UNI
FR

UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

eXascale Infolab

Geo-Social Media Analytics

Dingqi Yang
Senior Researcher
eXascale Infolab, University of Fribourg, Switzerland
Homepage: https://sites.google.com/site/yangdingqi/

Outline

- **Introduction to geo-social media**
- **Characteristics of geo-social media data**
- **Application of geo-social media analytics**
  - Personalized location based services
  - Computational social science
- **Concluding remarks**

Social Networks

"A social network is a social structure made up of individuals connected by one or more specific types of interdependency, such as friendship, common interests, and shared knowledge."

Social Networking Services

A social networking service builds on and reflects the real-life social networks among people through online platforms such as a website, providing ways for users to share ideas, activities, events, and interests over the Internet.

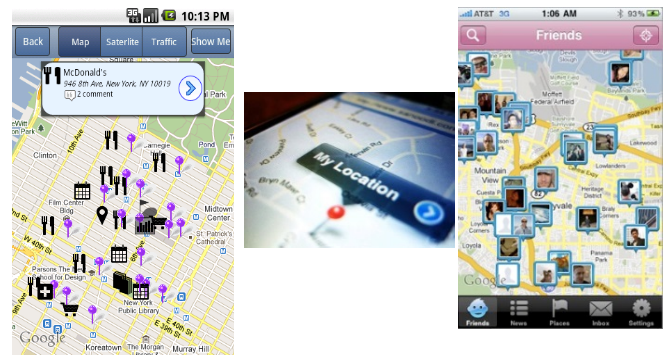
Social Network Services

- **Sharing**
  - Users can interact with each other via social network services by various means.

Geo-Social Media

- **Social network services + Smartphones**
  - What can those sensor-embedded smartphones bring us ?

Better Accessibility

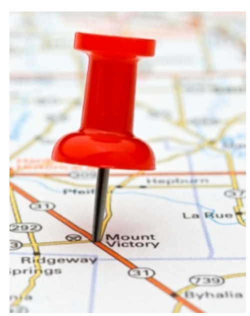


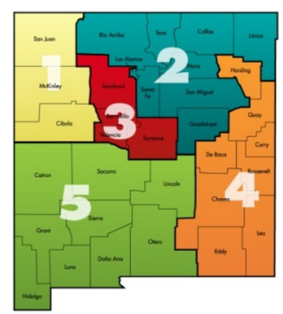
Geo-localization

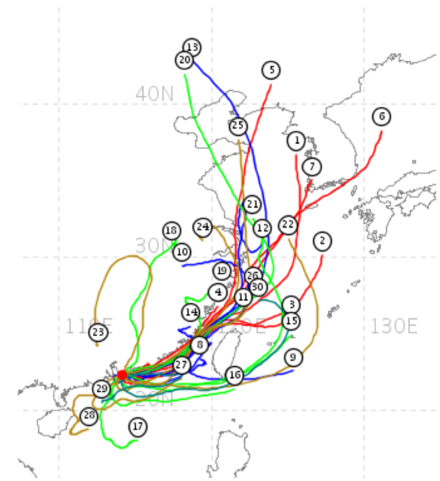
- **Location based social networks (LBSNs)**

Locations

- **Location-acquisition technologies**
  - Outdoor: GPS, GSM, CDMA, …
  - Indoor: Wi-Fi, RFID, supersonic, …
- **Representation of locations**
  - Absolute (latitude-longitude coordinates)
  - Relative (100 meters north of the Space Needle)
  - Symbolic (home, office, or shopping mall)
- **Forms of locations**
  - Point locations
  - Regions
  - Trajectories







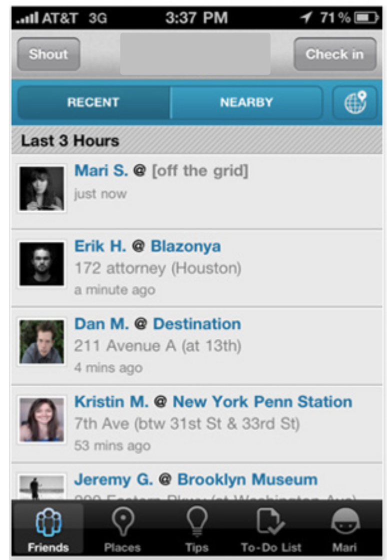
Locations + Social Networks

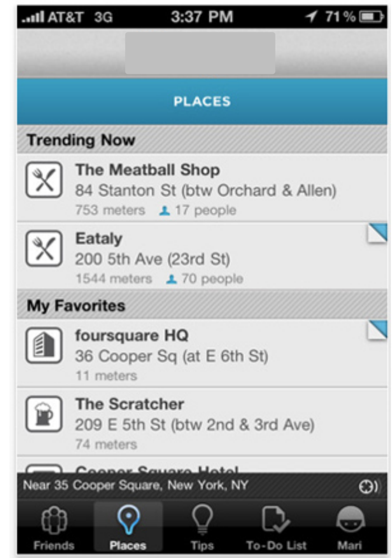
- **Add a new dimension to social networks**
  - Geo-tagged user-generated media: texts, photos, and videos, etc.
  - Recording location history of users
- **Location is a new object in the network**
- **Bridging the gap between the virtual and physical worlds**
  - Sharing real-world experiences online
  - Consume online information in the physical world

Examples

Virtual world

Sharing & Understanding





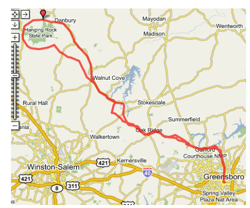
Interactions

Generating & Consuming

Physical world

Categories of LBSN Services

- ***Geo-tagged-media-based***
- ***Point-location-driven***
- ***Trajectory-centric***



LBSN Services	Focus	Real-time
Geo-tagged-media-based	Media	Normal
Point-location-driven	Point location	Instant
Trajectory-centric	Trajectory	Relatively Slow

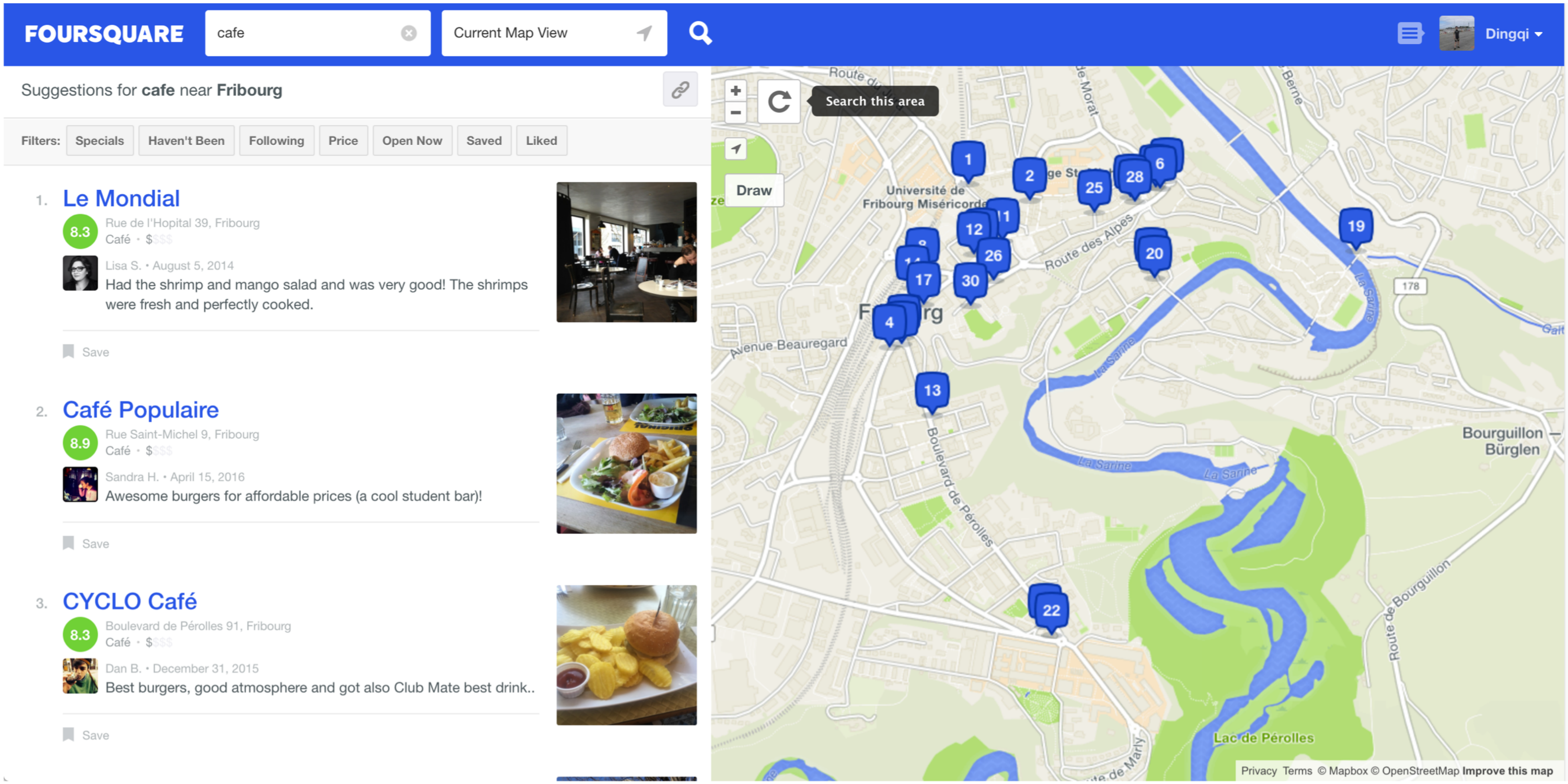
Location Based Social Networks-Foursquare

- **A location sharing service launched in 2009**

Check-in

Mayor

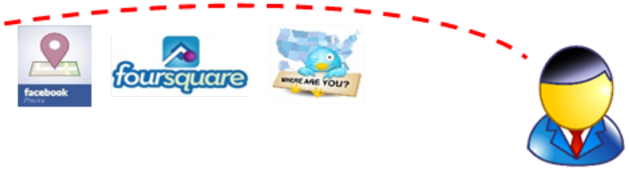
Location Based Social Networks-Foursquare



Location Based Social Networks-Foursquare

- **Active digital footprints of users**
  - A social representation of user daily activities.
  - Rich semantic information.

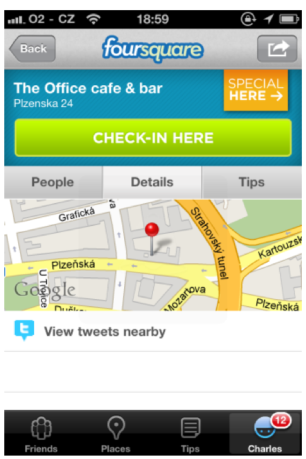
- **Big location-centric social media data**
  - By January 2014, Foursquare has attracted 45 million users, with 5 billion check-ins, with millions more check-ins every day.



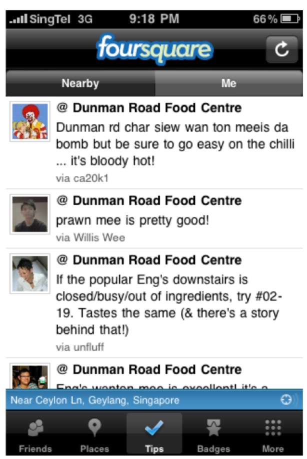
It's 20:20 pm, Jim is having his dinner in a French Restaurant at [40.71448, -74.00598]. He leaves a comment saying that "Good sea food here and reasonable price !"

Foursquare Data Collection

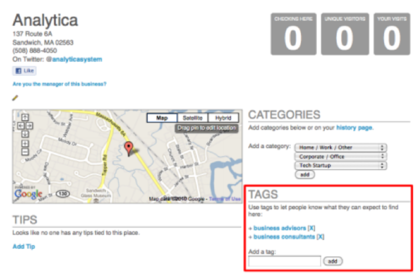
- **Foursquare**
  - Entities
    - User
    - Venue (i.e., Points of Interest)
    - Media content (e.g., status, tips, photos)
  - Active digital footprints



Check-ins



Leave tips

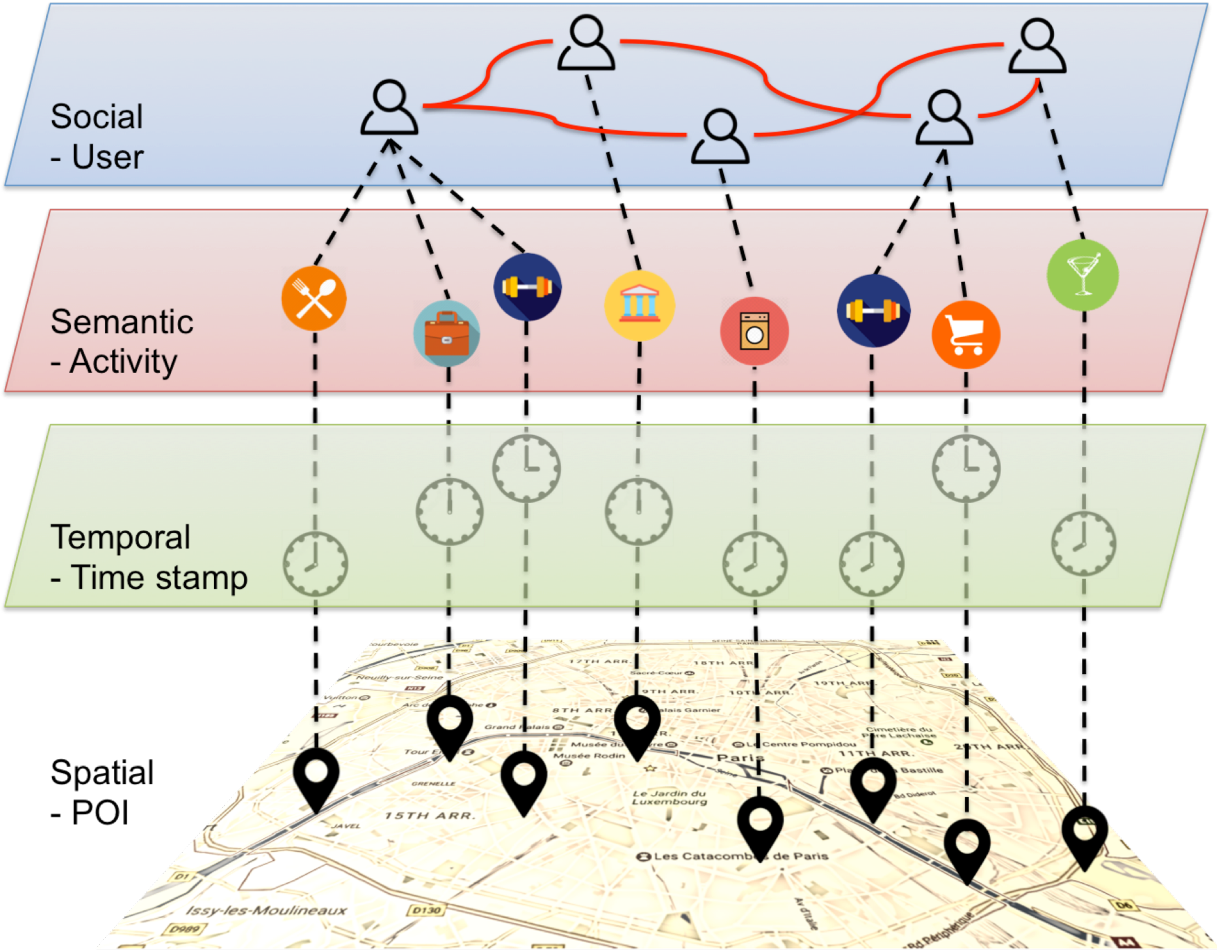


Add tags

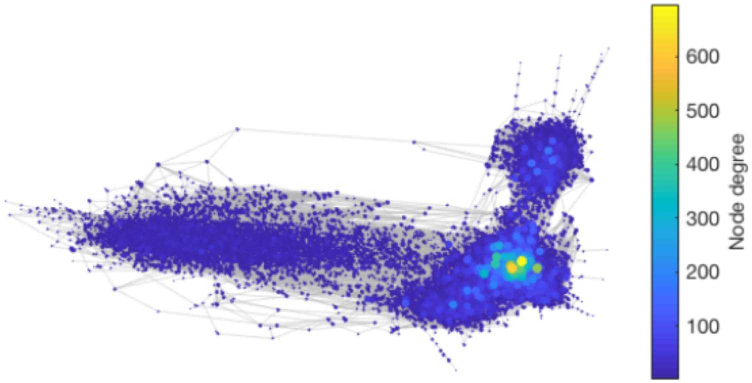
Statistics snapshot of collected data (about 18 months from April 2012 to September 2013)

2,418,223	users
10,428,709	venues
81,571,174	check-ins

Data Structure in Foursquare

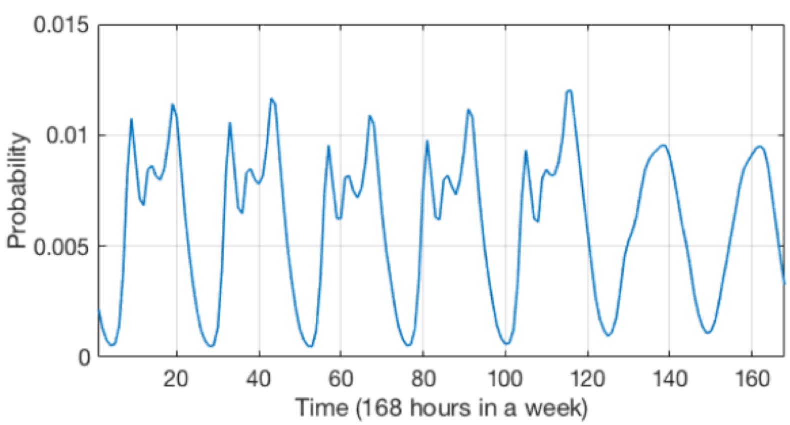


Visualization

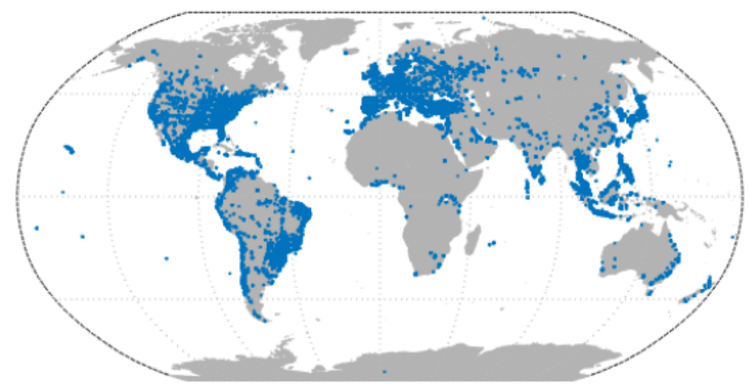


(a) Social network visualization

(b) Word cloud of check-in POI categories



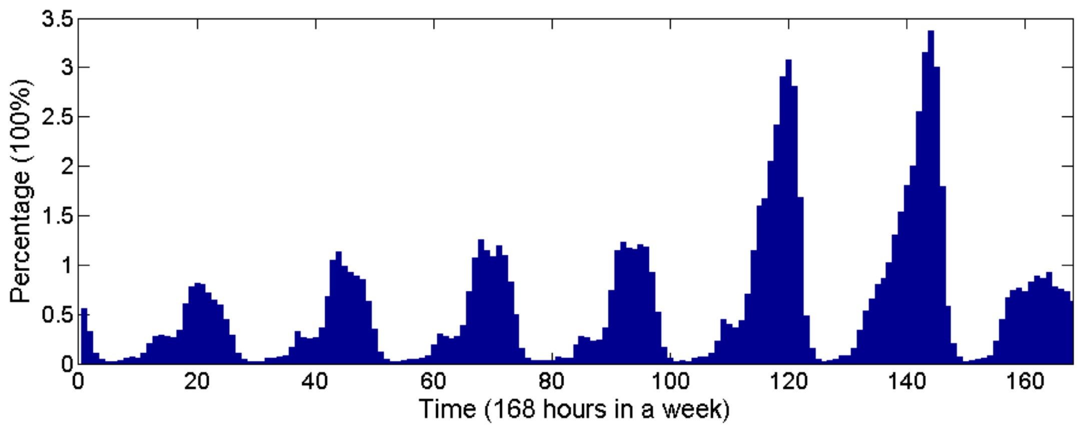
(c) Temporal traffic of check-ins

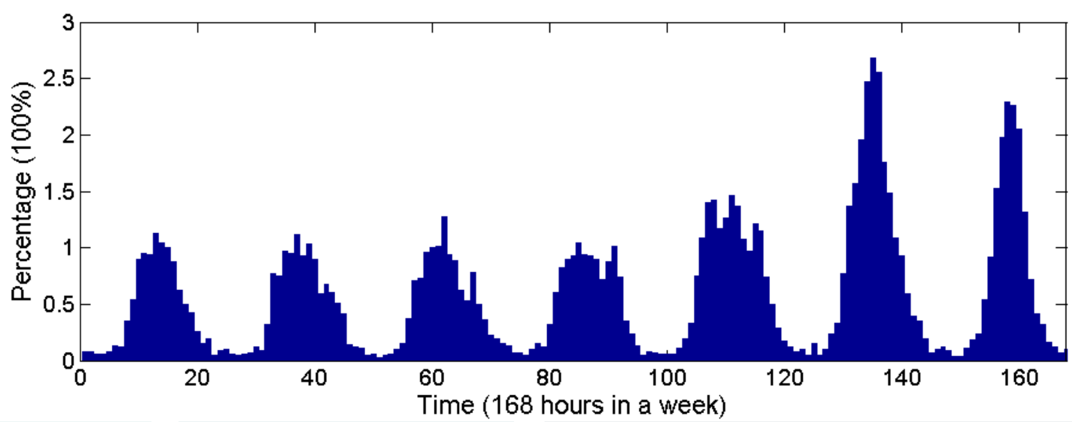


(d) World map of check-in POIs

Temporal Traffic Patterns

- **Traffic patterns of POIs with different categories**



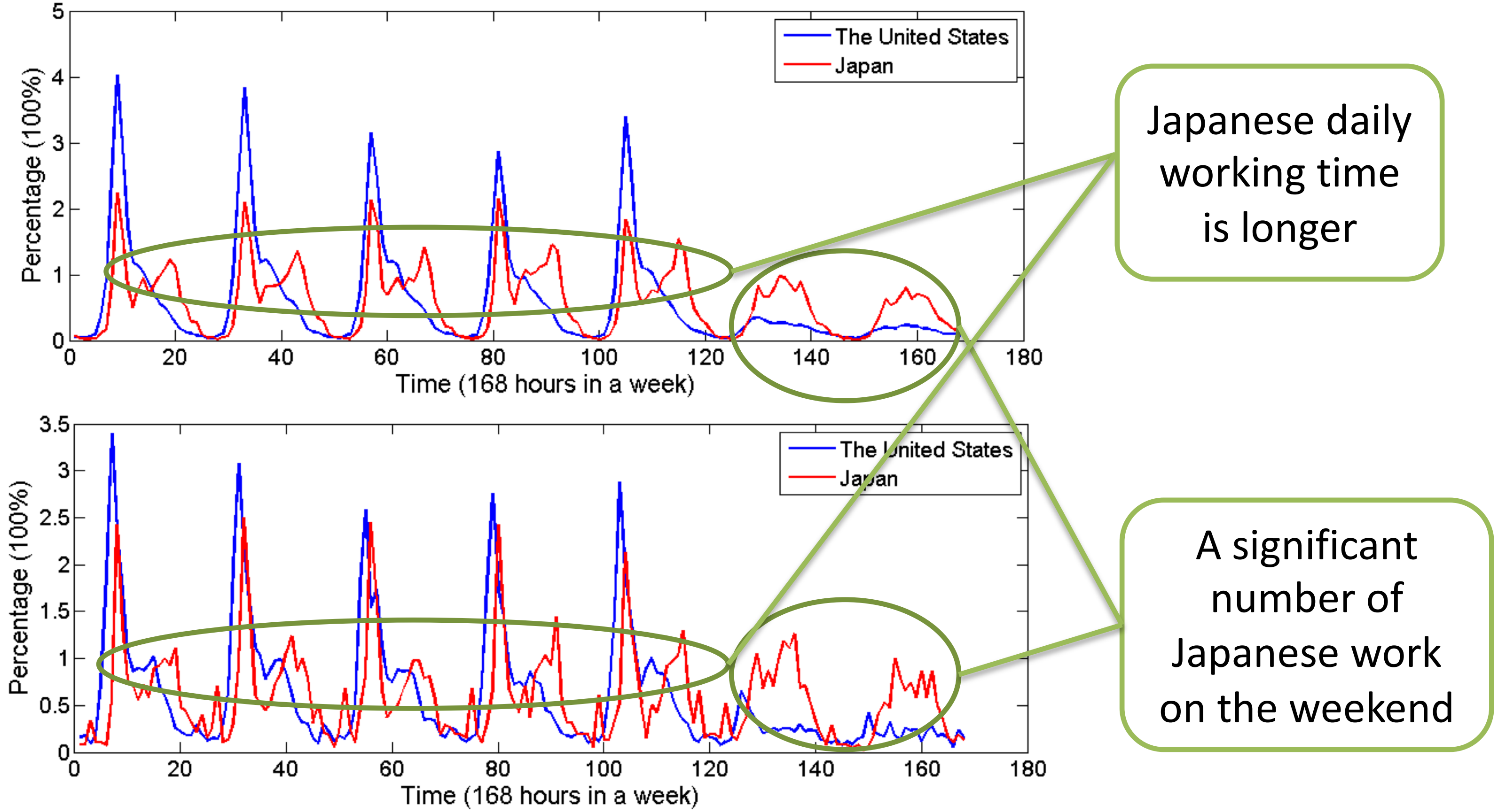


Temporal Traffic Patterns

- **Traffic patterns of POIs in different countries (US and JP)**

Office

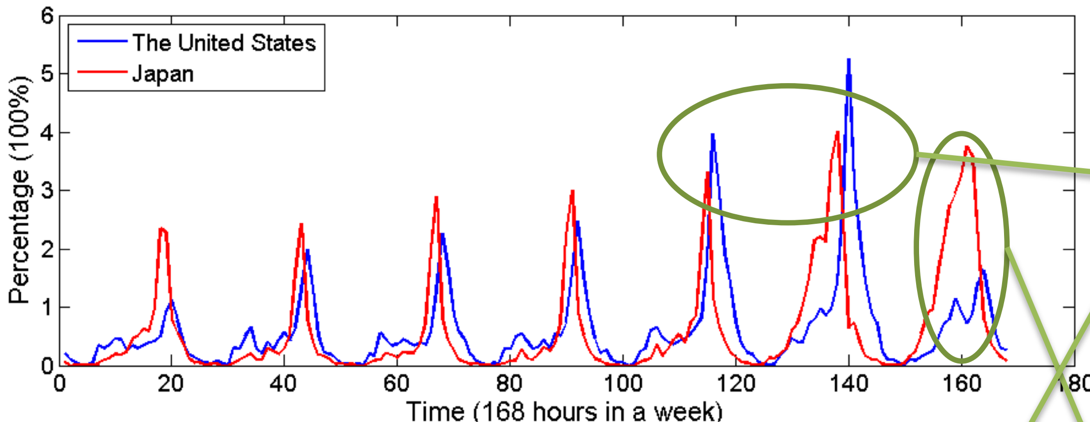
Factory



Temporal Traffic Patterns

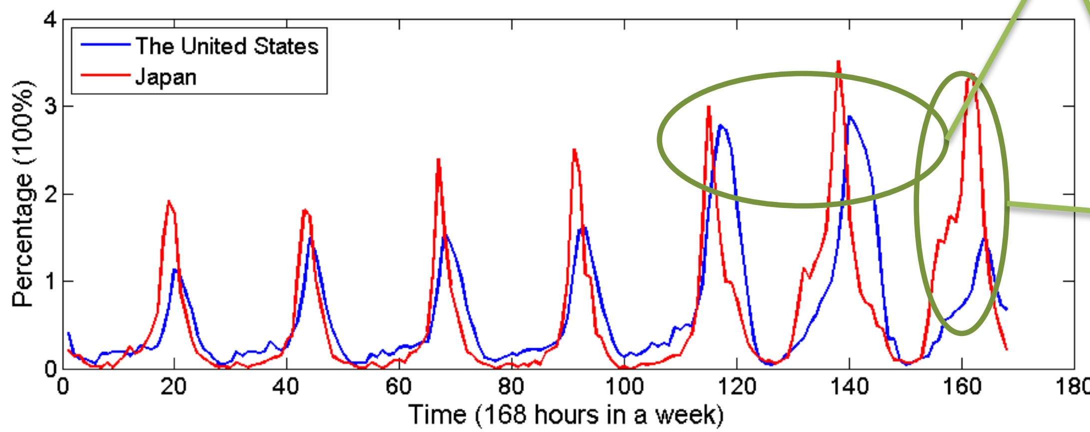
- Traffic patterns of POIs in different countries (US and JP)

Concert



The US' concert/music activities start later

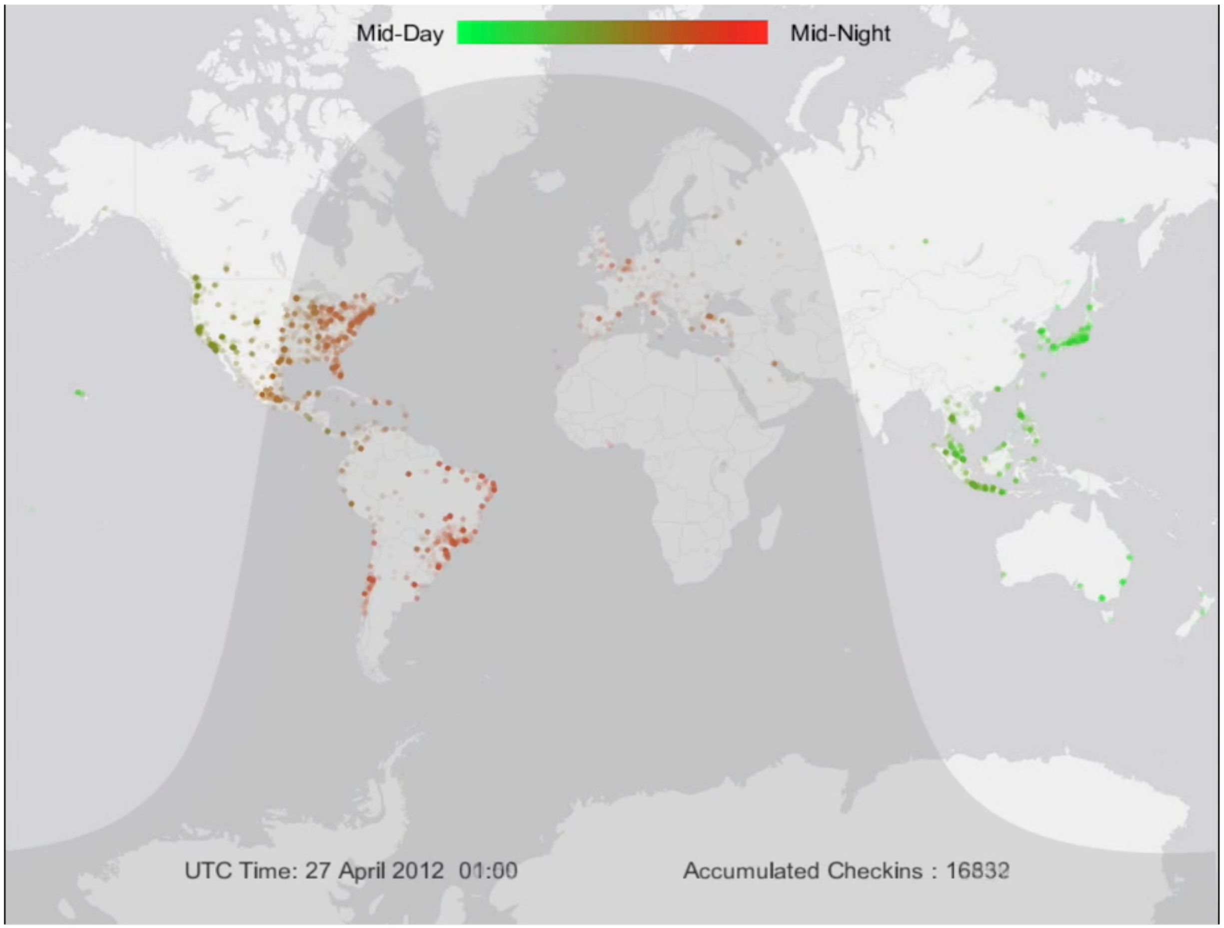
Music venue



There is significant less traffic on Sunday in the US

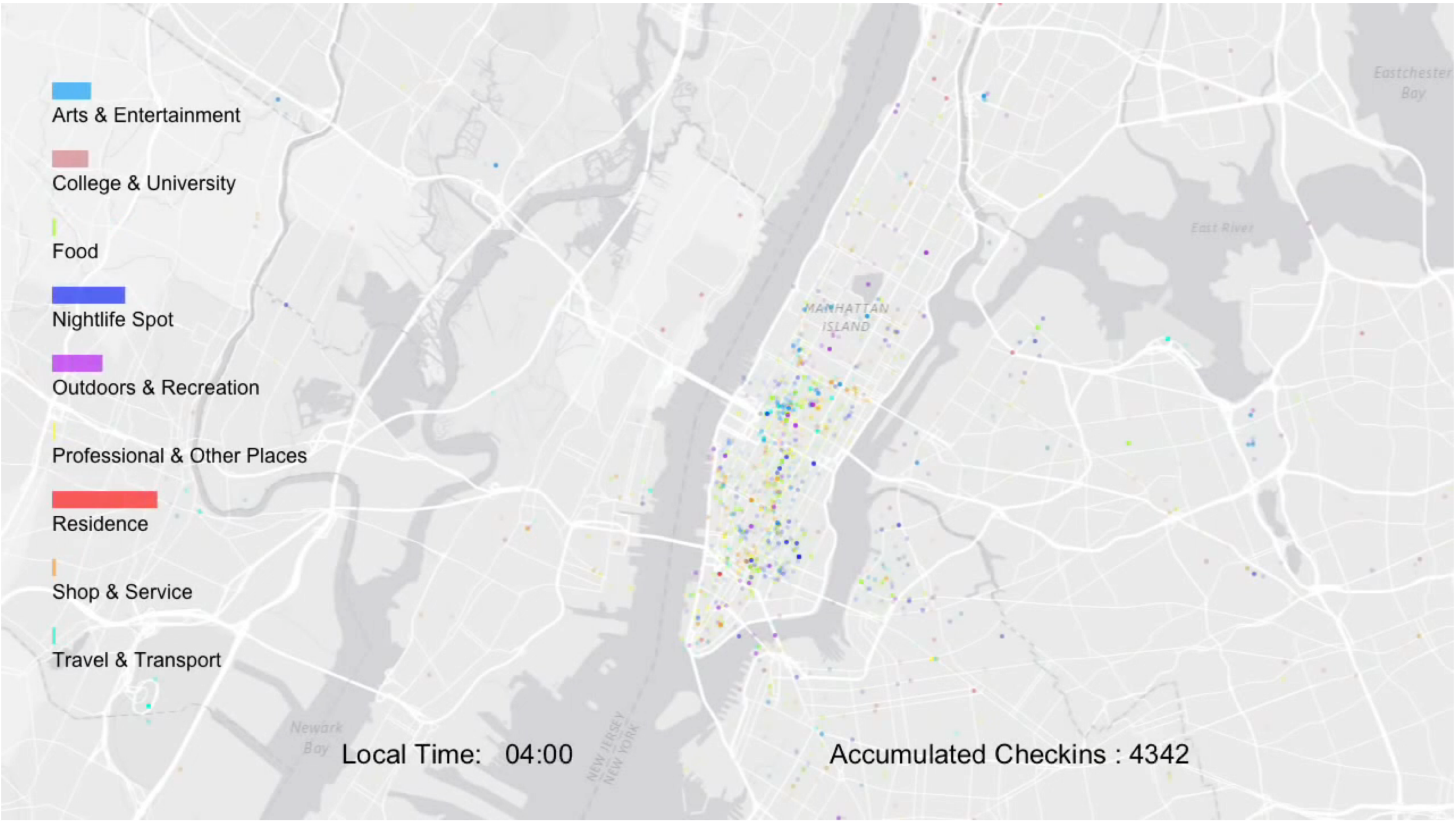
Data Visualization

- **One day Foursquare check-ins on a global scale**



Data Visualization

- **One typical day Foursquare check-ins in NYC**

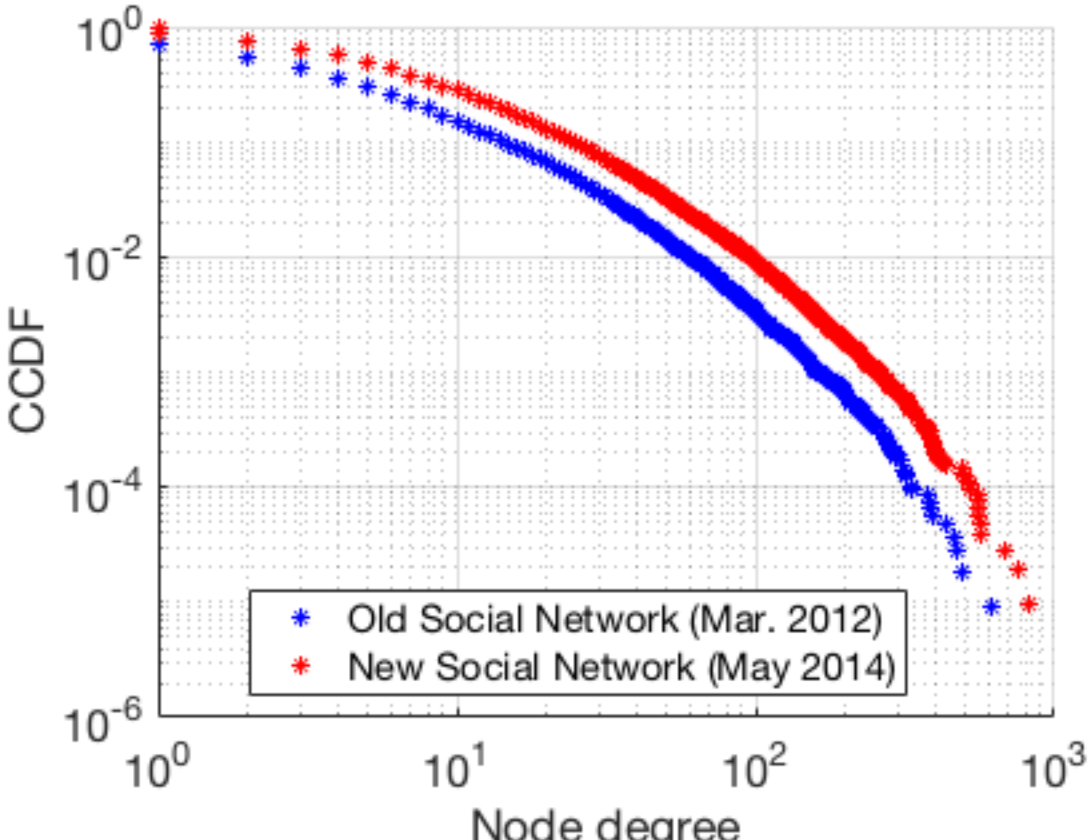


More Visualization

- **Foursquare check-ins show the pulse of cities**
  - https://foursquare.com/infographics/pulse

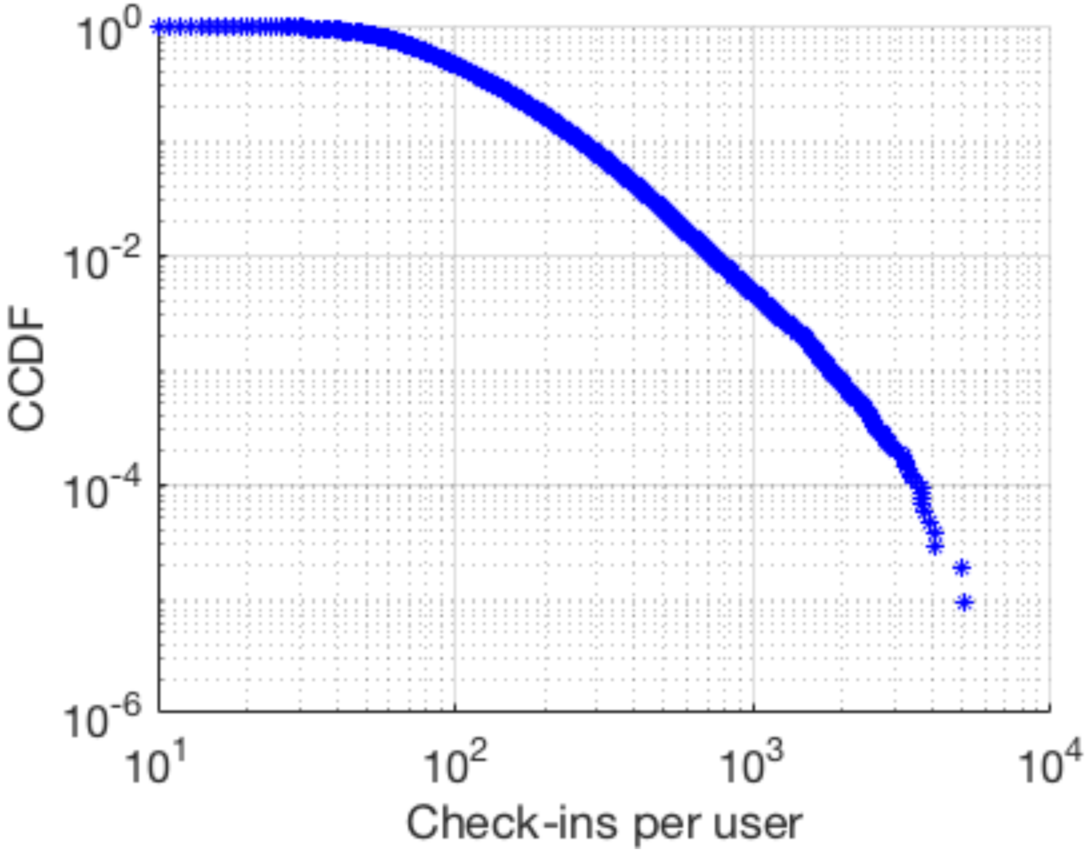
Statistical Characteristics

- **User (node) degree**



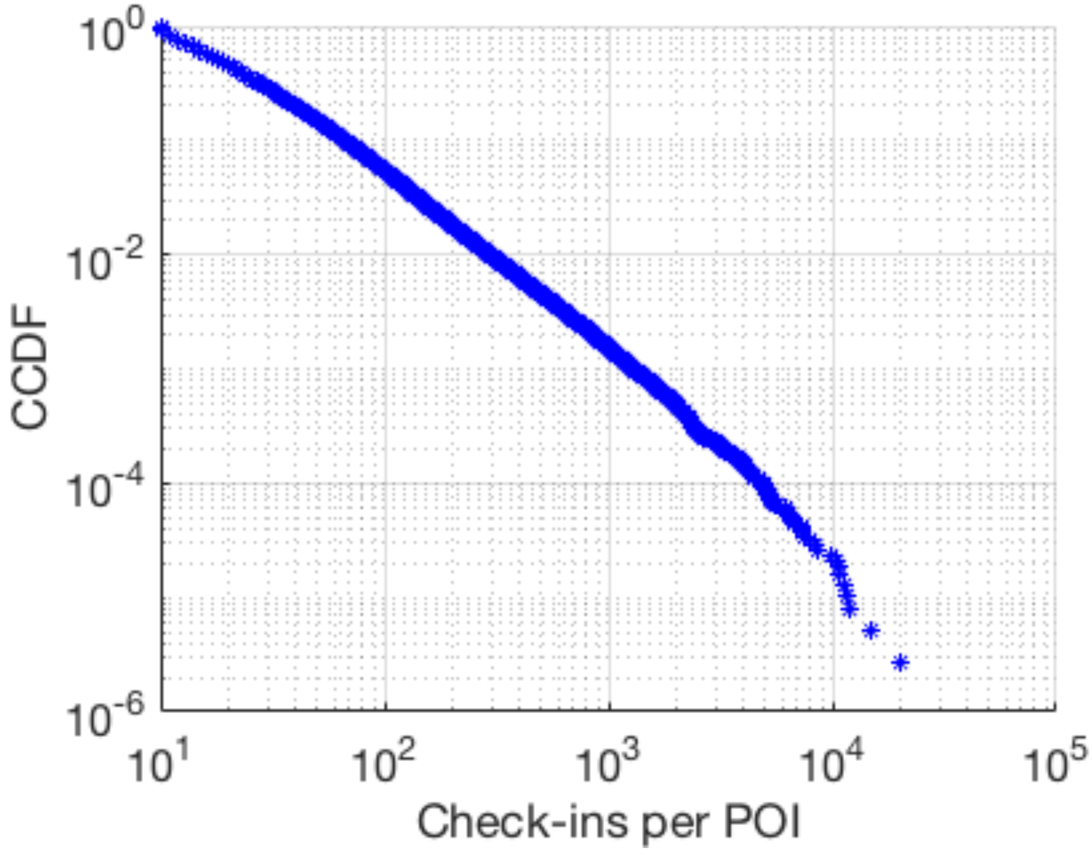
Statistical Characteristics

- **Number of check-ins per user**



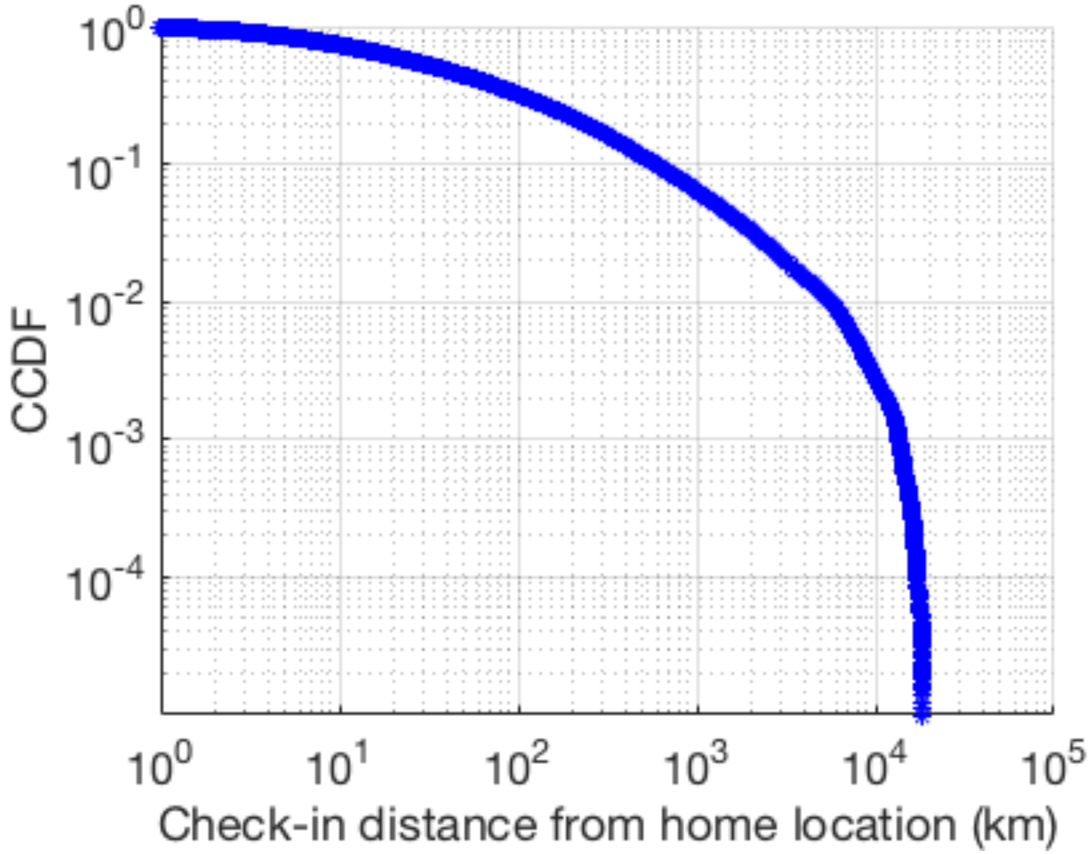
Statistical Characteristics

- **Number of check-ins per POI**

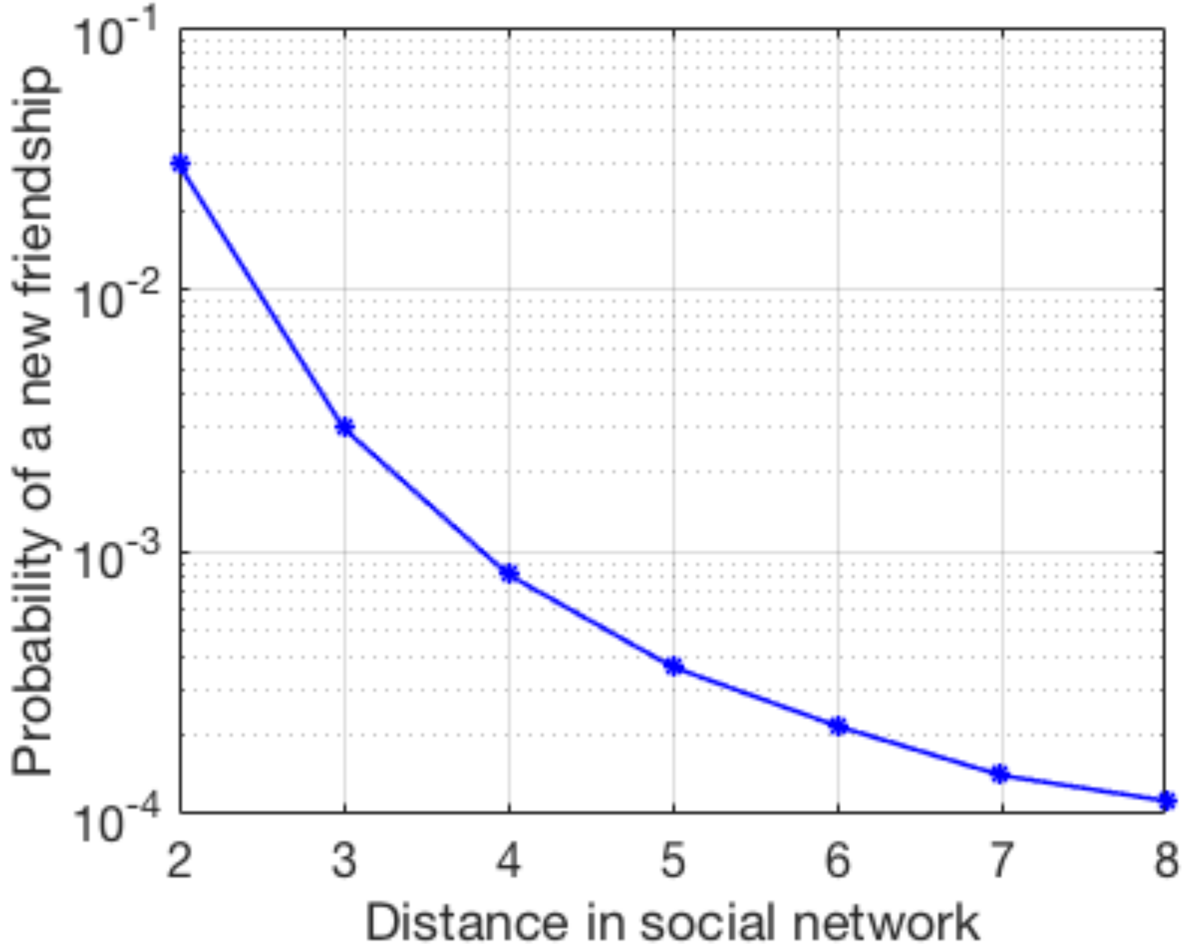


Statistical Characteristics

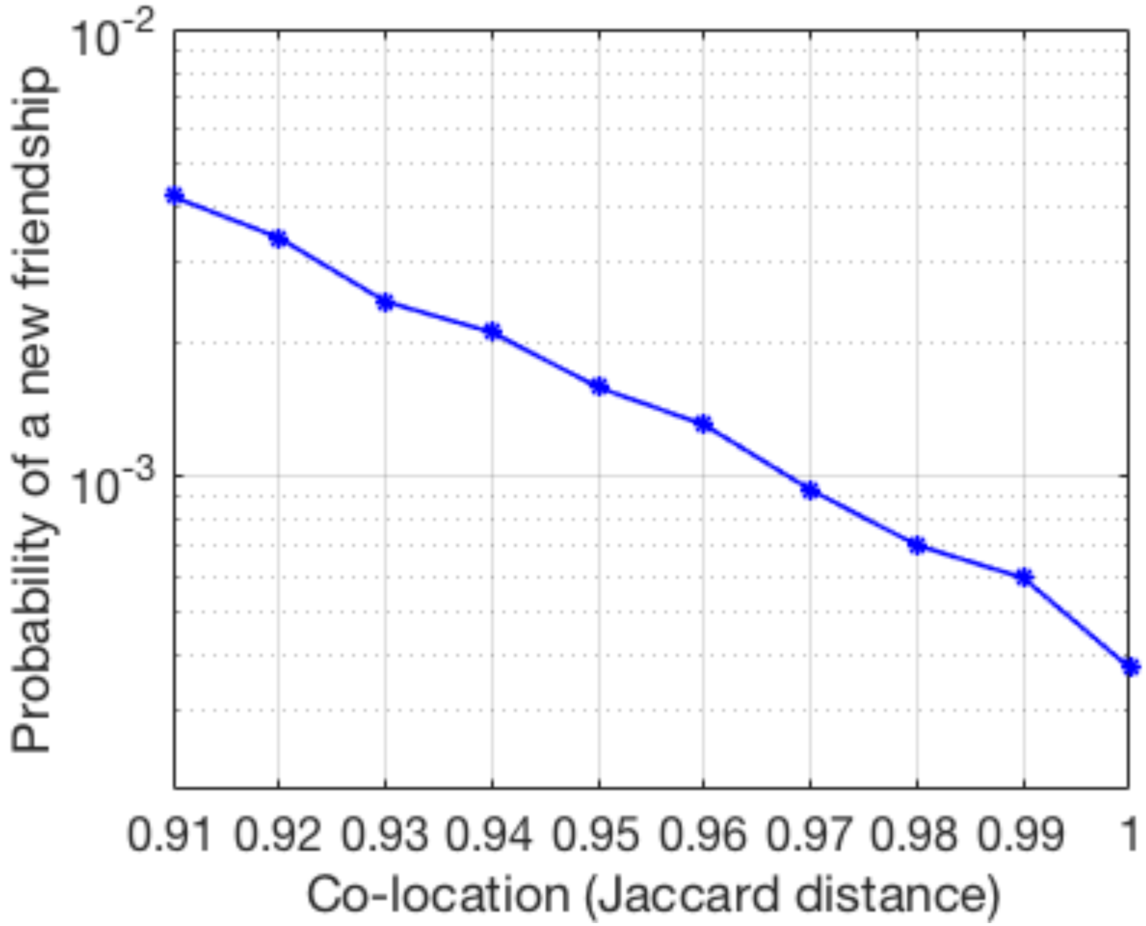
- **Check-in distance from home**



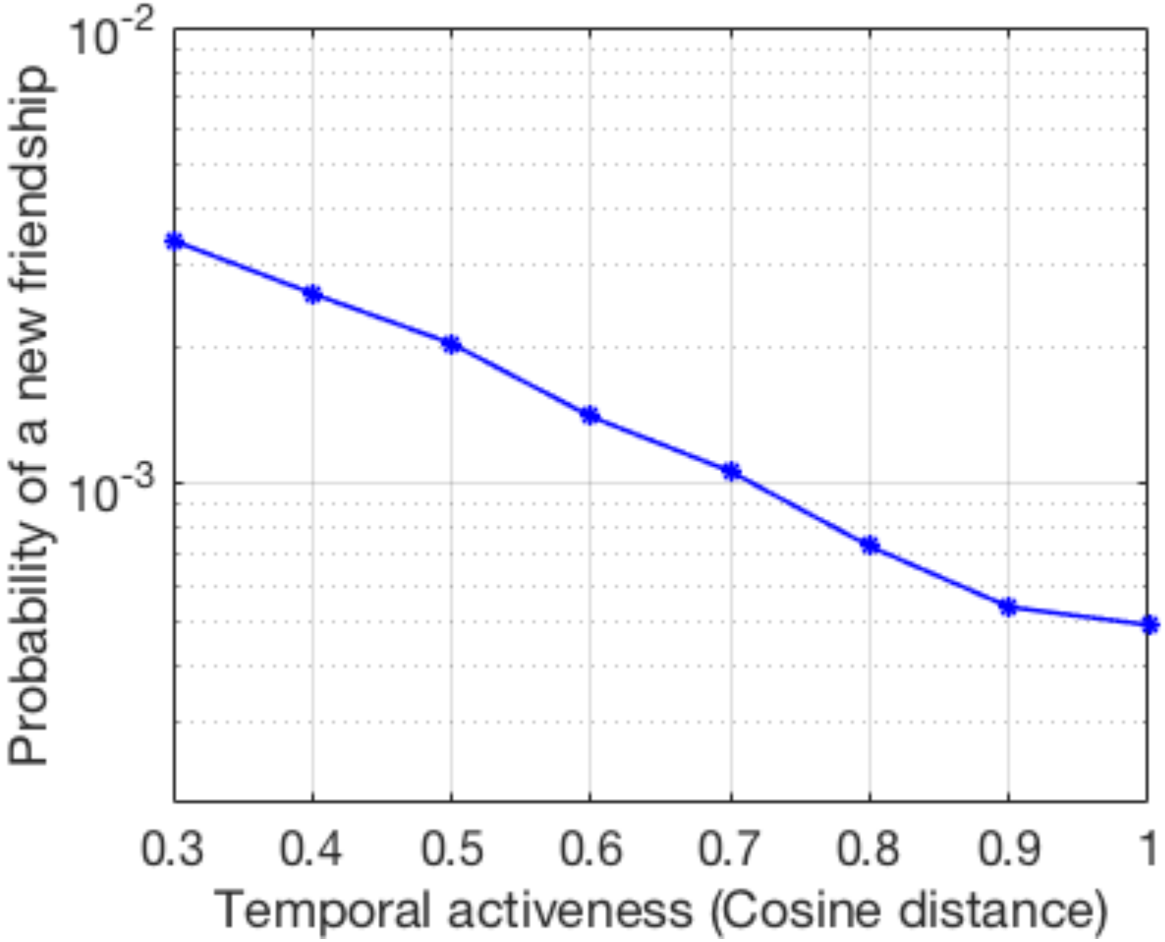
Predicting future social relationship



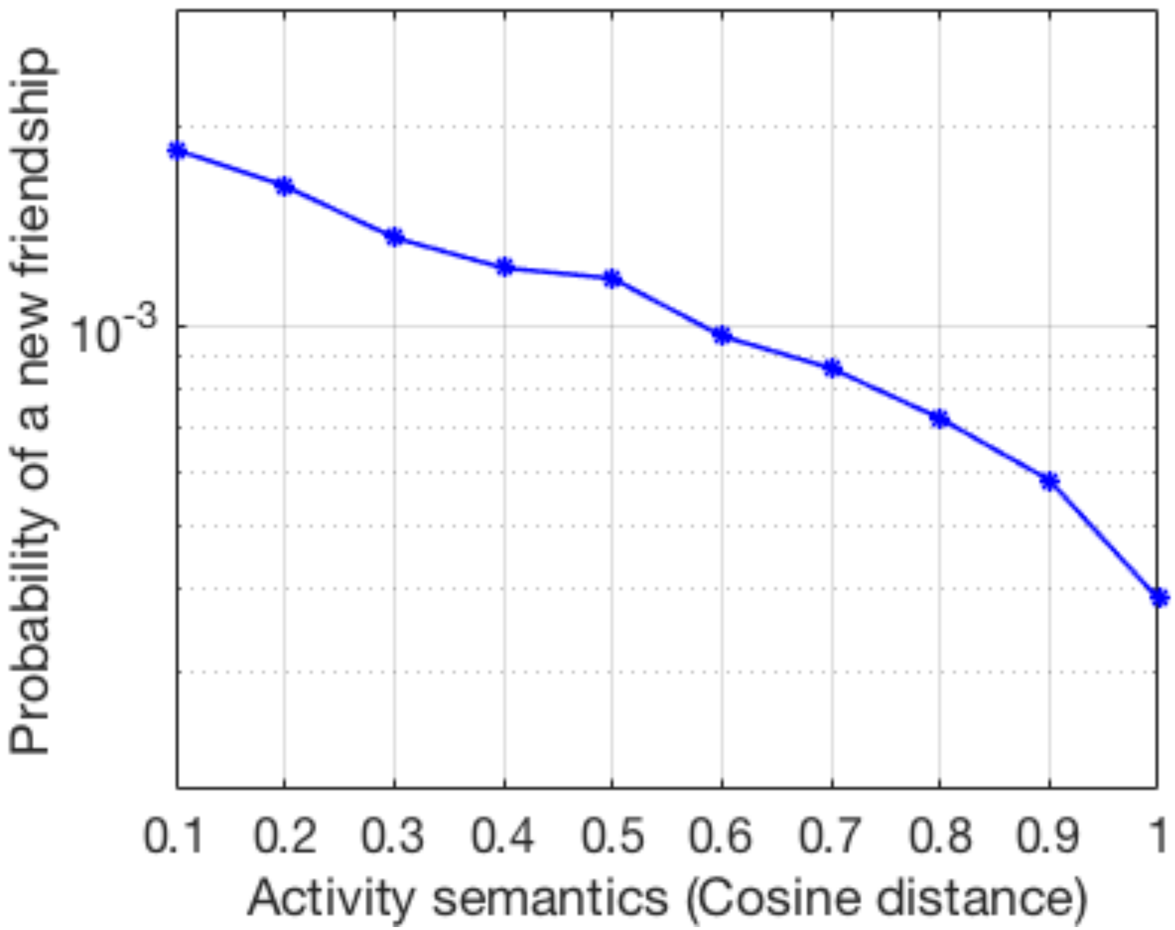
Impact of mobility on social relationship



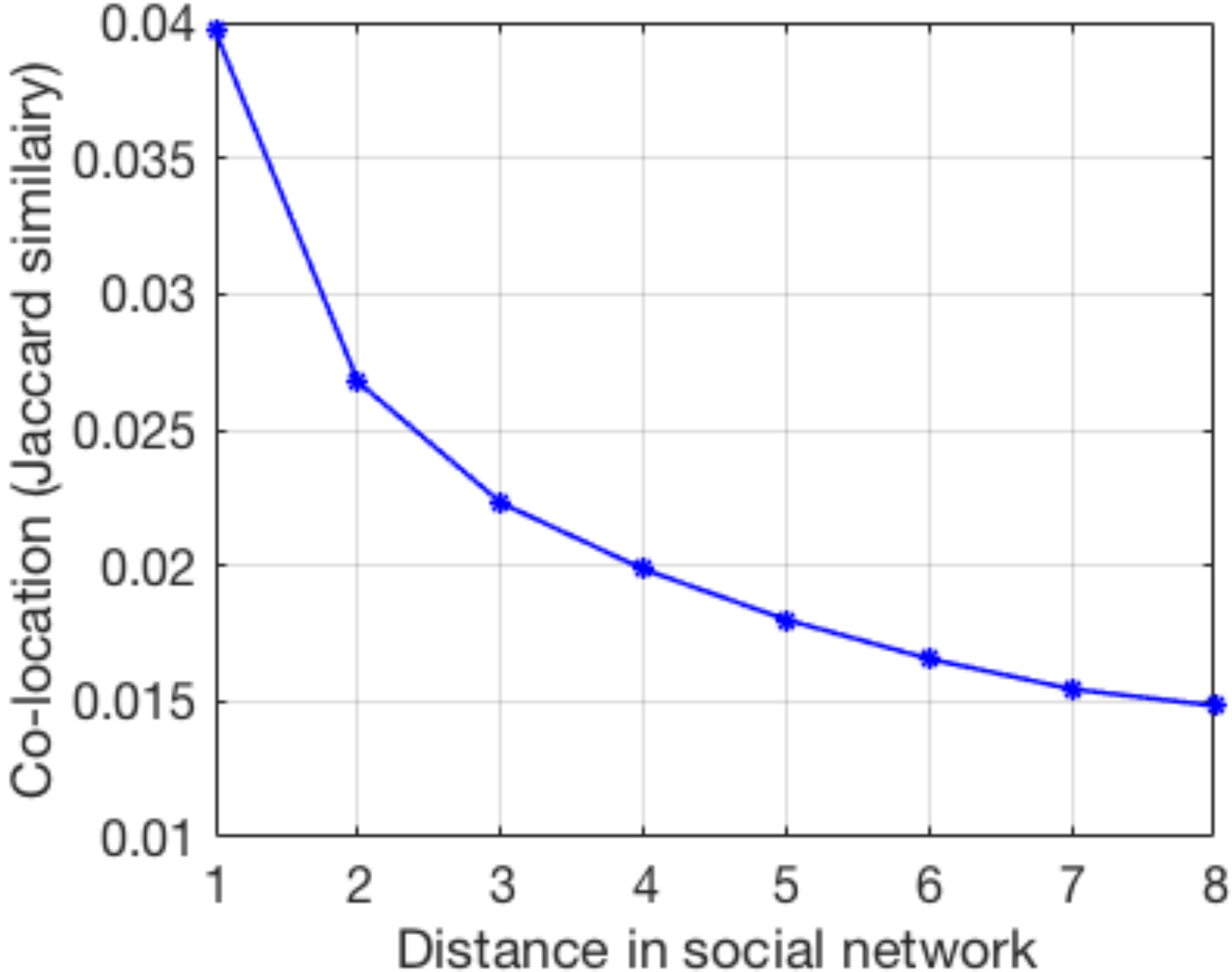
Impact of mobility on social relationship



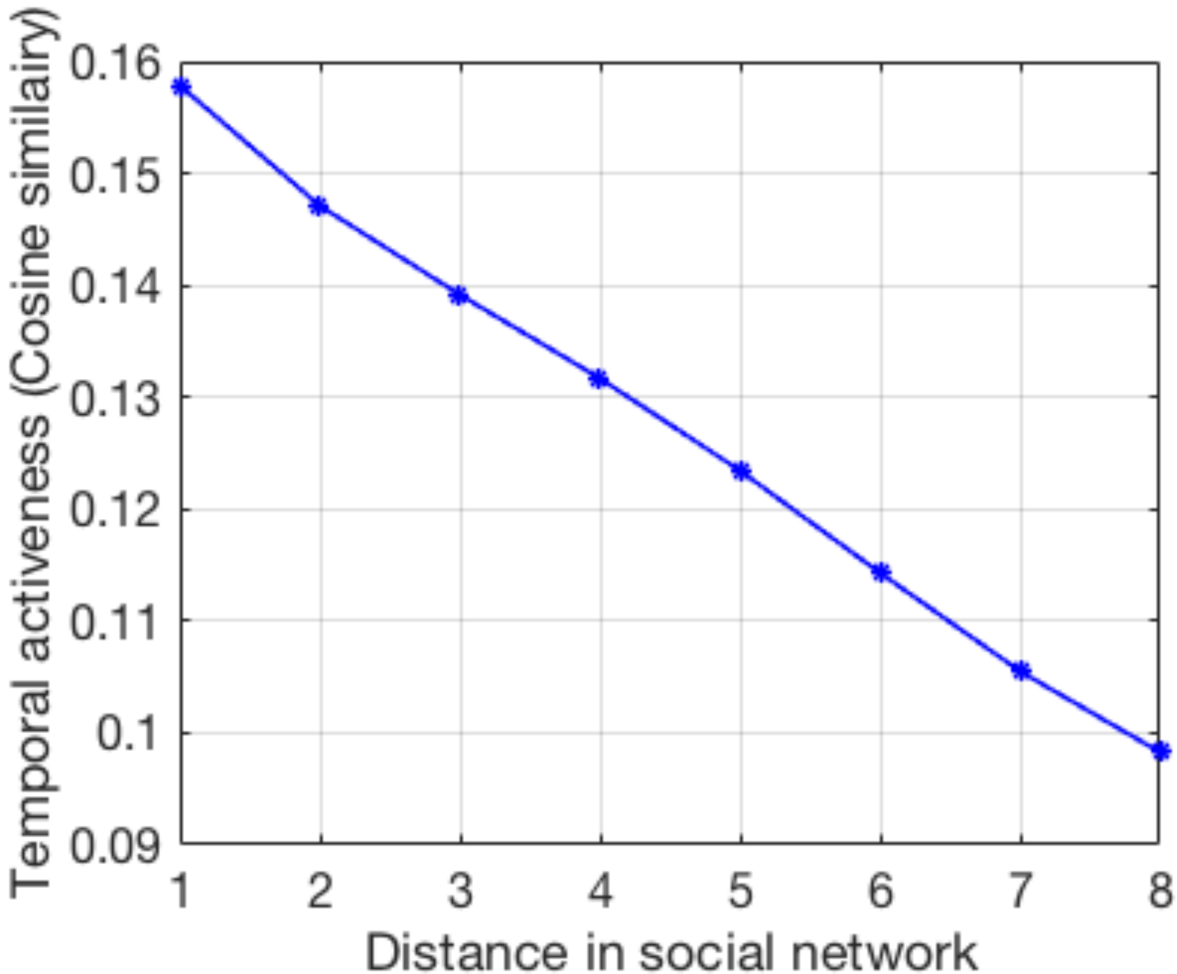
Impact of mobility on social relationship



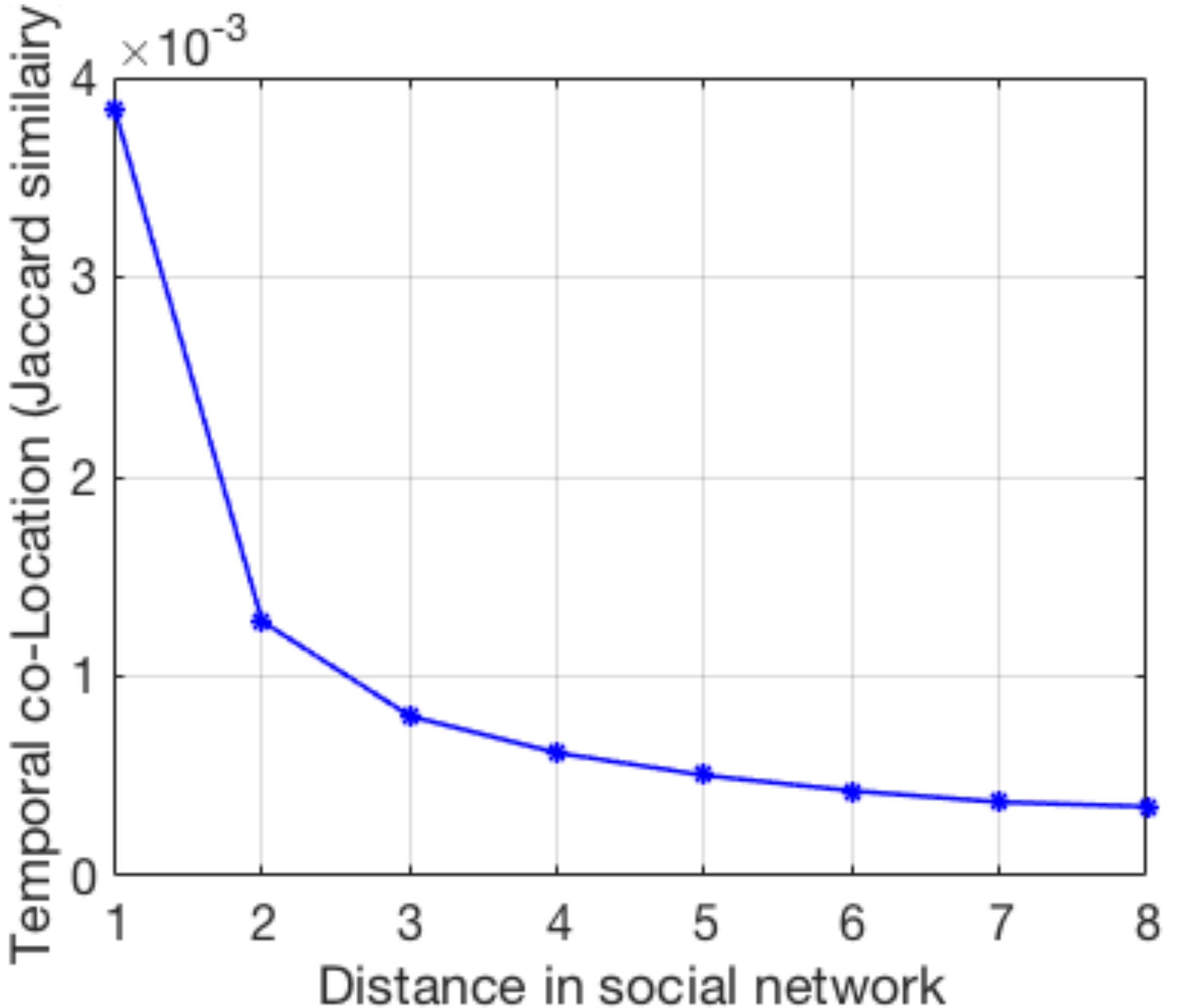
Impact of social network on mobility



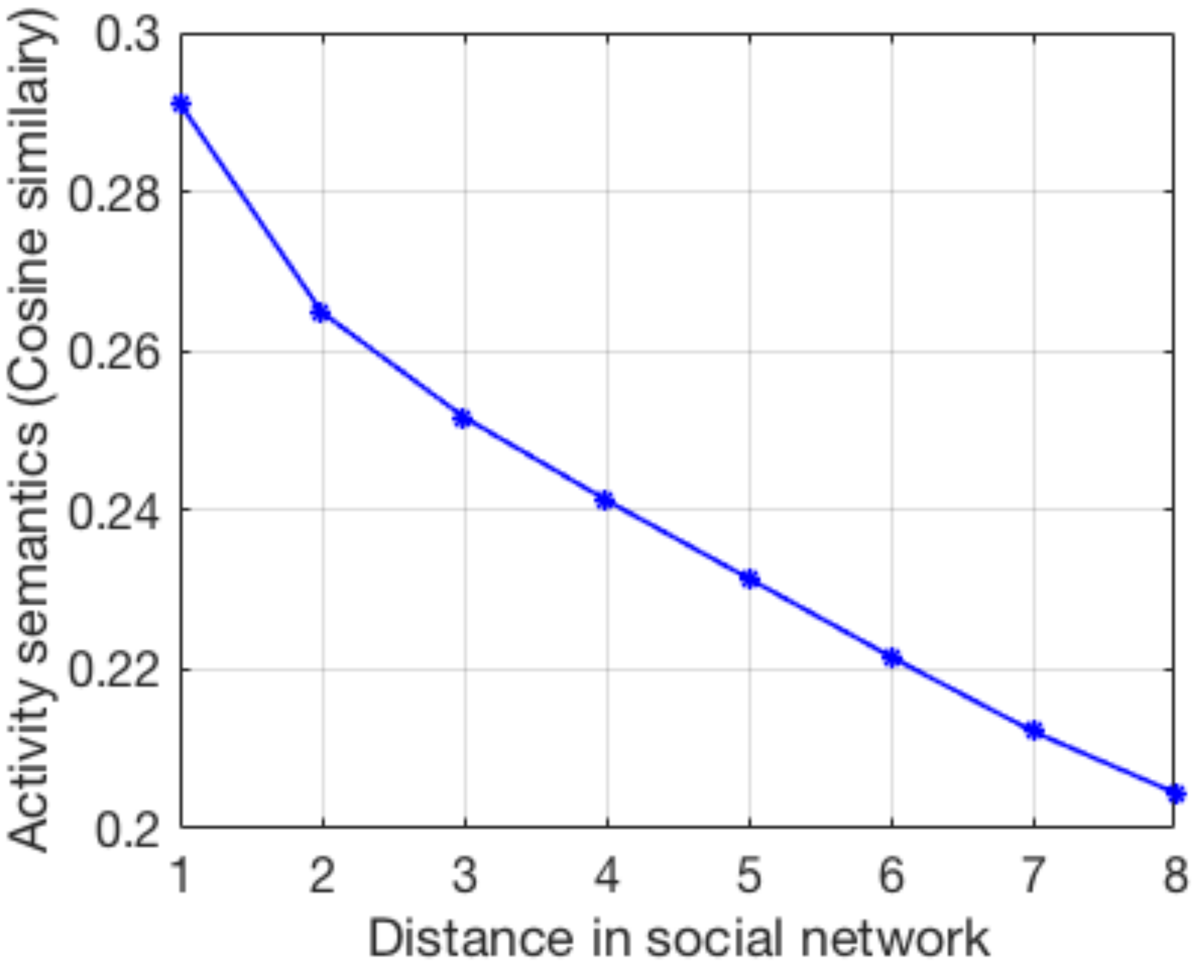
Impact of social network on mobility



Impact of social network on mobility



Impact of social network on mobility

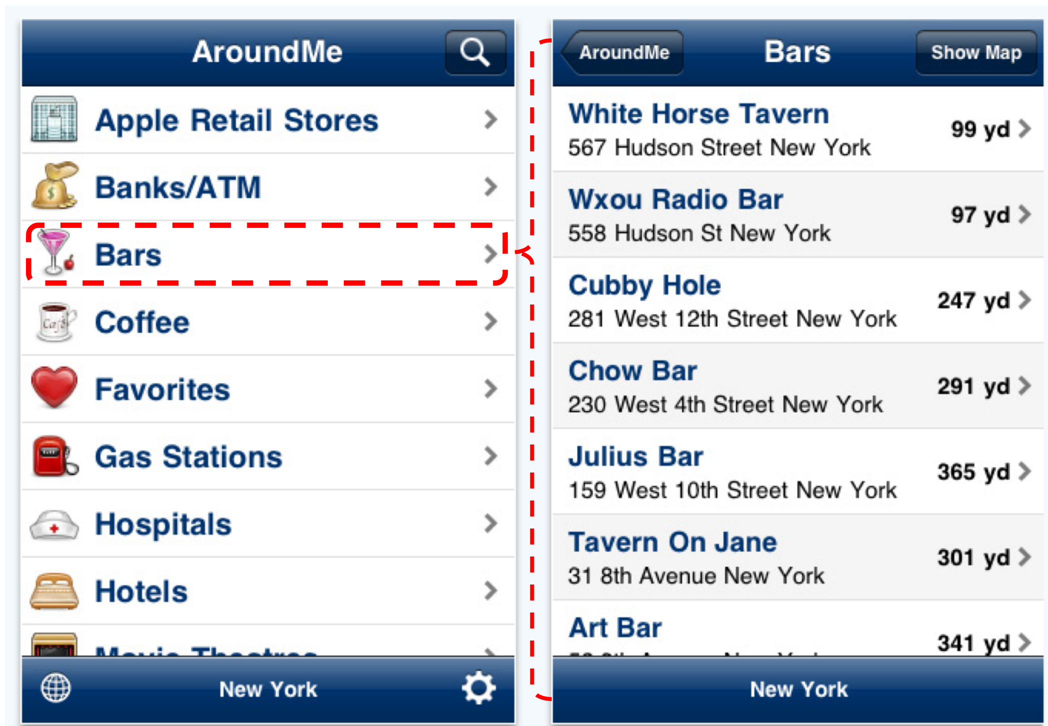


Application of Geo-Social Media Analytics

- **Personalized location based services**
  - Modeling spatial-temporal user activities
  - Understanding user preference on POIs
- **Computational social science**
  - Participatory cultural mapping

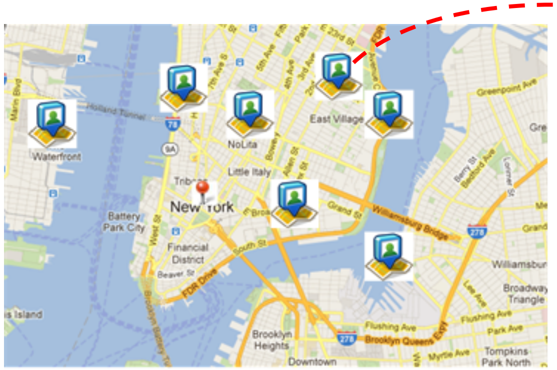
Personalized Location Based Services

- Modeling spatial-temporal user activities
- Understanding user preference on POIs



Modeling Spatial-Temporal User Activities

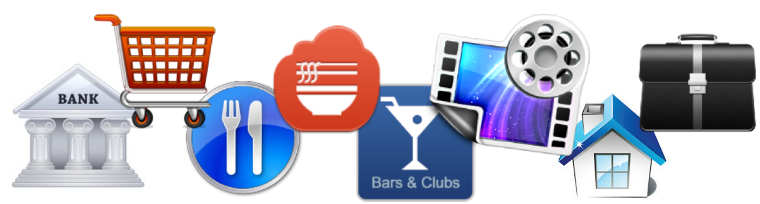
- **Objectives:**
  - Understanding spatial-temporal characteristics of user activities.
  - Inferring a user's activity preference under a given context, i.e., geo-location and time.



Context-aware LBS

It's Friday 20:00 pm, Jim is at [40.71448, -74.00598], which activity is he interested in?

Represented by 251 POI categories



Modeling Spatial-Temporal User Activities

- **Challenges**
  - High dimensionality of *user-location-time-activity quadruples*
    - Continuity of location dimension
    - Sparsity of check-in data

- **STAP model:**
  - To reduce the problem complexity, we separately consider spatial and temporal patterns of user activities
  - Two unique features of user activity patterns
    - ***Spatial Specificity***
    - ***Temporal Correlation***
  - Context-aware fusion framework for spatial-temporal activity inference

Spatial Characteristics of User Activities

- **Spatial Specificity**
  - Users' activities in LBSNs often show strong preference bias in their frequented regions.

Temporal Characteristics of User Activities

- **Temporal Correlation**
  - Users who share similar temporal preference on some activities are likely to have similar temporal preference on others.

The selection of these users as a group is based on the community detection method proposed in our previous work (Zhu Wang, et. al. Discovering and Profiling Overlapping Communities in Location Based Social Networks, IEEE Transactions on Systems, Man, and Cybernetics: Systems)

Modeling Spatial specificity

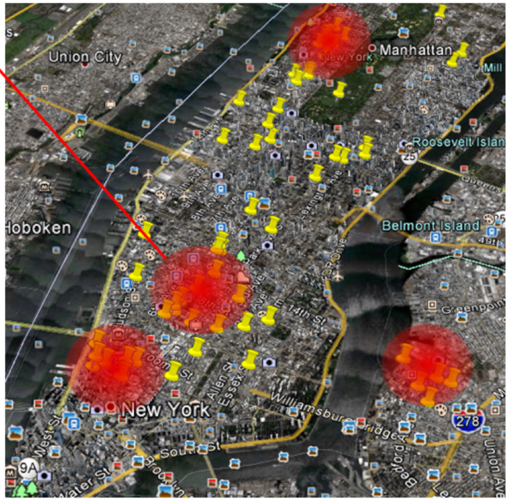
- **Spatial Specificity – Personal Functional Region**
  - A user frequented region with strong activity preference bias.

Where the user performs more than a certain percentage of all her check-ins.

Where the user has strong preference bias on certain activities (measured by the ratio of preference bias).

Nightlife Region

Category	Check-in count	Probability
Sports bars	15	0.5
Jazz bars	9	0.3
French Restaurant	3	0.1
Fast food restaurant	2	0.0667
Movie theater	1	0.0333



$$ratio_{PB} = 1 - \frac{H(\psi_{u,r})}{H_{max}(|C_l^d|)}$$

$$H(\psi_{u,r}) = -\sum_{c_i \in \psi_{r_u}} P(c_i)\log_2 P(c_i) = 1.78$$

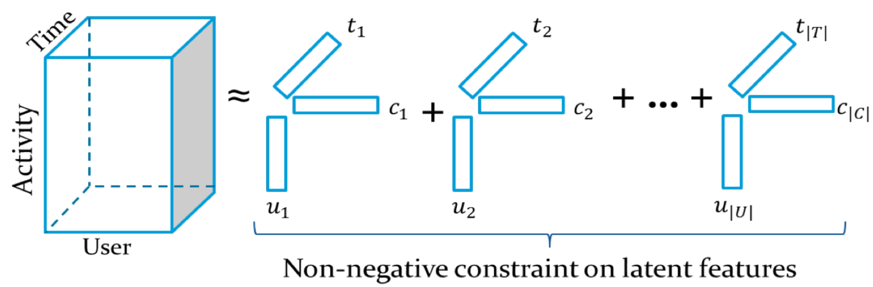
$$H_{max}(|C_l^d|) = \log_2 |C_l^d| = 4.32$$

$$ratio_{PB} = 1 - \frac{H(\psi_{u,r})}{H_{max}(|C_l^d|)} = 0.59$$

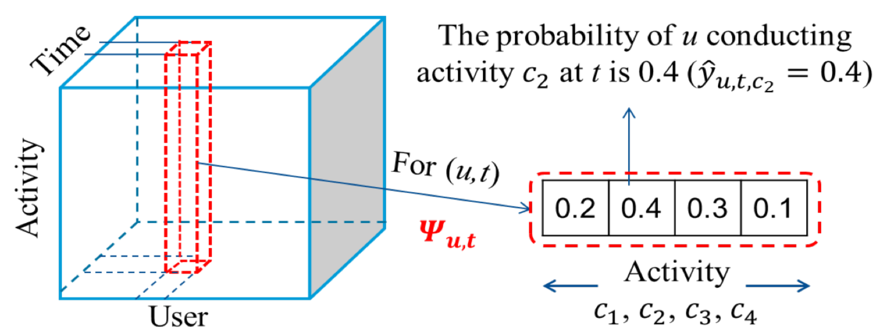
Modeling Temporal Correlation

Temporal Correlation – Tensor Factorization

- User-Time-Activity correlation captured by leveraging non-negative tensor factorization with Tucker model.

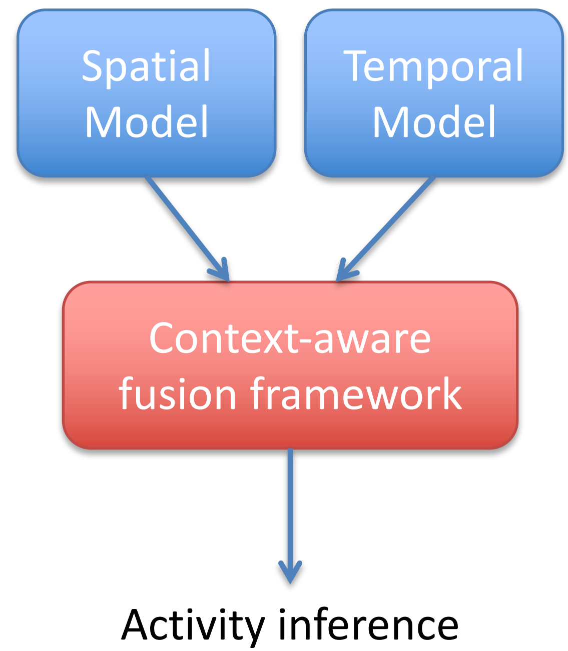


- Recovered tensor contains the probability that a user is interested in a specific activity at a given time.



Context-aware User Activity Inference

- **Context-aware fusion framework**
  - ***Learning*** : Based on a validation dataset, for each user, we evaluate the effectiveness of individual spatial and temporal models by calculating their accuracy over different contexts (i.e. location and time).
  - ***Inferring*** : Given a user's current context (i.e., location and time), it simply uses the model with higher accuracy to infer the user's current activity preference.

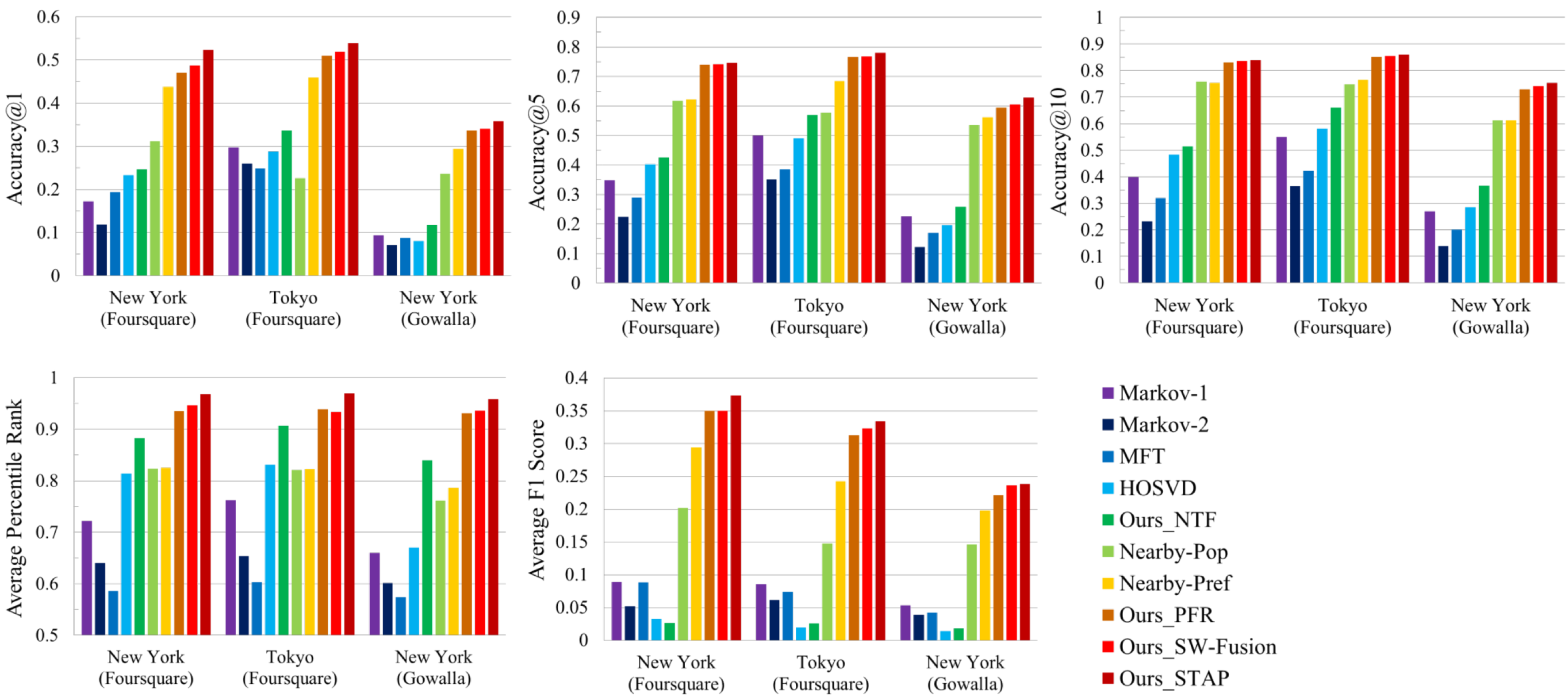


Evaluation

- **Dataset**
  - Foursquare Dataset.
    - Foursquare check-ins from 12 April 2012 to 16 February 2013 in New York and Tokyo
  - Gowalla Dataset
    - Gowalla check-ins from January 2010 to October 2010 in New York
  - 251 activity category
- **Metrics**
  - Top K inference accuracy (Accuracy@K), Average Percentile Rank (APR), Precision/Recall/F1score
- **Baselines**
  - Sequential pattern mining approaches
  - Temporal based approaches
  - Spatial based approaches
  - Static weighted spatial-temporal fusion approach

Activity Inference Performance

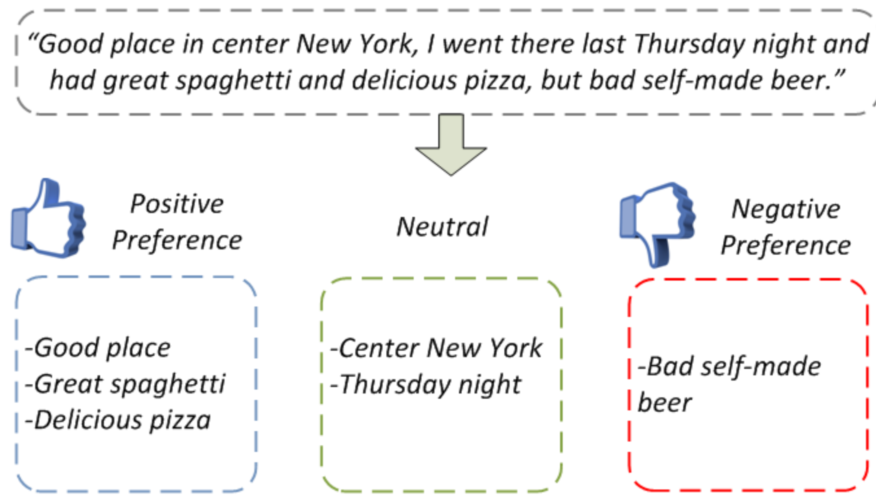
- The results show that the STAP model outperforms various baseline approaches in inferring user activity preference using different evaluation metrics (Accuracy@K, Average percentile Rank, F1 Score).



Understanding User Preference on POIs

- **User activities in LBSNs massively imply**
  - *Coarse-grained* user preference on POIs
  - *Fine-grained* user preference on POI-associated entities (e.g., specific dishes in a restaurant)

Check-ins – "Foot voting"

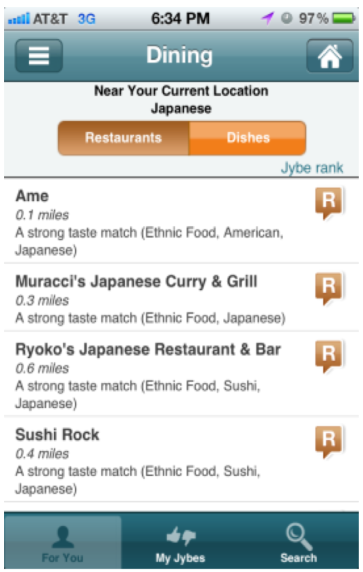


Tips – User reviews

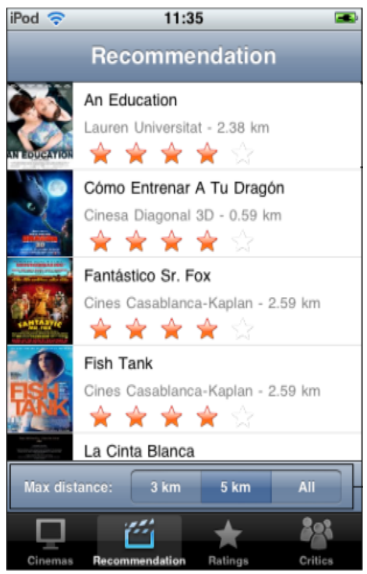
Preference-aware LBS

Use case I: Personalized POI recommendation

– Suggesting user interesting places to visit



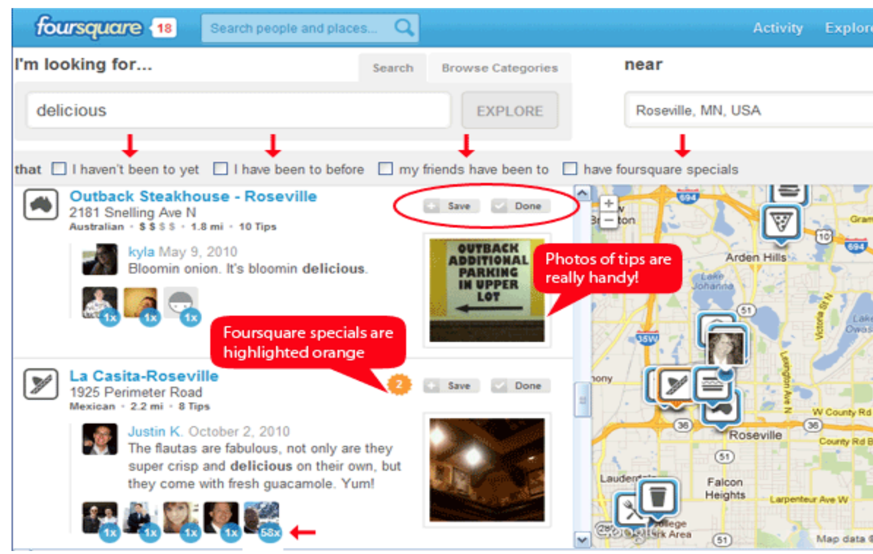
Restaurant recommendation



Cinema recommendation

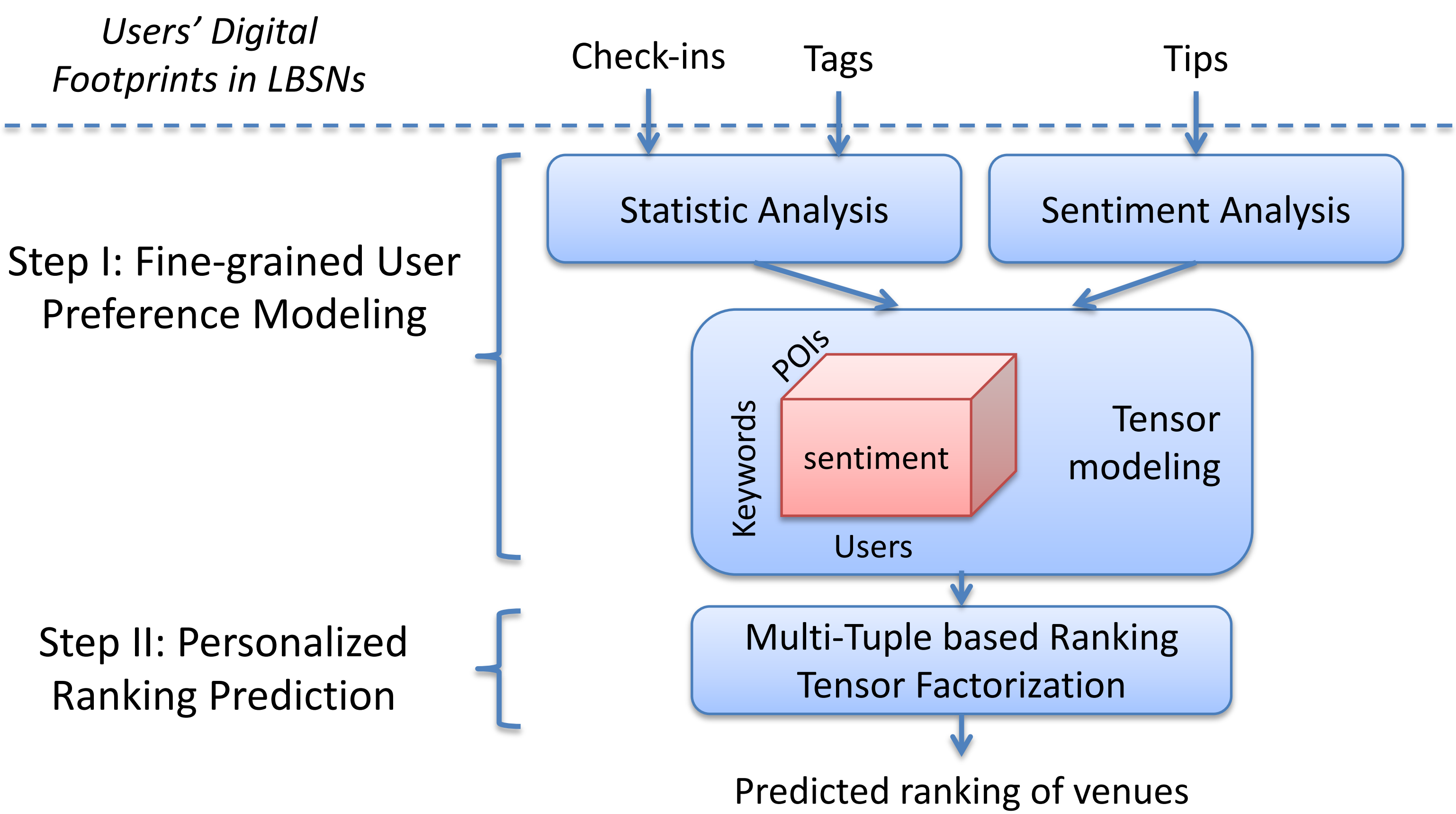
Use case II: Personalized POI search

– For given keywords, providing pertinent local search results according to user preference



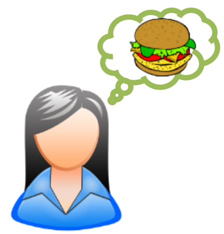
Use case: Personalized POI Search

- **Fine-grained Preference-Aware POI Search**



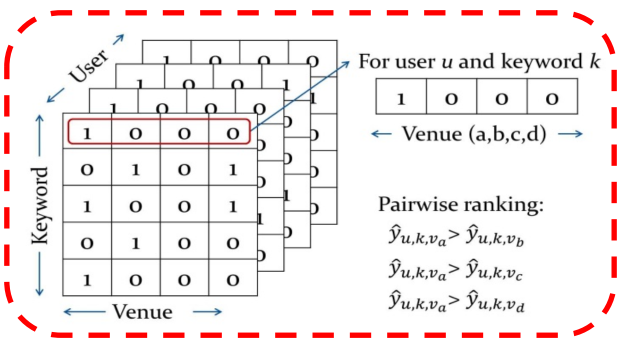
Use case: Personalized POI Search

- **Beyond state-of-the-art:**
  1. Fine-grained user preference modeling from heterogeneous user data

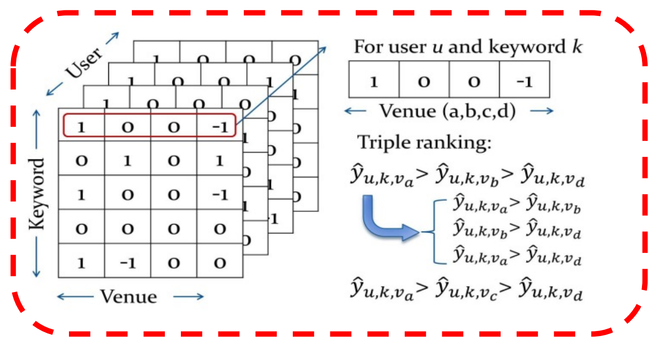


Jane likes burger restaurants

  2. Multi-Tuple Ranking Tensor Factorization:
     - Fine-grained user preference includes both positive and negative feedbacks
     - Bootstrap ranking tuple sampling and stochastic gradient descent optimization



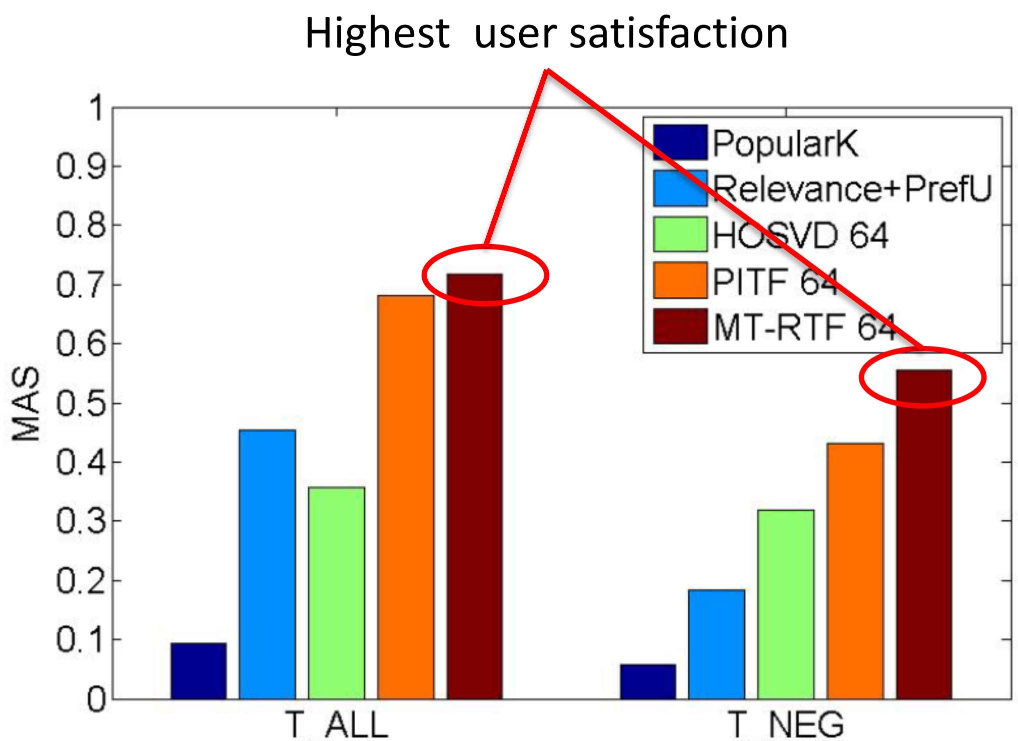
Pairwise ranking relation



Triple ranking relation

Use case II: Personalized POI Search

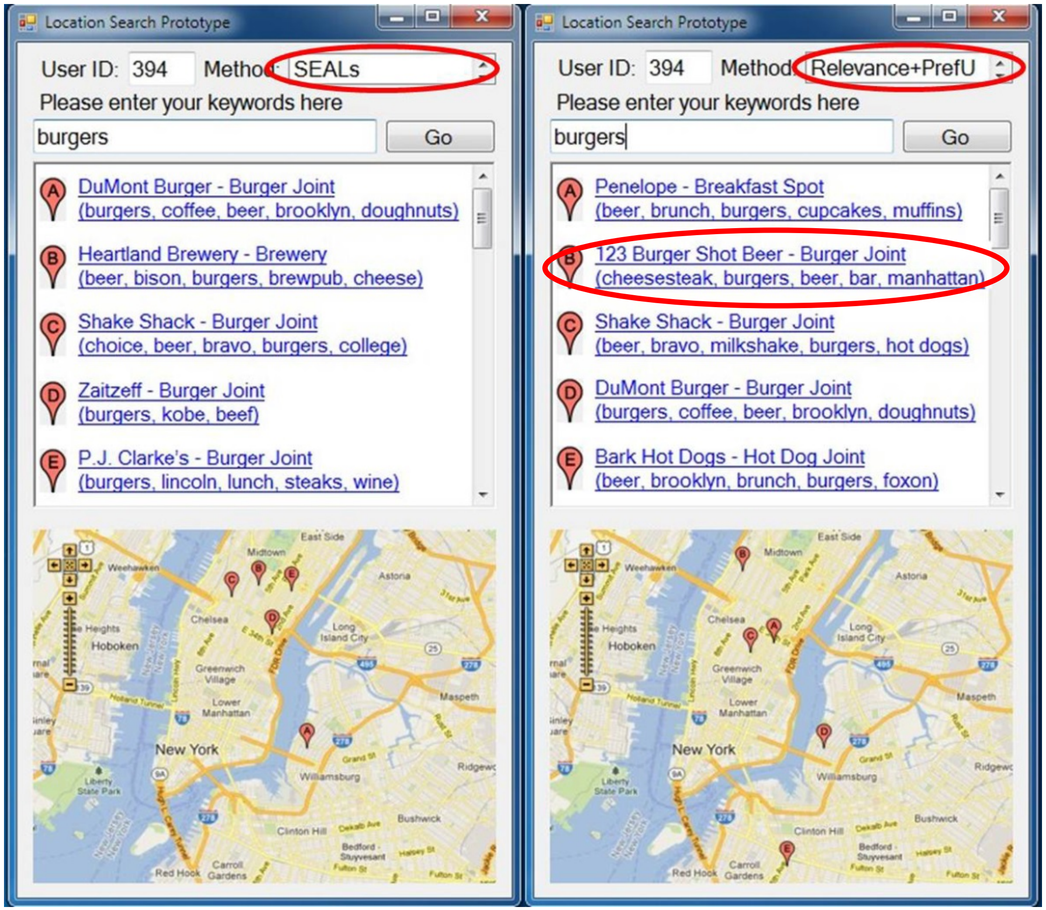
- Our approach outperforms other state-of-the-art preference-aware search methods



Use case: Personalized POI Search

SEALs Prototype & Case Study

Fine-grained preference *Coarse-grained preference*

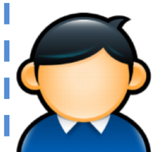


How fine-grained preference helps improve personalized location search ?

Case study:

- User 394
- “burgers” as keywords

In fact, this user’s fine-grained preference for “123 burger shot beer” is :



a) “pizza” ☺
b) “beer” ☺
c) “***burger***” ☹

Coarse-grained preference consider his overall preference for “123 burger shot beer” is positive

Application of Geo-Social Media Analytics

- **Personalized location based services**
  - Modeling spatial-temporal user activities
  - Understanding user preference on POIs
- **Computational Social Science**
  - Participatory cultural mapping

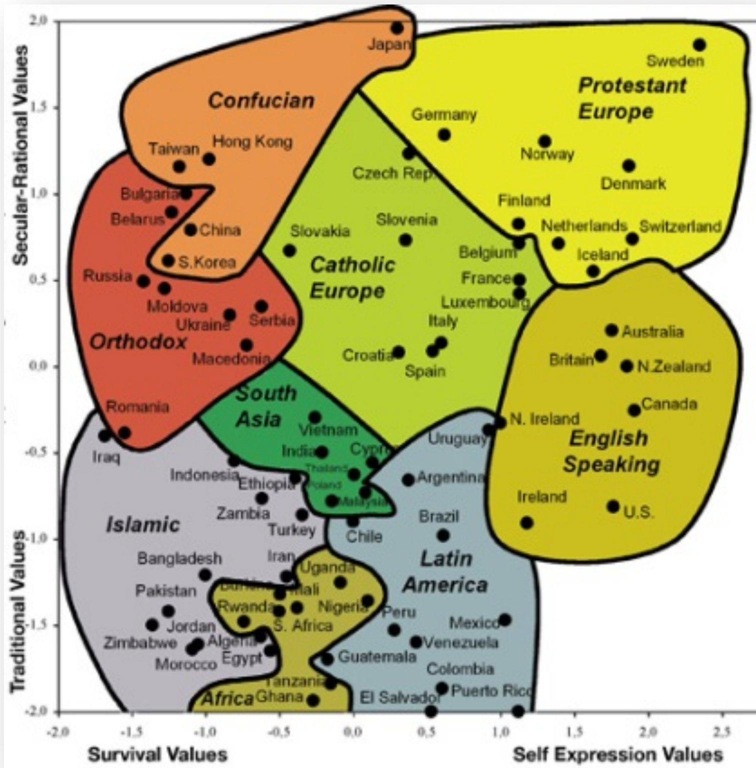
Global Cultures and Collective Behavior

"Culture is an integrated system of learned behavior patterns which are characteristic of the members of a society,…"

—E. Adamson Hoebel (1906–1993) , American anthropologist

- **Cultural Mapping**
  - Recognized by the United Nations Educational, Scientific and Cultural Organization (UNESCO) as a crucial tool and technique to visualize cultural differences on the map.
  - The goal is to create the map representation of different cultures and their boundaries, from the perspectives of ***indigenous and local people*** with respect to ***specific cultural aspects***.

Cultural Map Examples



Moral value cultural map
(World Value Survey)

World religion cultural map

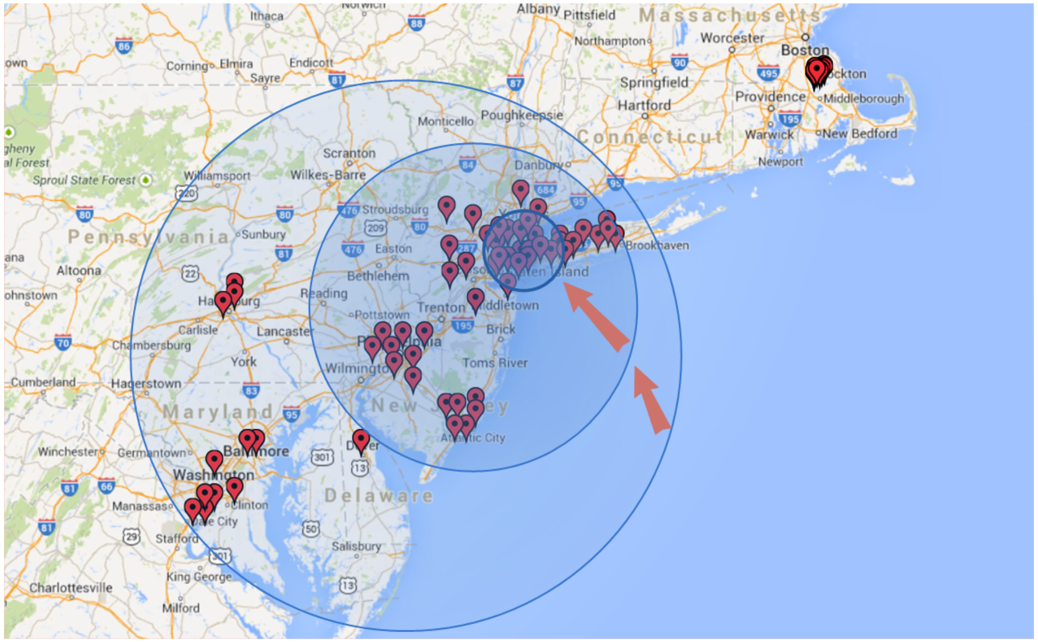
Cultural Mapping

- **Traditional Cultural Mapping**
  - Data collection via large-scale surveys
  - Significant cost of both human resources and time
  - Cultural Mapping from the psychological perspective (e.g., moral value)

- **Participatory Cultural Mapping: Beyond state-of-the-art**
  - Participatory sensed user behavior from LBSNs
  - Progressive local user identification
  - Cultural feature extraction (daily activity pattern, inter-city mobility, linguistic aspects)
  - Cultural clustering based on spectral clustering techniques

Local User Selection

- **Progressive “home” location identification**
  - “Home” **city** identification
  - Search iteratively the most-frequented circular regions with the decreasing radius



eXascale Infolab

Cultural Feature Extraction

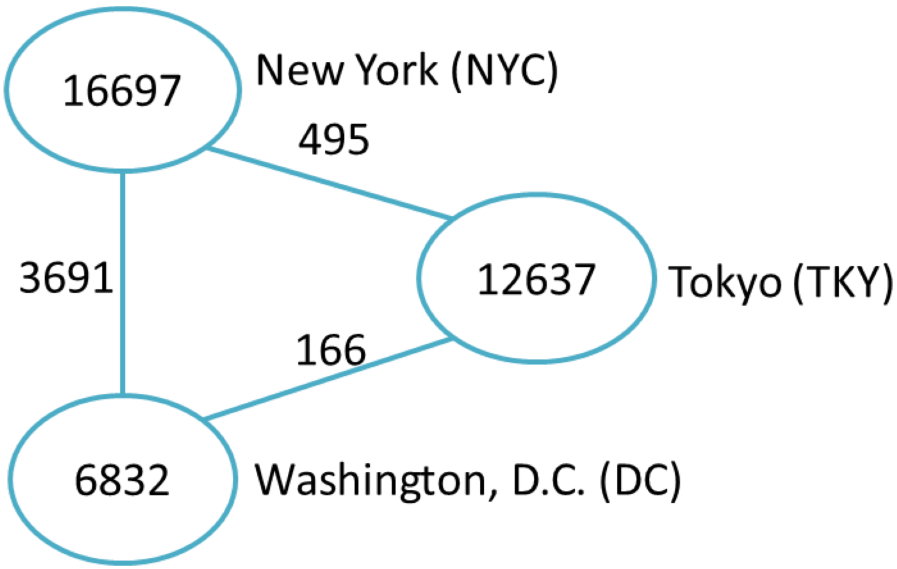
- **Daily Activity Pattern**
  - Motivation: Common culture implies similar lifestyle (in LBSNs)
  - Modeling: Categorical distribution of collective activities

New York

Tokyo

Cultural Feature Extraction

- **Inter-city Mobility**
  - Motivation: Human as culture carrier
  - Modeling: Ratio of users who travel between cities



$$Sim_{Mob}(NYC, DC) = 0.3456$$

$$Sim_{Mob}(NYC, TKY) = 0.0341$$

$$Sim_{Mob}(DC, TKY) = 0.0179$$

Cultural Feature Extraction

- **Linguistic Characteristics**
  - Motivation: Language expresses, embodies and symbolizes human culture
  - Modeling: language distribution of user status

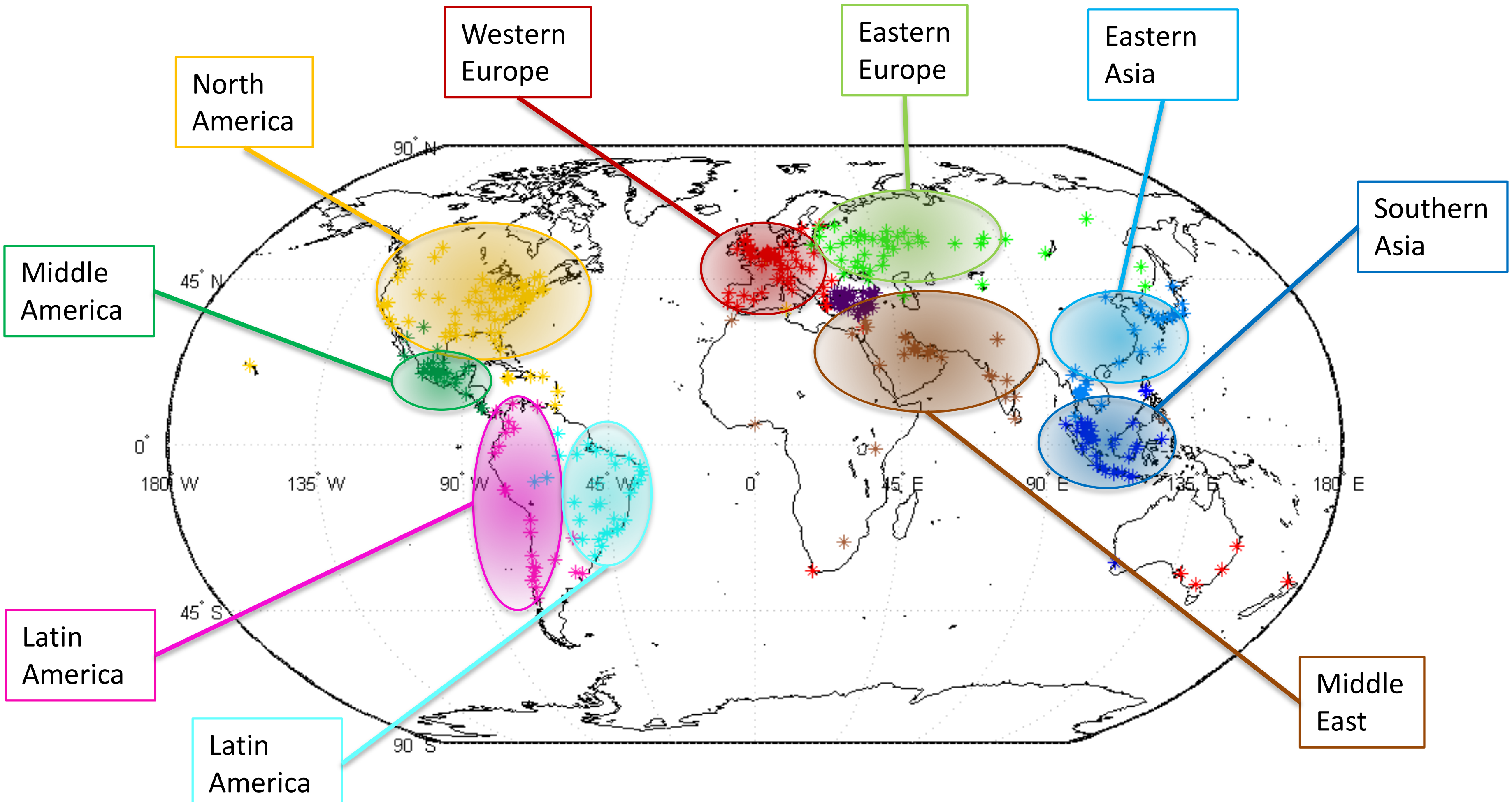
Mexico City

Rio de Janeiro

Evaluation

- **Dataset**
  - Global check-ins from April 2012 to September 2013
  - 266,909 users/33,263,233 check-ins/3,680,126 venues
  - 415 cities located in 77 countries

Created Cultural Map (Qualitative study)



Cultural Mapping (Quantitative study)

- **Comparison with traditional cultural mapping**
  - Two existing sets of cultural clusters (at the country level)
    - WVS: World Value Survey
    - GLOBE: Global Leadership and Organizational Behavior Effectiveness research)

Table 7.2: Cultural clusters and their city numbers.

Cultural Clusters of WVS [57]		Cultural Clusters of GLOBE [52]	
Cluster Name	Number of cities	Cluster Name	Number of cities
English Speaking	83	Anglo	85
Catholic Europe	17	Latin Europe	15
Protestant Europe	21	Nordic Europe	3
Orthodox	39	Germanic Europe	19
Latin America	75	Eastern Europe	38
Islamic	40	Arab	67
South Asia	57	Southern Asia	56
Confucian	22	Confucian Asia	23
		Middle East	42

Cultural Map (Quantitative study)

- **We calculate the purity of each cluster (C1－C8) w.r.t. the traditional cultural clusters.**

Table 7.4: Comparison of individual cultural clusters (WVS dataset).

	C1	C2	C3	C4	C5	C6	C7	C8
English Speaking	0.02	0	**0.96**	0.03	0	0	0.21	0.06
Catholic Europe	0	0	0	0	0	0	0.28	0
Protestant Europe	0	0	0	0	0	0.03	**0.34**	0
Orthodox	0	0	0	0	0	**0.97**	0.03	0
Latin America	0	**1**	0.01	**0.97**	0	0	0	0
Islamic	0	0	0	0	**0.92**	0	0.08	0
South Asia	0.41	0	0.03	0	0.08	0	0.06	**0.94**
Confucian	**0.57**	0	0	0	0	0	0	0

Two Latin America clusters due to the linguistic features

Single clusters for Europe

Concluding remarks

- **Geo-social media**
  - A novel dimension: location
- **Characteristics of geo-social media data**
- **Application of geo-social media analytics**
  - Personalized location based services
  - Computational social science