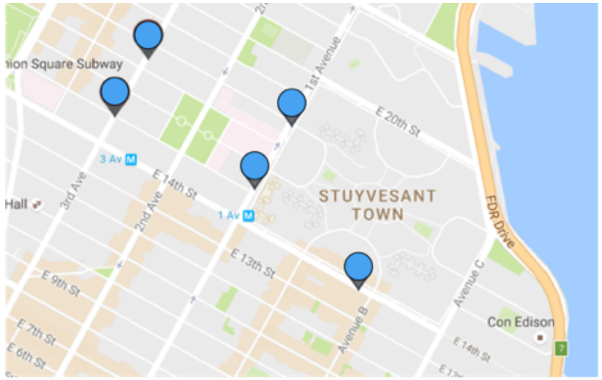
- Cluster: stations → clusters
- Know: demand in each cluster
- Act: deal with over-demand stations



over-demand station　　　　over-demand cluster

How To Group Stations into Clusters?

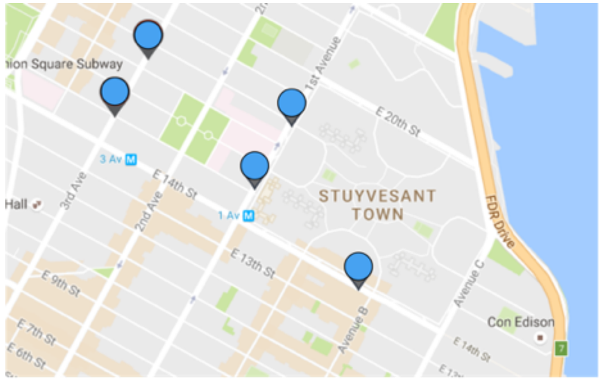
- Intuition: neighboring stations → static clusters
- However: bike demand varies in different contexts



a neighborhood in NYC

How To Group Stations into Clusters?

- **Intuition: neighboring stations → static clusters**
- **However: bike demand varies in different contexts**

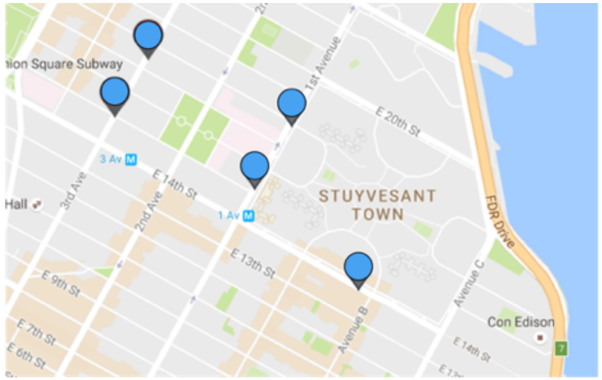


a neighborhood in NYC

the morning rush hours

How To Group Stations into Clusters?

- **Intuition: neighboring stations → static clusters**
- **However: bike demand varies in different contexts**



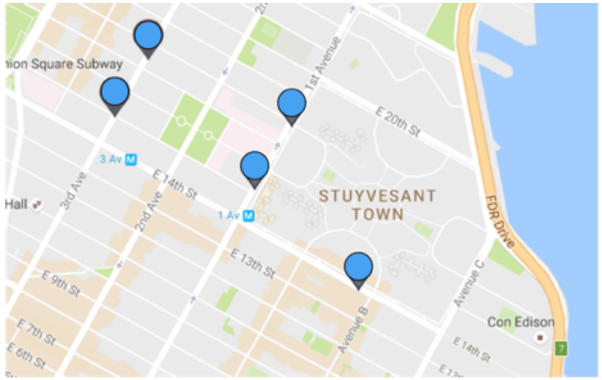
a neighborhood in NYC

the morning rush hours

a basketball game

How To Group Stations into Clusters?

- **Intuition: neighboring stations → static clusters**
- **However: bike demand varies in different contexts**



a neighborhood in NYC

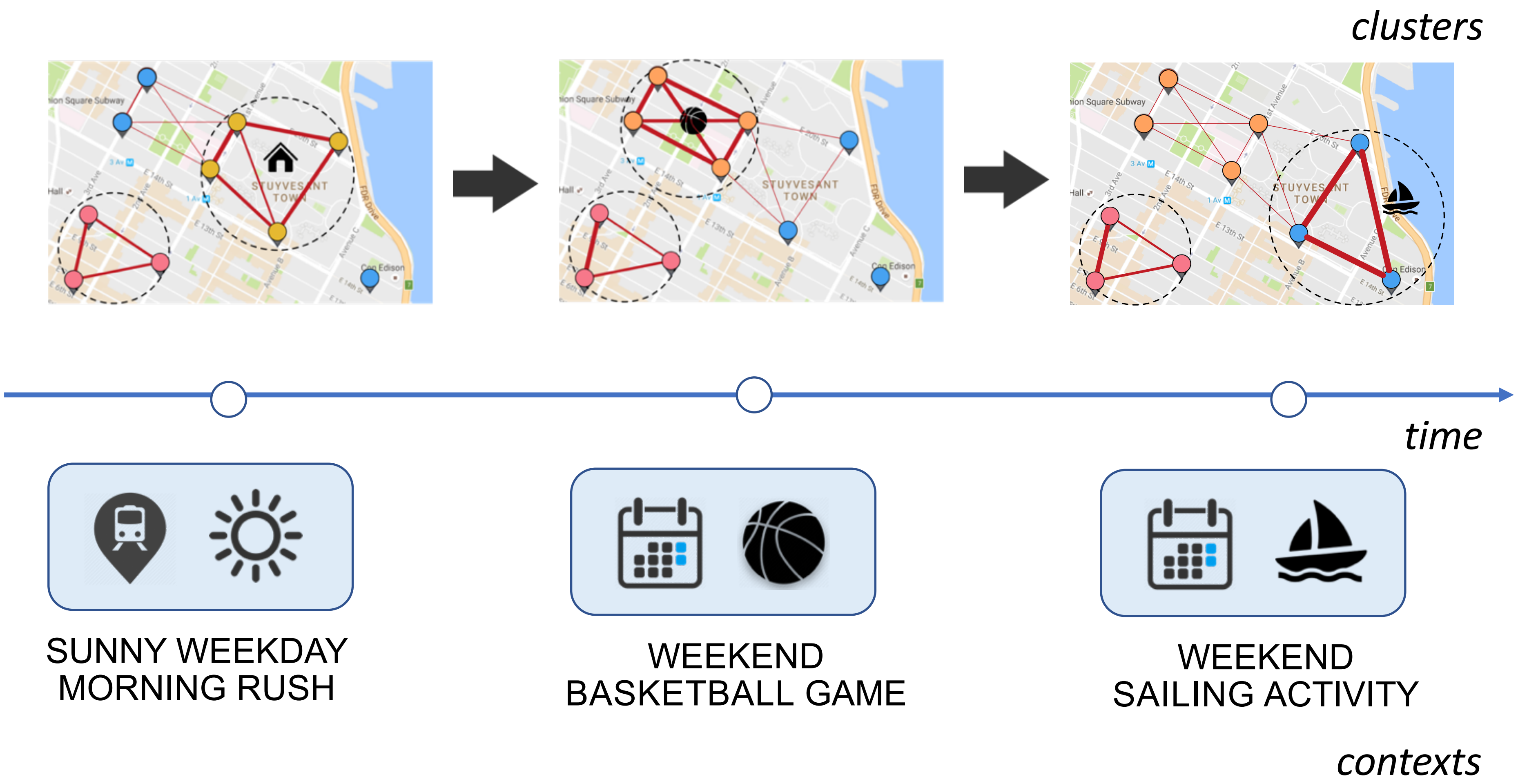
the morning rush hours

a basketball game

Cluster stations **dynamically** according to **contexts**

How To Group Stations into Clusters?

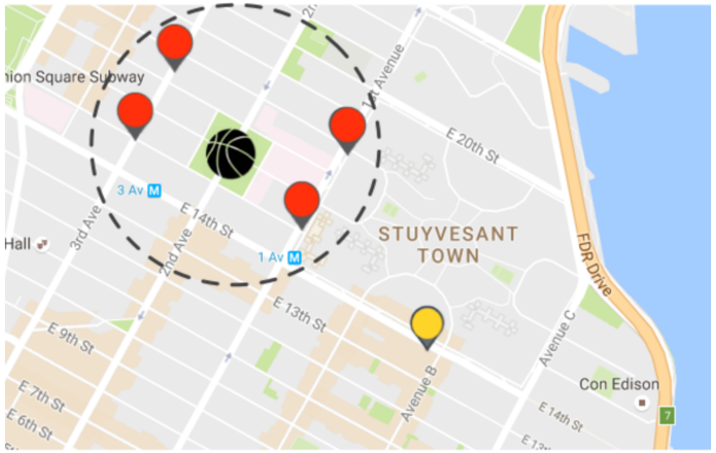
- **Clustering stations dynamically according to contexts**



Dynamic Station Clustering

- **Given different contexts, form different clusters**
  - Learn station similarity from historical data

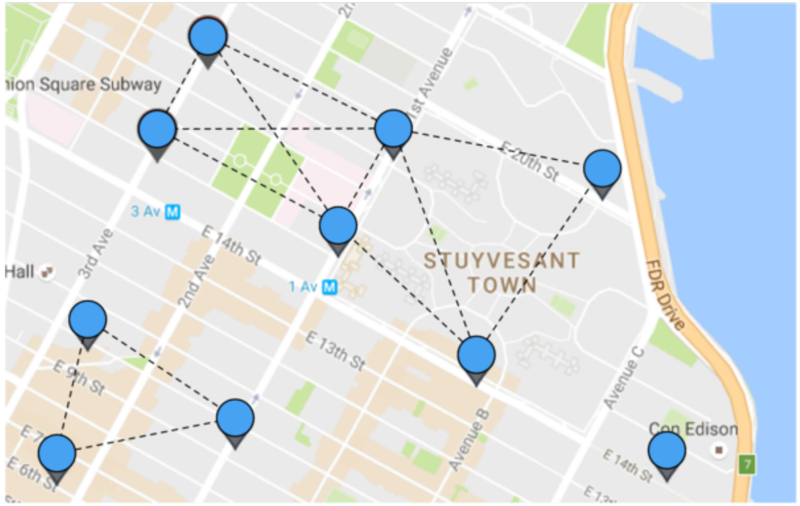
WEEKDAY MORNING, SUNNY



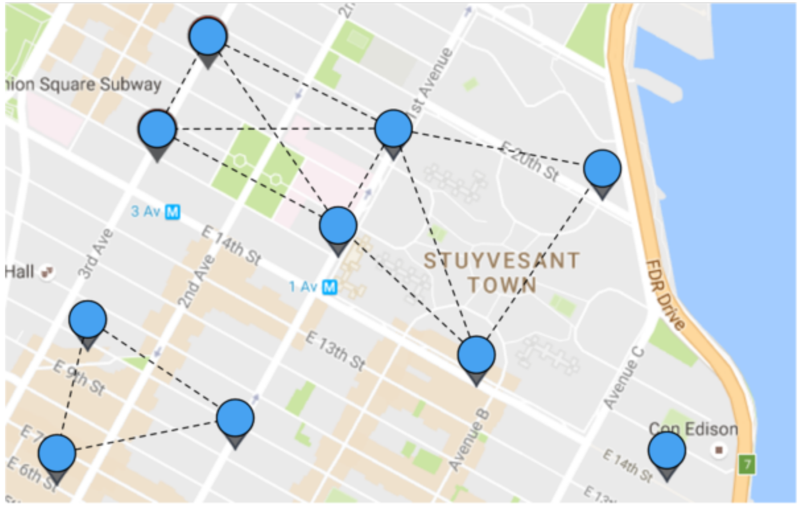
BASKETBALL GAME

Dynamic Station Clustering

- **Weighted Correlation Network-Based Clustering**
  - generate the network



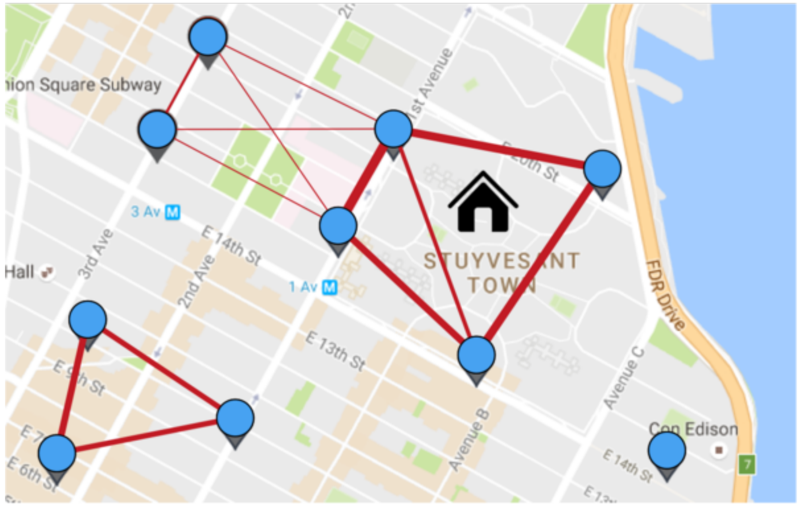
WEEKDAY MORNING, SUNNY



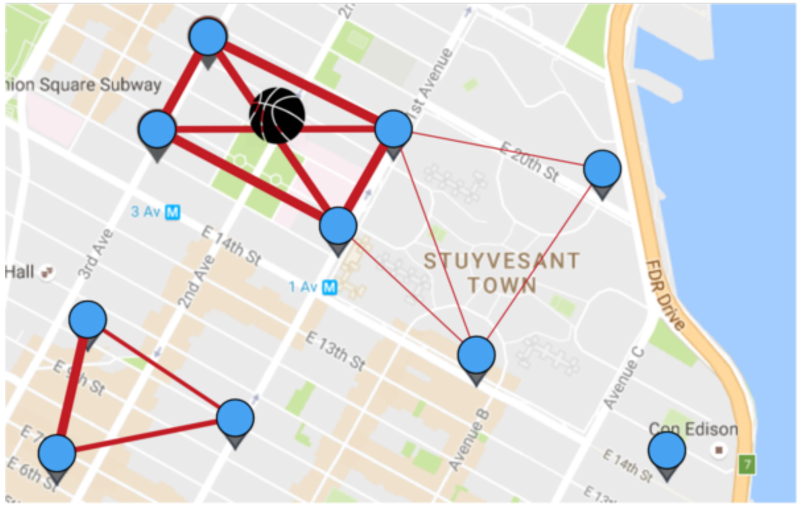
BASKETBALL GAME

Dynamic Station Clustering

- **Weighted Correlation Network-Based Clustering**
  - generate the network
  - determine weights based on traffic correlation



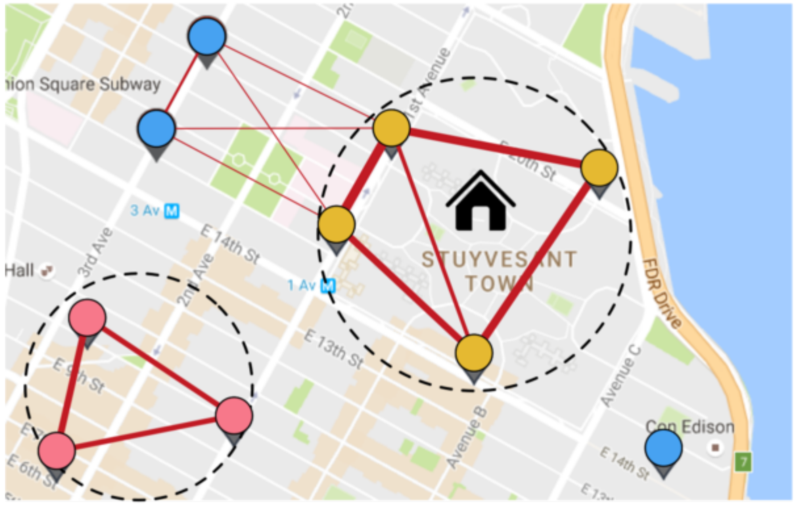
WEEKDAY MORNING, SUNNY



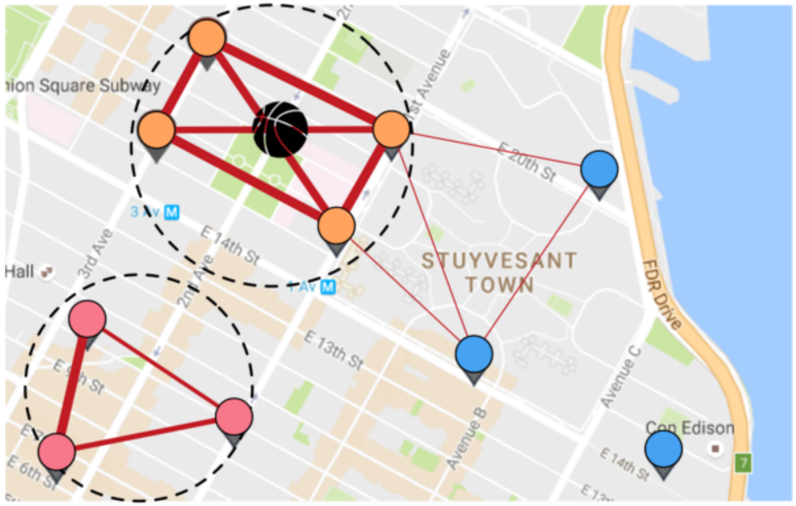
BASKETBALL GAME

Dynamic Station Clustering

- **Weighted Correlation Network-Based Clustering**
  - generate the network
  - determine weights based on traffic correlation
  - cluster correlated stations

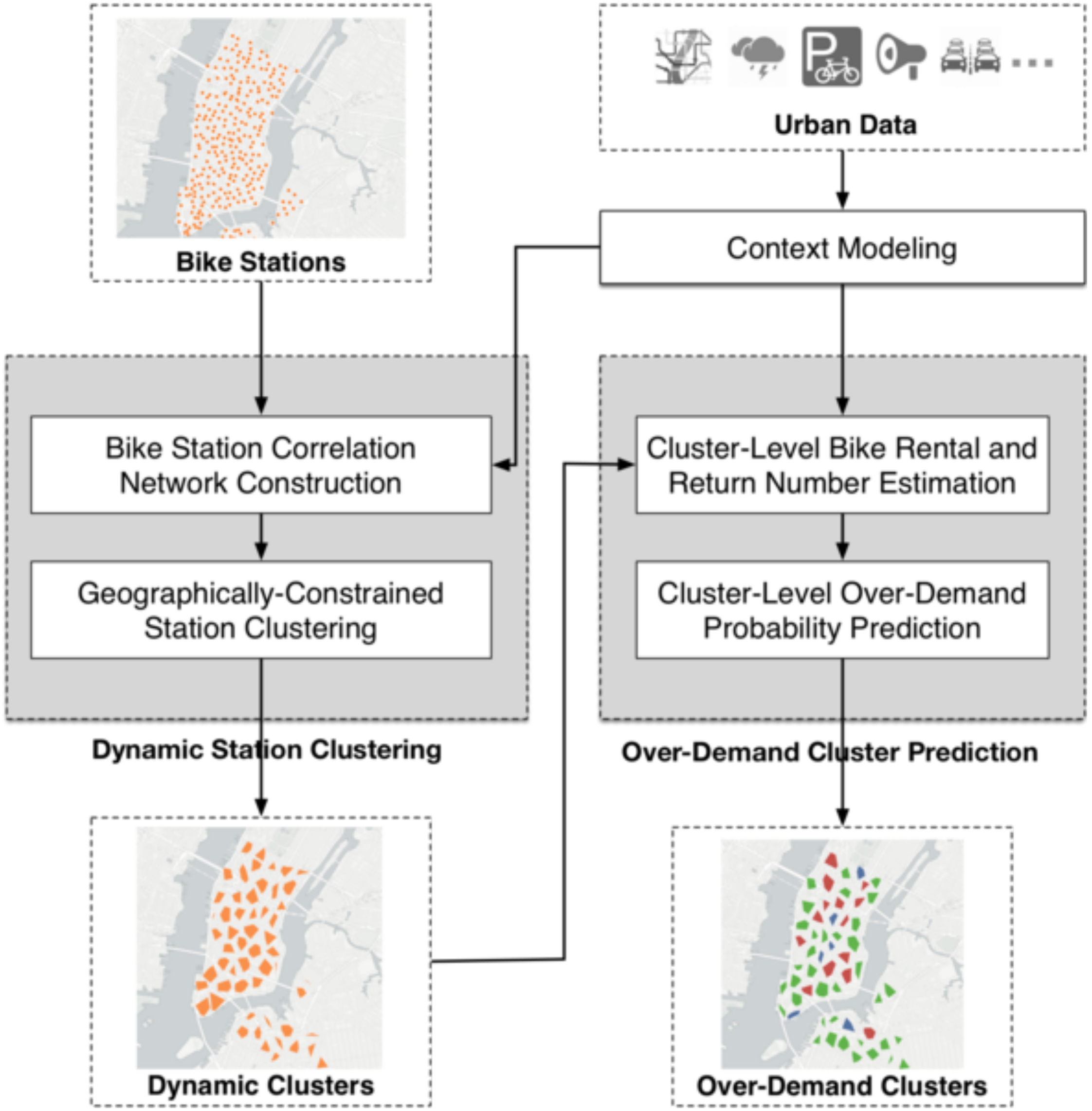


WEEKDAY MORNING, SUNNY



BASKETBALL GAME

Framework Overview



Evaluation

- **Datasets: NYC and DC, 2014**
  - bike trips + station status
  - meteorological: Weather Underground API
  - social event: Eventful API
  - traffic alert: 511 traffic feeds + Twitter accounts

Table 3. Summary of Datasets

Data type	Item	New York City	Washington, D.C
Bike sharing	# Stations	327	203
	# Bike trips	18,019,196	6,138,428
	# Station status	hourly	hourly
	# Over-demand	626,856	318,576
Contextual factors	# Weather forecast	hourly	hourly
	# Social events	435	329
	# Traffic events	958	745
Data collection period		01/01/2014–12/31/2015	

Evaluation

- **Baselines**
  - station-level over-demand prediction
  - cluster-level over-demand prediction

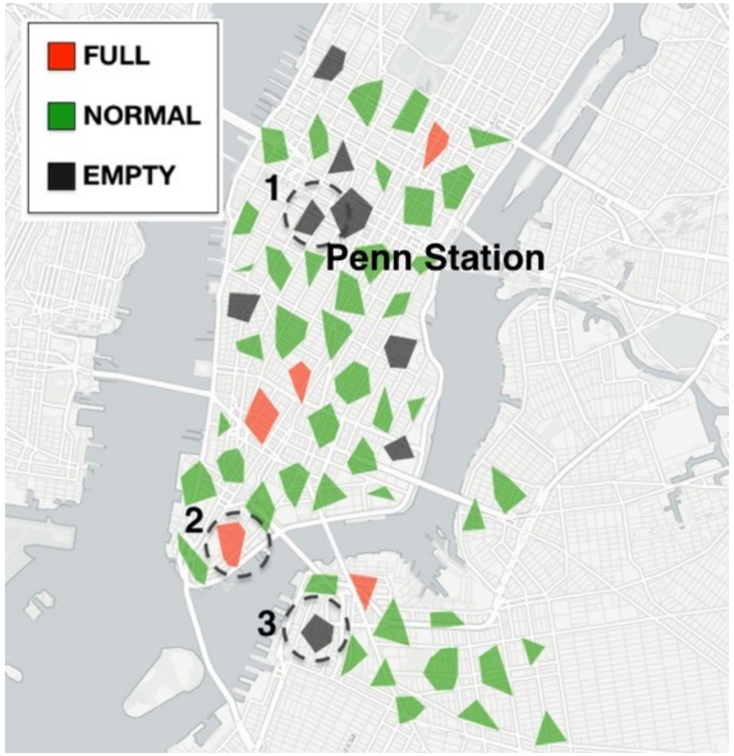
- **Results**

Table 5. Over-demand prediction results of different methods

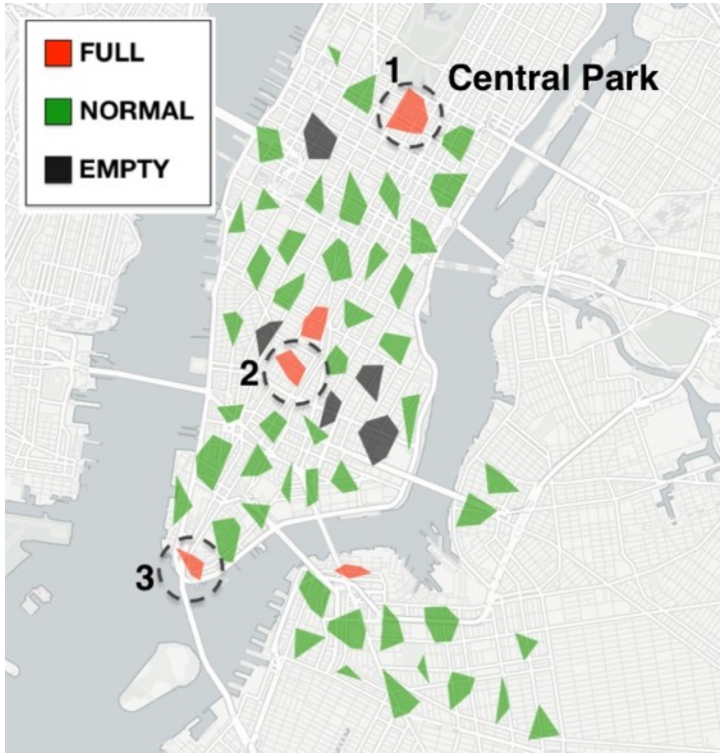
Methods	Precision	Recall	F1	Precision	Recall	F1
	NYC			DC		
ARIMA	0.548	0.506	0.526	0.520	0.541	0.530
B-MC	0.753	0.656	0.692	0.636	0.539	0.583
ANN-S	0.776	0.571	0.658	0.667	0.428	0.521
SC-MC	0.790	0.647	0.711	0.793	0.821	0.807
CCF-MC	0.833	0.832	0.828	0.815	0.880	0.846
ANN-C	0.673	0.852	0.752	0.857	0.600	0.706
WCN-MC	**0.882**	**0.938**	**0.909**	**0.857**	**0.923**	**0.889**

Case Study: Clusters in Different Contexts

- **Different clusters are formed**
- **Over-demand clusters are predicted in different areas**



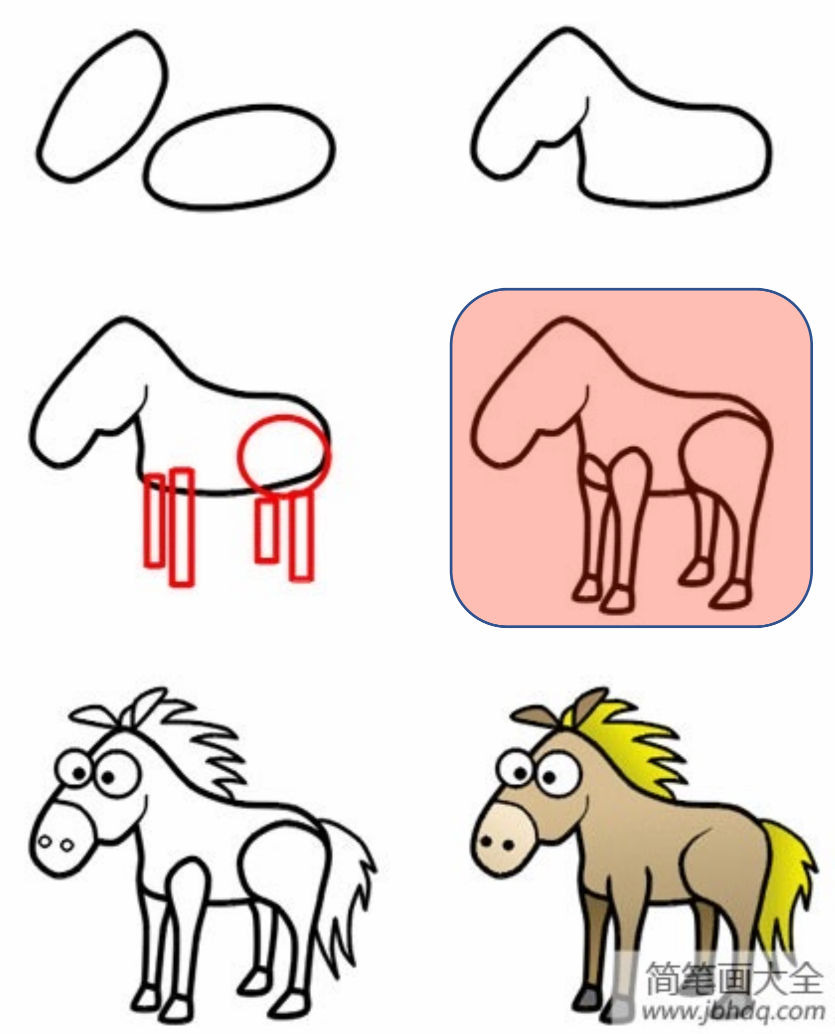
A rushing weekday morning



A sunny weekend afternoon

Factorization and Inference

Spatiotemporal Urban Data

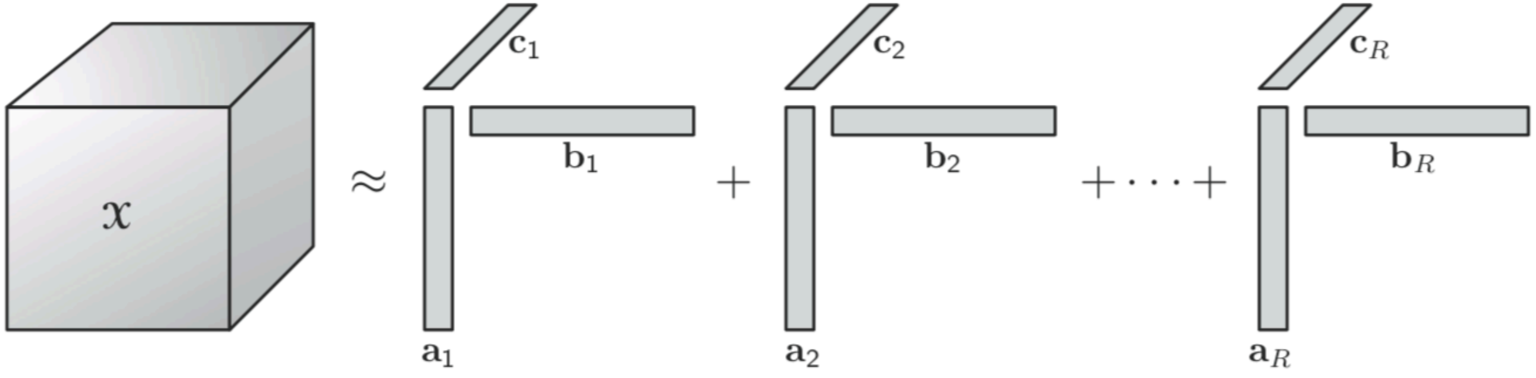


Spatiotemporal Tensor Decomposition: I

- **CP Decomposition**: factorizing a tensor into a sum of component rank-one tensors.

$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

- Algorithms: CANDECOMP, PARAFAC…



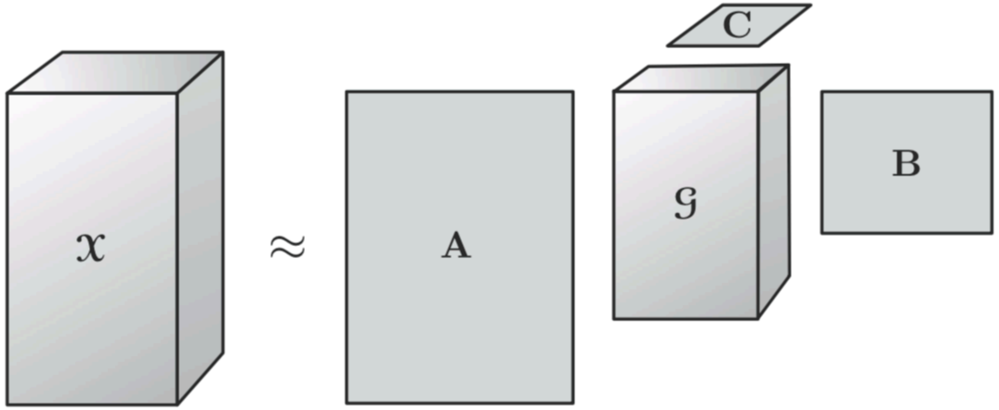
[Kolda and Bader, 2009]

Spatiotemporal Tensor Decomposition: II

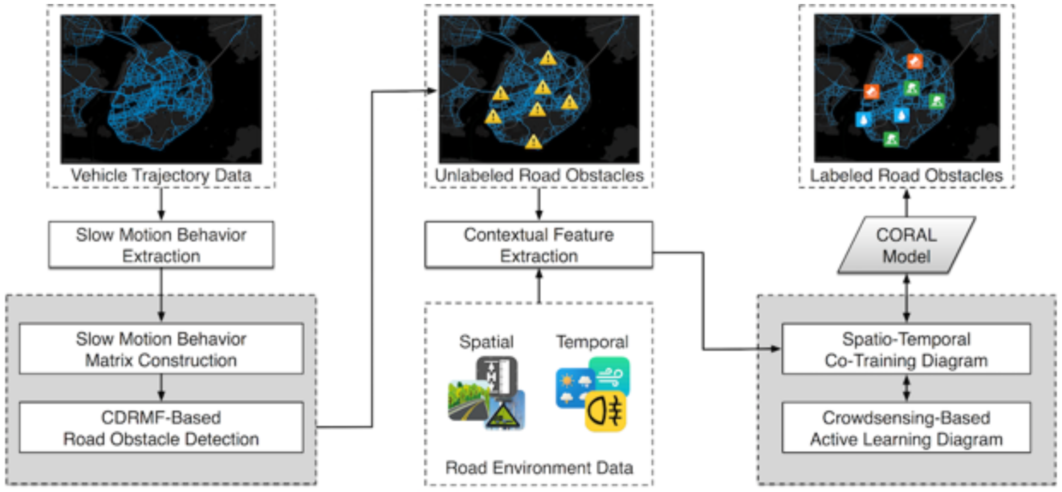
- **Tucker Decomposition**: factorizing a tensor into a core tensor multiplied (or transformed) by a matrix along each mode.

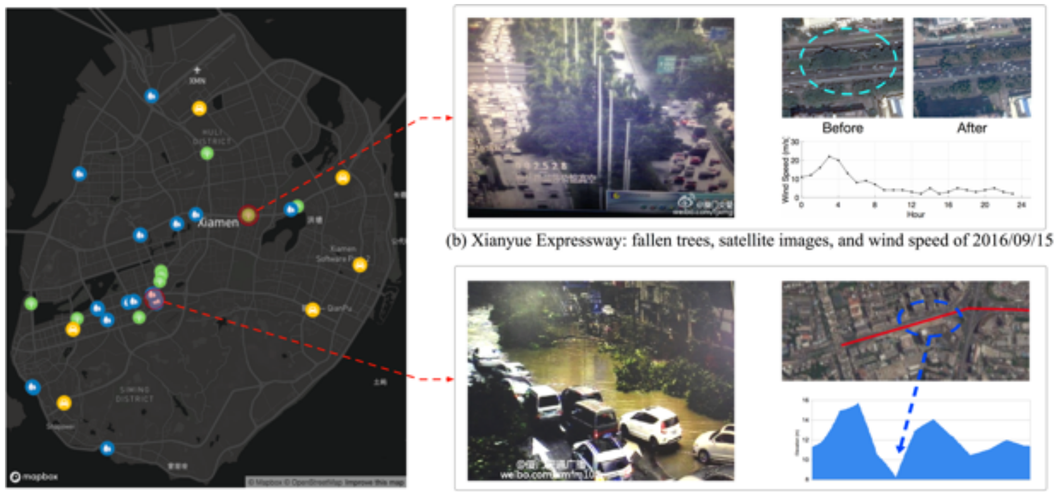
$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} \, \mathbf{a}_p \circ \mathbf{b}_q \circ \mathbf{c}_r = [\![\mathcal{G} \, ; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

- Algorithms: HOSVD, TUCKALS3, HOOI…



[Kolda and Bader, 2009]





(a) Identified road obstacles during Typhoon Meranti

(b) Xianyue Expressway: fallen trees, satellite images, and wind speed of 2016/09/15

(c) Hubin South Road: ponding water, satellite images, and elevation of the road segment

Factorization and Inference: Application

ACM International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp'18)

Urban Disasters

On September 15, 2016 3:00 AM, Typhoon Meranti landed in Xiamen, China, affecting **1.8 million** people, and caused **$4.8 billion** economic losses.

Urban Disaster Response

- **Restoring Road Transportation Network**
  - search and rescue, supply
  - first 24 hours: live hotline

Identifying Road Network Anomaly

- **Patrol and Report**
  - time and labor
  - potential risks
- **Road Surveillance Cameras**
  - slow and inaccurate
  - disaster-broken cameras
  - blind zones

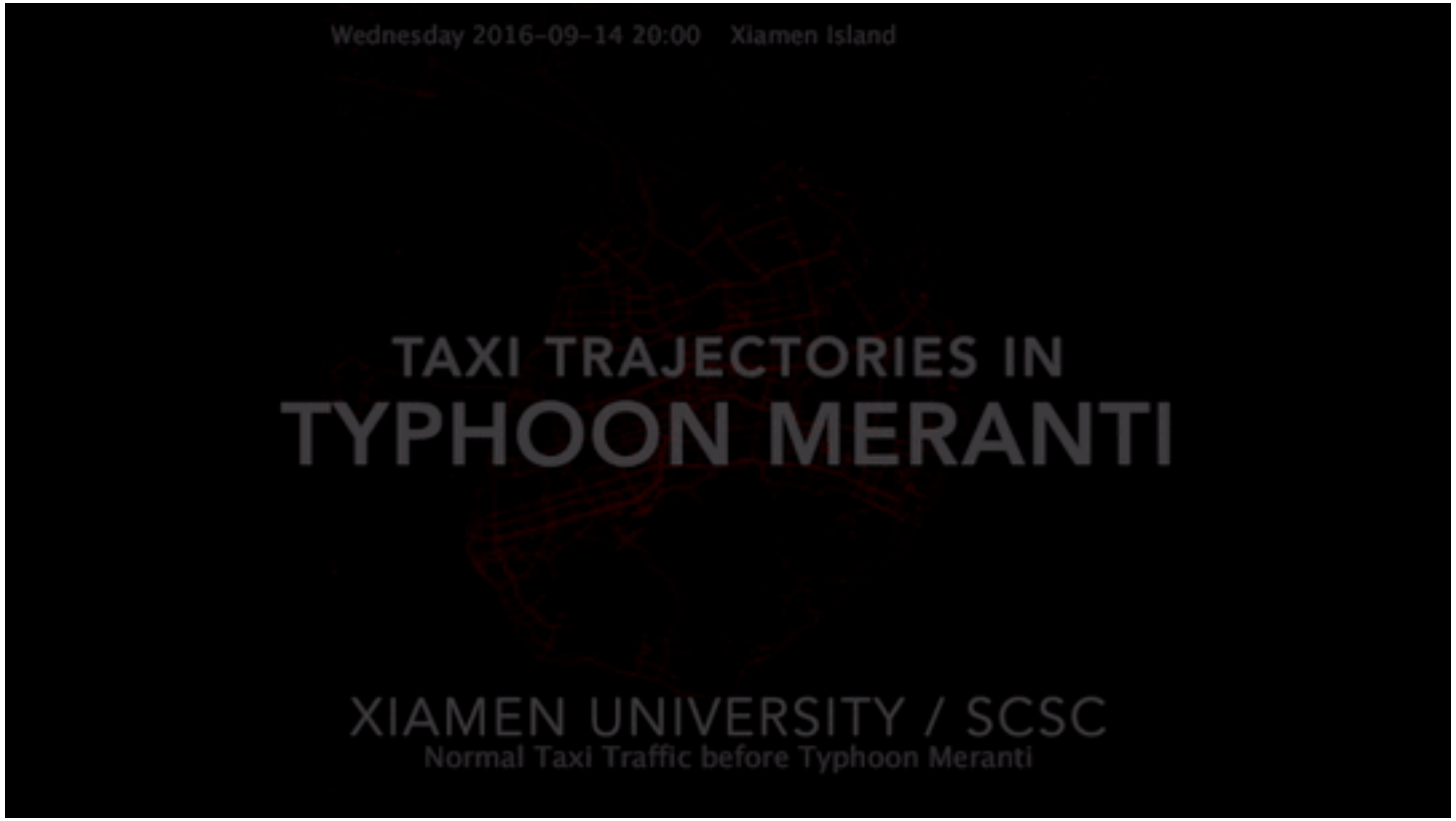
Detecting Road Network Anomaly

- **Patrol and Report**
  - time and labor
  - potential risks
- **Road Surveillance Cameras**
  - slow and inaccurate
  - disaster-broken cameras
  - blind zones

real-time, low-cost, comprehensive road network anomaly identification

Detecting Road Network Anomaly

- **Vehicles affected**



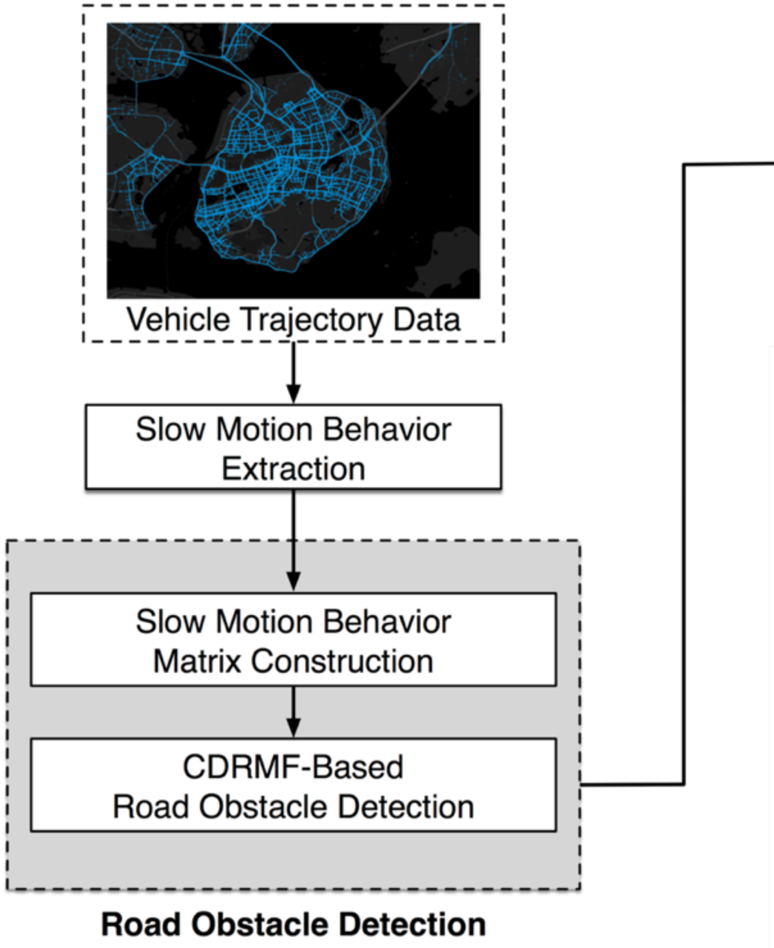
Typhoon landfall: vehicle mobility stopped

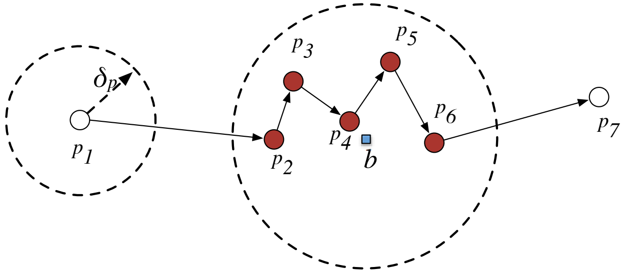
Mobility Big Data in Road Networks

- **Vehicle GPS trajectories (anonymized)**
  - taxi, bus, carshare…
  - frequency: every 30s

Step 1: Road Anomaly Detection

- **Slow motion behavior extraction**

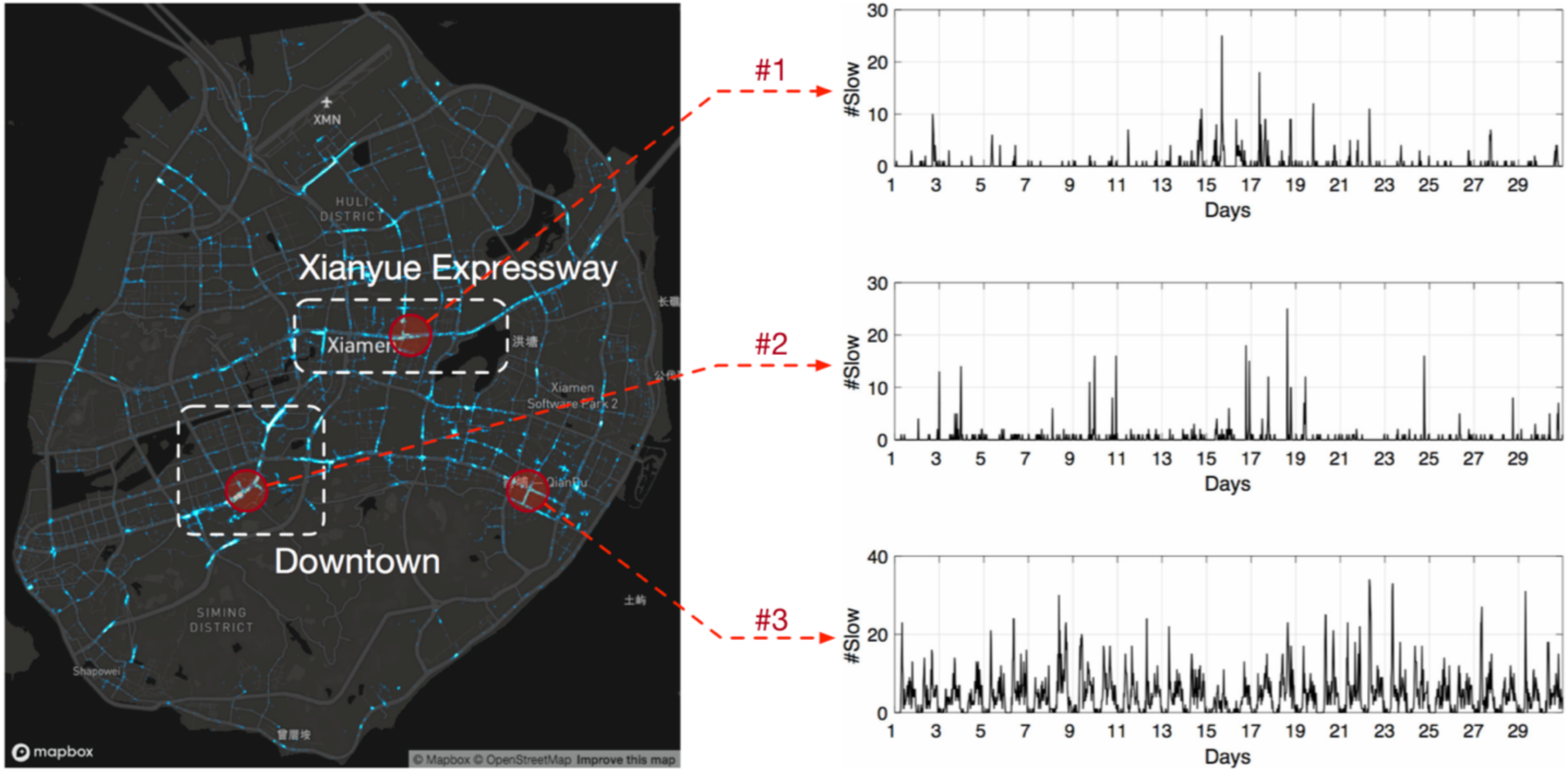




Slow motion behavior

Step 1: Road Anomaly Detection

- **Example: slow motion behaviors (2016/09)**

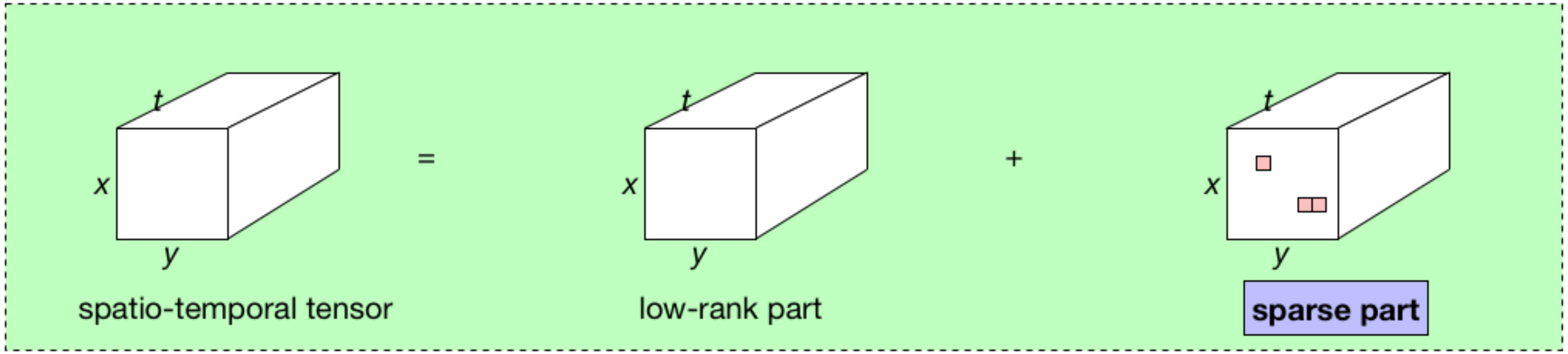


(a) Visualization of extracted slow motion behaviors

(b) Temporal variations of three typical road segments

Step 1: Road Anomaly Detection

- **Challenge: regular slow motion road anomaly**
  - traffic lights, road interactions, etc.
- **Algorithm: Robust Matrix Factorization (CDRMF)**
  - construct: slow motion behavior matrix
  - decompose: Overall (M) = Regular (L) + Abnormal (S)
  - pursue: sparse, clustered anomaly



Step 1: Road Anomaly Detection

- **Algorithm: Robust Matrix Factorization (CDRMF)**
  - construct: slow motion behavior matrix
  - decompose: Overall (M) = Regular (L) + Abnormal (S)
  - pursue: sparse, clustered anomaly

$$
\begin{aligned}
M &= L + S \\
s.t. \quad rank(L) &\le k, \\
card(S) &\le c, \\
outlier(S) &\le \epsilon \\
L \ge 0, S &\ge 0
\end{aligned}
$$

ALGORITHM 1: The CDRMF algorithm for collective anomaly detection

Input: M the slow motion behavior matrix
k the maximum rank
c the maximum cardinality
max_iter the maximum number of iterations

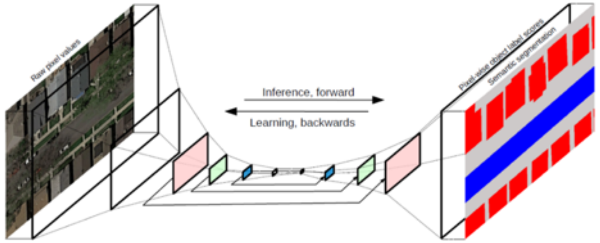
Output: L the low-rank component
S the sparse-and-clustered component

```
Initialize: S ← 0
while not converged and iteration < max_iter do
    a) Solve the low-rank approximation problem:
            L = arg min_L ||A − L||_F,    A = M − S
            s.t.    rank(L) ≤ k
    b) Solve the sparse-and-clustered optimization problem:
            S = arg min_S ||B − S||_F,    B = M − L
            s.t.    card(S) ≤ c,    outlier(S) ≤ ε
    error ← ||M − L − S||_F
end
```

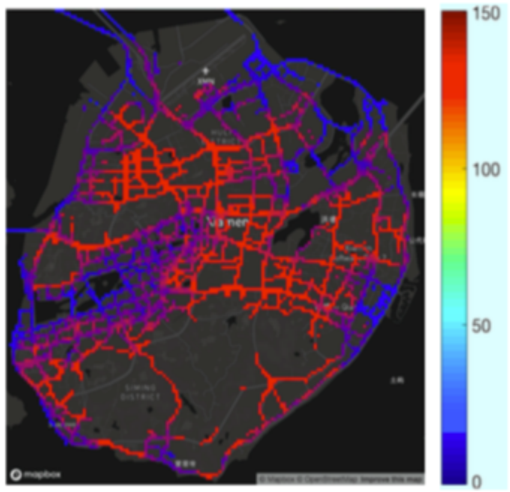
Step 2: Road Anomaly Identification

- **What kind of anomaly?**
  - fallen trees, ponding water…
  - route to urban departments
- **Context-aware inference**
  - flourish road + strong wind → fallen trees
  - low-lying road + heavy rain → ponding waters

Deep-learning-based
tree coverage classification



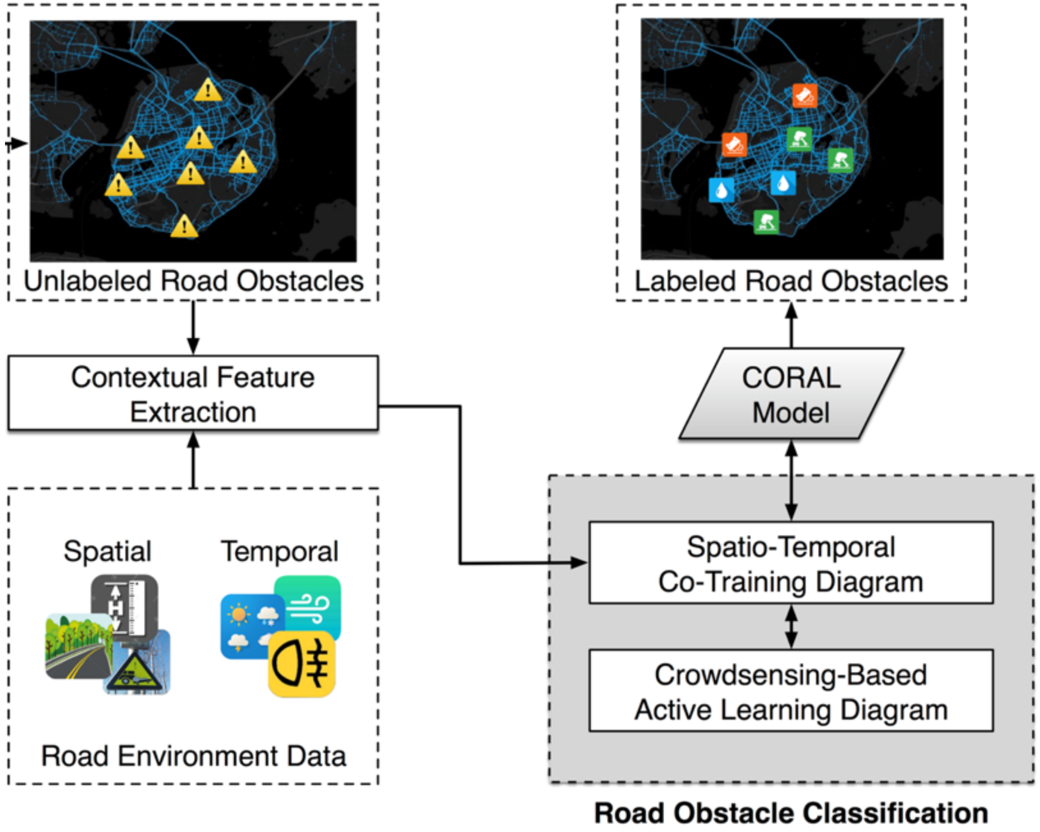
Tree coverage satellite image



Digital Elevation Map (DEM)

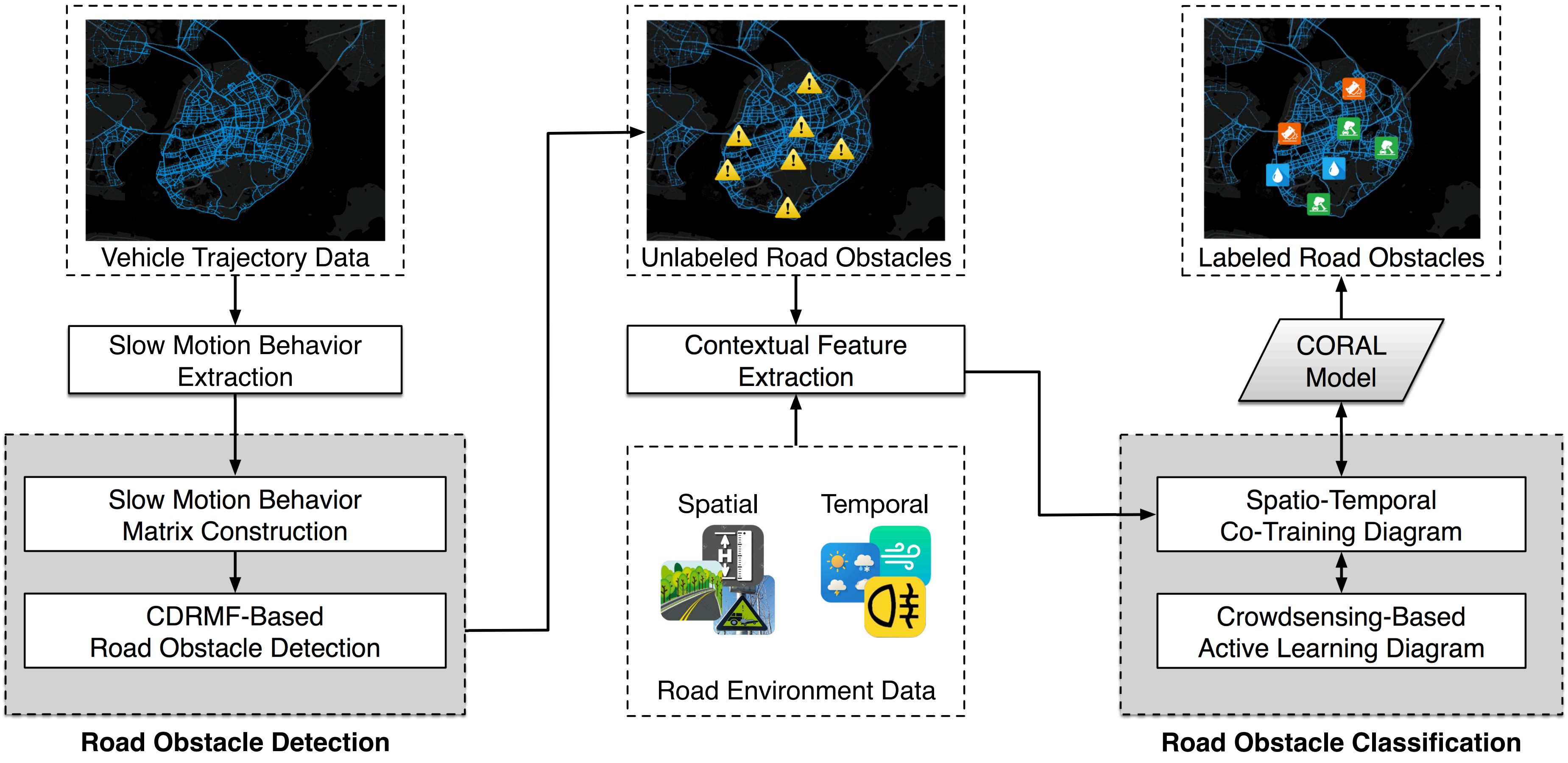
Step 2: Road Anomaly Identification

- **Context-Aware Road-Anomaly Classifcation**
  - Spatial: tree coverage, road elevation…
  - Temporal: weather, wind speed, visibility…



Framework Overview

- **Vehicle Mobility → Slow-Motion → Road Anomaly**



Evaluation in Xiamen

- **Taxi trajectories + urban sensing datasets**
  - 5,486 taxis and 3,928 road segments
  - tree coverage: Google Earth Images (2.5m resolution)
  - road elevation: Google DEM Dataset (30m resolution)
  - Meteorology: Weather Underground API (hourly)

Table 1. Summary of Datasets

Data type	Item	Value
Vehicle trajectory data	# Taxis	5,486
	Sampling rate	every minute
Road environment data	# Road segments	3,928
	Satellite image resolution	2.5 meter
	Elevation resolution	30 meter
	Meteorology data	every hour
Road Obstacles	# Fallen trees	71
	# Ponding water	54
	# Congested and crashed vehicles	34
Data collection period	07/01/2016 00:00–12/31/2016 23:59	
Geographic coverage area	Southwest: [24.423250, 118.064743], Northeast: [24.561485, 118.198504]	

Results

- **Road Anomaly Detection Accuracy**

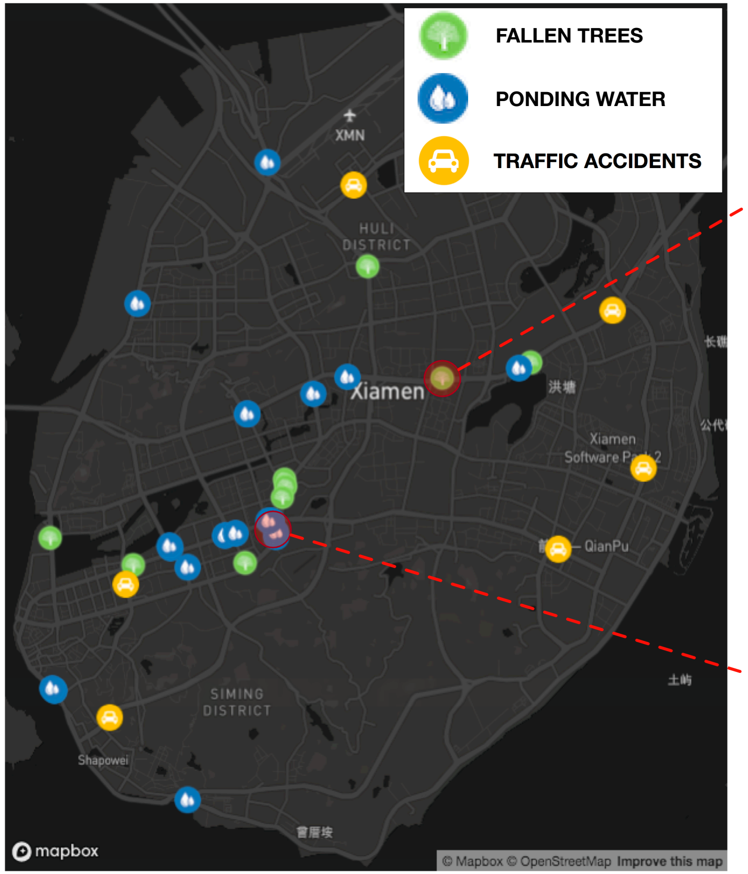
Methods	Precision	Recall	F1
TFBOY	0.503	0.489	0.496
ARIMA	0.872	0.697	0.775
DRMF	0.730	0.906	0.809
CDRMF	0.953	0.931	0.942

- **Road Anomaly Classification Accuracy**

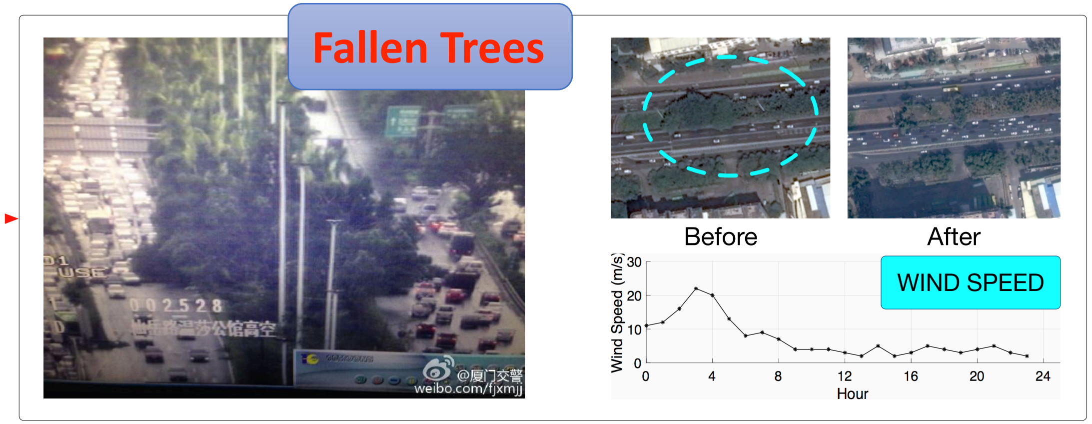
Methods	Precision	Recall	F1
ST-ANN	0.921	0.956	0.938
SCAL	0.772	0.738	0.755
TCAL	0.691	0.649	0.669
COTA	0.843	0.870	0.856
CORAL	0.902	0.931	0.916

Results

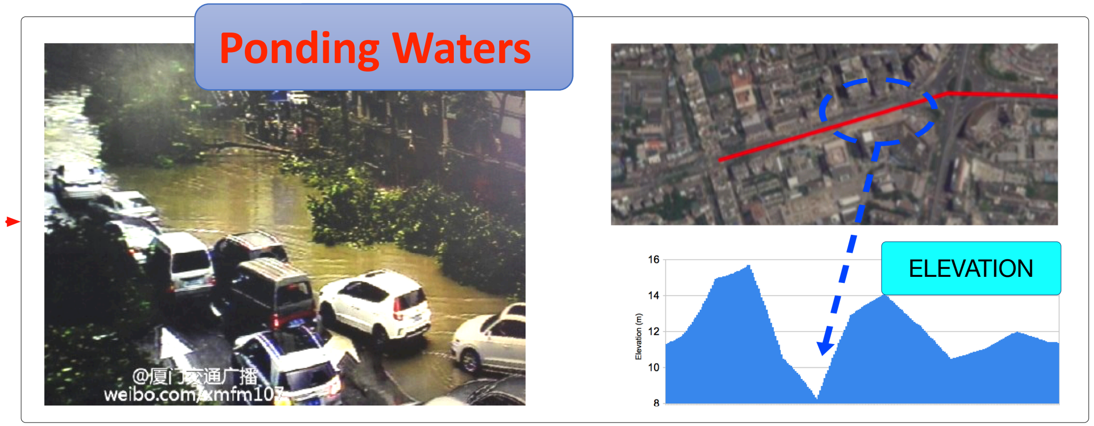
- **Case Study in Xiamen**



(a) Identified road obstacles during Typhoon Meranti



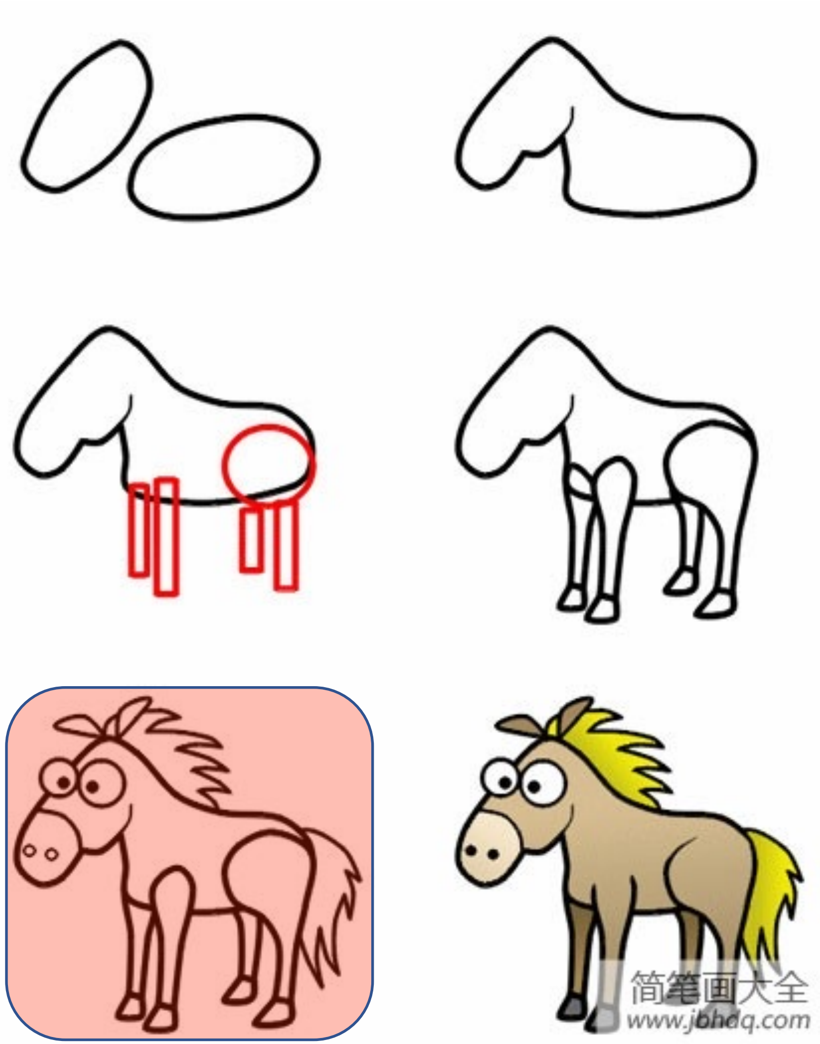
(b) Xianyue Expressway: fallen trees, satellite images, and wind speed of 2016/09/15



(c) Hubin South Road: ponding water, satellite images, and elevation of the road segment

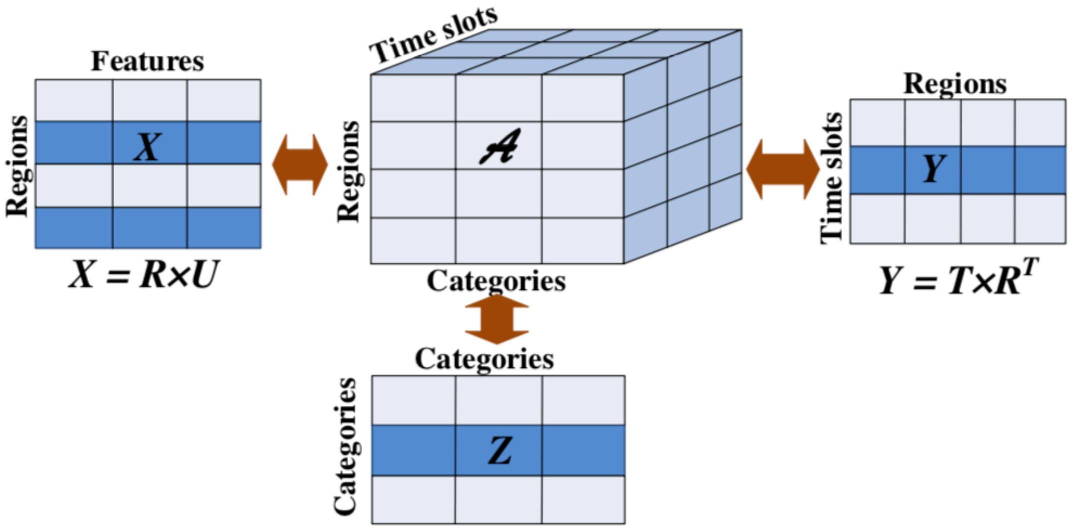
Future Directions

Spatiotemporal Urban Data



Contexts

- **Spatial contexts**
  - Nearby POI distributions
  - Geographic attributes
- **Temporal contexts**
  - Weekdays/weekends
  - AM/PM
- **A priori knowledge**
  - Social events
  - Weather conditions
- **How to model spatiotemporal contexts?**

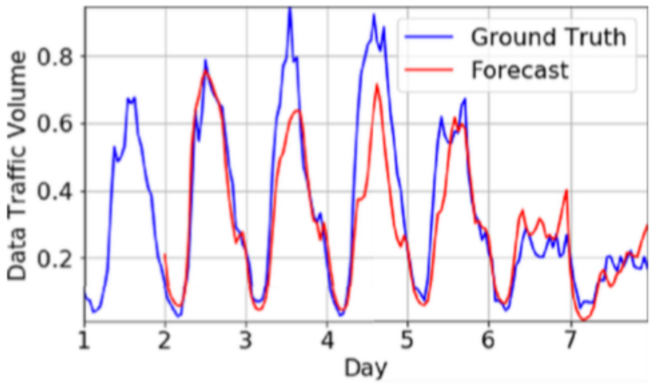


Context-aware tensor decomposition

[Zheng, 2014]

Deep Learning

- **Deep-learning-based modeling**
  - Convolutional Neural Networks
  - Recurrent Neural Networks
- **Prediction and Inference**



(a) MuLSTM forecast

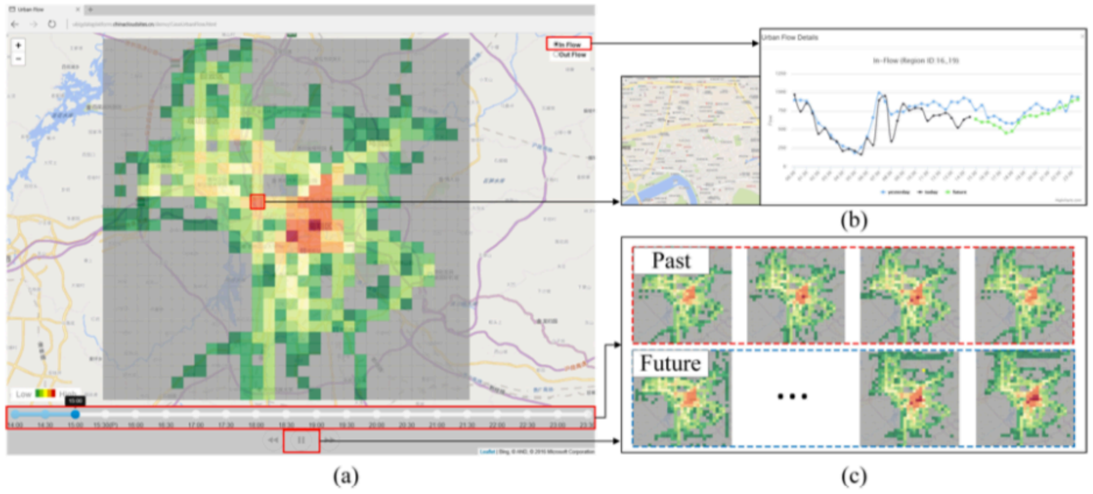


Figure 2: UrbanFlow: A real-time crowd flow forecasting system

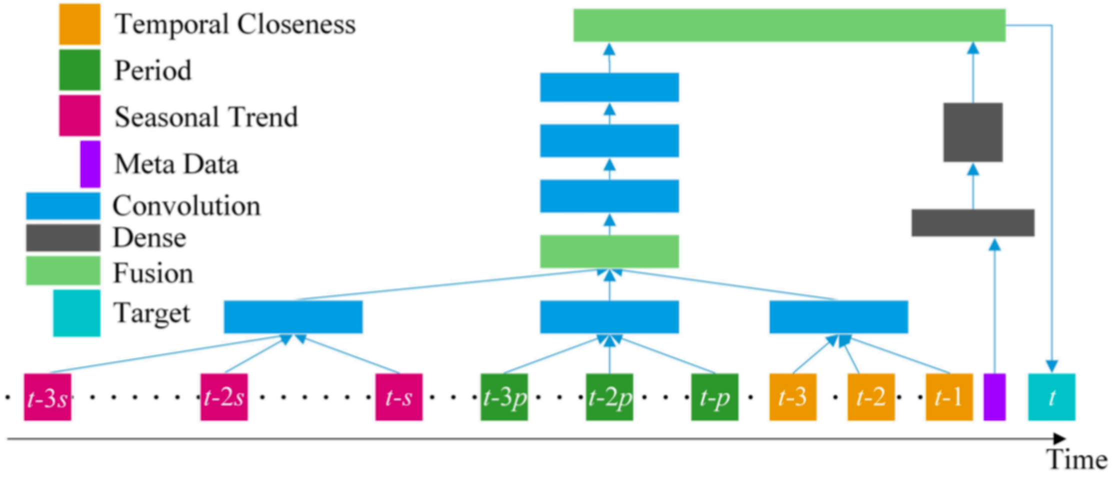
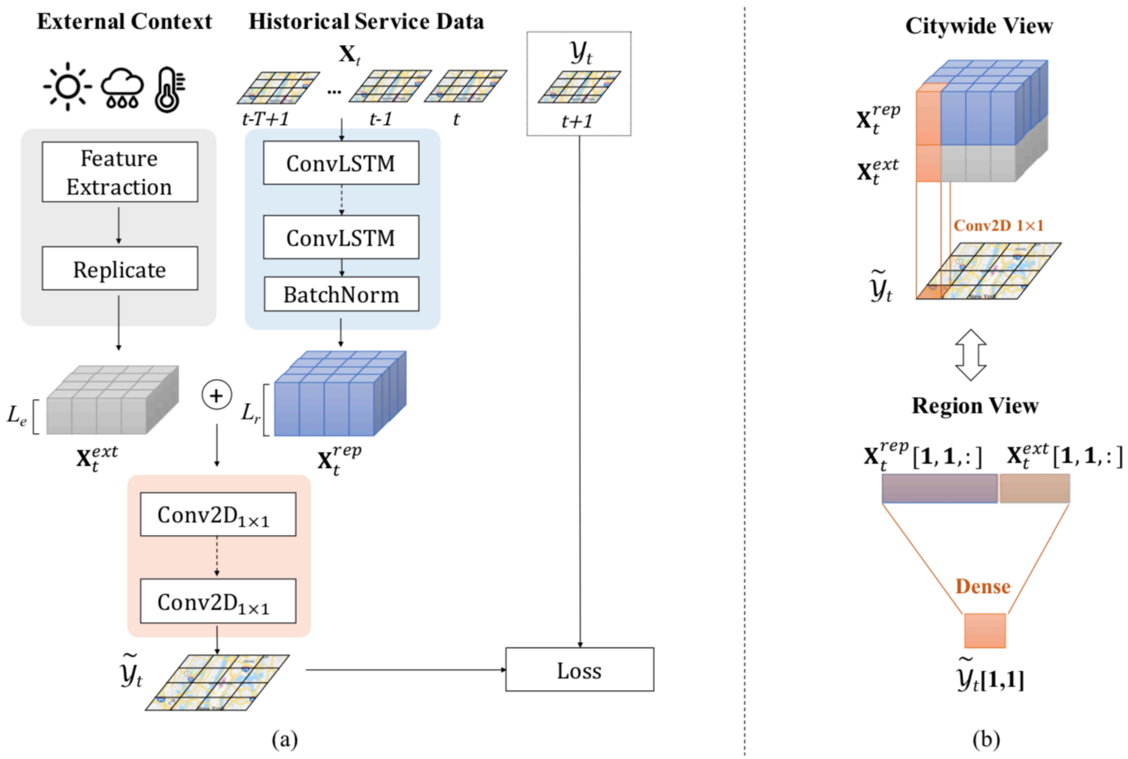


Figure 3: DeepST Architecture.

[Chen, 2018]

Transfer Learning

- **City Transfer**



[Wang, 2019]

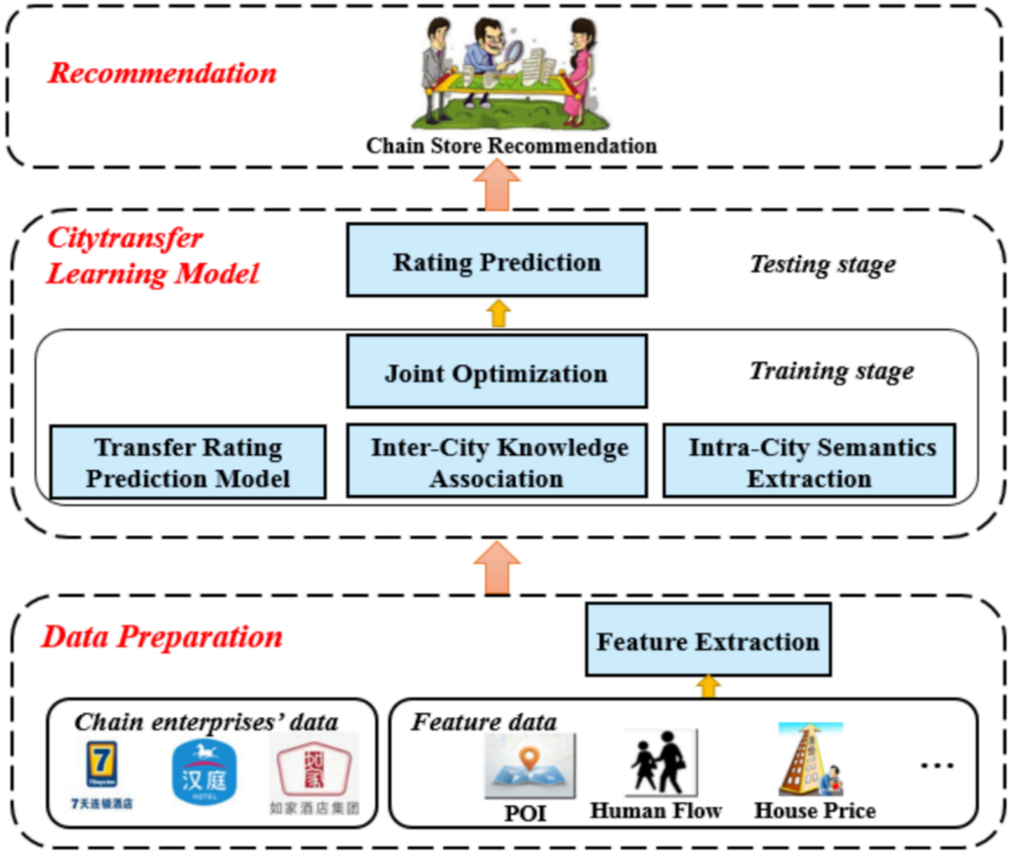


Fig. 2. The CityTransfer Framework.

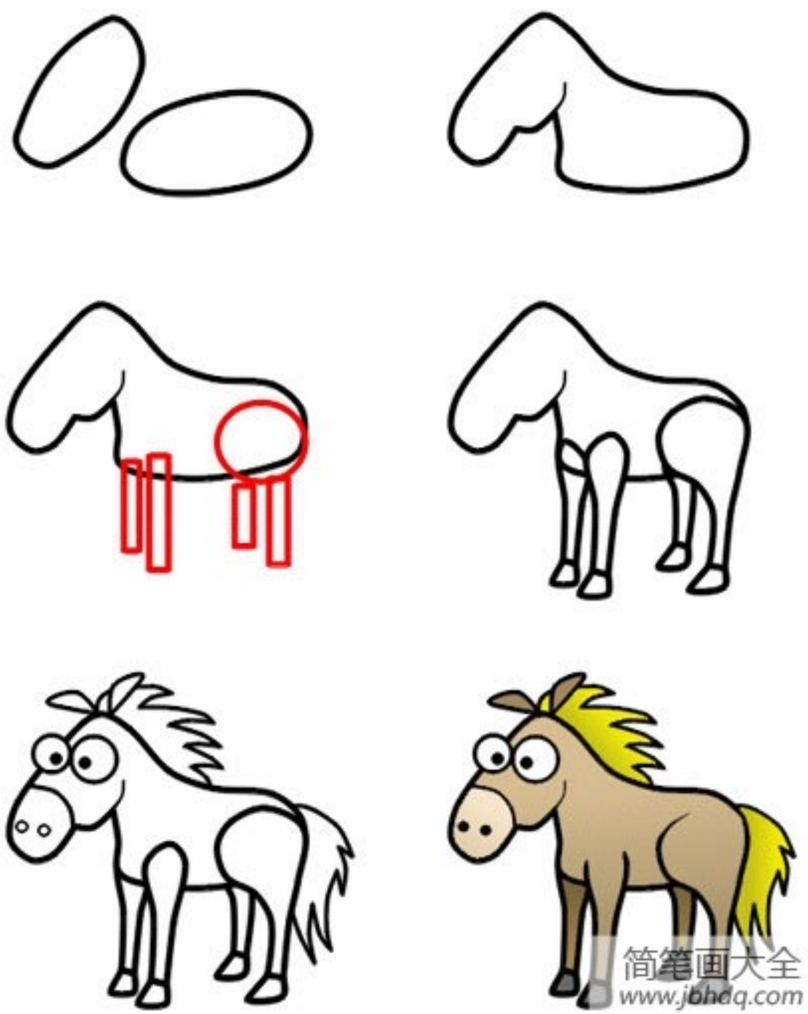
[Guo, 2018]

Take-Away Messages

- **Spatiotemporal Data + Urban Computing**
  - Sensing
  - Modeling
  - Clustering
  - Factorization

} **tensors**

- **Future: deep learning, transfer learning…**



Related Publications

1	**Longbiao Chen, Leye Wang, Daqing Zhang, Gang Pan, Cheng Wang, et al.** RADAR: Road Obstacle Identification for Disaster Response Leveraging Cross-Domian Urban Data	**UBICOMP 2018**	CCF A
2	**Longbiao Chen, Leye Wang, Daqing Zhang, Gang Pan, et al.** Dynamic Cluster-Based Over-Demand Prediction in Bike Sharing Systems	**UBICOMP 2016**	CCF A
3	**Longbiao Chen, Dingqi Yang, Daqing Zhang, Gang Pan, et al.** Bike Sharing Station Placement Leveraging Heterogeneous Urban Open Data	**UBICOMP 2015**	CCF A
4	**Longbiao Chen, Leye Wang, Chao Chen, Daqing Zhang, Gang Pan, et al.** Container Throughput Estimation Leveraging Ship GPS Traces and Open Data	**UBICOMP 2014**	CCF A
5	**Longbiao Chen, Leye Wang, Daqing Zhang, Gang Pan, et al.** Container Port Performance Measurement and Comparison Leveraging Ship GPS Traces and Maritime Open Data	**T-ITS 2016**	JCR II
6	**Longbiao Chen, Dingqi Yang, Daqing Zhang, Gang Pan, et al.** Fine-Grained Urban Event Detection and Characterization Based on Tensor Co-Factorization	**T-HMS 2017**	JCR III
7	**Longbiao Chen, Dingqi Yang, Daqing Zhang, et al.** NationTelescope: Monitoring and Visualizing Large-Scale Collective Behavior in LBSNs	**Elsevier JNCA**	JCR III
8	**Longbiao Chen, Daqing Zhang, Gang Pan, Jérémie Jakubowicz, et al.** Understanding Bike Trip Patterns Leveraging Bike Sharing System Open Data	**Springer FCS**	JCR IV

Total: **32** CCF-A: **4** SCI-II: **4** Citation: **393**

Thank you!

- Longbiao CHEN (陈龙彪) PhD, Assistant Professor
- Xiamen University, Xiamen, China
- longbiaochen@xmu.edu.cn