Human Mobility Synchronization and Trip Purpose Detection with Mixture of Hawkes Processes

(ACM SIGKDD 17)

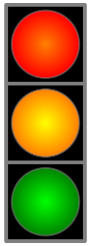
Yanjie Fu

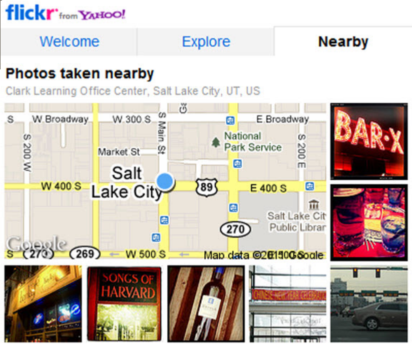
Outline

- **Background and Motivation**
- Problem Statement and Research Insights
- Methodology
- Evaluation
- Conclusions

Pervasive Sensing for Human Movements

IoT, GPS, wireless sensors, mobile Apps

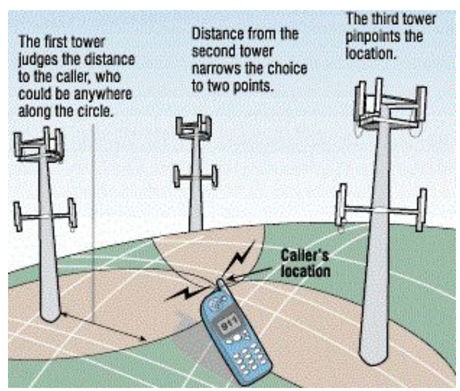




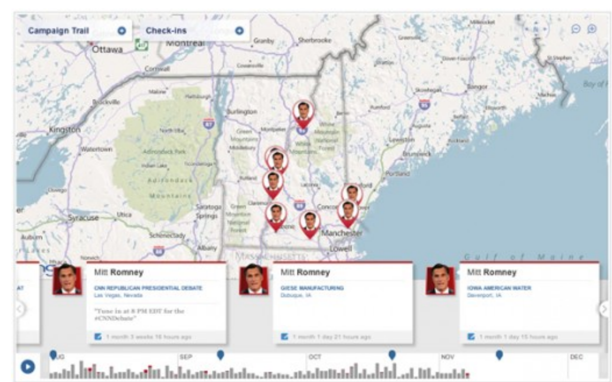
Human Mobility Data

- **Human mobility data are people's movement trajectories which can be**
  - Phone, WIFI, or network station traces
  - Trajectories of driving routes (citibike, taxicabs, buses, subways, lightrails)
  - Sequences of posts (geo-tweets, geo-tagged photos, or check-ins)

Taxicab GPS Traces



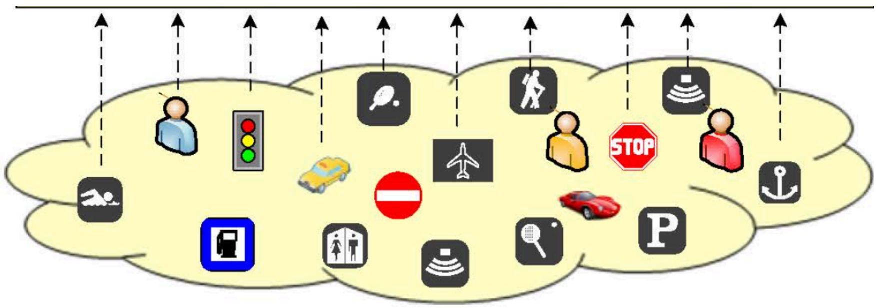
Phone Traces



Mobile Checkins

Unique Characteristics of Human Mobility Data

- **Multi-source, multi-dimensional, multi-domain, multi-format, semantically-rich, collectively-related data environments**
  - Devices, e.g., smart phones, smart watches
  - Vehicles, e.g., taxicabs, buses, subways, city bikes
  - Sensors, e.g., GPS, satellite remote sensing
  - Buildings, e.g., banks, shopping malls, restaurants
  - Human in location based services, e.g., Foursquare, Flickr, Tweeter, Facebook, Google+, Yelp



Making Sense of Human Mobility Data

A Closed-Loop Data Science System

- Control
- Action recommendation
- Feedback systems
- Incentive systems

Decision Support

Sensing

- Sensing systems
- Sensing strategies
- Data collection

Big and heterogeneous spatiotemporal textual data environments

Collective and semantic exploration

Modeling

- Predictive analysis
- Causal analysis

Representation

- Data cleaning, transformation, reduction
- Explicit feature extraction
- Latent representation learning

Collective and Semantic Exploration of Human Mobility Data (1)

- **The ultimate goal**
  - Understand the nature of human mobility by making it trackable and predictable

Collective and Semantic Exploration of Human Mobility Data (2)

- Primary focus areas
  - **Collective Modeling**
    - Geographic co-location: documents and words
    - Graph structure: dynamic graphs over time
    - Spatial diffusion: stochastic processes
    - Collaborative correlation: tensors and factorization
  - **Semantic Augmentation**
    - Trajectories: what (trip purposes), where (destinations), when (trip time), who (out-of-town travelers or local residents)
    - Users: user demographics, profiles, daily activities, preferences, social groups
    - Events: spatiotemporal event detections (e.g., protests, incidents)
    - Regions: important locations, spatial configurations, urban functions
  - **Human-Community Interactions**
    - Human-Transportation-Systems interactions
    - Human-Food-Services interactions

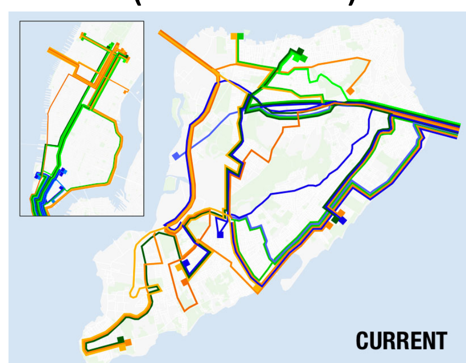
Preliminary Studies

Inferring Trip Purposes and Destinations (KDD17, SDM18, TIST)

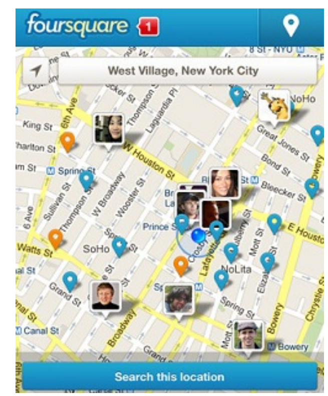
Gas Refilling Event Detection and Gas Station Site Selection (DASFAA16)

Bike Station Site Selection and Rebalancing (ICDM15)

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



Crowdedness-aware Route Optimization (TIST)



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

Self-Optimized Networks for Huawei

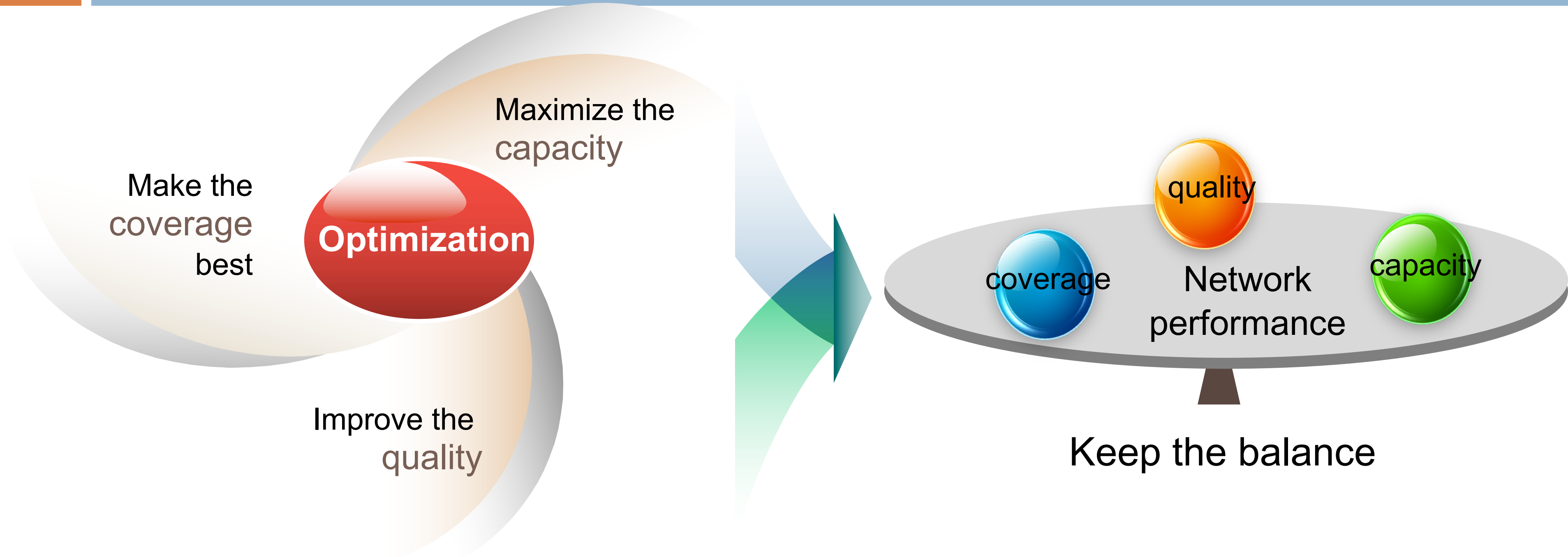
Top of cellular base station
tower with 3 sector antennas

Coverage:
1, Horizontal Azimuth
2, Vertical Tilt

Capacity:
1, Pilot Power

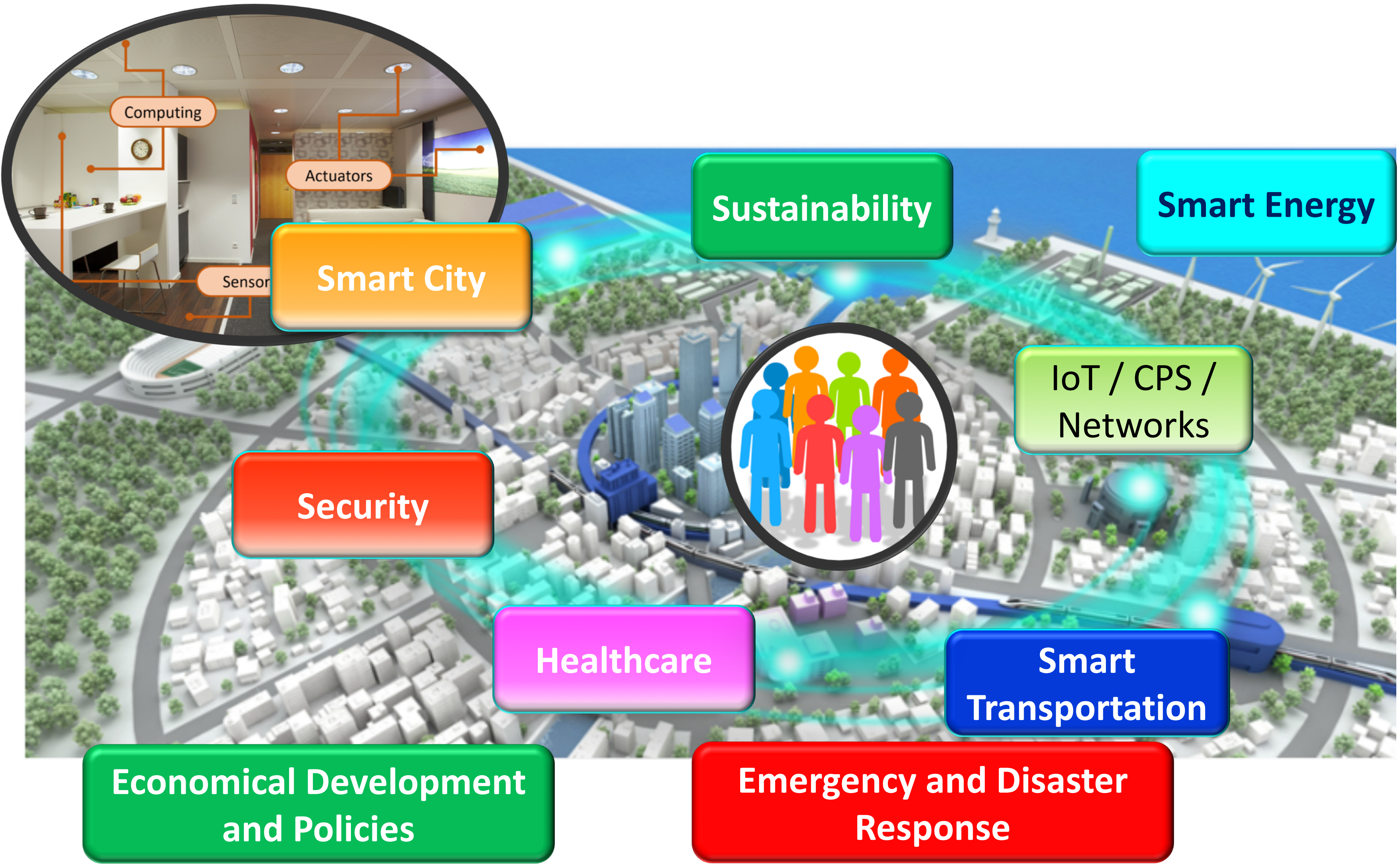
Quality
1, network resources such as bandwidth, SINR

Self-Optimized Networks for Huawei



- Automatically and jointly optimize the coverage, capacity and quality of networks using the data from network monitoring (devices), human mobility (physical), and social networks data streams (cyber)
- Decision rule based methods (white box) and reinforcement learning (black box)

More Applications



Mobility Research in Nature (1)

Letter

Nature **453**, 779-782 (5 June 2008) | doi:10.1038/nature06958; Received 19 December 2007; Accepted 27 March 2008

There is an Addendum (12 March 2009) associated with this document.

Understanding individual human mobility patterns

Marta C. González[1], César A. Hidalgo[1,2] & Albert-László Barabási[1,2,3]

1. Center for Complex Network Research and Department of Physics, Biology and Computer Science, Northeastern University, Boston, Massachusetts 02115, USA
2. Center for Complex Network Research and Department of Physics and Computer Science, University of Notre Dame, Notre Dame, Indiana 46556, USA
3. Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA

- Understanding individual human mobility patterns, by Barabasi 2009
  - Most individuals travel only over short distances, but a few regularly move over hundreds of kilometers
  - Travel distances vary over times
  - Usually return to a few highly frequented locations

Mobility Research in Nature (2)

Article | OPEN

Returners and explorers dichotomy in human mobility

Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti & Albert-László Barabási

Nature Communications **6**,
Article number: 8166 (2015)
doi:10.1038/ncomms9166
Download Citation
Theoretical physics

Received: 15 December 2014
Accepted: 24 July 2015
Published online: 08 September 2015

- **Returners and explorers dichotomy in human mobility, by Barabasi 2015**
  - Returners and explorers play a distinct quantifiable role in spreading phenomena
  - A correlation exists between the mobility patterns and social interactions of returners and explorers

Mobility Research in Nature (3)

nature.com > scientific data > comment > article

SCIENTIFIC DATA

Comment | OPEN

WorldPop, open data for spatial demography

Andrew J. Tatem

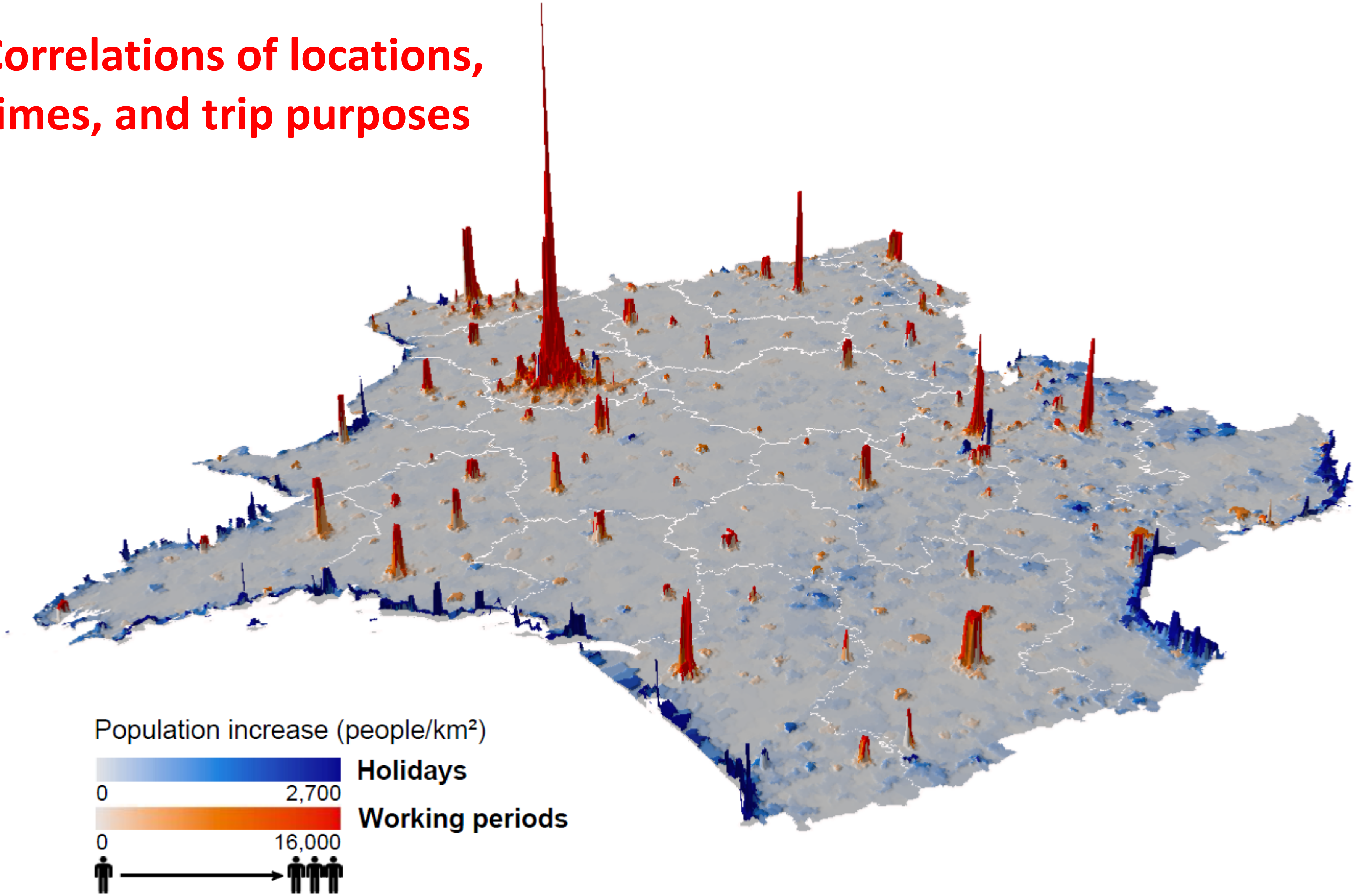
- Using WorldPop to
  - Map poverty
  - Understand maternal health
  - Understand the correlations of population dynamics, property, and health

Outline

- Background and Motivation
- **Problem Statement and Research Insights**
- Methodology
- Evaluation
- Conclusions

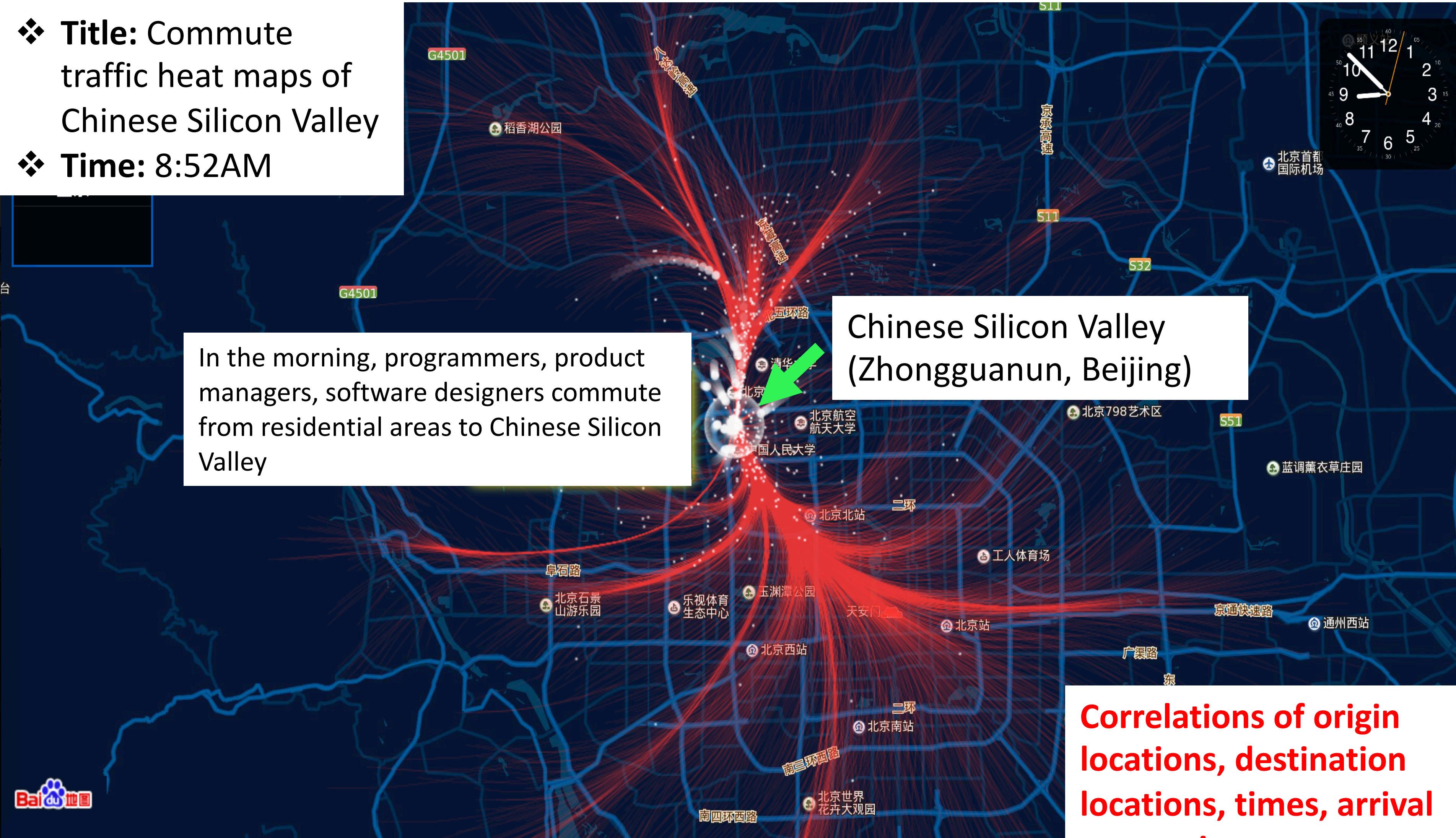
Population Distribution Changes in France between Holidays and Workdays

Correlations of locations, times, and trip purposes



IT Worker Movements of Morning Rush Hours in Chinese Silicon Valley

- **Title:** Commute traffic heat maps of Chinese Silicon Valley
- **Time:** 8:52AM



Chinese Silicon Valley (Zhongguanun, Beijing)

In the morning, programmers, product managers, software designers commute from residential areas to Chinese Silicon Valley

Correlations of origin locations, destination locations, times, arrival rates, trip purposes

Research Questions

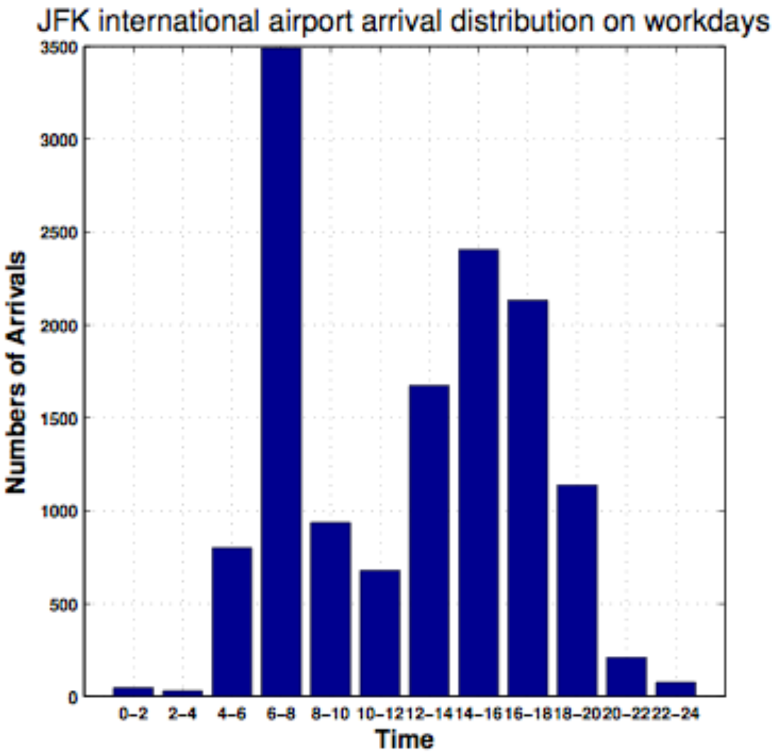
- **Understanding**
  - What is the nature of the spatial diffusion of human mobility across regions of different urban functions in different time periods?
- **Modeling**
  - How to infer and classify the trip purposes of human mobility trajectories?

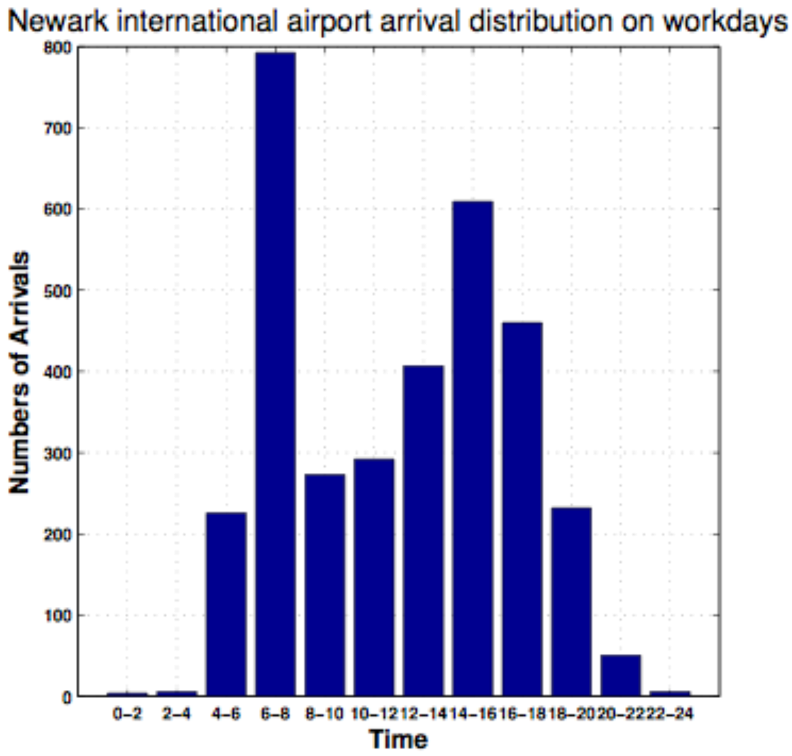
Synchronization in Human Behaviors

Applauds of audiences

Opinion agreement of meeting attendees

Human Mobility Synchronization



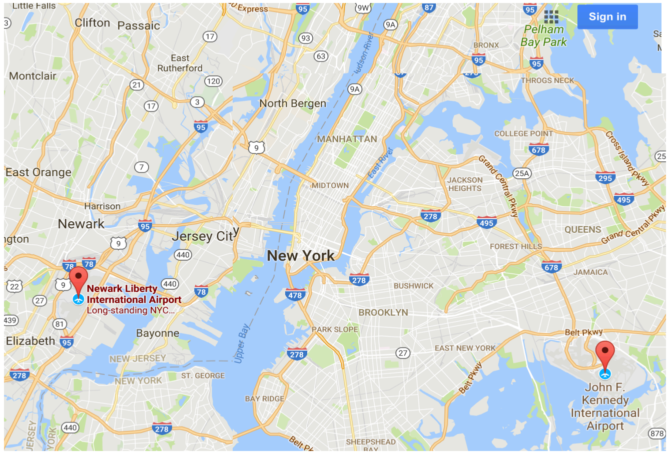


Taxi arrival distributions of JFK Airport and Newark Airport over 24 hours

What Drives Mobility Synchronization

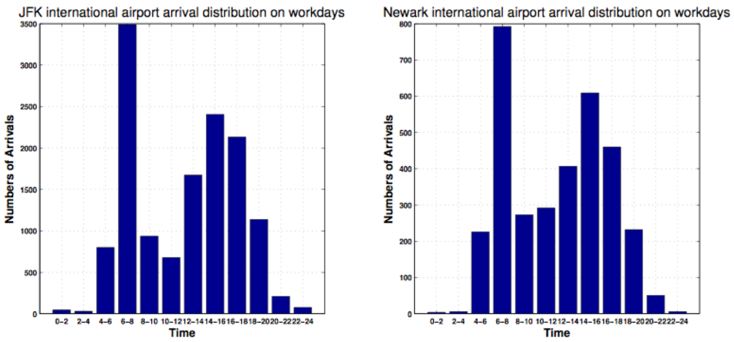
Factor 1: Trip purpose

Why go to JFK/NWK?



Factor 2: Arrival patterns

When go to JFK/NWK?

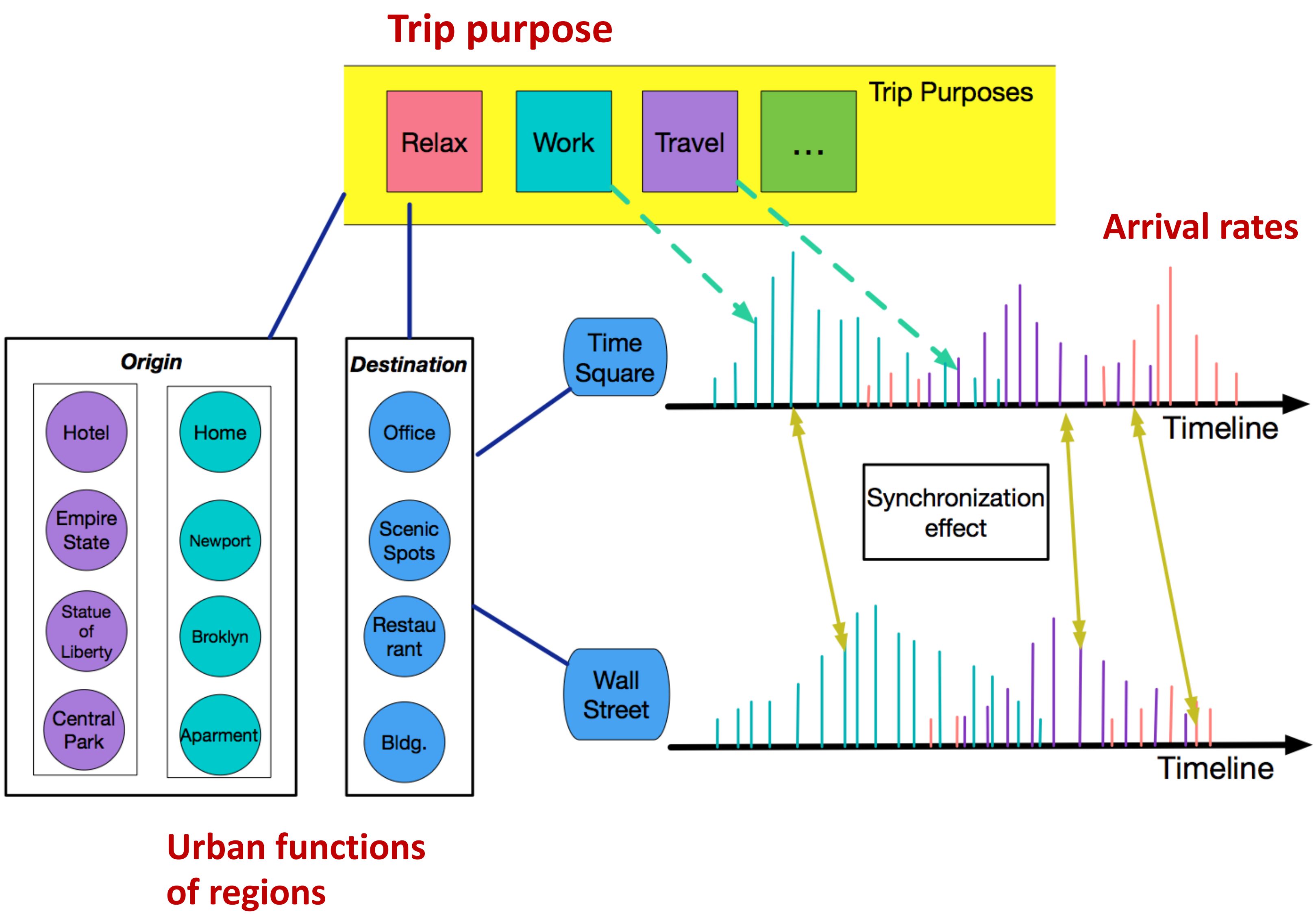


Factor 3: Urban functions of regions

What are JFK/NWK used for?

Two regions show similar arrival patterns in particular time periods if they share similar urban functions

Linking Arrivals, Regions and Purposes



Outline

- Background and Motivation
- Problem Statement and Research Insights
- **Methodology**
- Evaluation
- Conclusions

Framework Overview

Modeling the arrivals of a single region for single trip purpose

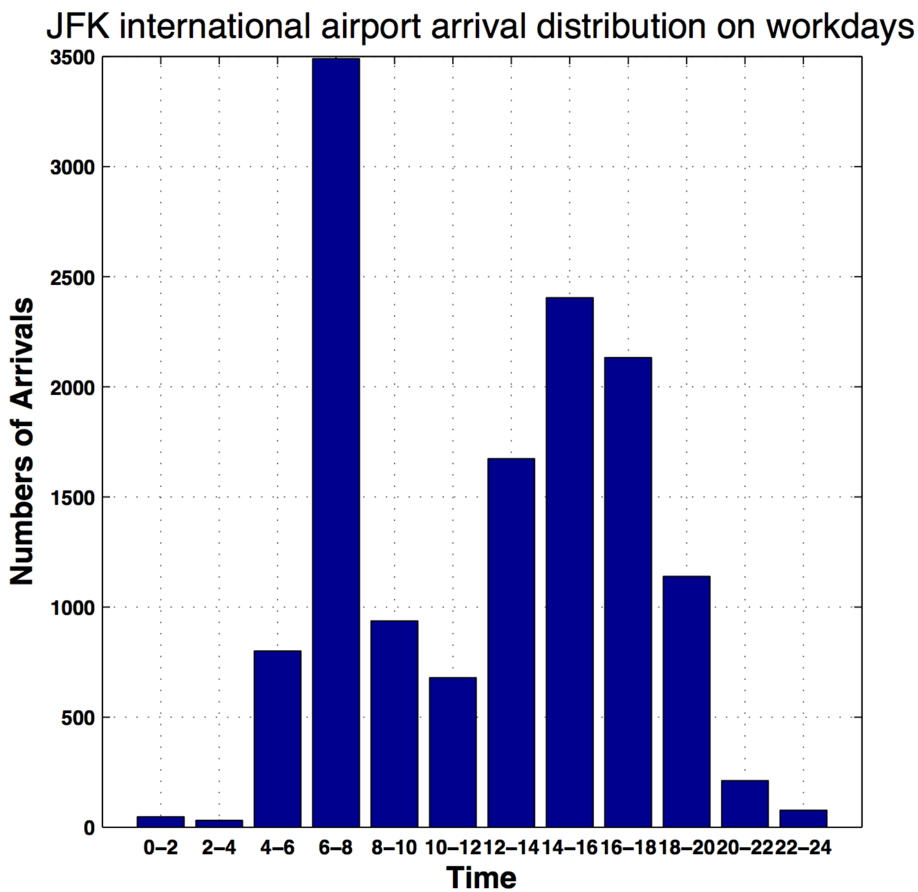
Modeling the arrivals of a single region for multiple trip purposes

Modeling the arrivals of multiple regions for multiple trip purposes

Incorporating human mobility synchronization effects

Adding explanations via modeling urban functions of origin and destination regions

Human Mobility Data as Arrival Events

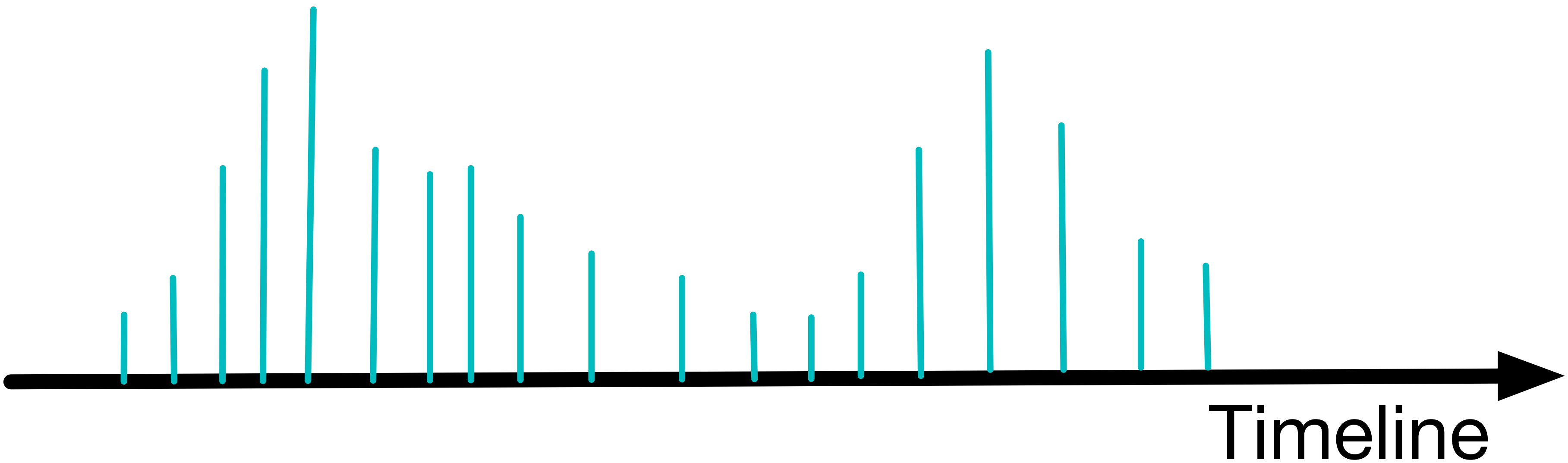


- For each region, we organize taxi trajectories as a sequence of arrivals: $E = \{E_1, E_2, \ldots, E_N\}$

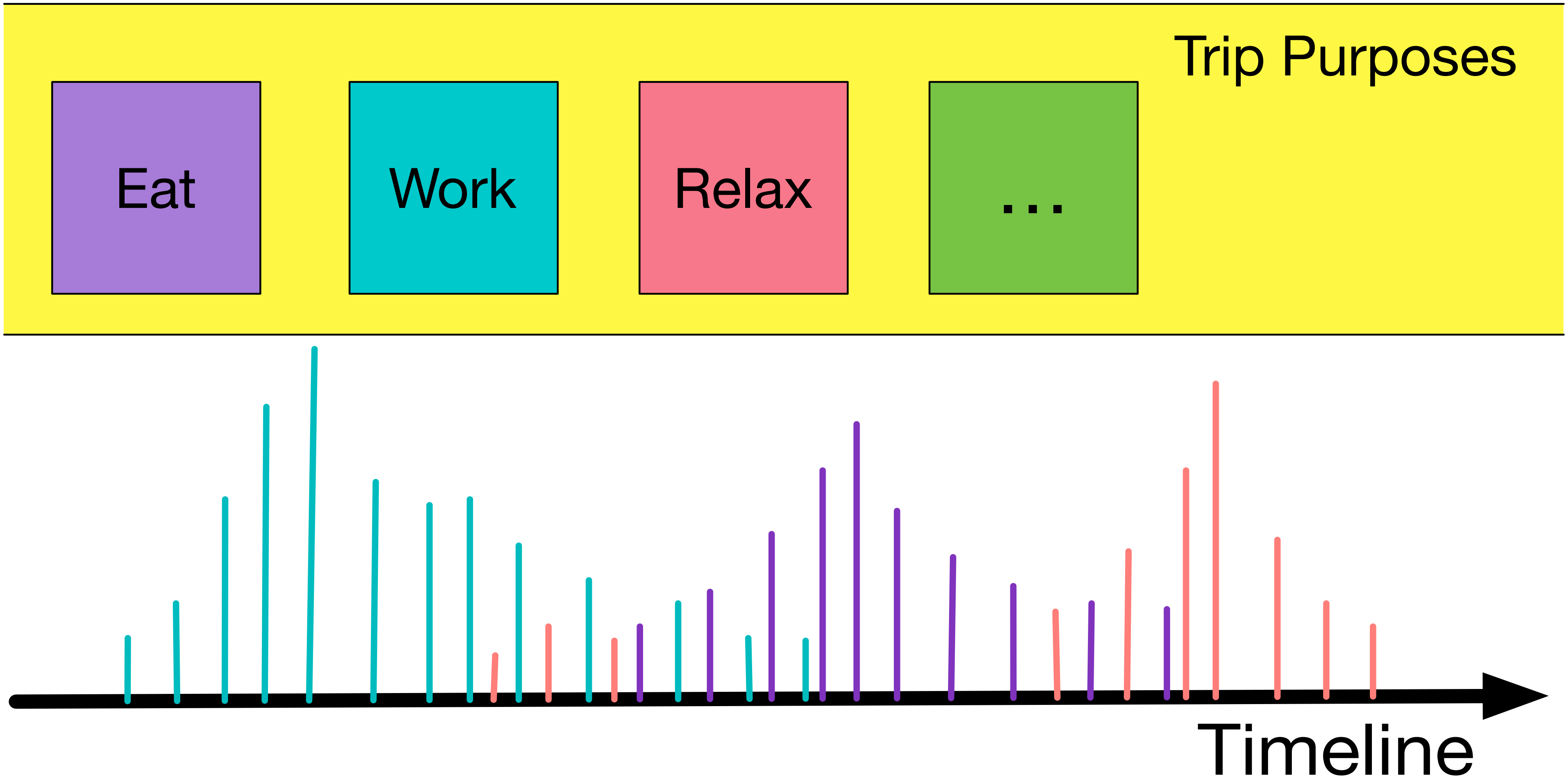
- Each event is a three-element tuple: $E_n = \{g_n, t_n, w_n^d, w_n^o\}$
  - g_n: trip purpose
  - t_n: timestamp of the n-th arrival
  - w_n^d: POIs of destination region
  - w_n^o: POIs of origin region

Modeling Arrivals of Single Region for A Particular Trip Purpose

- **Modeling mobility arrivals as a stochastic point process**
  - Hawkes Process: $\lambda(t) = \mu(t) + \int_{-\infty}^{t} g(t-s)dN(s)$



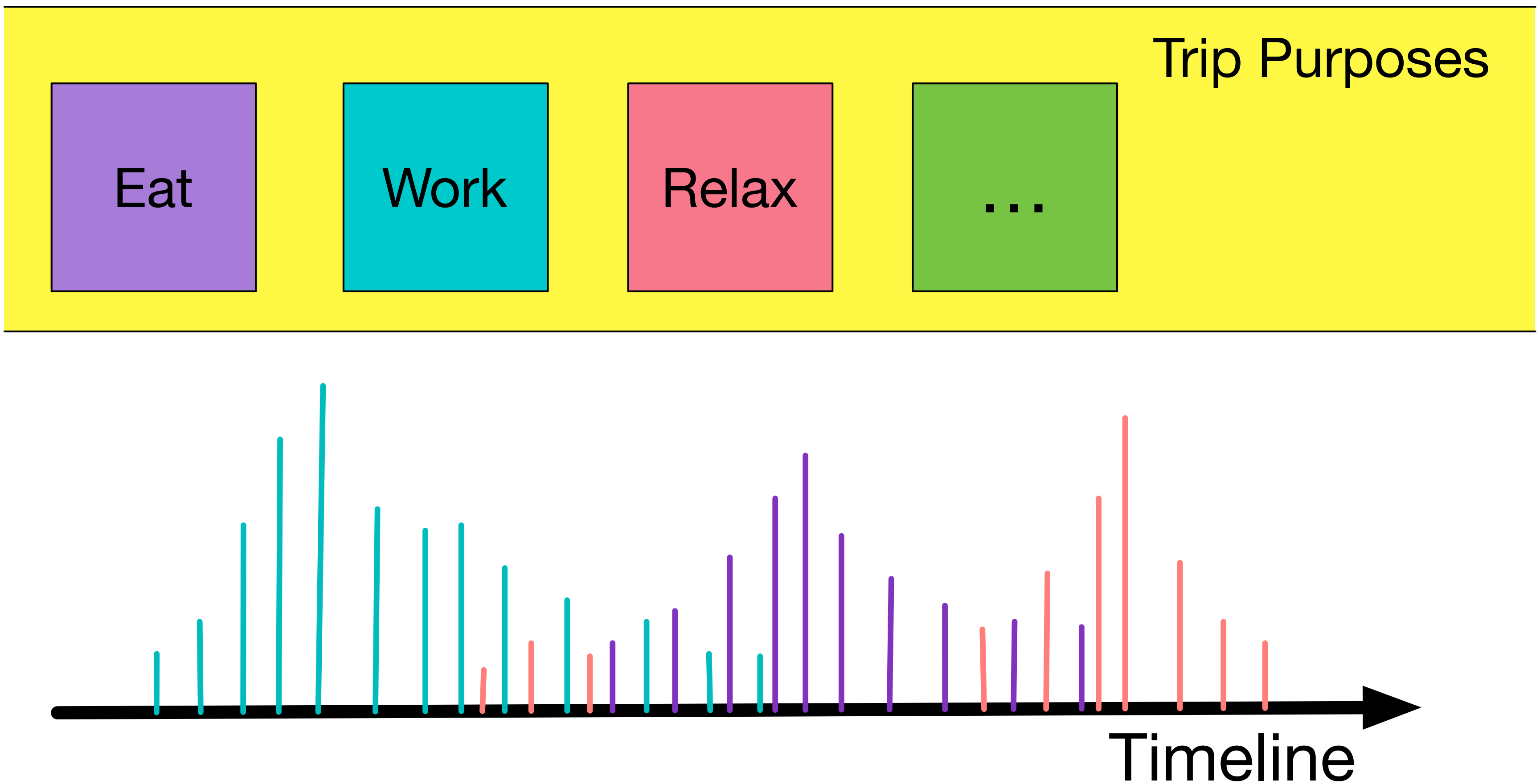
Modeling Arrivals of Single Region for Multiple Trip Purposes (1)



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

- **Mixture of multiple arrival sequences with respect to different trip purposes**
  - Mixture Hawkes Processes
  - $\lambda_{i,m}(t) = \mu_{i,m} + \int_{-\infty}^{t} g(t-s)dN(s) = \mu_i * \gamma_m + \int_{-\infty}^{t} g(t-s)dN(s)$
    - i: the i-th region
    - m: the m-th trip purpose
    - $\mu_{i,m}$: the 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 decading function

Modeling Arrivals of Single Region for Multiple Trip Purposes (3)

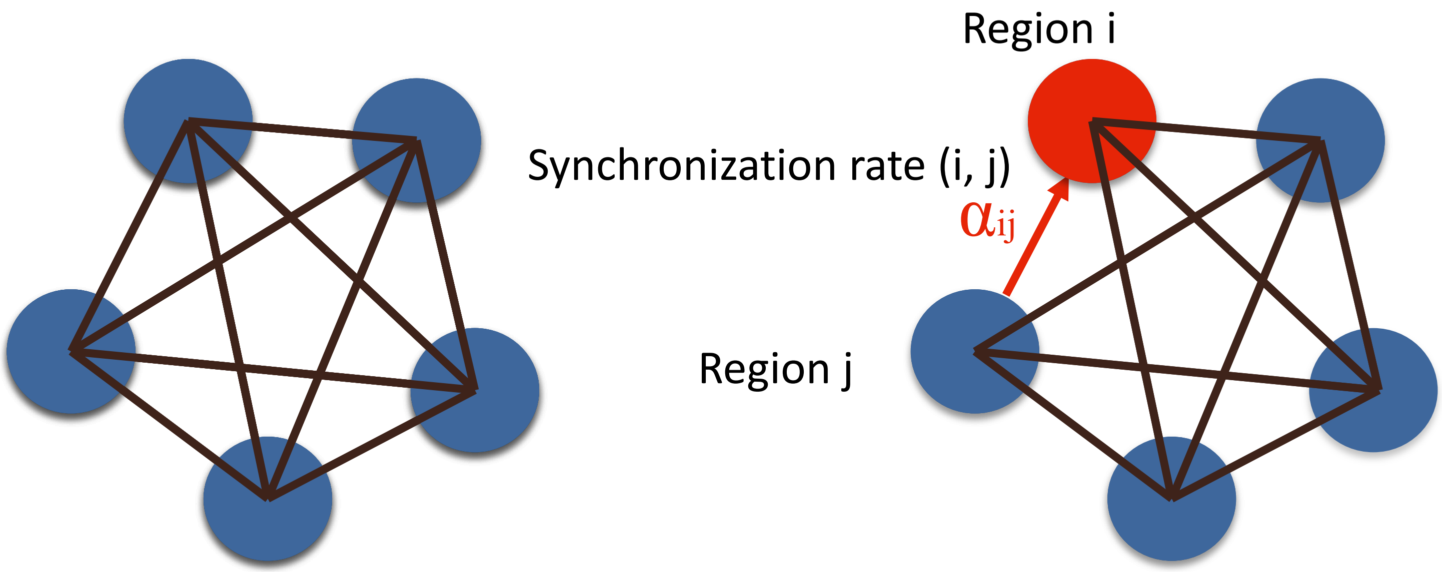


Mobility arrivals in the i-th region :

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

Synchronization Effect Across Regions (1)

- Road networks: graph
- Region: nodes in the graph
- Synchronization rate between two regions: similarity (edge connectivity) between two nodes



Synchronization Effect Across Regions (2)

- **Enhancing the modeling of mobility arrivals by combining the synchronization effects across regions into mixture Hawkes processes**
  - $\lambda_{i,m}(t) = \mu_i * \gamma_m + \sum_{j=1}^{I} \alpha_{ji}^{m} \int_{-\infty}^{t} g(t-s)dN(s)$
  - A joint arrival process of **self-exciting** within a region and **mutual-exciting** across multiple regions
  - Self-exciting: individual dependency in terms of urban functionalities and spatial configurations
  - Mutual-exciting: peer dependency in terms of the similarity among similar regions

Incorporating Urban Functions of Origin and Destination Regions

- **Weakness:** improve explanations and interpretations
- **Trip Purposes** are semantically embedded in the neighborhood buildings of the origin and destinations
- A region as a document, a building (POI) as a word, urban functions as latent topics

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

Origin

Working purpose

Destination

Collective Topic Modeling of Origin and Destination

- **The generative process of POIs in origin and destination regions**
  - Draw a trip purpose for each trip
  - Draw POIs of origin region from a trip purpose
  - Draw POIs of destination region from a 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 Optimization Problem

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

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

- Variational inference

$$\mathfrak{L} = \sum_{n=1}^{N}\sum_{m=1}^{M} \phi_{nm}(\log \pi_m + \mathbf{E}_q[\log \lambda_{i_n,m}(t_n)] + \mathbf{E}_q \log p_m(W_n)) - \sum_{i=1}^{I}\sum_{m=1}^{M} \int_0^T \mathbf{E}_q[\lambda_{i,m}(s)]ds + \mathcal{E}[q]\,.$$

$$\begin{aligned}
\phi_{nm} \propto\ & \pi_m \ : \text{prior} \\
\times & \prod_{r_o=1}^{R} (\zeta^o_{mr^o})^{\zeta^o_{mr^o}} \prod_{r^d=1}^{R} (\zeta^d_{mr^d})^{\zeta^d_{mr^d}} \ : \text{POI topics} \\
\times & \prod_{r^o=1}^{R}\prod_{c^o=1}^{C} (\beta^o_{r^o c^o})^{\zeta^o_{mr^o} W^o_{nc^o}} \prod_{r^d=1}^{R}\prod_{c^d=1}^{C} (\beta^d_{r^d c^d})^{\zeta^d_{mr^d} W^d_{nc^d}} \ : \text{POIs} \\
\times & (\gamma_m \mu_{i_n})^{\eta^m_{nn}} \ : \ \text{self triggering} \\
\times & \prod_{l=1}^{n-1} (\alpha^m_{i_l i_n} g(t_n - t_l))^{\phi_{lm}\eta^m_{ln}} \ : \ \text{influences from the past} \\
\times & \prod_{l=n+1}^{N} (\alpha^m_{i_n i_l} g(t_l - t_n))^{\phi_{lm}\eta^m_{nl}} \ : \ \text{influences to the future} \\
\times & \exp\left(-\mathcal{G}(T - t_n, t_n) \sum_{i=1}^{I} \alpha^m_{i_n i}\right) \ : \ \text{influences by trip purpose.}
\end{aligned}$$

$$\zeta^o_{m,r} \propto \prod_{c=1}^{C} (\beta_{rc})^{\epsilon W^o_{nc}}$$

$$\zeta^o_{m,r} \propto \prod_{c=1}^{C} (\beta_{rc})^{\epsilon W^d_{nc}}$$

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

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

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

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

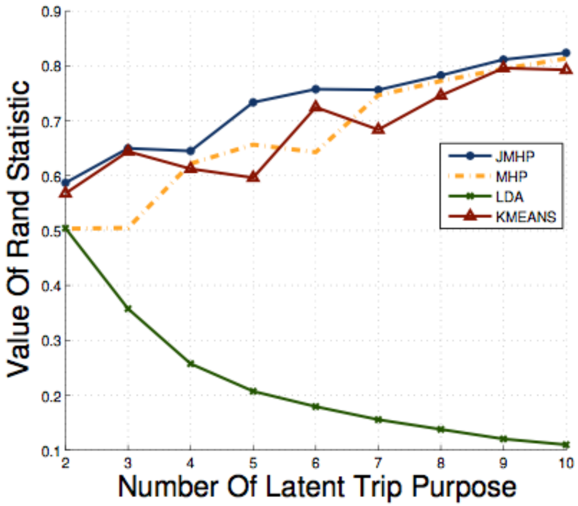
$$\beta_{rc} \propto \sum_{n}\sum_{m} \phi_{nm}(\zeta^o_{m,c} W^o_{m,c} + \zeta^d_{m,c} W^d_{m,c}).$$

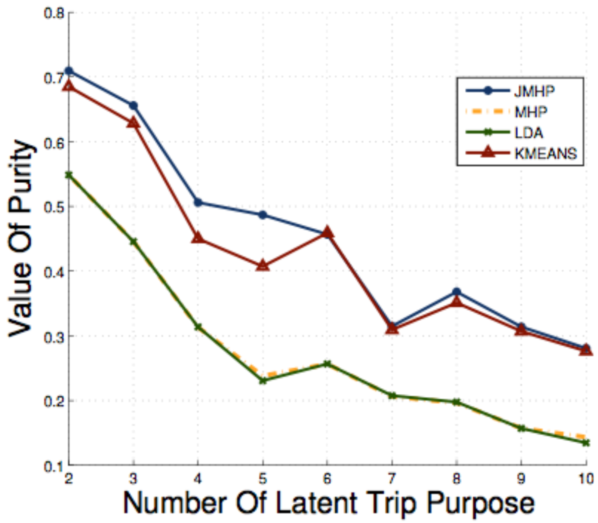
Outline

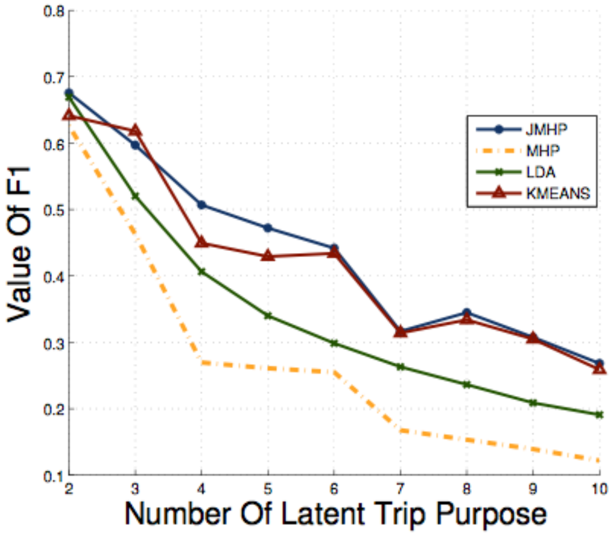
- Background and Motivation
- Problem Statement
- Methodology
- **Evaluation**
- Conclusion and Future Work

Experiment on Synthetic Data

- 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







Experiments on Real World Data

- NYC Taxi trips: 7-day taxi trips, millions of GPS trajectories, 152 valid regions
- POIs of NYC

Identified trip purposes

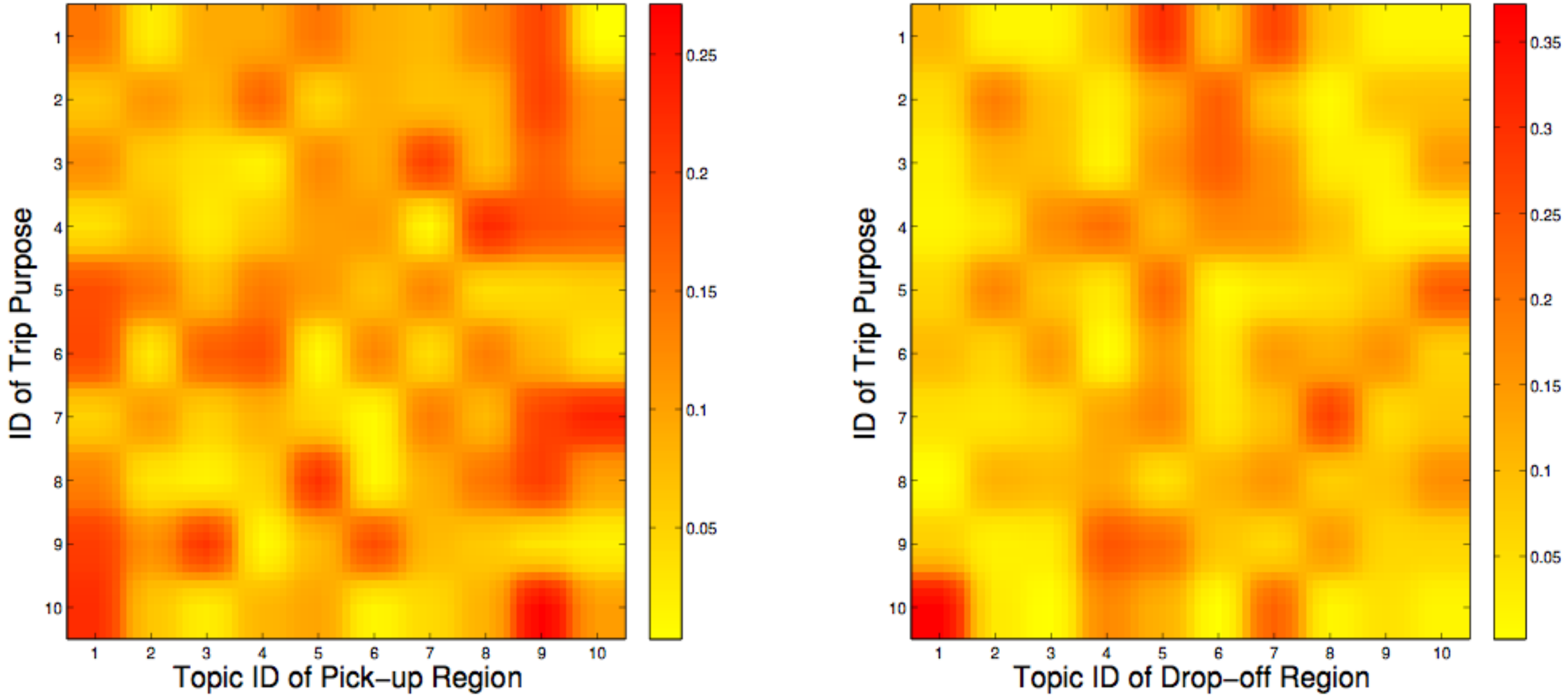
nightlife dining work shoping

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

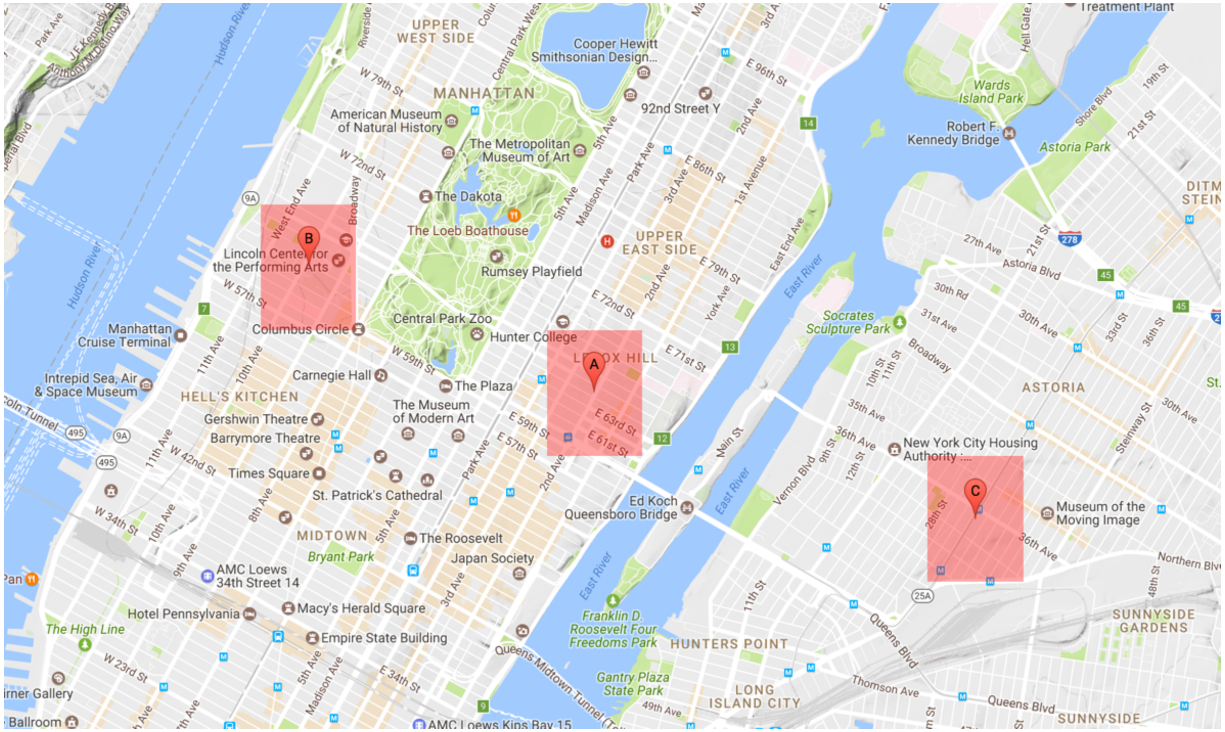
school sightseeing home

Experiments on Real World Data

- POI topic distribution over latent trip purposes for origin and destination



Synchronization Effect



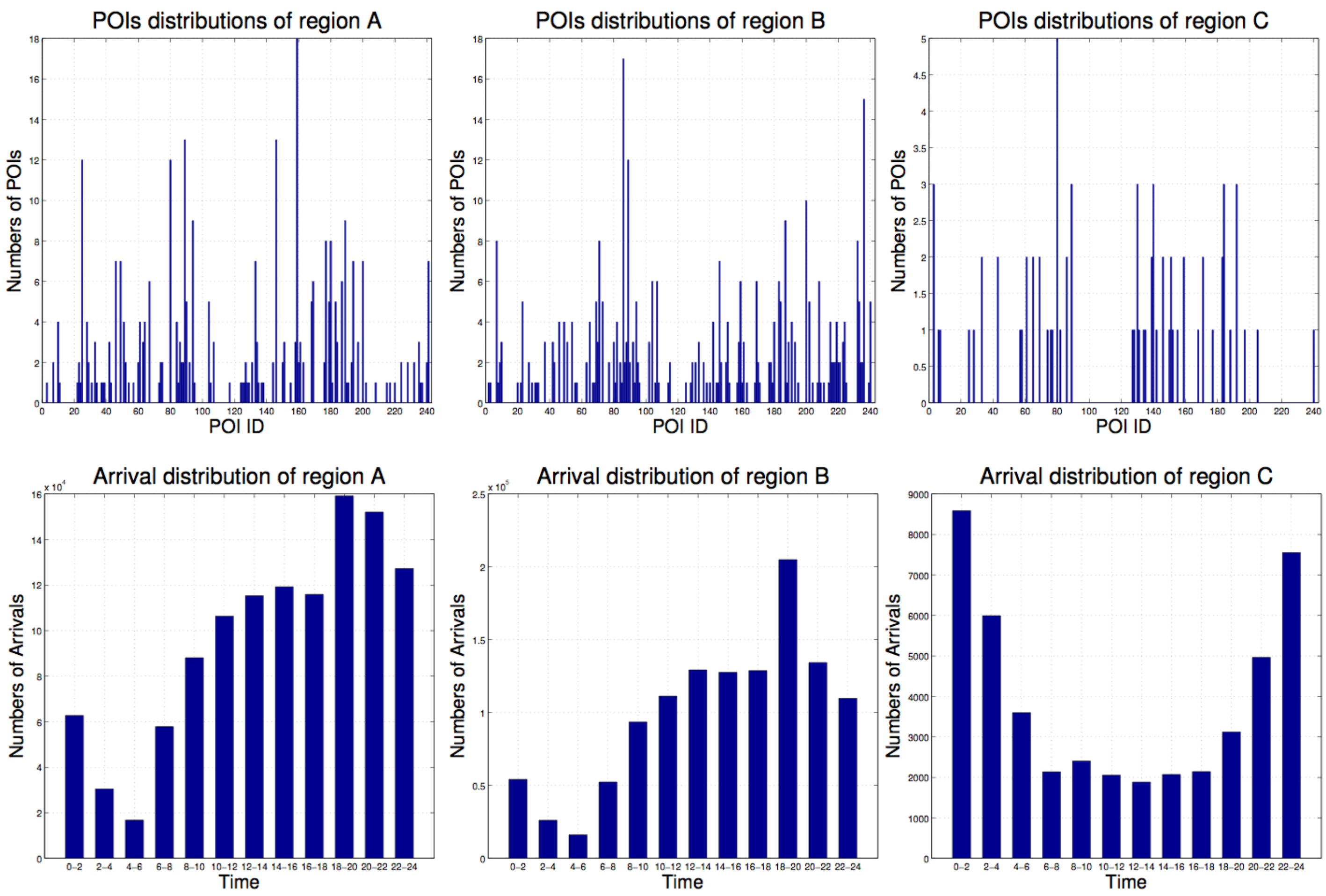
$\alpha_{AB} = 8.27066832$

$\alpha_{AC} = 0.00711464$

A and B have a higher synchronization rate

A and C have a lower synchronization rate

Synchronization Effect



Conclusion Remarks

- **Problem**
  - Human mobility modeling
- **Two research questions**
  - What is the nature of the spatial diffusion of human mobility across functional regions?
  - How to spot and trace the trip purposes of trajectories?
- **Property (provide in-depth understanding)**
  - Identify the synchronization property of human mobility
- **Modeling (make it predictable and traceable)**
  - Provide a unique perspective of modeling human mobility as stochastic point processes
  - Use human mobility synchronization property to link mobility arrivals, functional regions, and trip purposes

Questions?

Looking Forward to Future Collaboration

THANK YOU

Homepage: www. yanjiefu.com