Collective and Semantic Exploration of Human Mobility Data

—Modeling, Representation, and Applications

Yanjie Fu

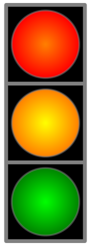
MISSOURI S&T

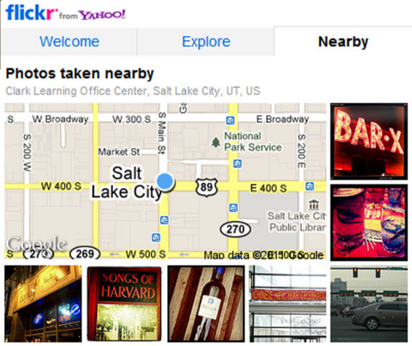
Outline

- **Background and Motivation**
- Modeling Spatiotemporal Dynamics
- Collective Representation Learning
- Applications
- Conclusion and Future Work

Pervasive Sensing for Human Movements

IoT, GPS, wireless sensors, mobile Apps





airbnb

foursquare
CHECK-IN HERE

Human Mobility Data

- **Human mobility data are people's movement trajectories which can be the traces of**
  - devices: phones, WIFIs, network stations, RFID
  - vehicles: bikes, taxicabs, buses, subways, light-rails
  - location based services: geo-tweets (Facebook, Twitter), geo-tagged photos (Flickr), check-ins (Foursquare, Yelp)

Taxicab GPS Traces

Bus Traces

Phone Traces

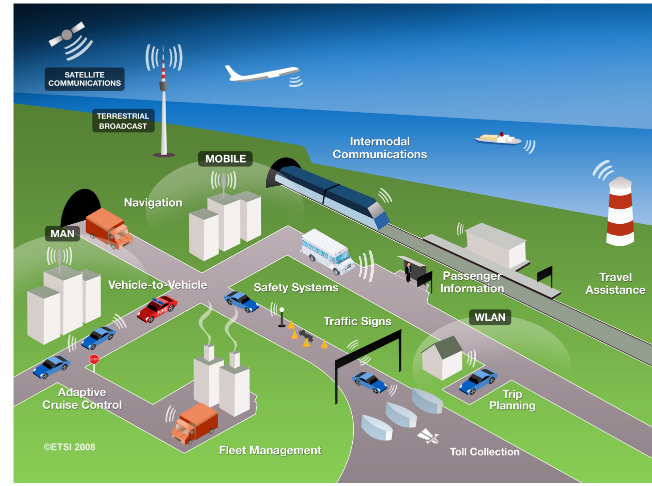
Mobile Check-ins

Represent the spatial, temporal, social, and semantic contexts of dynamic human behaviors within and across regions

Important Applications of Human Mobility

Understand Human Movement Patterns

Automated User Profiling



Intelligent Transportation Systems

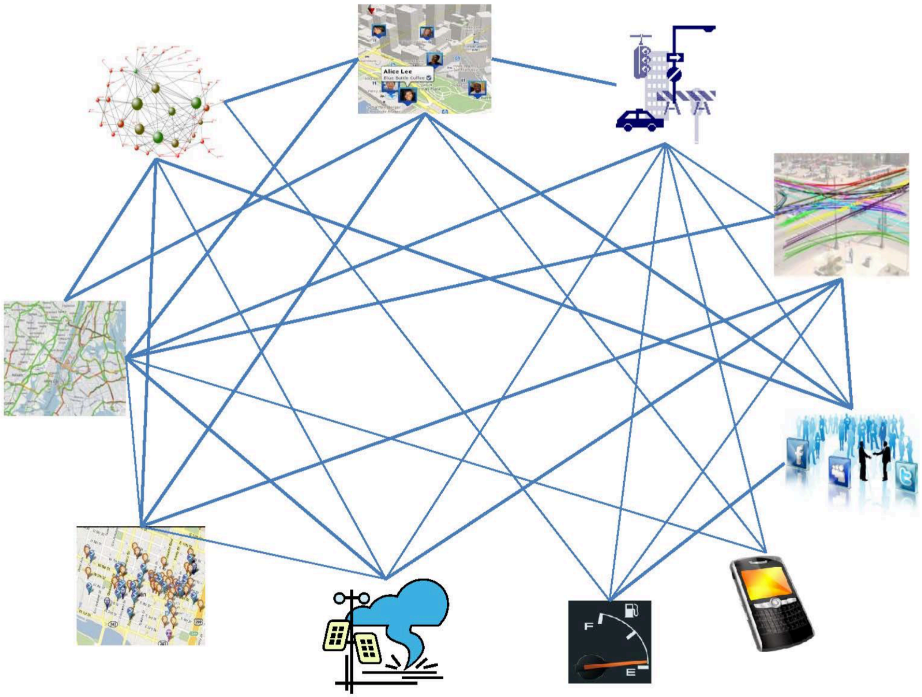
Smart and Connected Communities

Smart Heath Care

City Governance and Emergency Management

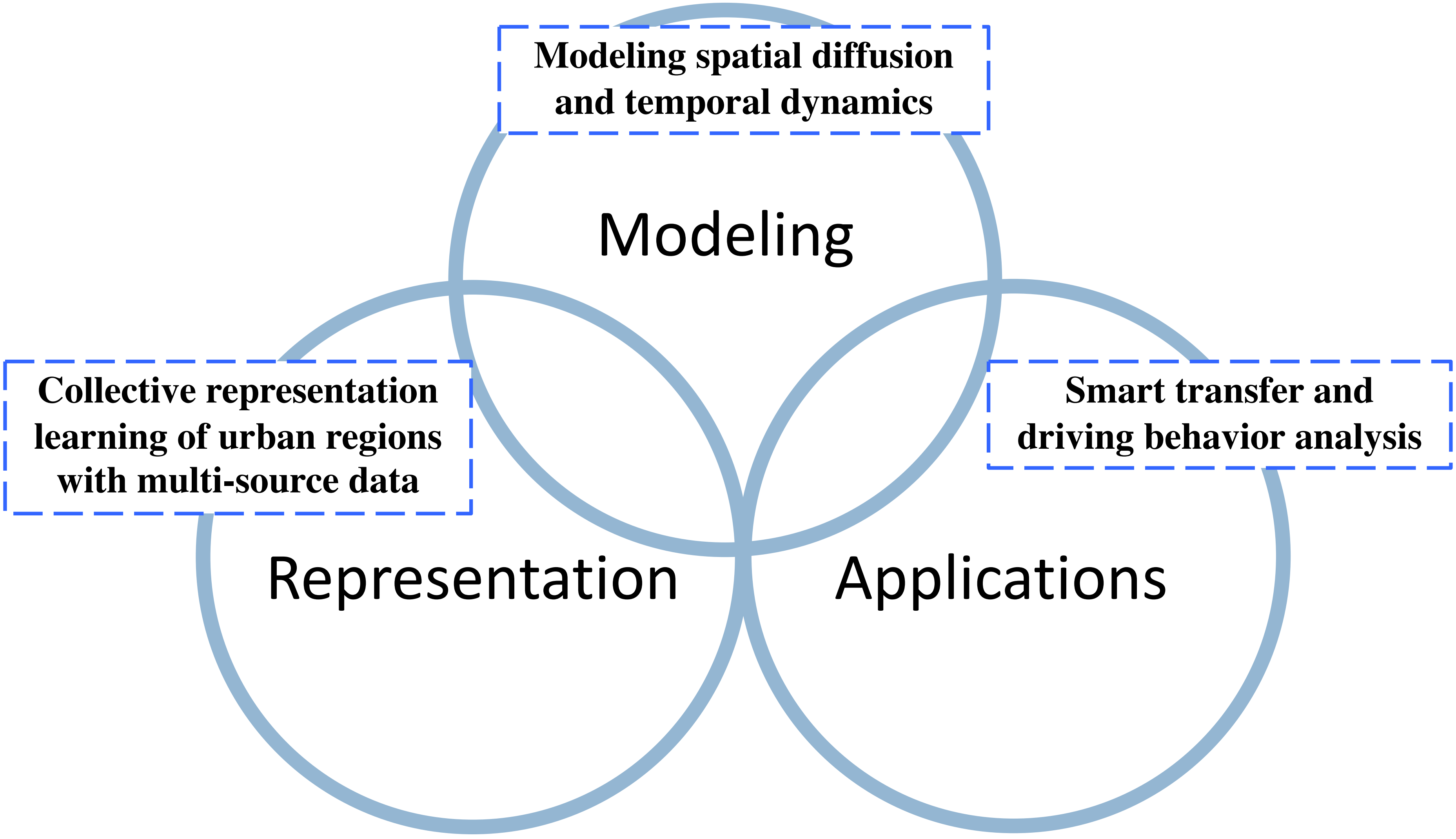
Unprecedented and Unique Complexity

- Spatio-temporal-textual
- Networked
  - Collectively-related
- Heterogeneous
  - Multi-source
  - Multi-domain
  - Multi-format
- Semantically-rich
  - Trip purposes
  - User profiles
  - Outlier events/incidents
  - Spatial configuration and urban functions of regions



The Overview of The Talk

Collective and Semantic Exploration



Outline

- Background and Motivation
- **Modeling Spatiotemporal Dynamics**
- Collective Representation Learning
- Applications
- Conclusion and Future Work

Spatiotempral Dynamics Modeling

f() =

Past Current Future

	t_1	t_2							t_k
R1	0	0	0	0	0	0	4	0	0
R2	0	0	0	10	0	0	0	0	0
R3	0	0	0	0	0	30	0	0	0
R4	0	0	0	8	0	0	0	0	5
R5	0	6	0	0	0	0	0	0	0
R6	0	0	0	0	0	0	15	0	0
R7	0	0	12	0	0	0	0	0	0
R8	0	0	0	0	20	0	0	0	0

Where?
Location

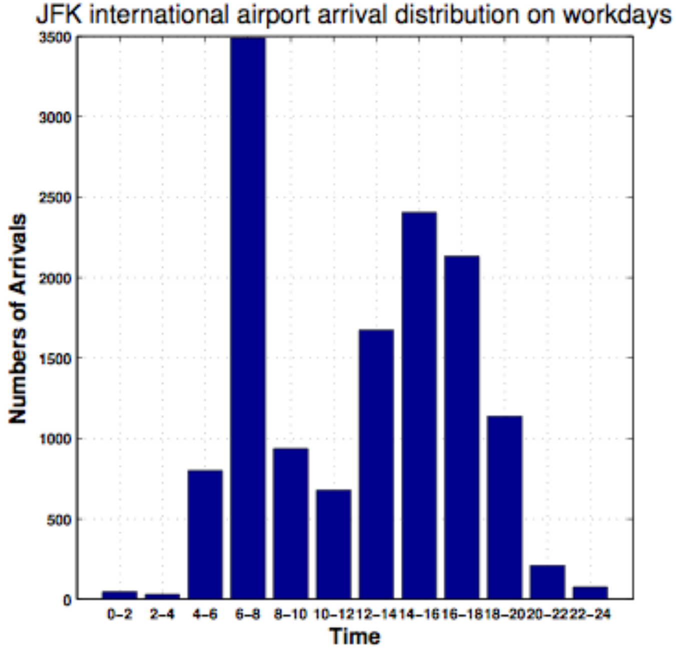
When?
Time

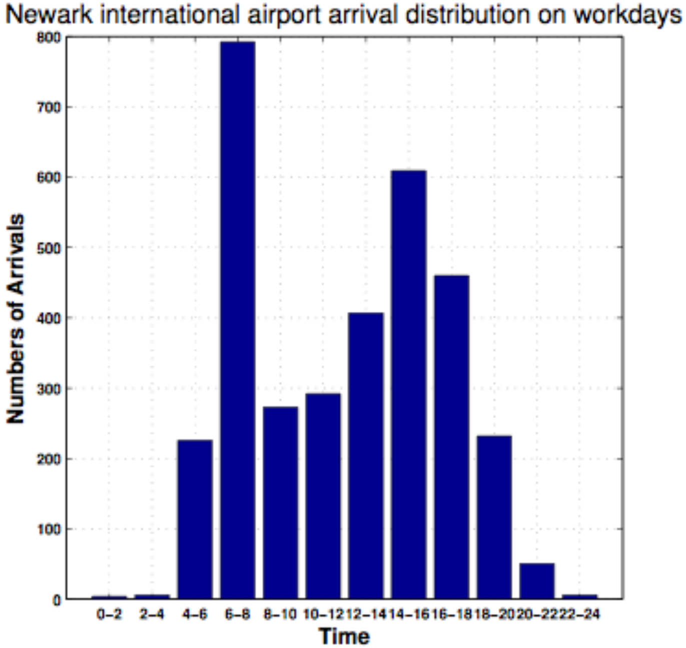
What to do?
Trip Purpose

Learn the patterns of spatiotemporal arrival matrix, and forecast 3W (where, when, what) of future arrivals

Human Mobility Synchronization

Taxi arrival distributions of JFK Airport and Newark Airport

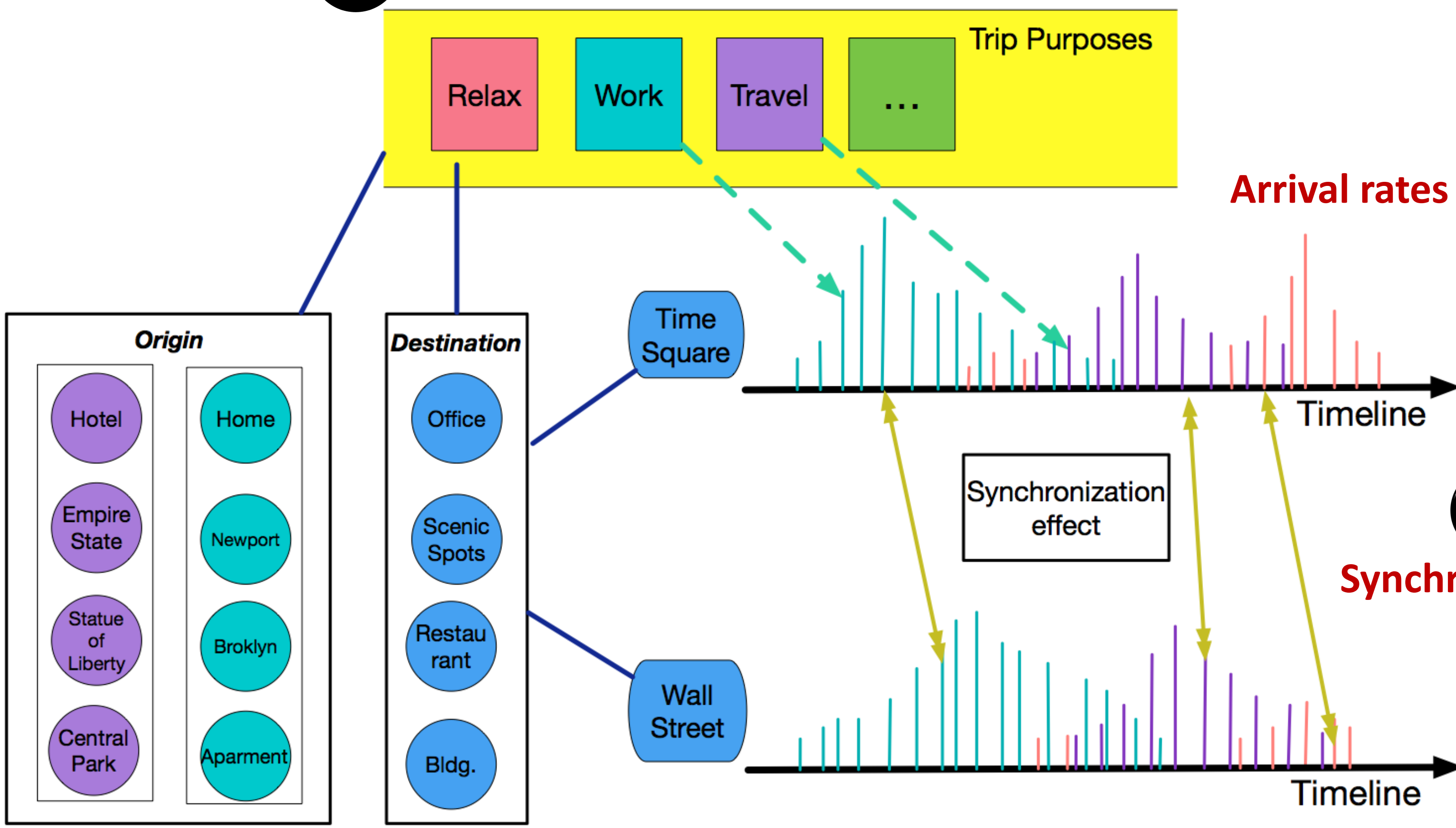




If two regions are similar in urban functions, they show similar arrival patterns

Linking Arrivals, Regions and Purposes

2 **Trip purposes**

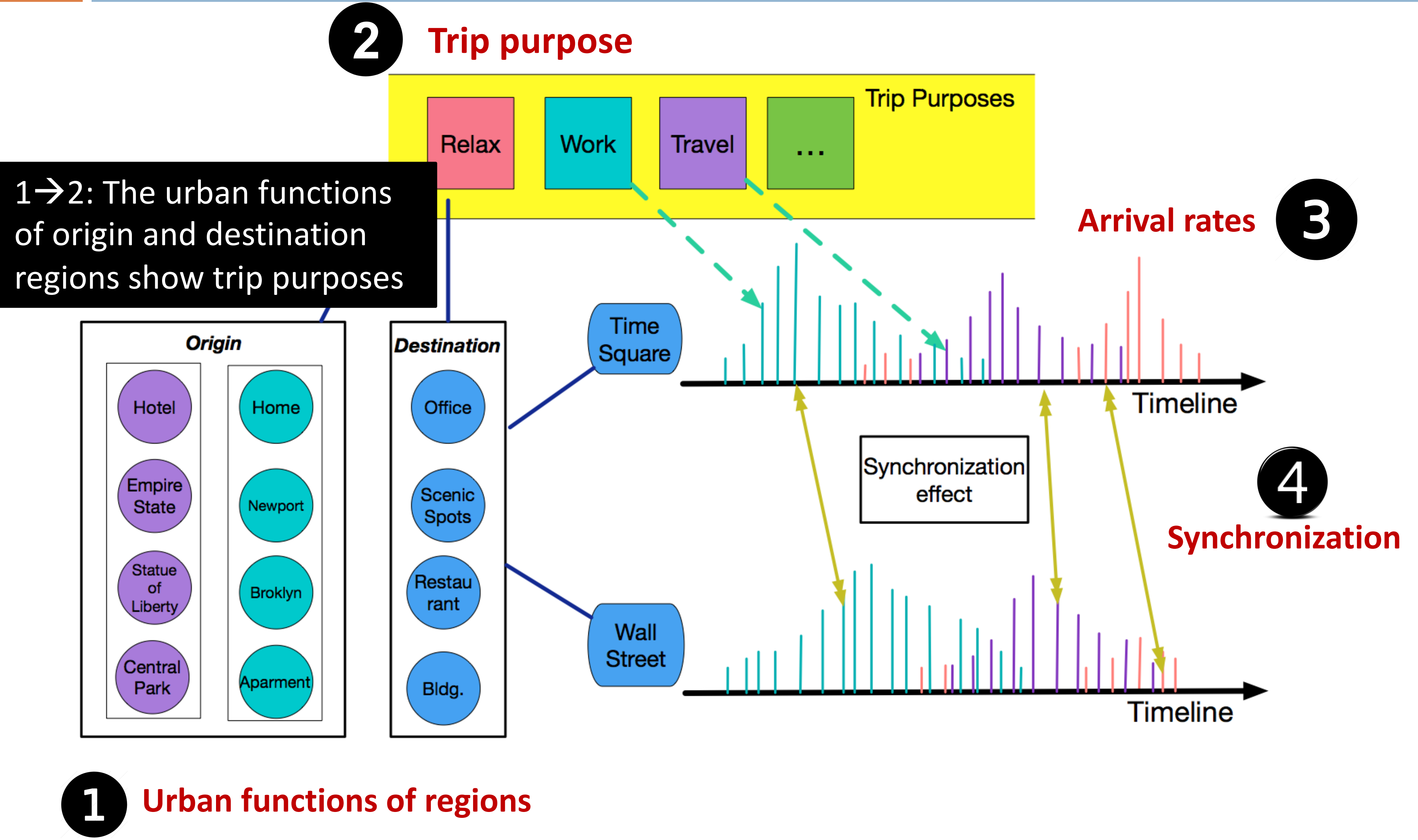


3 **Arrival rates**

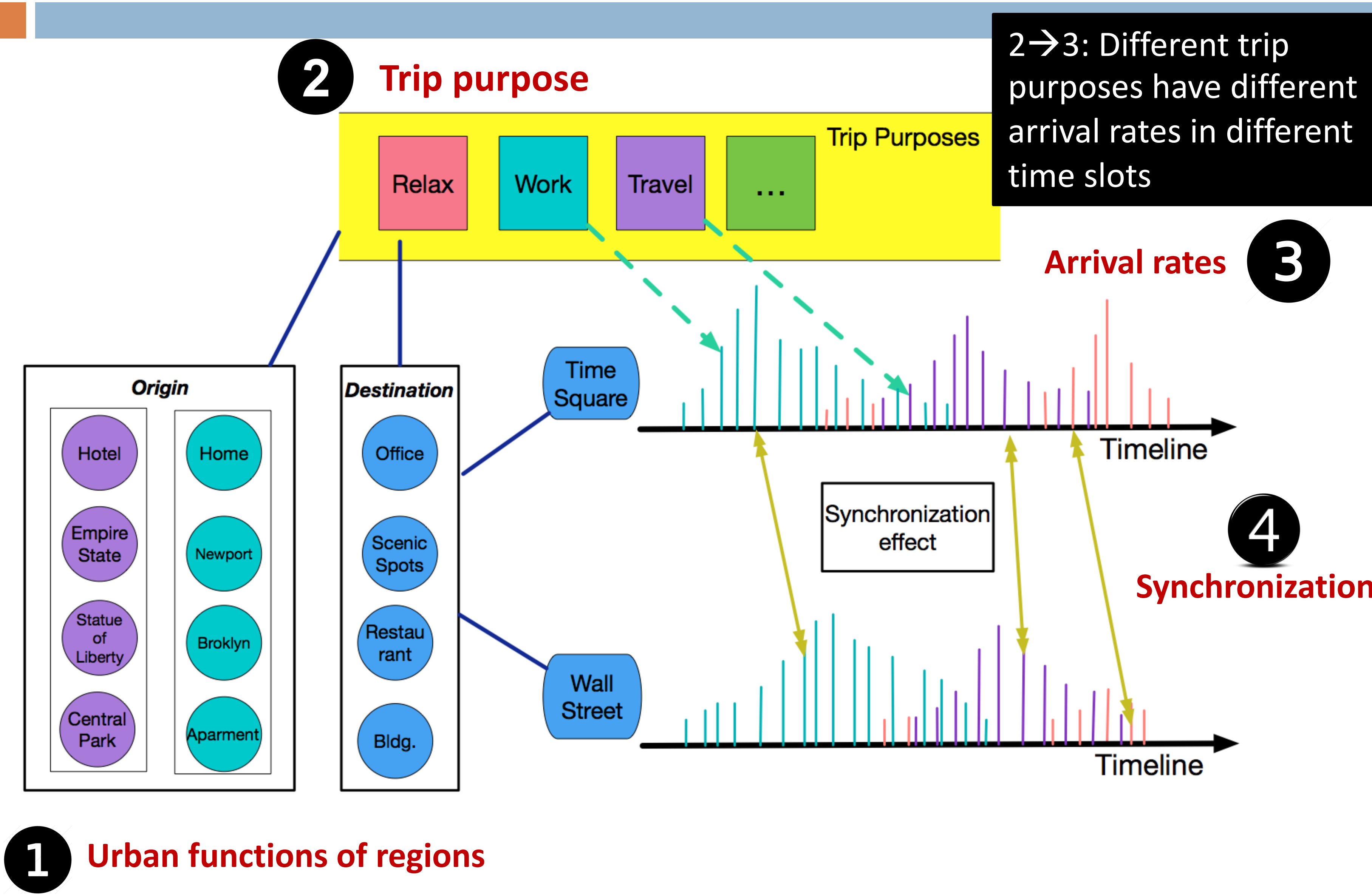
4 **Synchronization**

1 **Urban functions of regions**

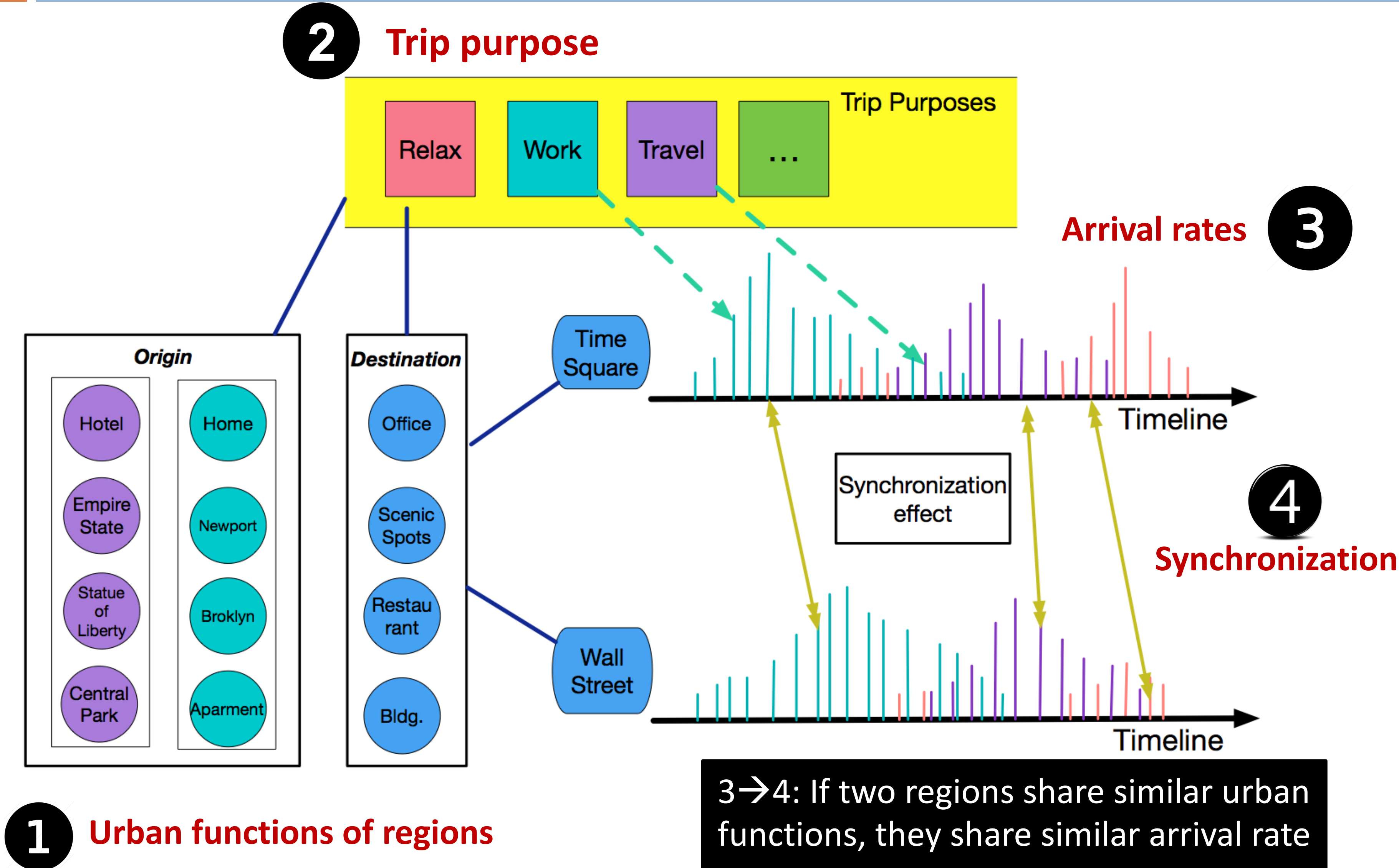
Linking Arrivals, Regions and Purposes



Linking Arrivals, Regions and Purposes



Linking Arrivals, Regions and Purposes



3→4: If two regions share similar urban functions, they share similar arrival rate patterns

Framework

Modeling the arrivals of a single region for single trip purpose

Modeling the arrivals of a single region for multiple trip purposes

Modeling the arrivals of multiple regions for multiple trip purposes

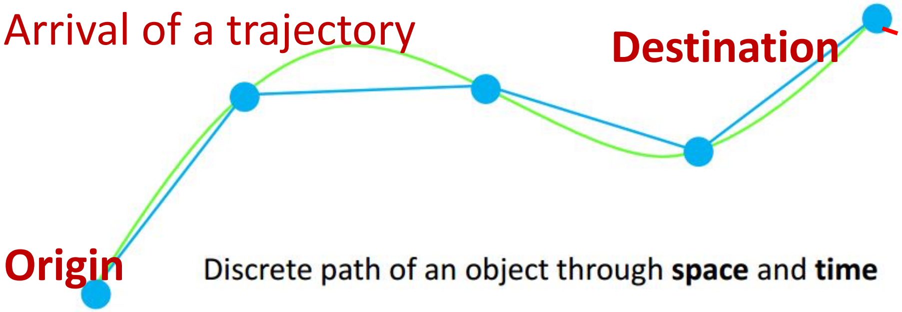
Incorporating human mobility synchronization effects

Incorporating the modeling of origin and destination regions

Integrating human knowledge

Convert Trajectories To Arrival Events

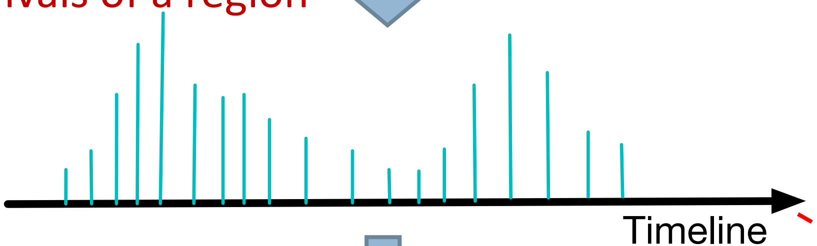
Arrival of a trajectory



Discrete path of an object through **space** and **time**

$$T = < (\boldsymbol{P_1}, t_1), (\boldsymbol{P_2}, t_2), \ldots, (\boldsymbol{P_n}, t_n) >$$

Arrivals of a region



Arrivals of a map

	t_1	t_2							t_k
R1	0	0	0	0	0	0	4	0	0
R2	0	0	0	10	0	0	0	0	0
R3	0	0	0	0	0	30	0	0	0
R4	0	0	0	8	0	0	0	0	5
R5	0	6	0	0	0	0	0	0	0
R6	0	0	0	0	0	0	15	0	0
R7	0	0	12	0	0	0	0	0	0
R8	0	0	0	0	20	0	0	0	0

- Each trajectory is a five-element arrival event is : $E_n = \{g_n, z_n, t_n, w_n^d, w_n^o\}$
  - g_n: the trip purpose of the n-th arrival
  - t_n: the timestamp of the n-th arrival
  - w_n^d: POIs of destination region
  - w_n^o: POIs of origin region
- For each region, we organize trajectories as a sequence of arrivals: $E = \{E_1, E_2, \ldots, E_N\}$
- Benefits: support multi-source mobility data, e.g., trajectories, check-ins

Modeling Arrivals of Single Region for A Single Trip Purpose

- **Modeling mobility arrivals as a stochastic point process**

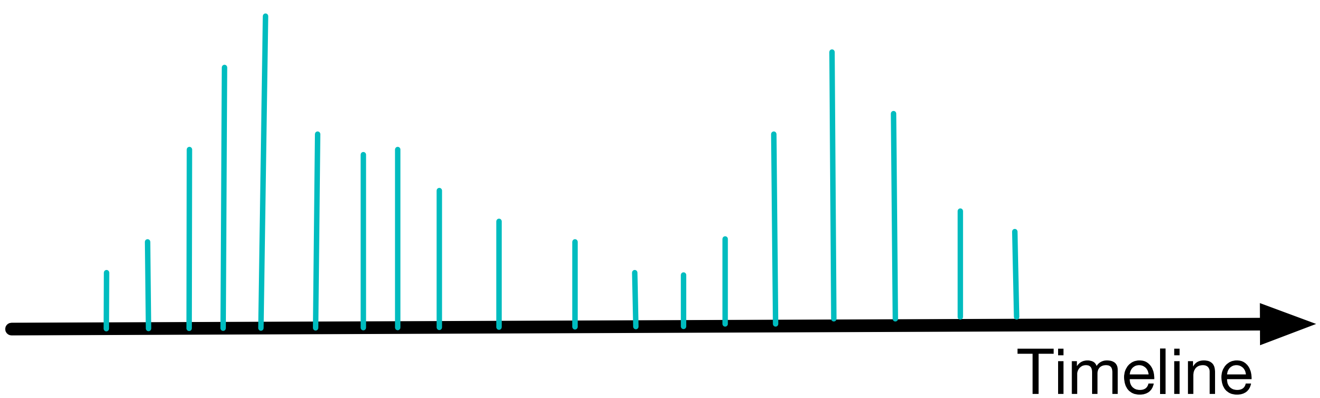
Base arrival rate

Self-exciting effect via backward memory using decay function

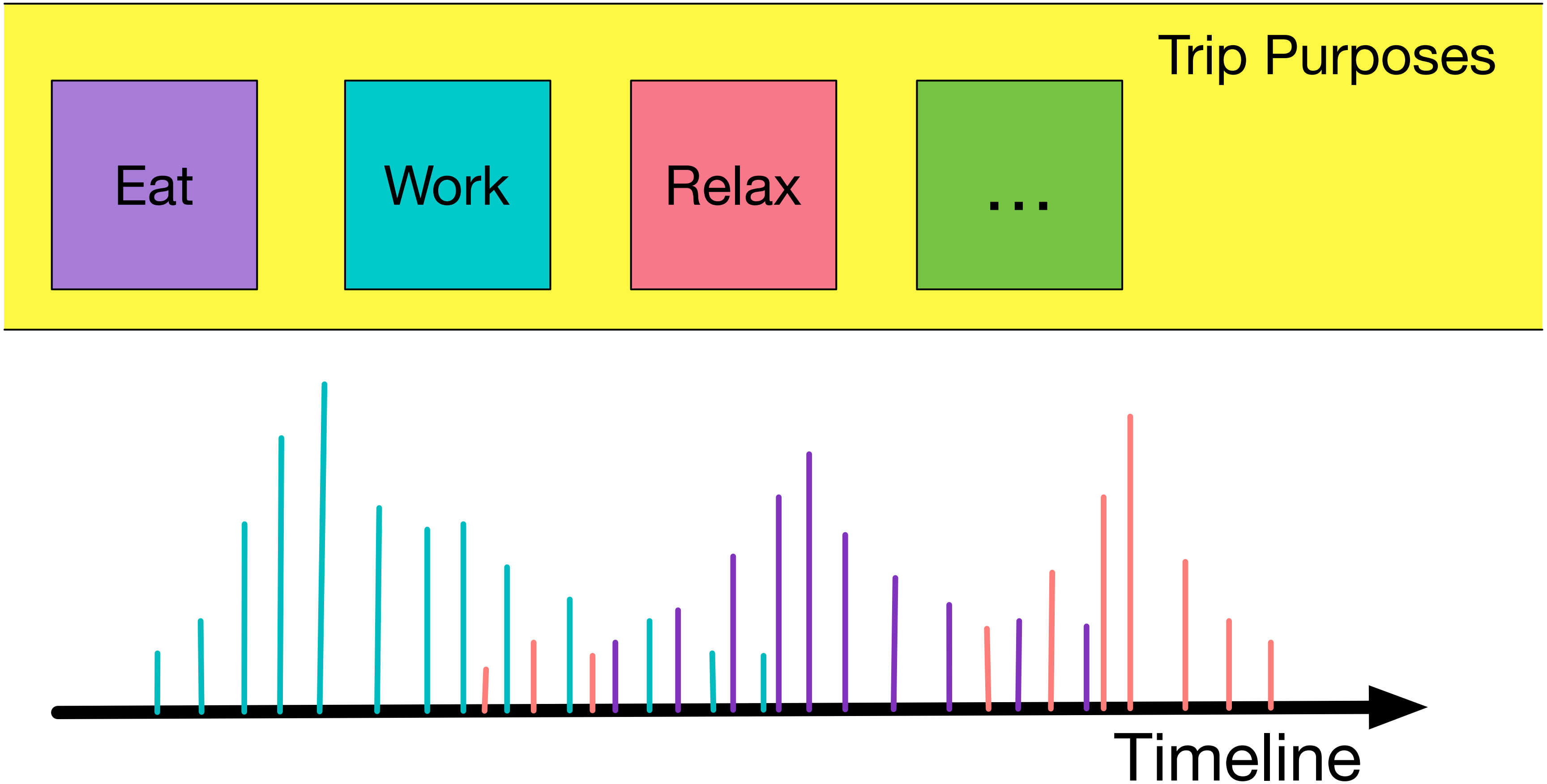
  - Hawkes Process: $\lambda(t) = \mu + \boxed{\int_{-\infty}^{t} g(t-s)dN(s)}$

Current-past temporal dependency

  - Self-exciting for multi-peak gradually-excited human activities
    - The to-work arrivals at 9am are self-excited by the increasingly intensive to-work arrivals at 8am



Modeling Arrivals of Single Region for Multiple Trip Purposes (1)



Mobility arrivals in the i-th region :

$$\lambda_i = \lambda_{i,eat}(t) + \lambda_{i,work}(t) + \lambda_{i,relax}(t) + \cdots$$

Modeling Arrivals of Single Region for Multiple Trip Purposes (2)

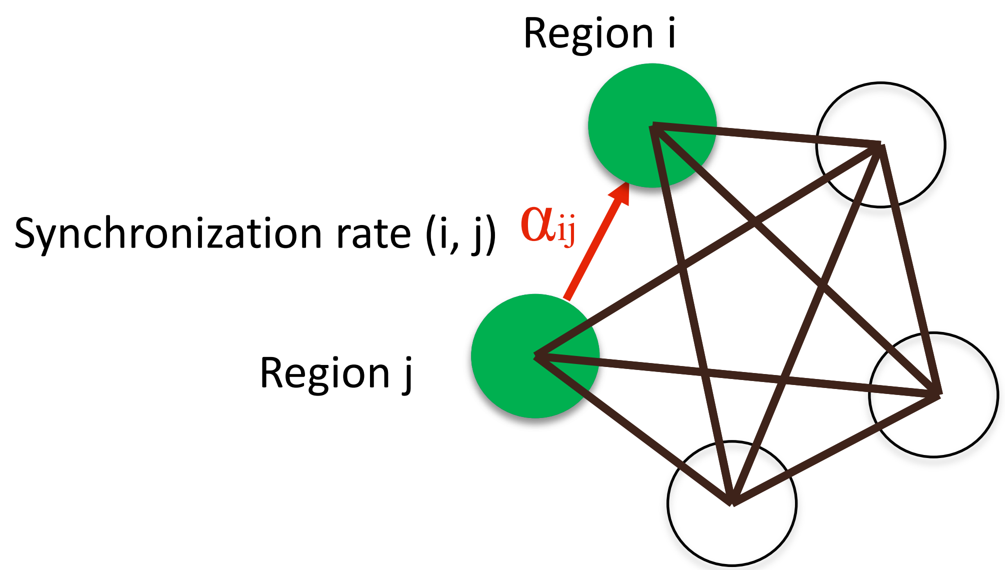
- **Mixture Hawkes processes with respect to different trip purposes**
  - $\lambda_{i,m}(t) = \mu_{i,m} + \int_{-\infty}^{t} g(t-s)dN(s) = \mu_i * \gamma_m + \int_{-\infty}^{t} g(t-s)dN(s)$

    Decouple the base rates of location and trip purpose to reduce the number of parameters

    - i: the i-th region
    - m: the m-th trip purpose
    - $\mu_{i,m}$: the base rate that region i get visited with trip purpose m
    - μ_i: the base visit rate of region i
    - γ_m: the base visit rate of trip purpose m
    - $g(t-s)$: memory decay function

Synchronization Effect Across Regions

- Region synchronization graph
  - Road networks as graph
  - Regions as nodes in the graph
  - Synchronization rate between two regions as the edge weight between two nodes



If Region(i) and Region(j) are both office areas, and many to-work arrivals are observed in Region(j), then it is likely to observe many to-work arrivals in Region(i)

Modeling Synchronization Effect Across Regions in Mixture Hawkes Processes

- **Integrating the synchronization effects across regions into mixture Hawkes processes**
  - $\lambda_{i,m}(t) = \mu_i * \gamma_m + \boxed{\sum_{j=1}^{I} \alpha_{ji}^{m}} \int_{-\infty}^{t} g(t-s)dN(s)$

Base arrival rate

Sync effect when j!=i
(region-region peer dependency)

Self-exciting effect when j==i
(past-current temporal dependency)

- Synchronization (Mutual-exciting)
  - The arrivals are not just self-excited by previous arrivals within a region, but also excited by the arrivals of peer regions
  - Example: The to-work arrivals of the i-th region at 9am are excited by the to-work arrivals of the j-th similar region at 9am

Enhance Trip Purpose Labeling via Modeling Origin and Destination Regions

Origin (residential)

Working purpose

Destination (office)

The urban functions of origin and destination regions can jointly show trip purposes

Incorporating the Joint Modeling of Origin and Destination Regions

Analogies between region modeling and textual mining

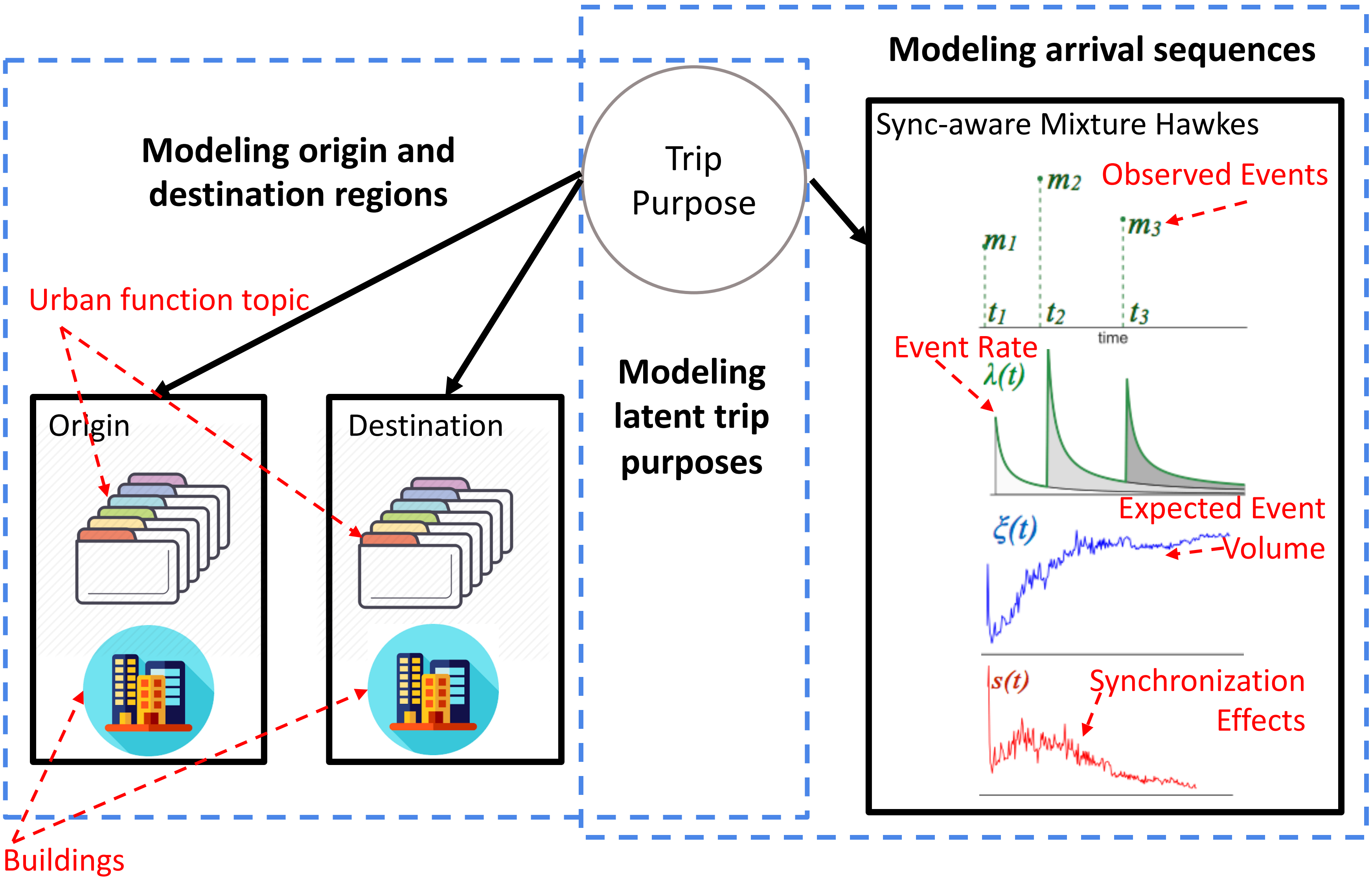
Region-Building	Document-Word
Region	Document
Building category	Word
Urban function	Topic

Topic Modeling of Origin and Destination

- **Probabilistic generative model of buildings in origin and destination regions**
  - Draw a trip purpose for each trip
  - Draw buildings of origin region from the trip purpose
  - Draw buildings of destination region from the trip purpose

- Generate a purpose $m \sim \text{Multi}(\pi)$
- Generate the POI Topic for the origin $z_o \sim Multi(\Phi_{mz})$
  - For each POI w^o in the origin neighborhood
  - Generate the POI $w^o \sim Multi(\beta_{zw})$
- Generate the POI Topic for the origin $z_d \sim Multi(\Phi_{mz})$
  - For each POI w^d in the origin neighborhood
  - Generate the POI $w^d \sim Multi(\beta_{zw})$

Solving the Co-optimization (1)



Solving the Co-optimization (2)

1. Training Data

Trip purposes | Urban function topics | Time stamp | Origin and destination regions

$$(G, z, t, W) = \{(G_n, z_n, t_n, W_n)\} \text{ with } t_0 = 0 \text{ and } t_N = T$$

2. Likelihood Function

Modeling trip purposes | Modeling origin and destination regions | Modeling arrival sequences

$$L(G, t, \mathbf{W}) = \prod_{n=1}^{N} p(G_n)p(\mathbf{W}_n^o, \mathbf{W}_n^d | G_n)p(t_n | G_n)$$

3. A Lower Bound

$$\mathcal{L}(t, W) \equiv \log\left(\int_{\{(G,z)\}} L(G, z, t, W) d\{(G, z)\}\right)$$

$$\geq \int_{\{(G,z)\}} \log\left(L(G, z, t, W)\frac{d\{(G, z)\}}{dq(\{(G, z)\})}\right) dq(\{(G, z)\})$$

$$= \mathbf{E}_q[\mathcal{L}(G, z, t, W)] + \mathcal{E}[q] \equiv \mathfrak{L} ,$$

Surrogate Function of Likelihood

4. Parameter Update Rules

$$\zeta_{m,r}^o \propto \prod_{c=1}^{C} (\beta_{rc})^{\epsilon W_{nc}^o}$$

$$\zeta_{m,r}^o \propto \prod_{c=1}^{C} (\beta_{rc})^{\epsilon W_{nc}^d}$$

$$\beta_{rc} \propto \sum_n \sum_m \phi_{nm}(\zeta_{m,c}^o W_{m,c}^o + \zeta_{m,c}^d W_{m,c}^d).$$

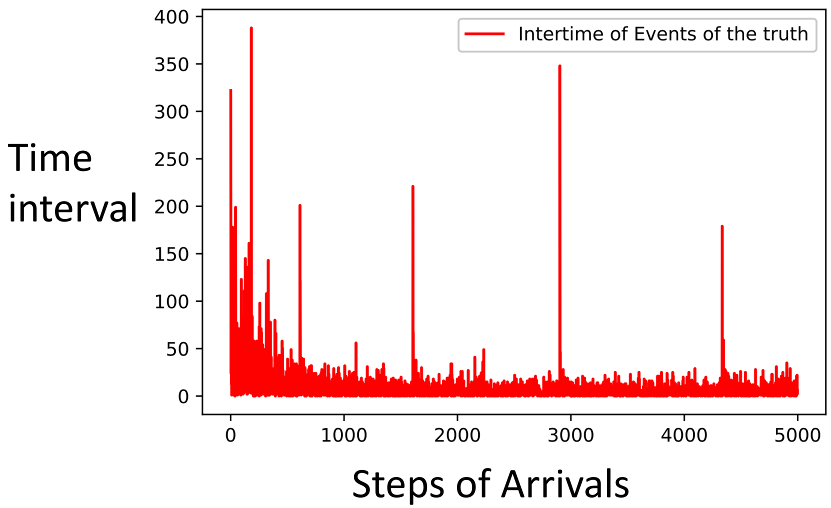
$$\pi_m \propto \sum_{n=1}^{N} \phi_{nm} ,$$

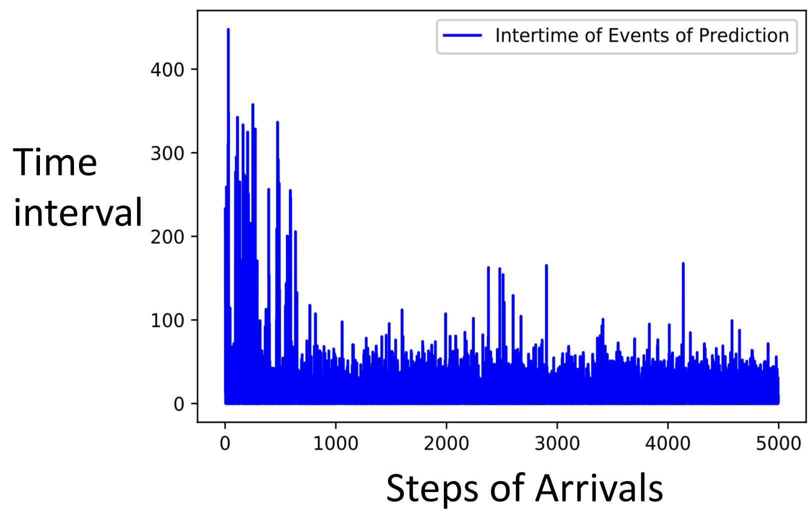
$$\mu_i \propto \frac{\sum_{m=1}^{M} \phi_{nm} \sum_{n=1}^{N} \delta_{i_n,i}\eta_{nn}^m}{\sum_{m=1}^{M} \gamma_m T} ,$$

$$\gamma_m \propto \frac{\sum_{n=1}^{N} \phi_{nm}\eta_{nn}^m}{\sum_{i=1}^{I} \mu_i T} ,$$

$$\alpha_{ij}^m = \frac{\sum_{n=1}^{N} \sum_{l=1}^{n-1} \phi_{nm}\phi_{lm}\eta_{ln}^m \delta_{i_l i}\delta_{i_n j}}{\sum_{n=1}^{N} \mathcal{G}(T - t_n)\phi_{nm}\delta_{i_n i}} ,$$

Study of Forecasting Next Arrivals



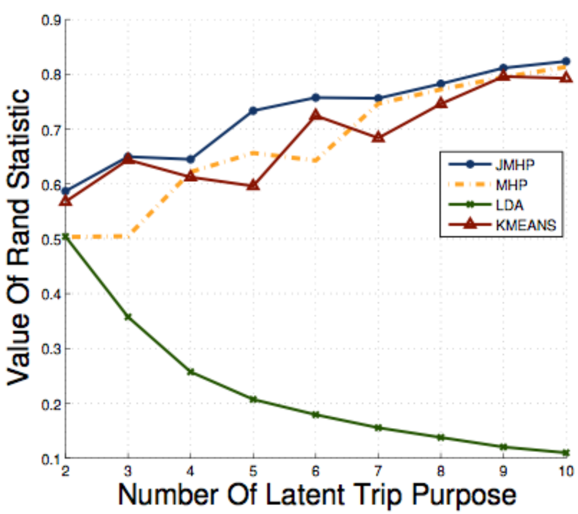


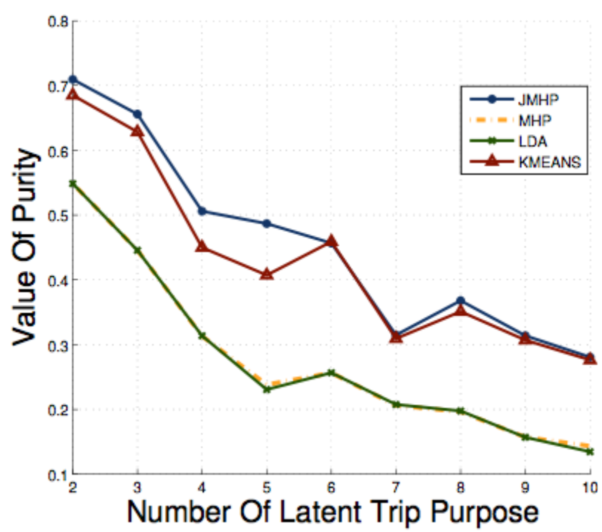
Benchmark time intervals of every two arrival events

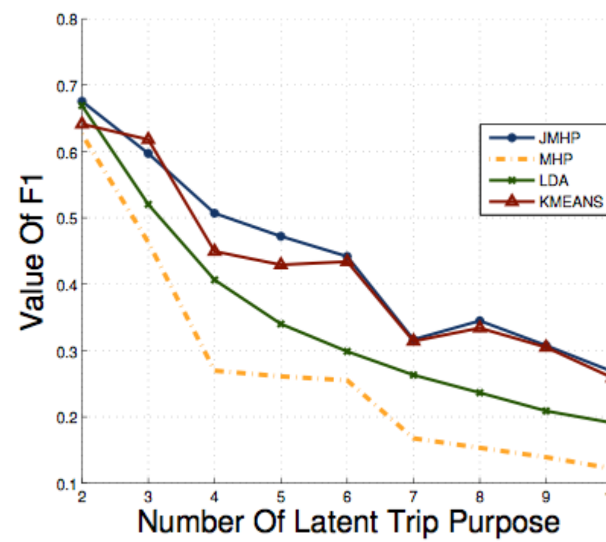
Predicted time intervals of every two arrival events

Study of Trip Purpose Clustering

- Experiments on synthetic data: validate the identified trip purposes
- Synthetic data generation: Ogata's modified thinning algorithm for sampling arrival sequences
- Task: Clustering the trajectories based on the inferred trip purposes
- Baseline methods: MHP, LDA, K-means
- Metrics: purity, F1-Measure, Rand Statistics







Study of Trip Purpose Interpretation

- Data
  - Taxi trips of NYC: 7-day taxi trips, millions of GPS trajectories, 152 valid regions
  - Point of Interests data of NYC

Identified trip purposes

nightlife dining work shopping

TOPIC 1	prob.	TOPIC 2	prob.	TOPIC 3	prob.	TOPIC 4	prob.	TOPIC 5	prob.
Bar	0.1884	Chinese Rest.	0.1286	Bar	0.0933	Office	0.3331	Clothing Store	0.0995
Home	0.0953	Italian Rest.	0.0913	Italian Rest.	0.0565	General Entertain	0.1035	Cafe	0.0693
Nightclub	0.0571	Asian Rest.	0.0541	American Rest.	0.0442	Hotel	0.1023	Office	0.0574
Event Space	0.0495	Tea Room	0.0481	Wine Bar	0.0373	Building	0.0869	Coffee Shop	0.0535
Cocktail Bar	0.0495	Bar	0.0472	Sushi Rest.	0.0319	Event Space	0.0593	Cosmetics Shop	0.0419
Lounge	0.0495	Spa or Massage Parlor	0.0416	Mexican Rest.	0.0306	Sandwich Place	0.0376	General Entertain	0.0408
Speakeasy	0.0471	Salon or Barbershop	0.0403	Lounge	0.0297	Hotel Bar	0.0342	French Rest.	0.0406
Breakfast Spot	0.0382	Vietnamese Rest.	0.039	Pizza Place	0.0278	Lounge	0.0342	High Tech Outlet	0.0388
French Rest.	0.0334	Art Gallery	0.0342	Coffee Shop	0.0256	Other Outdoors	0.0298	Salon or Barbershop	0.0368
Boat or Ferry	0.0316	Cocktail Bar	0.0316	Salon or Barbershop	0.0256	Performing Arts Venue	0.0289	Miscellaneous Shop	0.0331

TOPIC 6	prob.	TOPIC 7	prob.	TOPIC 8	prob.	TOPIC 9	prob.	TOPIC 10	prob.
College Acad.	0.0808	Park	0.1343	Art Gallery	0.2773	American Rest.	0.1023	Home	0.2005
Food Truck	0.0756	Other Outdoors	0.1	Park	0.1021	Deli or Bodega	0.0619	Building	0.0591
University	0.0653	Scenic Lookout	0.0767	Other Outdoors	0.0892	Office	0.0569	Deli or Bodega	0.0471
College Library	0.0639	General Travel	0.0753	Cafe	0.0555	Pizza Place	0.0464	Pizza Place	0.0442
General College/University	0.0573	Building	0.074	Playground	0.049	Bar	0.0448	Laundromat or Dry Cleaner	0.0342
College Dorm	0.0565	Airport	0.074	Automotive Shop	0.0386	Food Truck	0.0434	Coffee Shop	0.0317
Cafe	0.0499	Harbor or Marina	0.0616	Event Space	0.033	Sandwich Place	0.0392	Drugstore or Pharmacy	0.0291
Plaza	0.0485	Taxi	0.0534	Strip Club	0.0265	Coffee Shop	0.0346	Chinese Rest.	0.0256
Park	0.0382	Government Building	0.048	Sculpture Garden	0.0241	Burger Joint	0.0326	Mexican Rest.	0.0236
College Classroom	0.0374	Seafood Rest.	0.0343	Plaza	0.0233	Cafe	0.0307	Apartment Building	0.0206

schooling sightseeing home

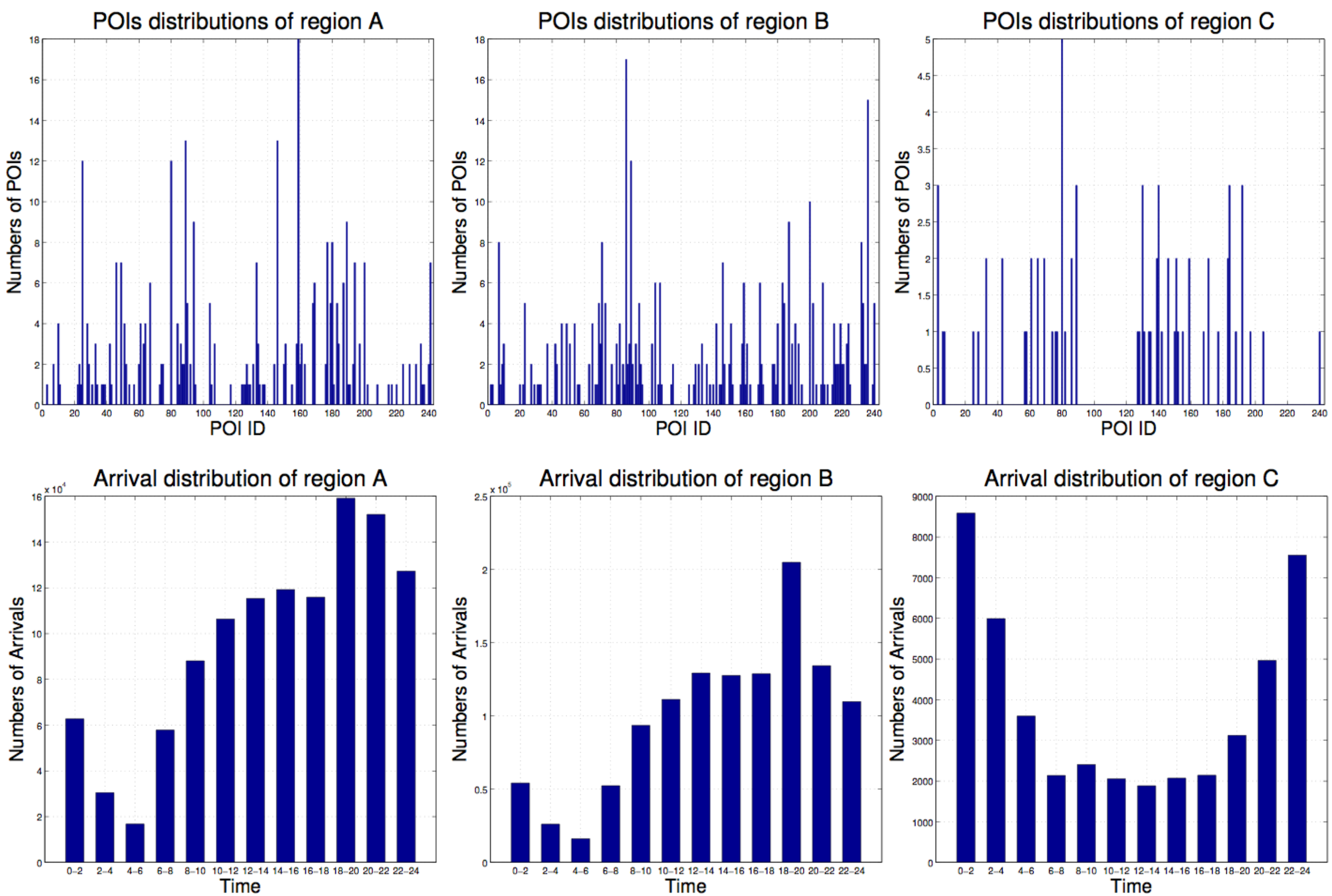
Study of Synchronization Effect

α_{AB} = 8.27066832 α_{AC} = 0.00711464

A and B have a higher synchronization rate

A and C have a lower synchronization rate

Study of Synchronization Effect



The POI and arrival distributions of A,B, C are consistent with the pairwise sync rates of A, B, C

Summary

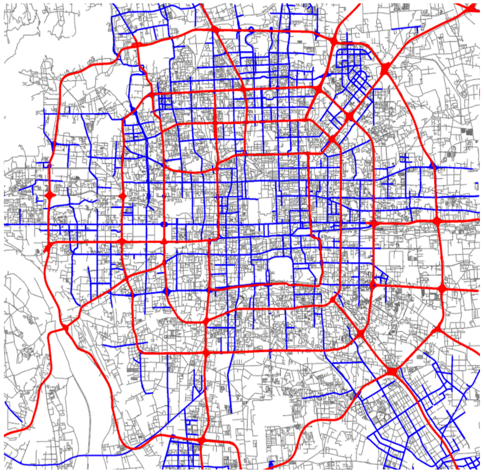
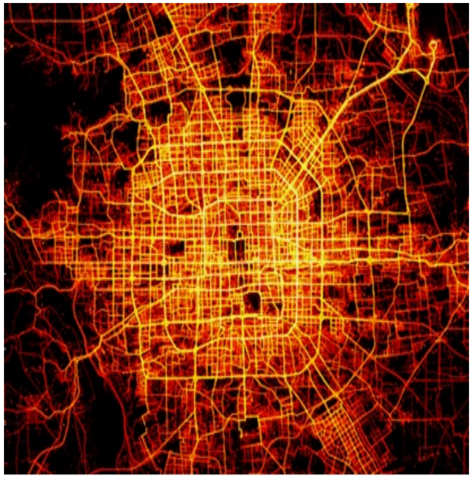
- **Task**
  - Modeling spatial diffusion and temporal dynamics of human mobility data
- **Property (provide in-depth understanding)**
  - Identify the synchronization property of human mobility
- **Modeling (make it predictable and traceable)**
  - Model human mobility as stochastic point processes
  - Develop a synchronization-aware mixture Hawkes model to jointly capture synchronization effects, mobility arrivals, urban regions, and trip purposes
  - Unify mobility arrival forecasting and trajectory semantic annotation

Outline

- Background and Motivation
- Modeling Spatiotemporal Dynamics
- **Collective Representation Learning**
- Applications
- Conclusion and Future Work

Spatial Representation Learning

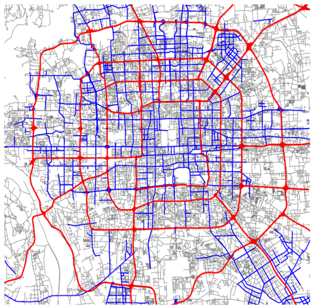
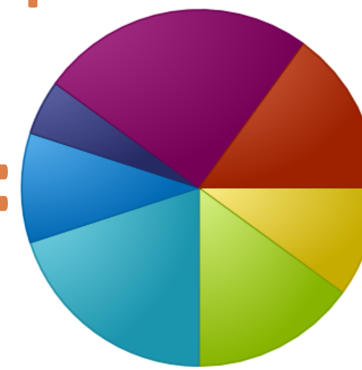
Spatial Objects (e.g., Regions) | **Single-source Human Mobility** | **Vector Representations**

f( , ) = 

- Given: urban regions, single-source human mobility data
- Objective: learn the vector representations of regions in a latent feature space
- Constraints: **similar regions share similar representations**

Collective Representation Learning with Multi-source Mobility Data

Spatial Objects (e.g., Regions) | **Multi-Source Human Mobility Data** | **Vector Representations**

f( ,  , ) = 

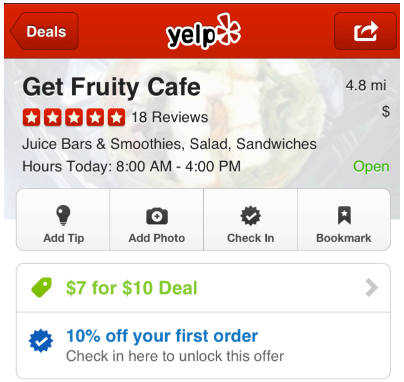
- Given: urban regions, multi-source human mobility data
- Objective: learn the vector representations of regions in a latent feature space
- Constraints
  - Similar regions share similar representations
  - **Integrate the mutual validation of multi-source human mobility patterns**

Why Collective Representation Learning?

- **Automated representation learning from widely-available data without domain experts**
  - Non-automated: Find domain experts, design variables, and extract vector representations
- **Automated fusion of multi-source unbalanced data**
  - Non-automated: Design features, select features, weigh features, weighted combination of features
- **Enable the availability of existing algorithms**
  - Enable classification, ranking, clustering, outlier detection for spatial contexts

The Patterns of Three Mobility Events

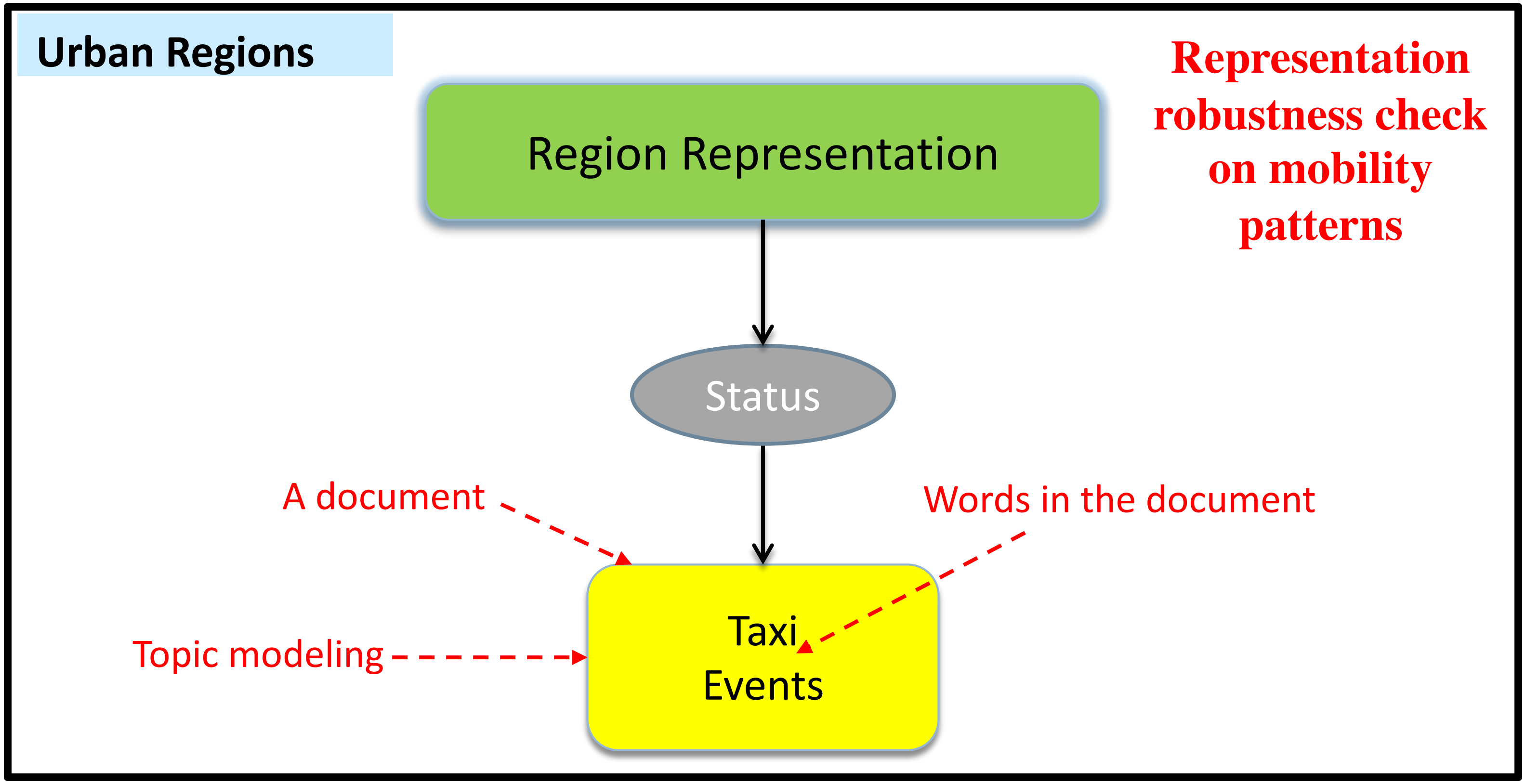
- **Checkin mobility pattern**
  - <day, hour, location category> of a checkin event



- **Taxi mobility pattern**
  - <day, hour, leaving or arriving> of a taxi pickup or delivery event

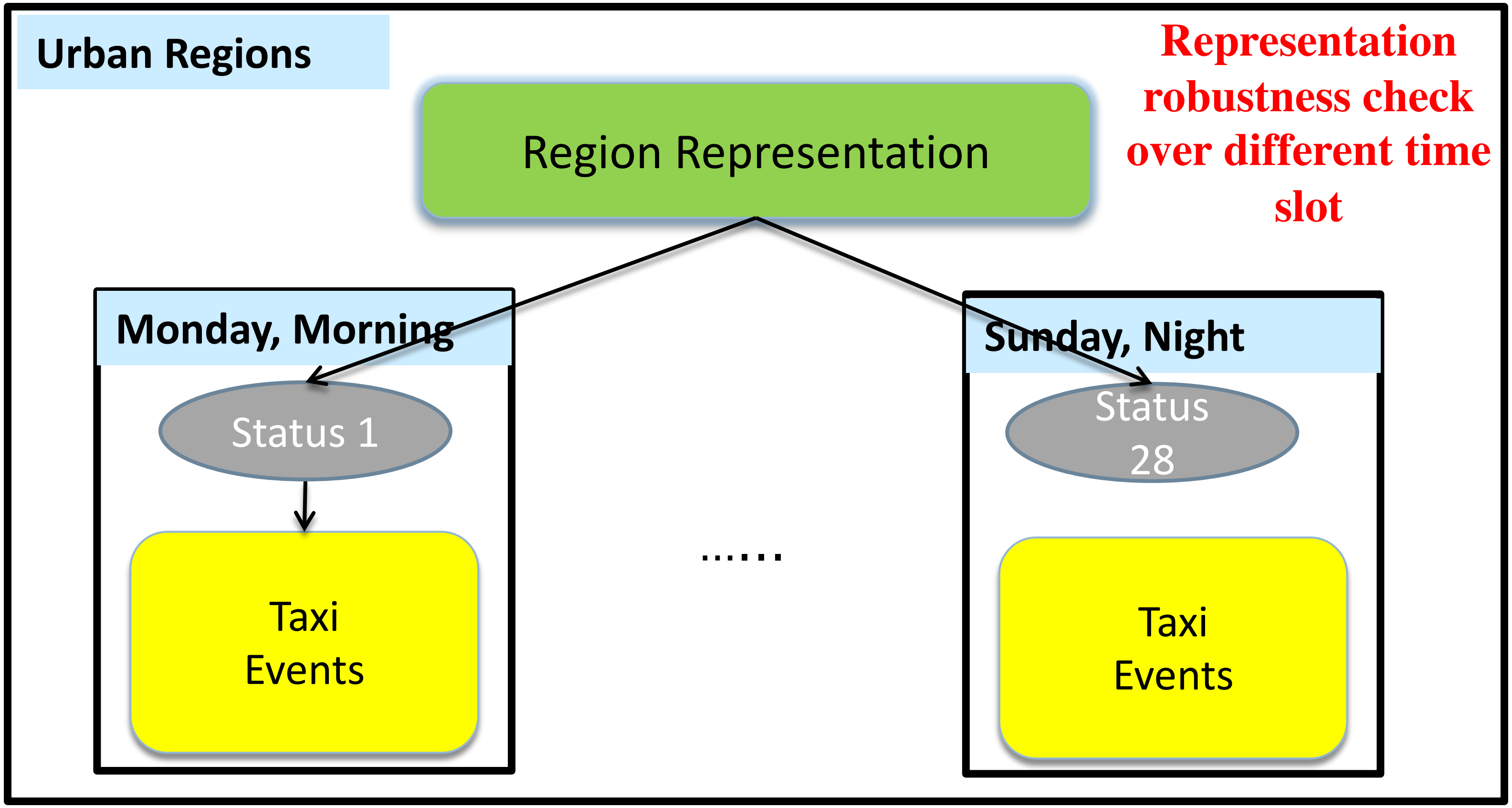
- **Bus mobility pattern**
  - <day, hour, leaving or arriving> of a bus pickup or delivery event

Learning Representation with Robustness Guarantee (1)



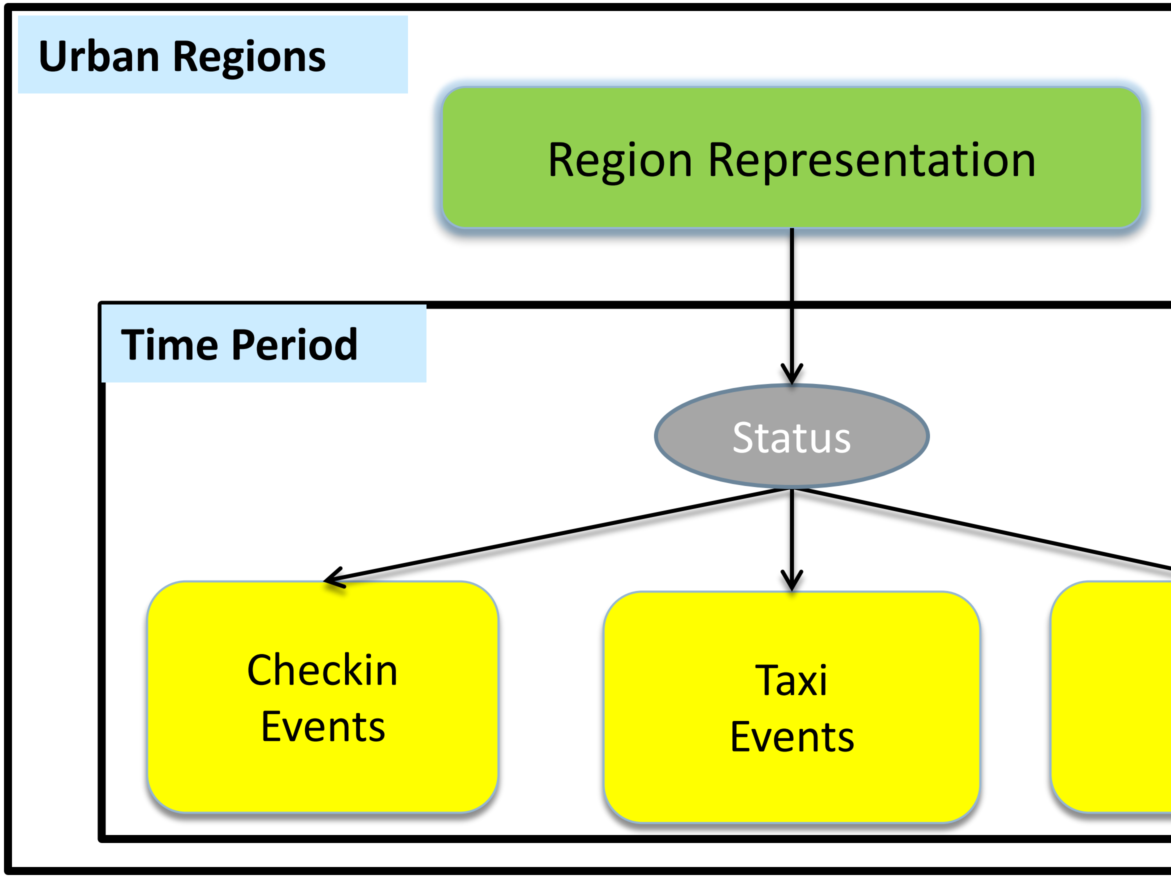
If the representations of two regions are similar, the mobility patterns are similar

Learning Representation with Robustness Guarantee (2)



If the representations of two regions are similar, the mobility patterns in different time slots are similar

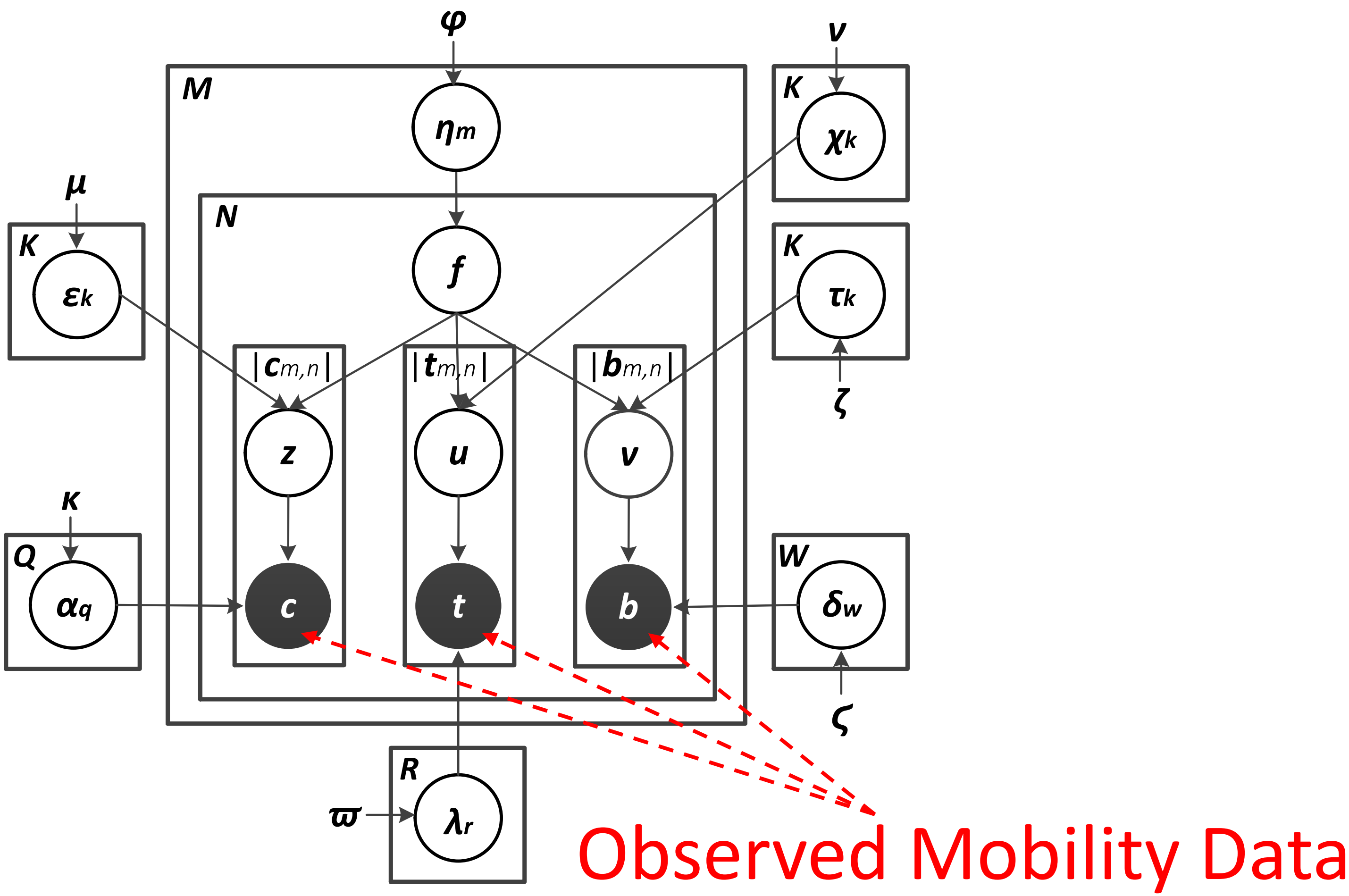
Learning Representation with Robustness Guarantee (3)



Representation robustness check over different time slots and multi-source data

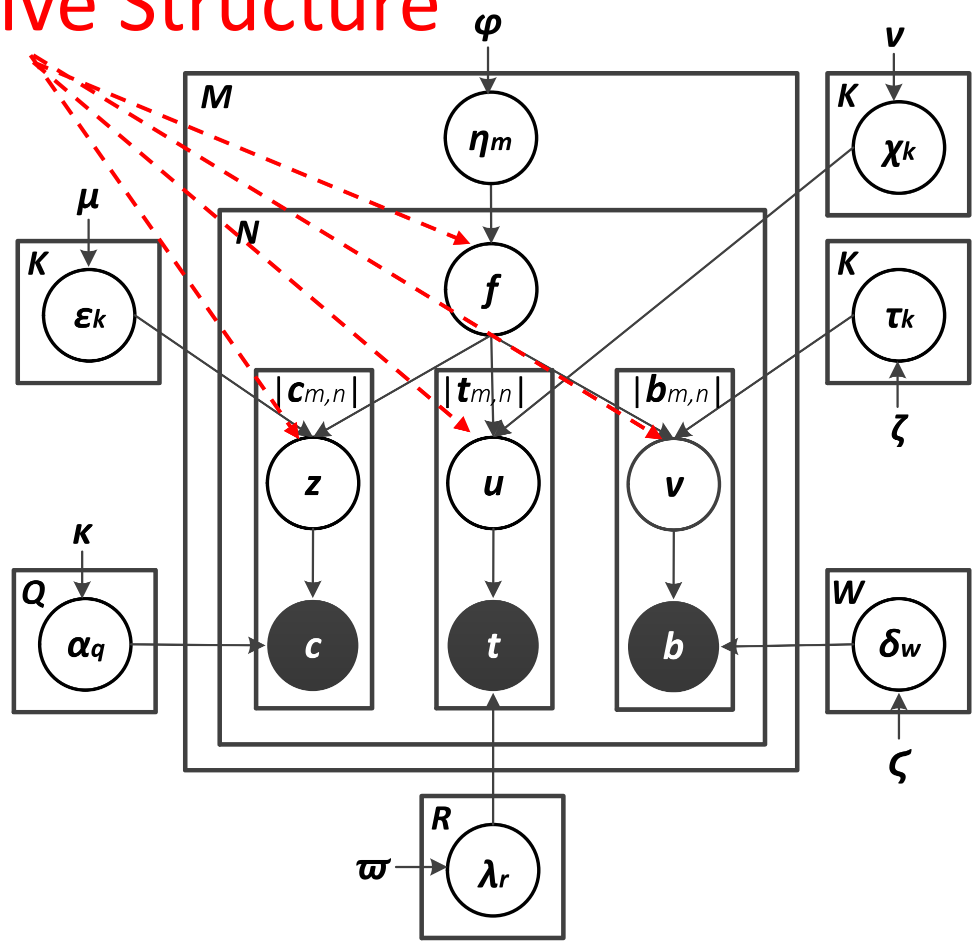
If the representations of two regions are similar, the mobility patterns in different time slots and of different sources are similar

A Probabilistic Hierarchical Model for Collective Representation Learning

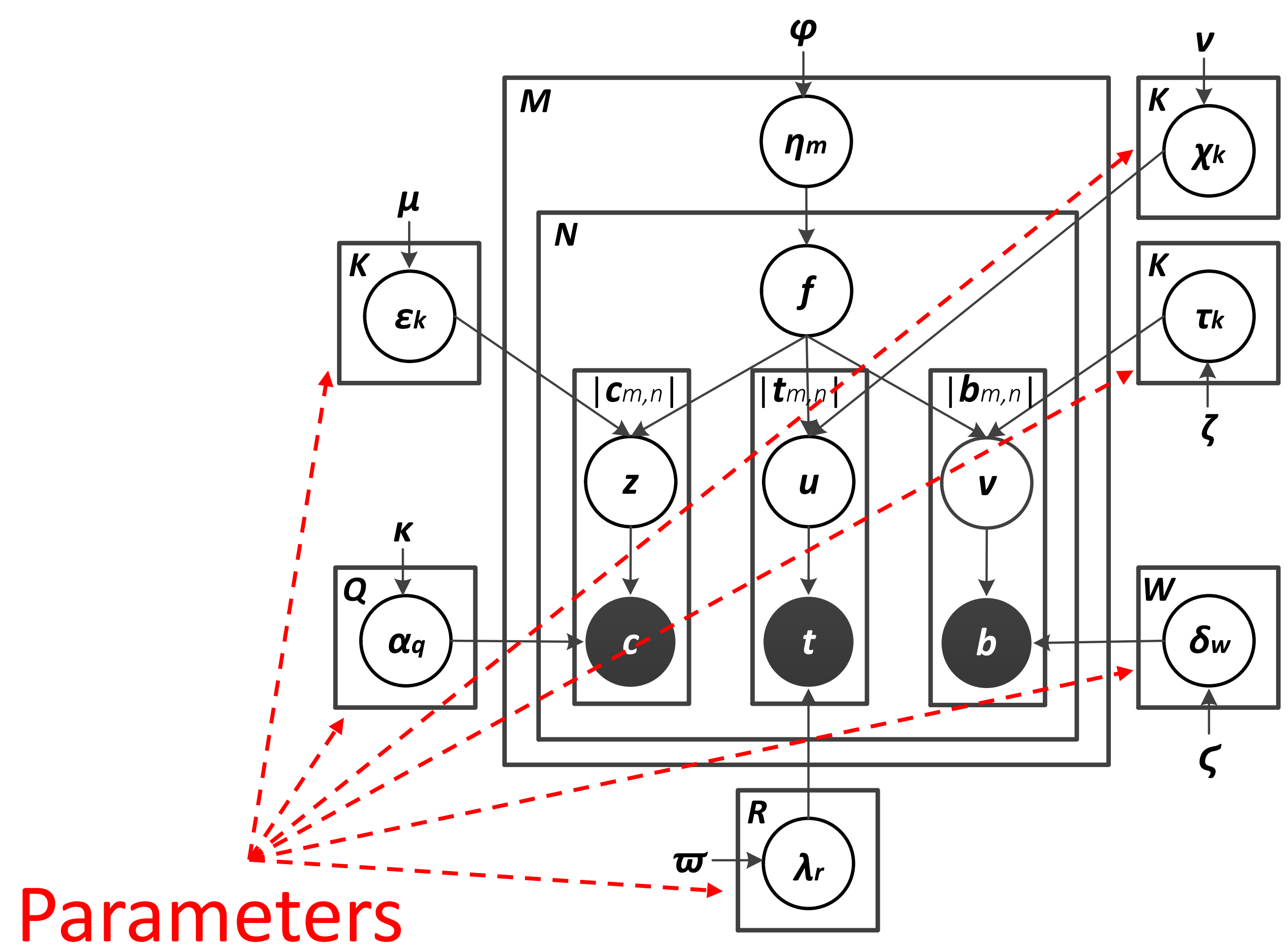


A Probabilistic Hierarchical Model for Collective Representation Learning

Generative Structure

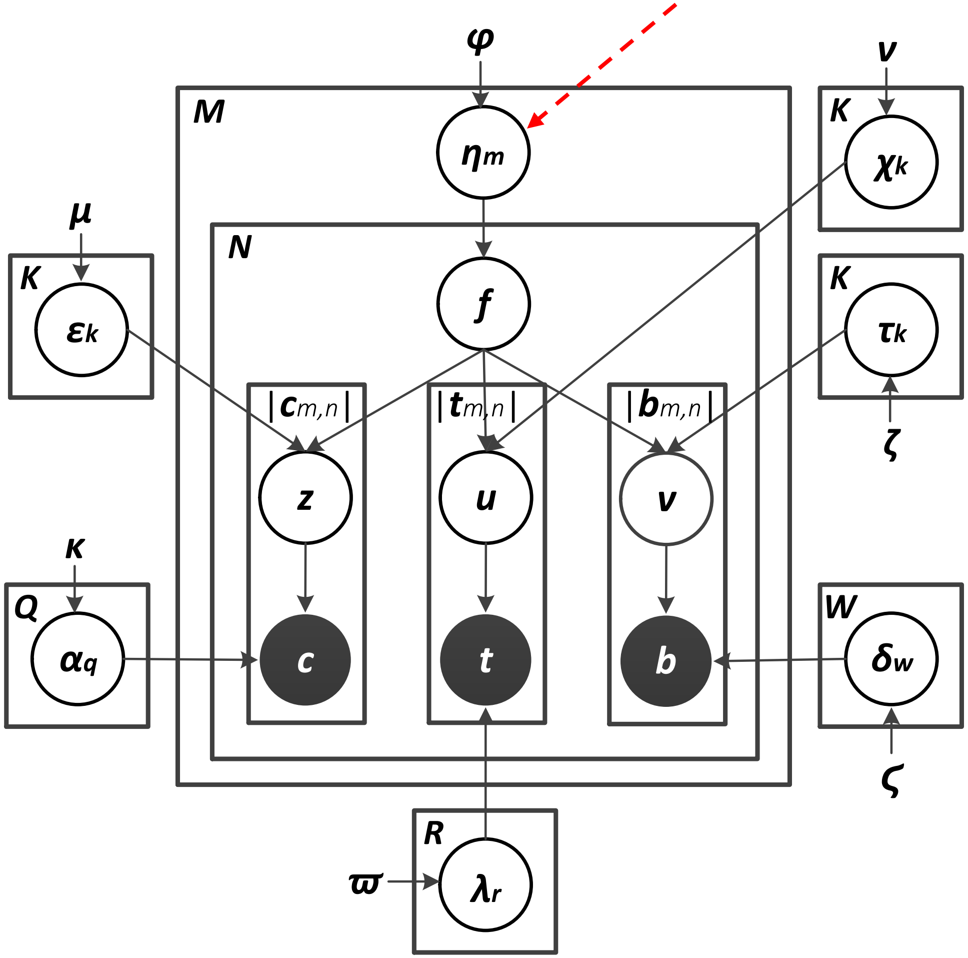


A Probabilistic Hierarchical Model for Collective Representation Learning



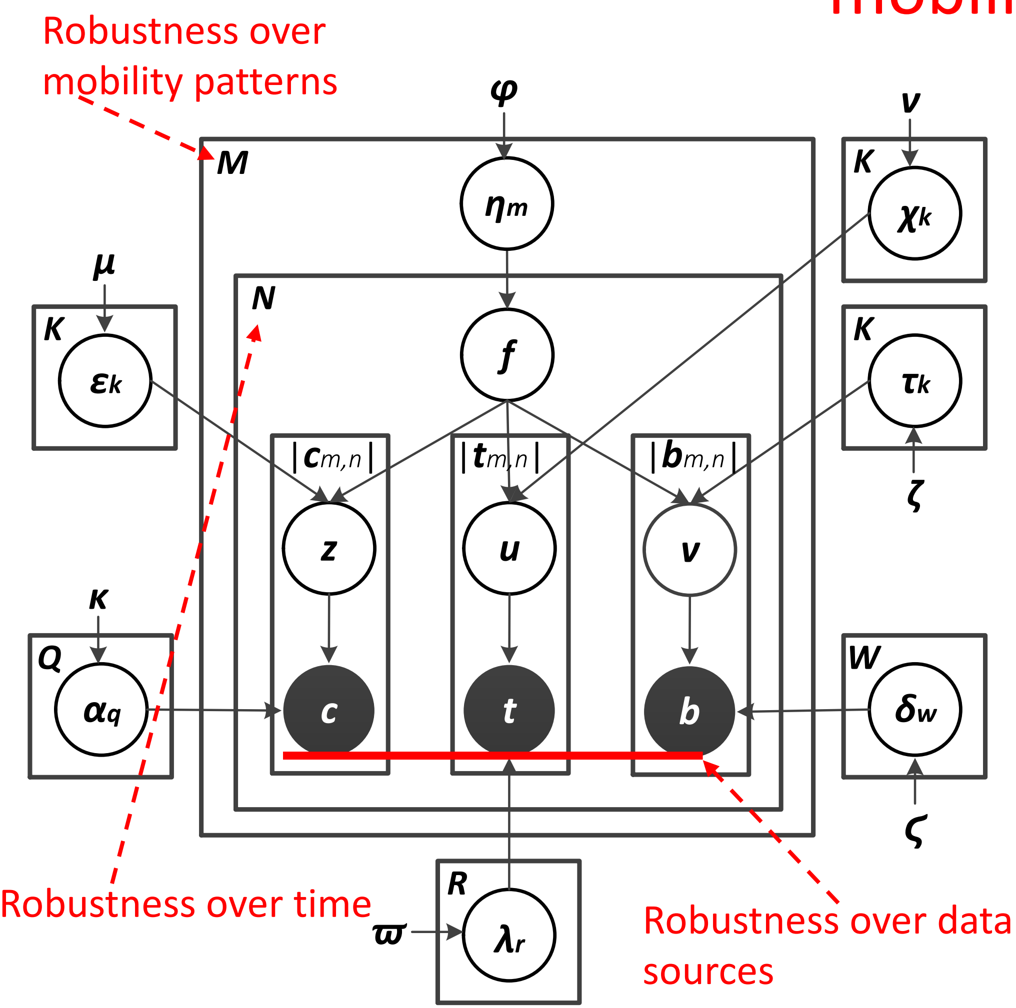
A Probabilistic Hierarchical Model for Collective Representation Learning

Region Representation



A Probabilistic Hierarchical Model for Collective Representation Learning

(M regions for N time periods on K hidden status with C/T/B mobility)



- The region m is represented by a latent probabilistic vector η_m
- The hidden status f of the region m changes over time
- In a period, a region shows checkin (C), taxi (T), and bus (B) clusters of mobility patterns that reflect the hidden status f
- A cluster of mobility patterns = a document
- A mobility event= a word
- Model doc-word with topic modeling

Solving the Optimization Problem

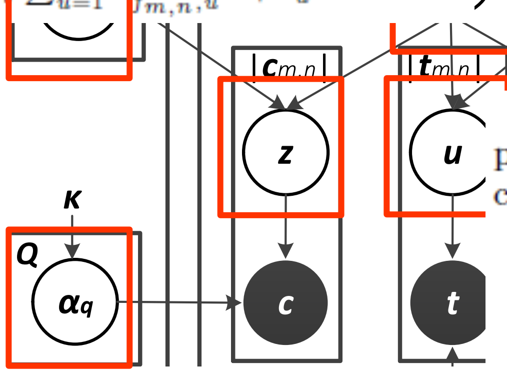
Collapsed Gibbs Sampling to Solve Probabilistic Hierarchical Model

For the i-th taxi pattern $t_{m,n,i} \in \boldsymbol{t}_{m,n}$, the conditional posterior for its latent taxi topic is computed by

$$P(u_{m,n,i} = r|D, \Upsilon - u_{m,n,i}) = \frac{\mathbb{T}^{-(m,n,i)}_{r,t_{m,n,i}} + \varpi_{t_{m,n,i}}}{\sum_{t=1}^{|\boldsymbol{P}_t|} \mathbb{T}^{-(m,n,i)}_{r,t} + \varpi_t} \cdot \frac{\mathbb{U}^{-(m,n,i)}_{f_{m,n},r} + \nu_r}{\sum_{u=1}^{R} \mathbb{U}^{-(m,n,i)}_{f_{m,n},u} + \nu_u}. \quad (4)$$

For the i-th bus pattern $b_{m,n,i} \in \boldsymbol{b}_{m,n}$, the conditional posterior for its latent bus topic is computed by

$$P(v_{m,n,i} = w|D, \Upsilon - v_{m,n,i}) = \frac{\mathbb{B}^{-(m,n,i)}_{w,b_{m,n,i}} + \varsigma_{b_{m,n,i}}}{\sum_{b=1}^{|\boldsymbol{P}_b|} \mathbb{B}^{-(m,n,i)}_{w,b} + \varsigma_b} \cdot \frac{\mathbb{V}^{-(m,n,i)}_{f_{m,n},w} + \zeta_w}{\sum_{v=1}^{W} \mathbb{V}^{-(m,n,i)}_{f_{m,n},v} + \zeta_v}. \quad (5)$$



For the n-th mobility segment in estate m, the conditional posterior probability for its latent function assignment f is computed by

$$P(f_{m,n} = k|\mathcal{D}, \Upsilon - f_{m,n}) = \frac{\mathbb{F}^{-(m,n)}_{m,k} + \ldots}{\ldots}$$
$$\times \frac{\prod_{z=1}^{Q} \Gamma(\mathbb{Z}_{k,z} + \mu_z)\Gamma(\sum \ldots)}{\prod_{z=1}^{Q} \Gamma(\mathbb{Z}^{-(m,n)}_{k,z} + \mu_z) \ldots}$$
$$\times \frac{\prod_{u=1}^{R} \Gamma(\mathbb{U}_{k,u} + \nu_u)\Gamma(\sum \ldots)}{\prod_{u=1}^{R} \Gamma(\mathbb{U}^{-(m,n)}_{k,u} + \nu_u)\Gamma(\sum_{u=1} \ldots u)}$$
$$\times \frac{\prod_{v=1}^{W} \Gamma(\mathbb{V}_{k,v} + \zeta_v)\Gamma(\sum_{v=1}^{W} \mathbb{V}^{-(m,n)}_{k,v} + \zeta_v)}{\prod_{v=1}^{W} \Gamma(\mathbb{V}^{-(m,n)}_{k,v} + \zeta_v)\Gamma(\sum_{v=1}^{W} \mathbb{V}_{k,v} + \zeta_v)}. \quad (2)$$

Theory guarantee?

For the i-th checkin pattern $c_{m,n,i} \in \boldsymbol{c}_{m,n}$, the conditional posterior for its latent checkin topic is computed by

$$P(z_{m,n,i} = q|D, \Upsilon - z_{m,n,i}) = \frac{\mathbb{C}^{-(m,n,i)}_{q,c_{m,n,i}} + \kappa_{c_m}}{\sum_{c=1}^{|\boldsymbol{P}_c|} \mathbb{C}^{-(m,n,i)}_{q,c} \ldots}$$

After all the latent assignments ... update rules of the model paramet...

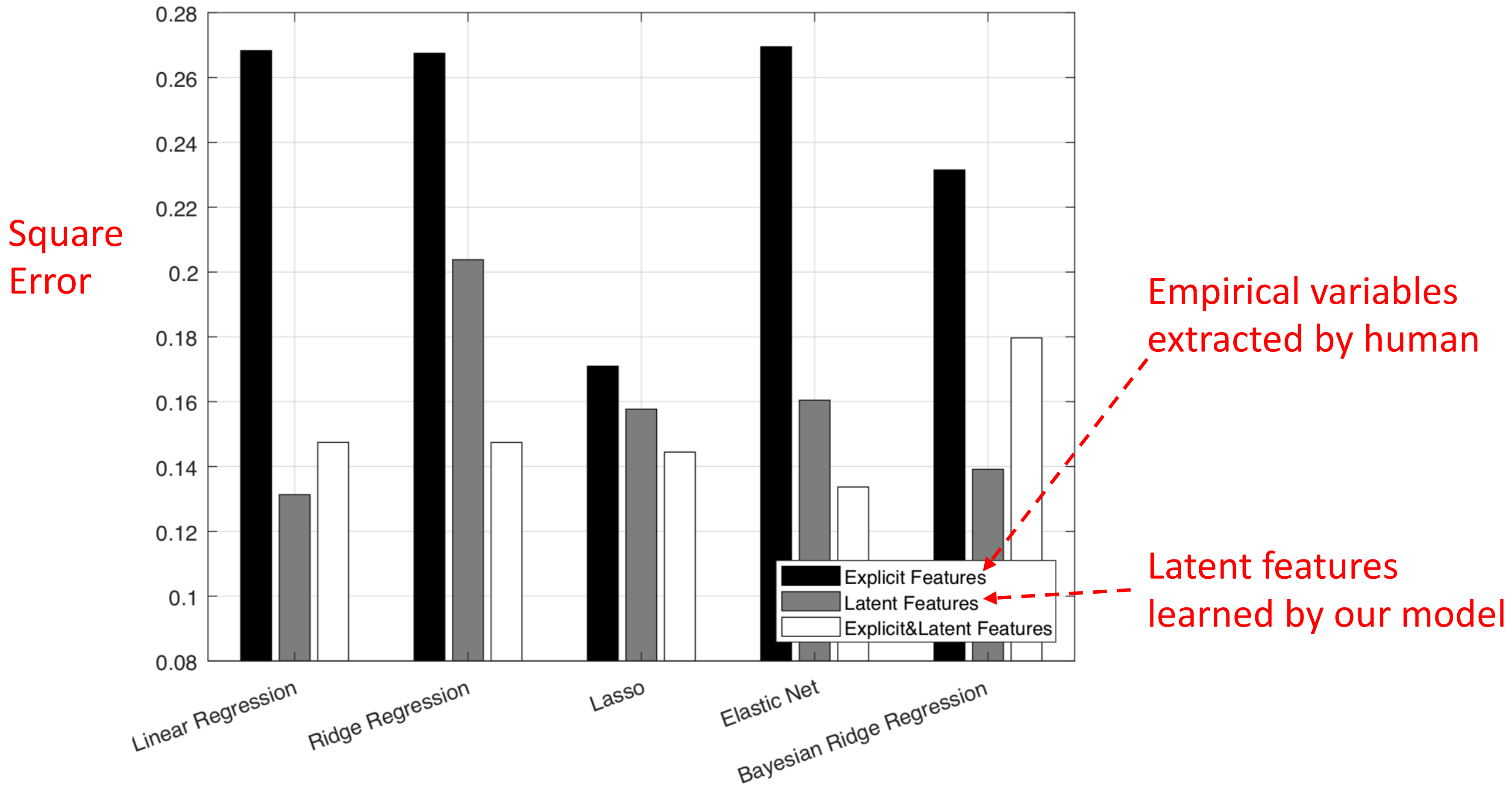
$$\epsilon_{f,z} = \frac{\mathbb{Z}_{f,z} + \mu_z}{\sum_{q=1}^{Q} \mathbb{Z}_{f,q} + \mu_q}, \chi_{f,u} = \frac{\mathbb{U}_{f,u} + \nu_u}{\sum_{r=1}^{R} \mathbb{U}_{f,r} + \nu_r}, \tau_{f,v} = \frac{\mathbb{V}_{f,v} + \zeta_v}{\sum_{w=1}^{W} \mathbb{V}_{f,w} + \zeta_w},$$
$$\alpha_{z,c} = \frac{\mathbb{C}_{z,c} + \kappa_c}{\sum_{p=1}^{|\boldsymbol{P}_c|} \mathbb{C}_{z,p} + \kappa_p}, \lambda_{u,t} = \frac{\mathbb{T}_{u,t} + \varpi_t}{\sum_{p=1}^{|\boldsymbol{P}_t|} \mathbb{T}_{u,p} + \varpi_p}, \delta_{v,b} = \frac{\mathbb{B}_{v,b} + \varsigma_b}{\sum_{p=1}^{|\boldsymbol{P}_b|} \mathbb{B}_{v,p} + \varsigma_p}.$$

hidden status, check ... c hierarchical model

for each mobility pat...

Study of Restaurant Popularity Prediction

Accuracy comparison of human-defined explicit features and machine-learned latent representations over different predictive models

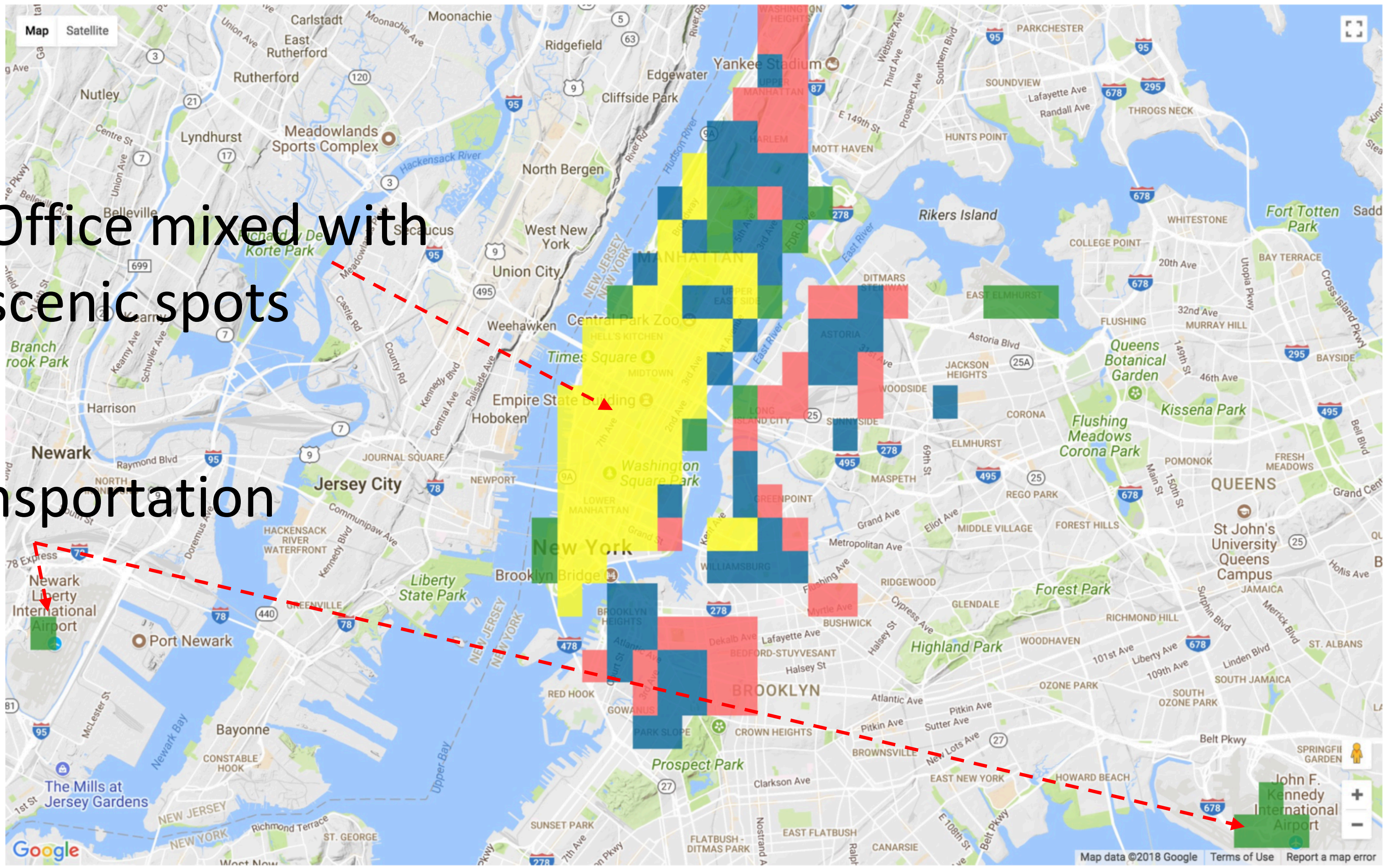


Annotating Regions of Urban Functions

Apply K-Means to cluster regions into 4 categories using the learned representations

Office mixed with scenic spots

Transportation

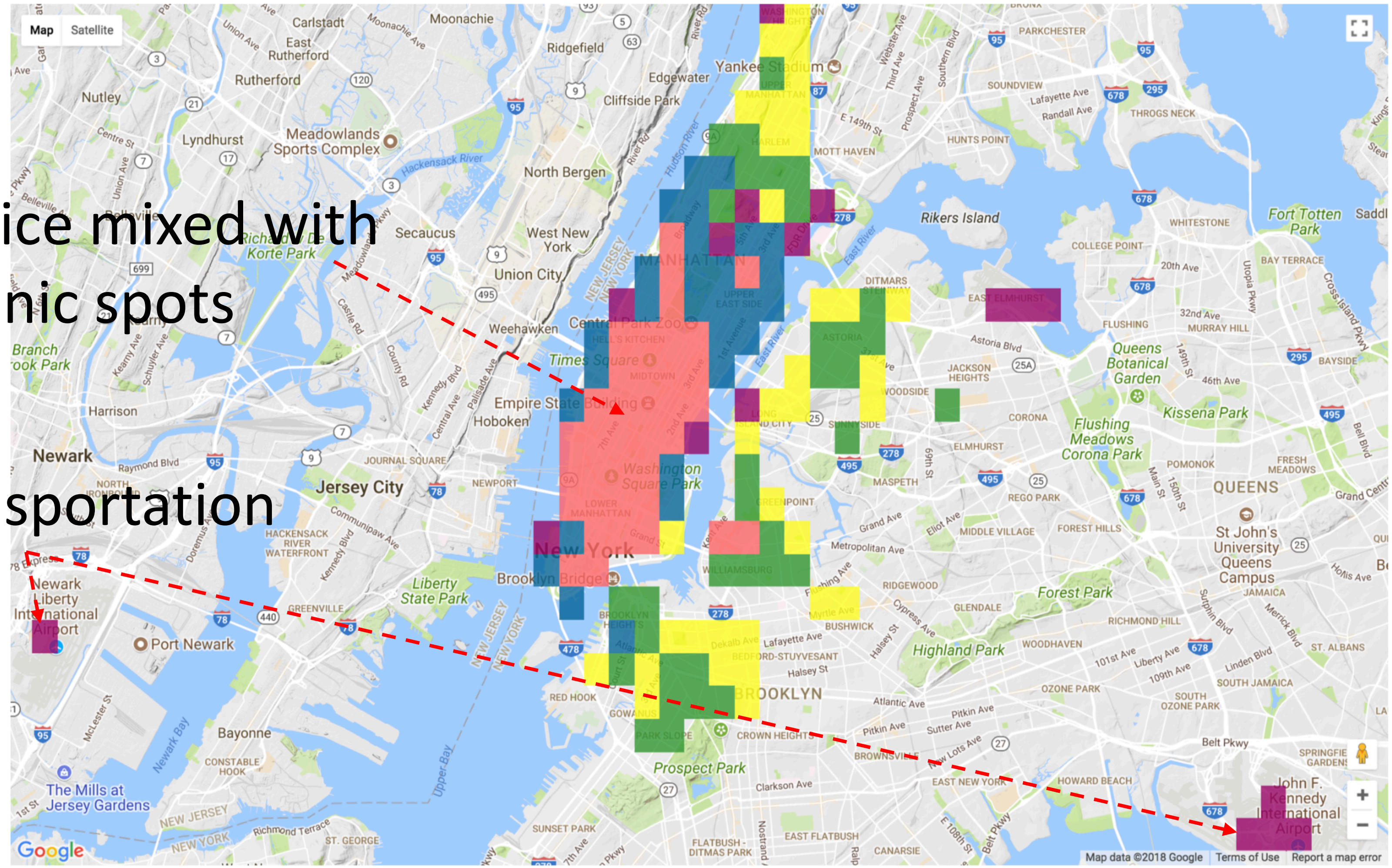


Annotating Regions of Urban Functions

Apply K-Means to cluster regions into 5 categories using the learned representations

Office mixed with scenic spots

Transportation

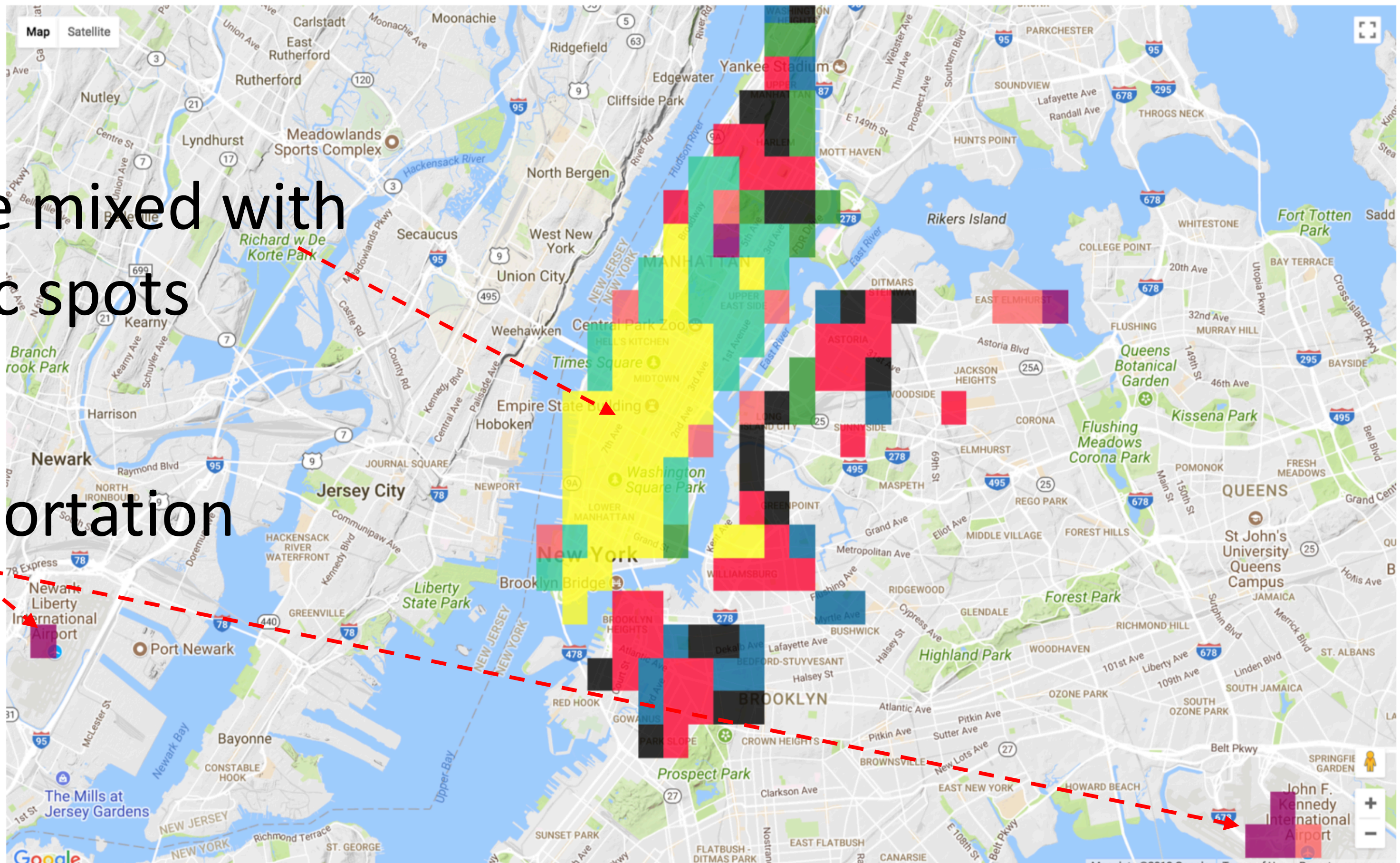


Annotating Regions of Urban Functions

Apply K-Means to cluster regions into 8 categories using the learned representations

Office mixed with scenic spots

Transportation



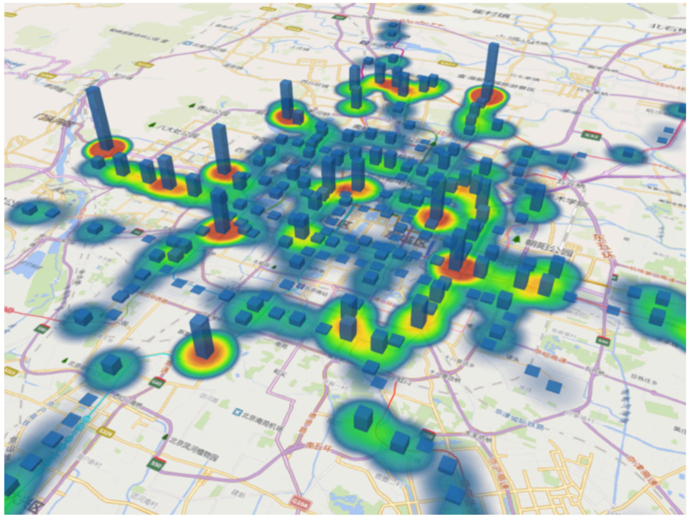
Outline

- Background and Motivation
- Modeling Spatiotemporal Dynamics
- Collective Representation Learning
- **Applications**
- Conclusion and Future Work

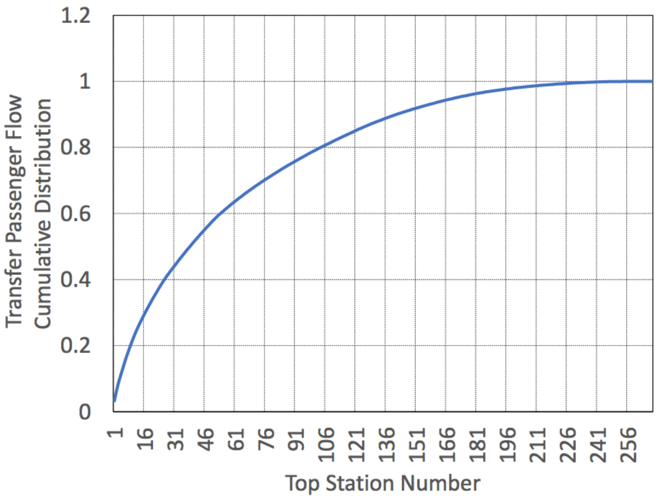
SmartTransfer: Modeling the Spatiotemporal Dynamics of Passenger Transfers for Crowdedness-aware Route Recommendations

Route and Transfer Recommendations in Public Transportation Systems

Fastest routes != best routes



Spatial distribution of transfer passenger flow

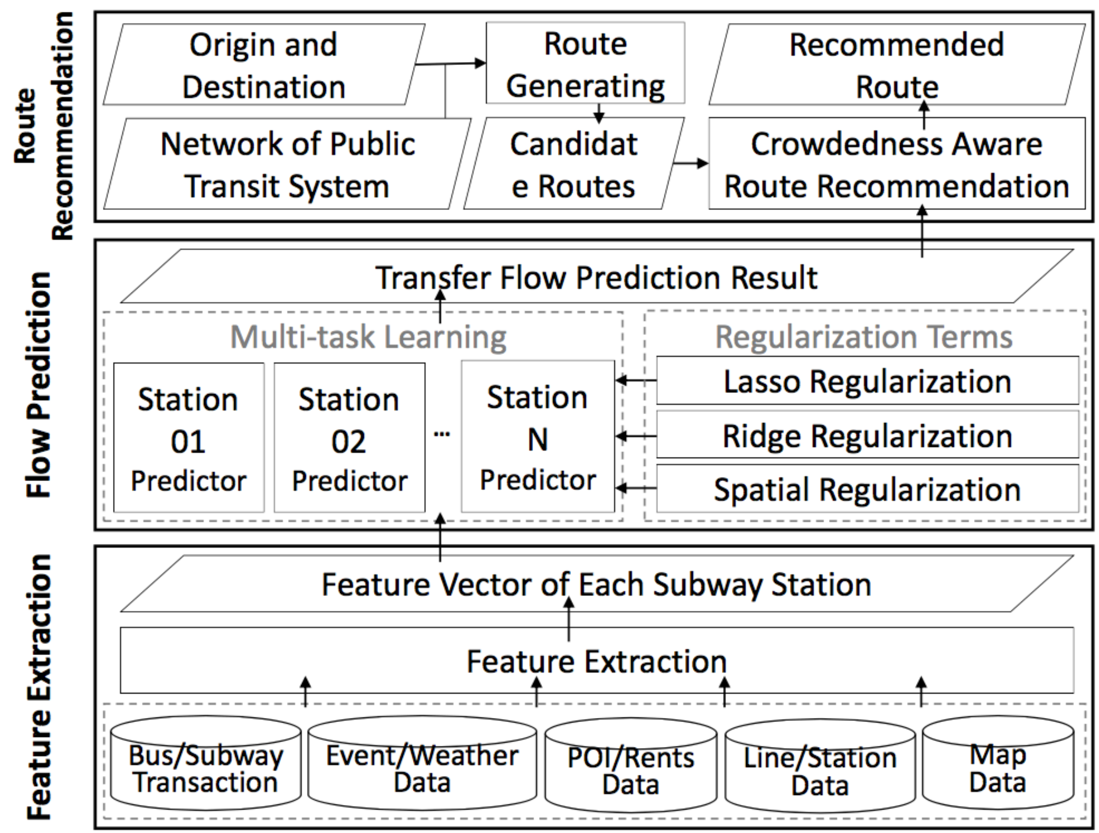


Cumulative distribution of transfer passenger flow for top 256 subway stations

Root Cause: Spatial-temporal unbalance of traffic demand and transportation capacity supply

Crowdedness-aware Route Recommendations

- Feature extraction of subway stations
- Predict the transfer demands of subway stations with spatial-temporal multi-task learning
- Given origin and destination, generate candidate routes from subway networks and bus networks
- Recommend routes based on potential time cost and crowdedness



You Are How You Drive: Peer and Temporal-Aware Representation Learning for Driving Behavior Analysis

Beyond Accidents: Vehicles as Weapons

Charlottesville, Virginia

A car plows into counterprotesters marching against white

Date of attack: August 12, 2017

Number of casualties: A 32-year-old woman was kil

Nice, France

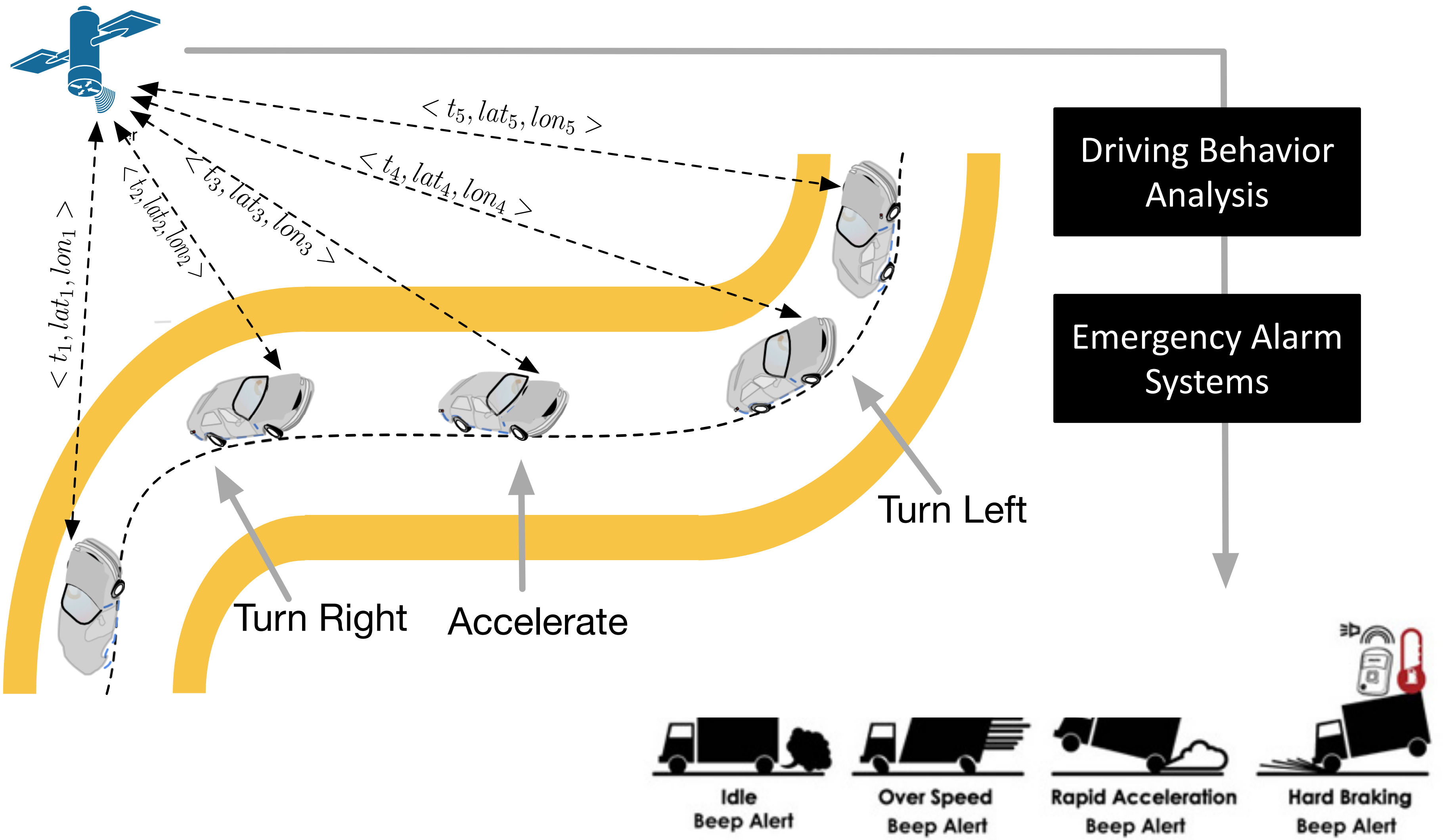
French citizens in mourning over Nice attack 02:19

Date of attack: July 14, 2016

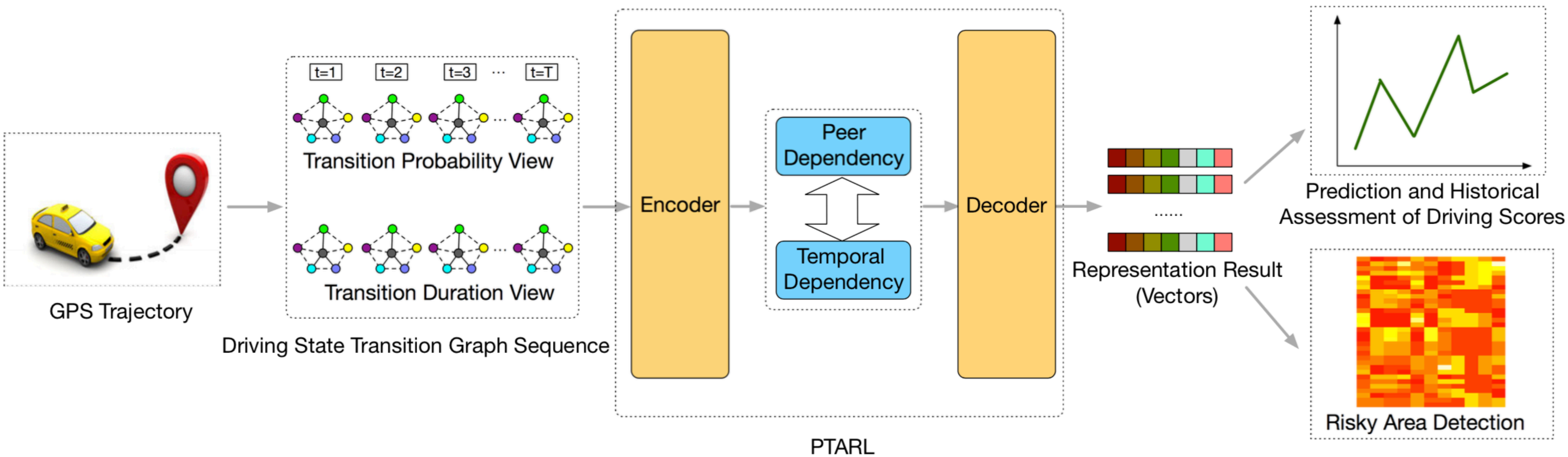
Number of casualties: Eighty-four people were killed and more than 200 wounded.

What can we do to protect human-transportation systems from vehicle-ramming attacks?

Toward Machine-Learning Based Driving Behavior Analysis



Driving Performance Scoring and Risky Area Detection



1. **Learn driving behavior profiles from driving state transition graphs with spatiotemporal representation learning**
2. **Exploit driving behavior profiles to automatically score driving performances and detect risky areas**

More Applications

Energy Consumption

Gas Refilling Event Detection and Gas Station Site Selection (DASFAA16)

Bike Station Site Selection and Rebalancing (ICDM15)

Mitigate Traffic Congestion

Community Planning

Residential Community Analysis for Affordable Housing (KDD14, KDD15, TKDD)

Point-Of-Interests Recommender Systems (KDD13, SDM14, ICDM16)

User Modeling

Outline

- Background and Motivation
- Modeling Spatiotemporal Dynamics
- Collective Representation Learning
- Applications
- **Conclusion and Future Work**

Conclusion Remarks

- Data Environments
  - Human mobility data
- Data Science Foundations
  - Modeling spatial diffusion and temporal dynamics as mixture stochastic point processes integrated with human knowledge
    - Generalized for ecommerce click rate data, online hospital comment data, network intrusion data, malware/disease infection data, paypal e-payment data
    - Spatiotemporal forecasting of 3W(when, where, what)
  - Collective representation learning with multi-source data
    - Generalized for automated heterogeneous data fusion and automated representation learning
    - Spatiotemporal embedding + semantic labeling
- Data Science Applications
  - Smart transfer systems
  - Driving behavior analysis