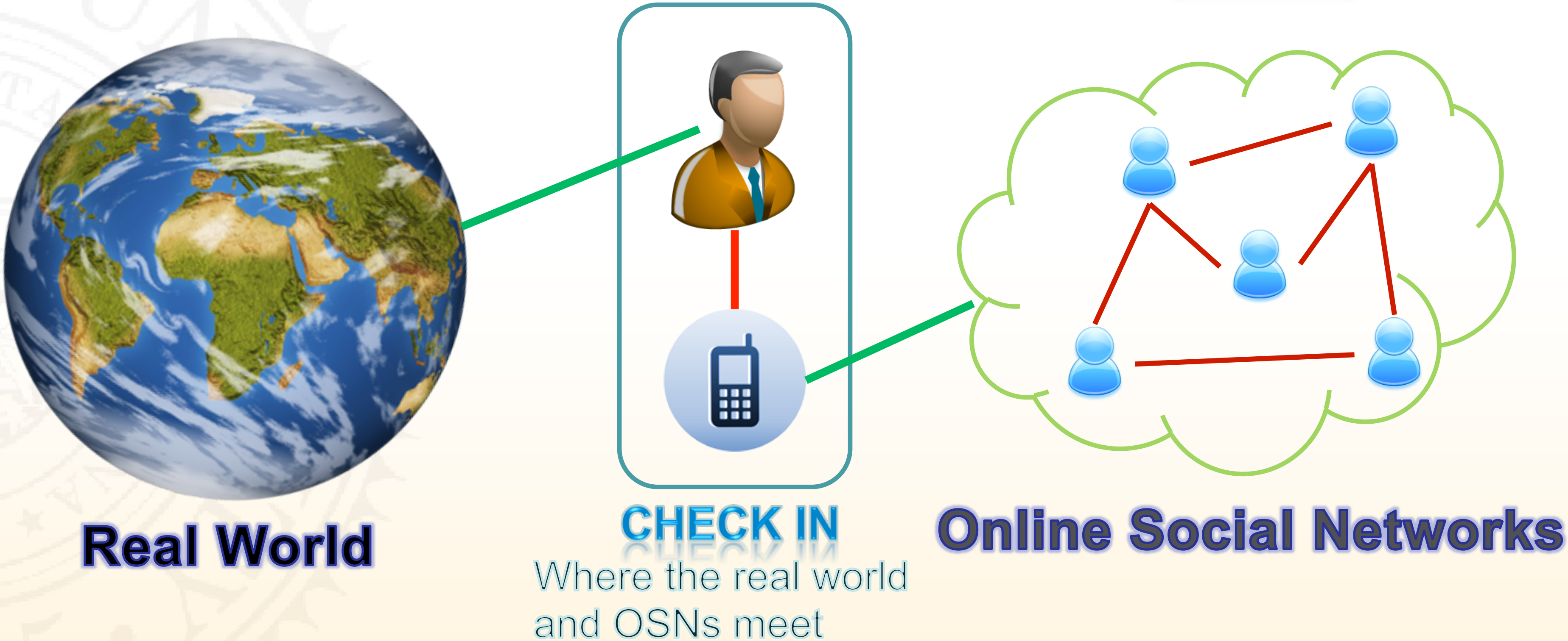
Toward Mobile Cloud Computing: Data Analysis with Location-Based Social Network

Huan Liu

Joint Work with Huiji Gao and Jiliang Tang

Location-Based Social Networks (LBSNs)

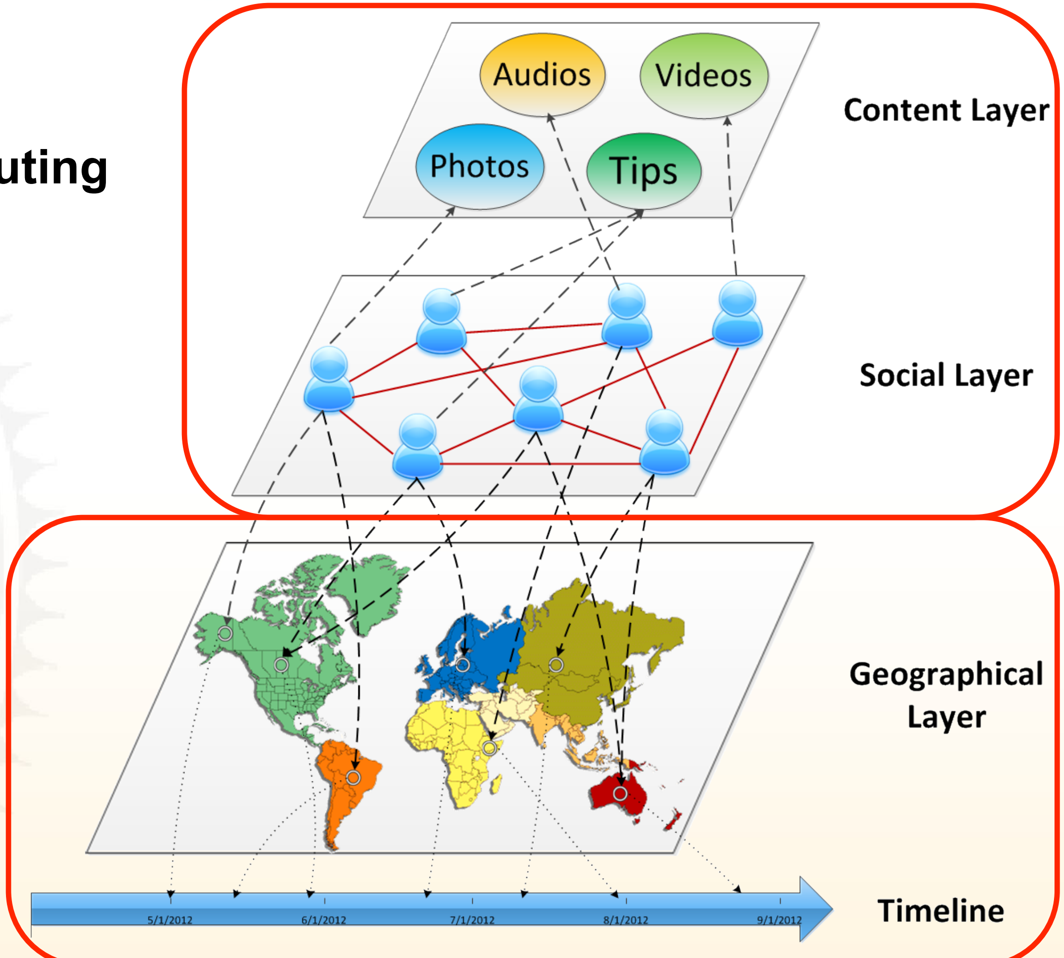
- Location-Based Social Networking Sites
  Foursquare, Facebook Places, Yelp



A Location-Based Social Network Framework

Social Computing

Traditional Mobile Computing



DMML

Arizona State University

Essential Data from LBSN

- Check-in history with time stamps
- Social networks derived from check-in locations
- User generated contents
- Interdependency of social networks and locations

Distinct Properties of LBSN Data

- Large-Scale Mobile Data
- Accurate Location Descriptions
- Explicit Social Friendships
- Significant Sparsity of Data

Research Opportunities

- Study a user's mobile behavior through both **real** and **virtual worlds** in spatial, temporal and social dimensions.
- Understand the role of **social networks** and **geographical properties** with large amounts of heterogeneous data
- Improve the development of **location-based services** such as mobile marketing, disaster relief, traffic forecasting, and etc.
- Mobile cloud computing

Some Challenges

- How to study human mobile behavior from high dimensional data from heterogeneous sources
- How to deduce human movement through sparse check-in data
- How to design location-based services to improve user's experience without sacrificing one's privacy

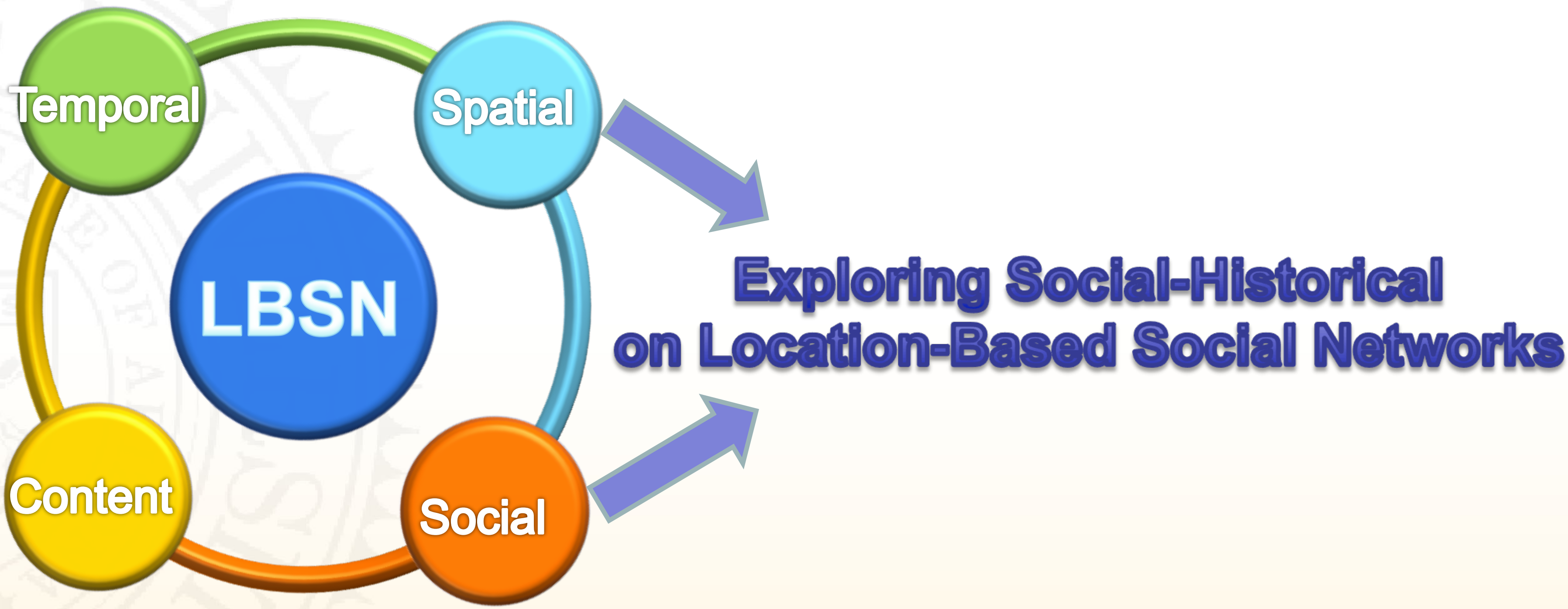
Potential Applications

- Disaster Relief/Crisis Response
- Mobile Search/Recommendation
- Location Prediction
- Recommendation Systems
- Mobile Community Detection
- Location Privacy Protection
- Mobile Marketing

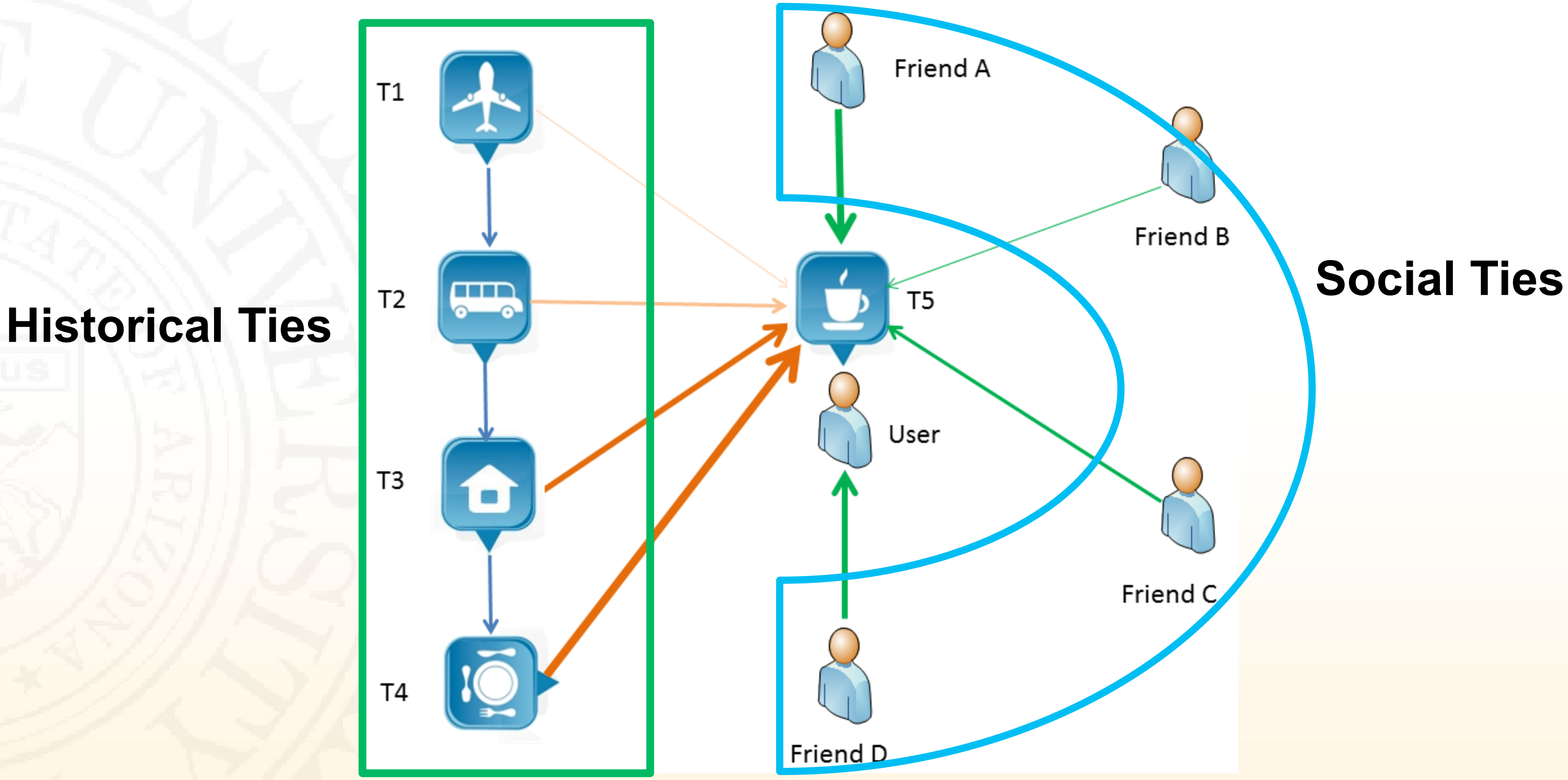
Some of Our Recent Findings

- Social-Historical Ties on Location-Based Social Networks (ICWSM'2012)
  - Are two types of ties equally important?
- Geo-Social Correlation (CIKM'2012)
  - Handling the Cold Start Problem
- Mobile Location Prediction in Spatio-Temporal Context in Next Location Prediction in 2012 *Nokia Mobile Data Challenge Workshop*, 3rd Prize
  - Together is better

Exploring Social-Historical on Location-Based Social Networks

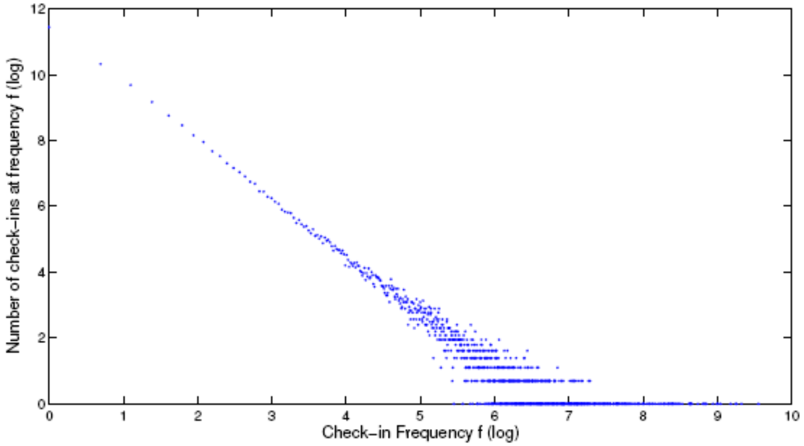


Social-Historical Effect of Online Check-ins



Why is the prediction hard

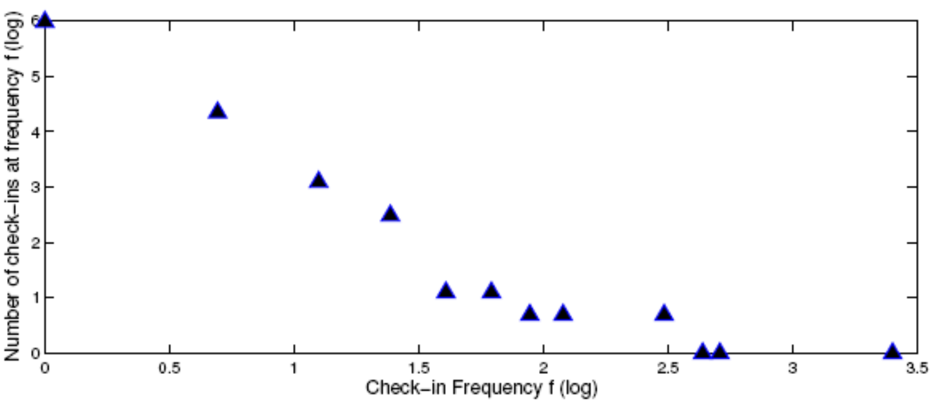
- **Power-law distribution**



(a) Power-law distribution of check-ins in whole dataset

Whole Dataset

Individual



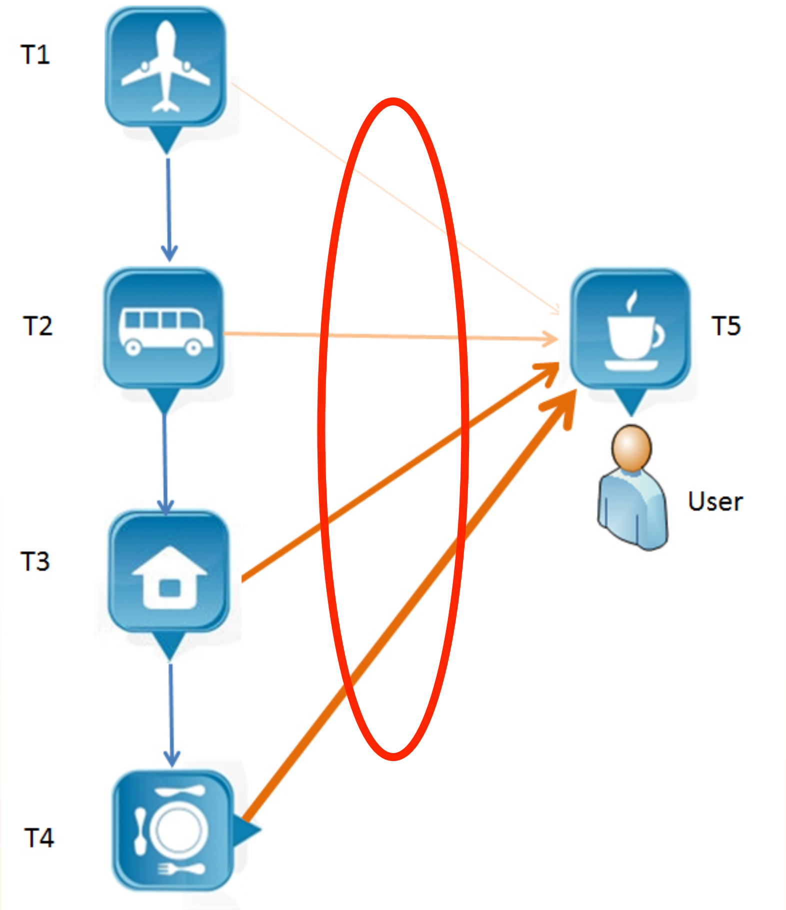
(b) Power-law distribution of individual check-ins

DMML

Arizona State University

Analyzing User's Historical Ties

- **Short Term Effect**
  - The historical tie strength decreases over time.
  - The historical ties of the previous check-ins at airport, shuttle stop, hotel and restaurant have different strengths to the latest check-in of drinking coffee.



Modeling User's Historical Ties

- **Correspondences between language and LBSN modeling**

Language Modeling		LBSN Modeling	
Corpus		Check-in collection	
Document		Individual check-ins	
Document Structure	Paragraph	Check-in Structure	Monthly check-in sequence
	Sentence		Weekly check-in sequence
	Phrase		Daily check-in sequence
	Word		Check-in location

- **Power-law distribution**
- **Short Term Effect**

HPY (Hierarchical Pitman-Yor) Language Model

Modeling User's Social Ties

❖ **Social Ties**

➢ **Common Check-ins**

	Common check-ins
between friends	11.8306
between strangers	4.3226

➢ **Check-in Similarities**

Users with friendship have higher check-in similarity than those without. Null hypothesis $H_0: S_F \le S_R$, rejected at significant level α = 0.001 with p-value of 2.6e-6.

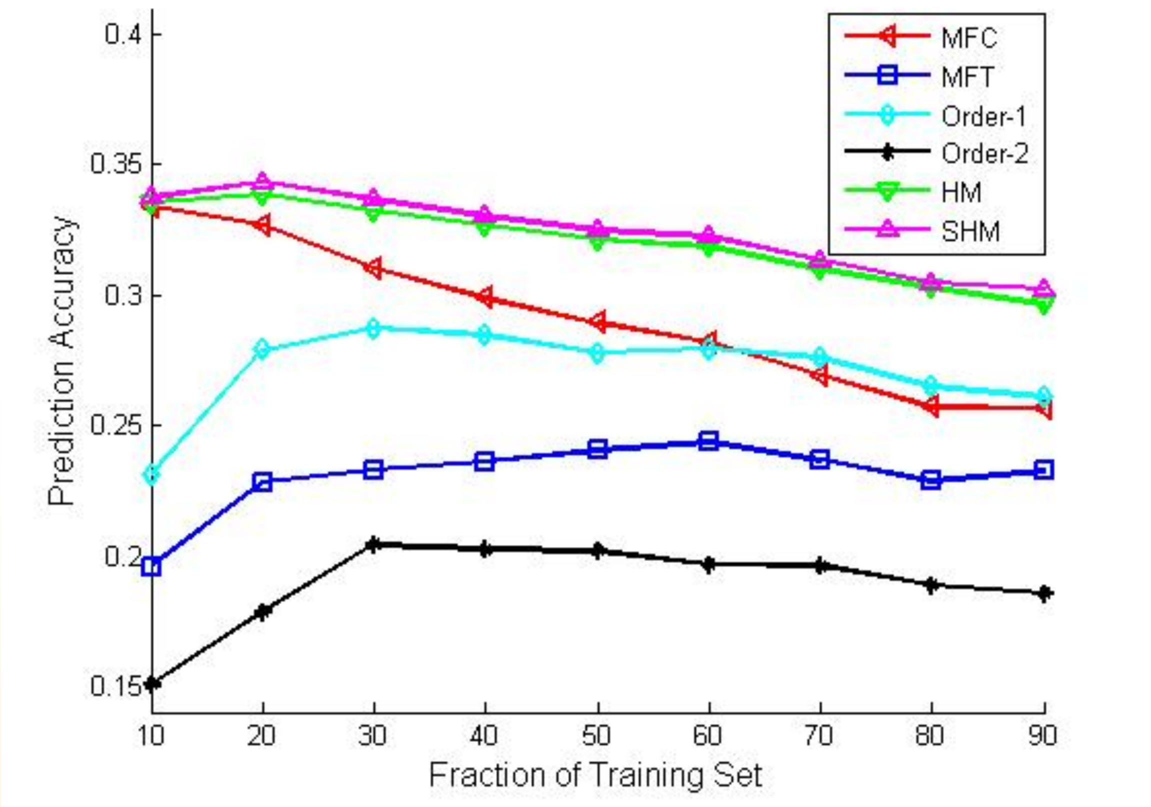
- **Friend Similarity**
- **Friends' Check-in Sequence**
- **HPY**

$\left.\begin{array}{l}\text{Friend Similarity}\\ \text{Friends' Check-in Sequence}\end{array}\right\}$ **Social Model**

$$p_{SH}^{i}(c_{n+1}=l)=\eta P_{H}^{i}(c_{n+1}=l)+(1-\eta)P_{S}^{i}(c_{n+1}=l)$$

Experiment Results for Location Prediction

- Experiment Results
  - **MFC**
  Most Frequent Check-in Model
  - **MFT**
  Most Frequent Time Model
  - **Order-1**
  Order-1 Markov Model
  - **Order-2**
  Order-2 Markov Model
  - **HM**
  Historical Model
  - **SHM**
  Social-Historical Model

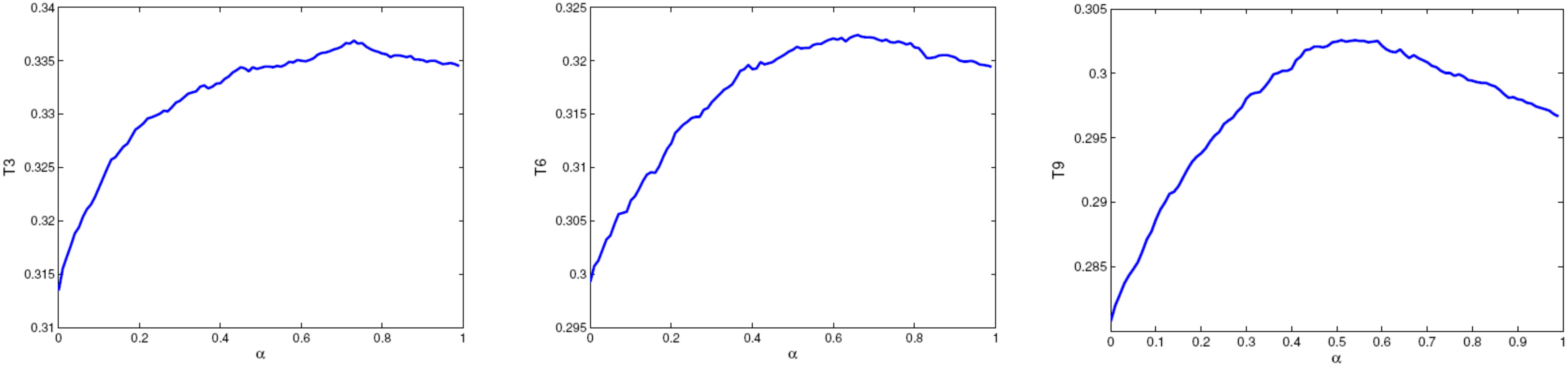


DMML

ARIZONA STATE UNIVERSITY

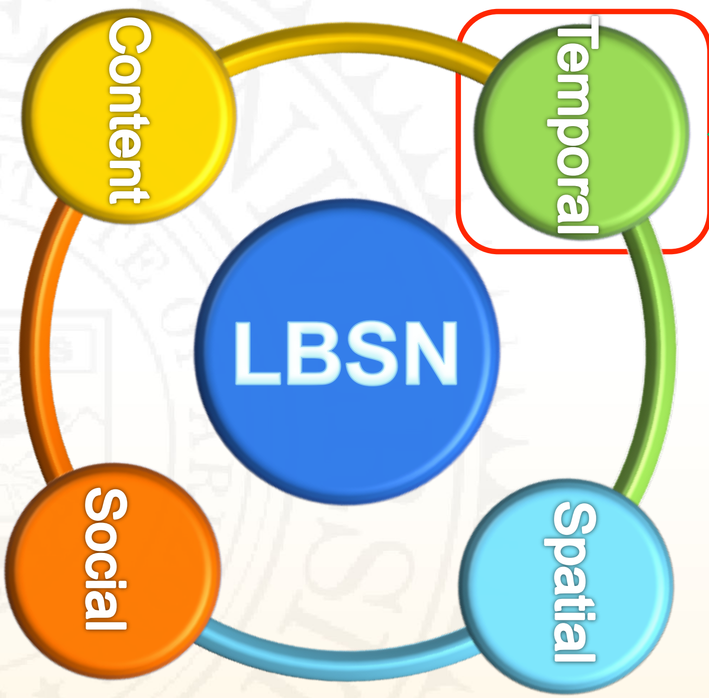
Social-historical Tie Effect w.r.t. η

- When no historical information is considered, the prediction performs worst, suggesting that considering social information only is not enough to capture the check-in behavior.
- By gradually adding the historical information, the performance shows the following pattern: first increasing, reaching its peak value and then decreasing. Most of the time, the best performance is achieved at around $\eta = 0.7$. A big weight is given to historical ties, indicating that historical ties are more important than social ties.

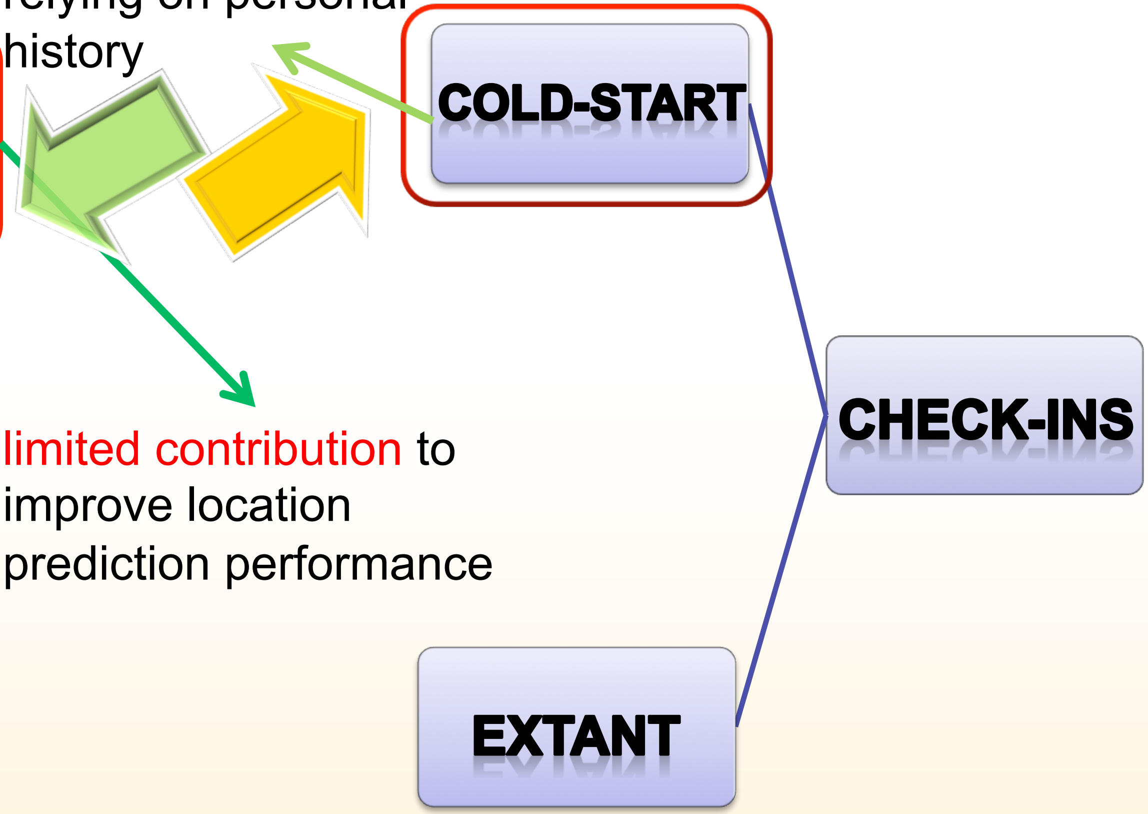


(a) T3 (b) T6 (c) T9

Predicting New Check-Ins

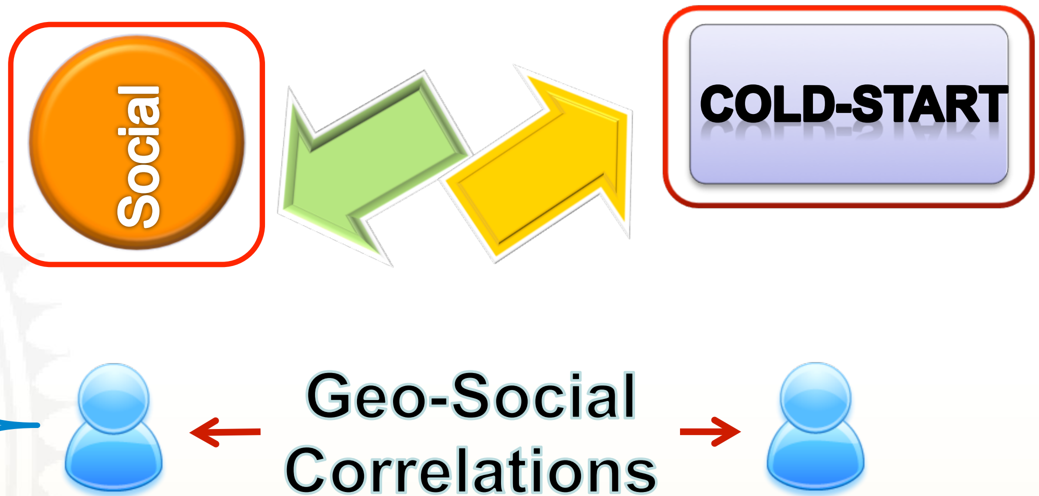


Impossible to predict relying on personal history



limited contribution to improve location prediction performance

Motivation

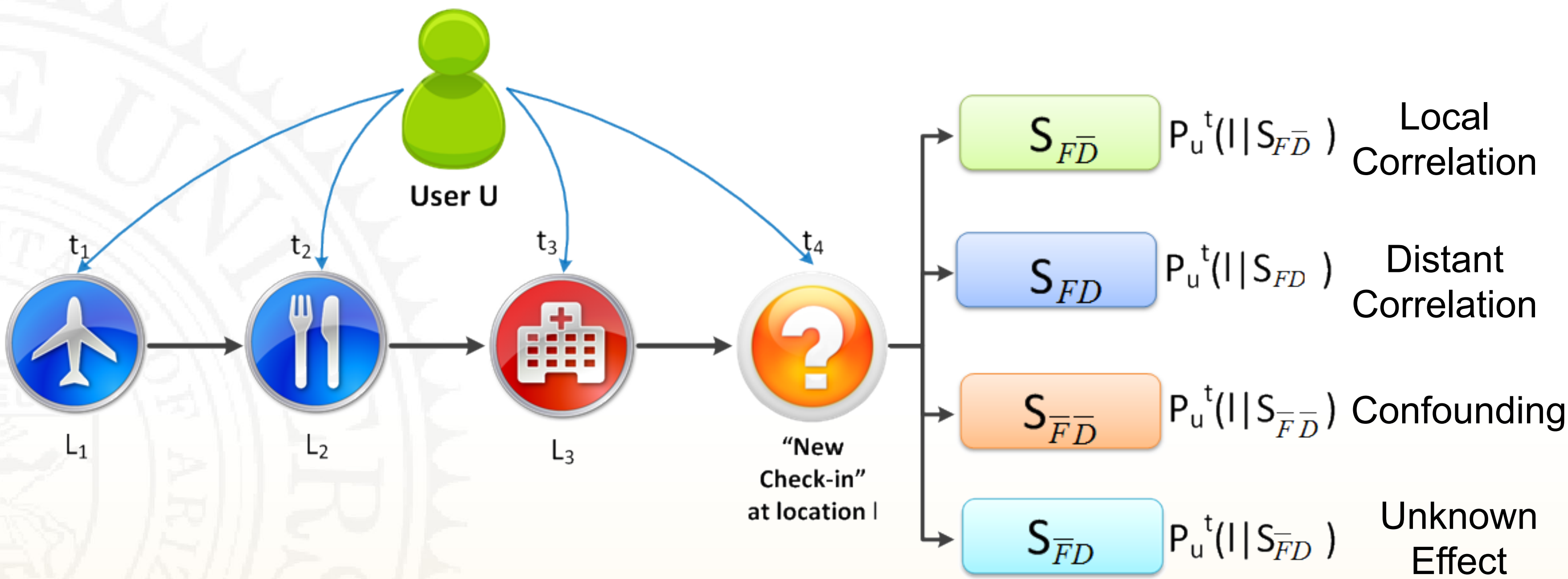


Geographical
Distance

F

	: Local Friends	: Local Non-friends
D	: Distant Friends	: Distant Non-friends

Geo-Social Correlations



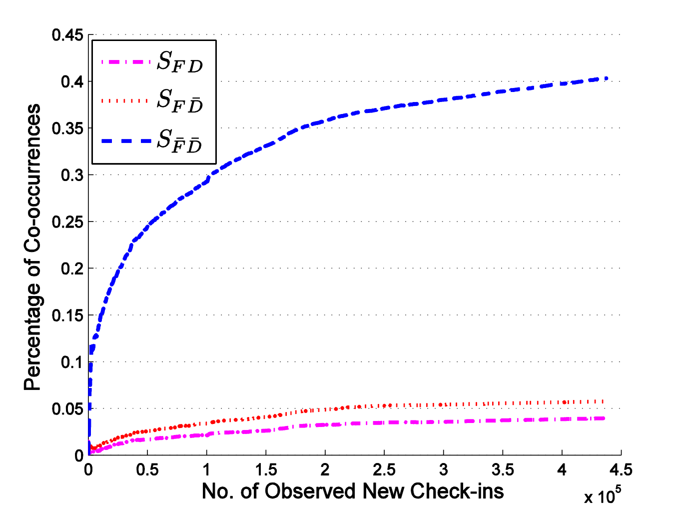
$$P_u^t(l) = \Phi_1 P_u^t(l|S_{\bar{F}\bar{D}}) + \Phi_2 P_u^t(l|S_{F\bar{D}})$$
$$+ \Phi_3 P_u^t(l|S_{FD}) + \Phi_4 P_u^t(l|S_{\bar{F}D}).$$

Modeling Geo-Social Correlations

- $P_u^t(l)$: the probability of a user *u* checking-in at a new location *l* at time *t*

$$P_u^t(l) = \Phi_1 P_u^t(l|S_{\bar{F}\bar{D}}) + \Phi_2 P_u^t(l|S_{F\bar{D}}) + \Phi_3 P_u^t(l|S_{FD}) + \Phi_4 P_u^t(l|S_{\bar{F}D}).$$

Geo-Social Correlation Strength



$$\Phi_1 = f(\mathbf{w}^T \mathbf{f}_u^t + b), \; 0 \leq \Phi_1 \leq 1$$

$$\Phi_2 = (1 - \Phi_1)\phi_1$$

$$\Phi_3 = (1 - \Phi_1)(1 - \phi_1)\phi_2$$

$$\Phi_4 = (1 - \Phi_1)(1 - \phi_1)(1 - \phi_2)$$

Modeling Geo-Social Correlations

- $P_u^t(l)$: the probability of a user *u* checking-in at a new location *l* at time *t*

$$P_u^t(l) = \Phi_1 \boxed{P_u^t(l|S_{\bar{F}\bar{D}})} + \Phi_2 \boxed{P_u^t(l|S_{F\bar{D}})}$$
$$+ \Phi_3 \boxed{P_u^t(l|S_{FD})} + \Phi_4 \boxed{P_u^t(l|S_{\bar{F}D})}.$$

Geo-Social Correlation Probability

- **Geo-Social Correlation Probability Measures:**

1. Sim-Location Frequency (S.Lf)

$$P_u^t(l|\mathcal{S}_x) = \frac{\sum_{v \in \mathcal{S}_x} s(u,v) N_v^t(l)}{\sum_{v \in \mathcal{S}_x} s(u,v) N_v^t}$$

2. Sim-User Frequency (S.Uf)

$$P_u^t(l|\mathcal{S}_x) = \frac{\sum_{v \in \mathcal{S}_x} \delta_v^t(l) s(u,v)}{\sum_{v \in \mathcal{S}_x} s(u,v)}$$

3. Sim-Location Frequency & User Frequency (S.Lf.Uf)

$$P_u^t(l|\mathcal{S}_x) = \frac{\sum_{v \in \mathcal{S}_x} s(u,v) N_v^t(l)}{\sum_{v \in \mathcal{S}_x} s(u,v) N_v^t} \frac{\sum_{v \in \mathcal{S}_x} \delta_v^t(l)}{N_{\mathcal{S}_x}}$$

DMML

ARIZONA STATE UNIVERSITY

Dataset

- Foursquare Dataset

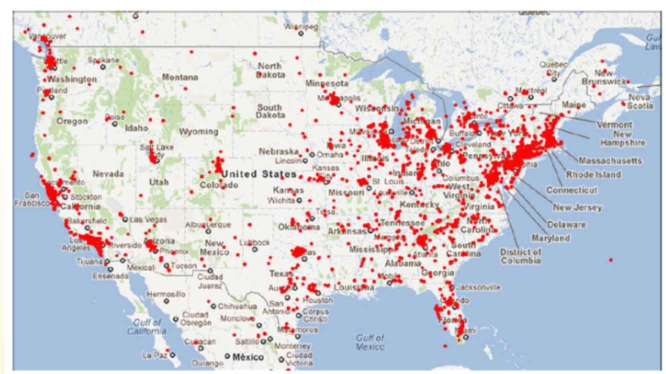
Table 2: Statistical information of the dataset

Duration	Jan 1, 2011-July 31, 2011
No. of user	11,326
No. of check-ins	1,385,223
No. of unique locations	182,968
No. of links	47,164

Table 3: Statistical information of the July data

Social Circle	No. of SCCs	Ratio
$S_{\bar{F}\bar{D}}$	34,523	44.50%
$S_{F\bar{D}}$	5,636	7.26%
S_{FD}	3,588	4.62%
$S_{\bar{F}D}$	39,423	50.82%
Others	1,672	2.2%
$S_{\bar{F}\bar{D}} \cup S_{F\bar{D}}$	35,277	45.47%
$S_{\bar{F}\bar{D}} \cup S_{FD}$	35,784	46.12%
$S_{F\bar{D}} \cup S_{FD}$	8,235	10.61%
$S_{\bar{F}\bar{D}} \cup S_{F\bar{D}} \cup S_{FD}$	36,486	47.03%

(a) The user distribution over the world.



(b) The user distribution over USA.

Experiments

- **Location Prediction Evaluation Metrics**

	Single Measure	Various Measures
Equal Strength	EsSm	EsVm
Random Strength	RsSm	RsVm
Various Strength	VsSm	gSCorr

- **Effect of Geo-Social Correlation Strength and Probability Measures**

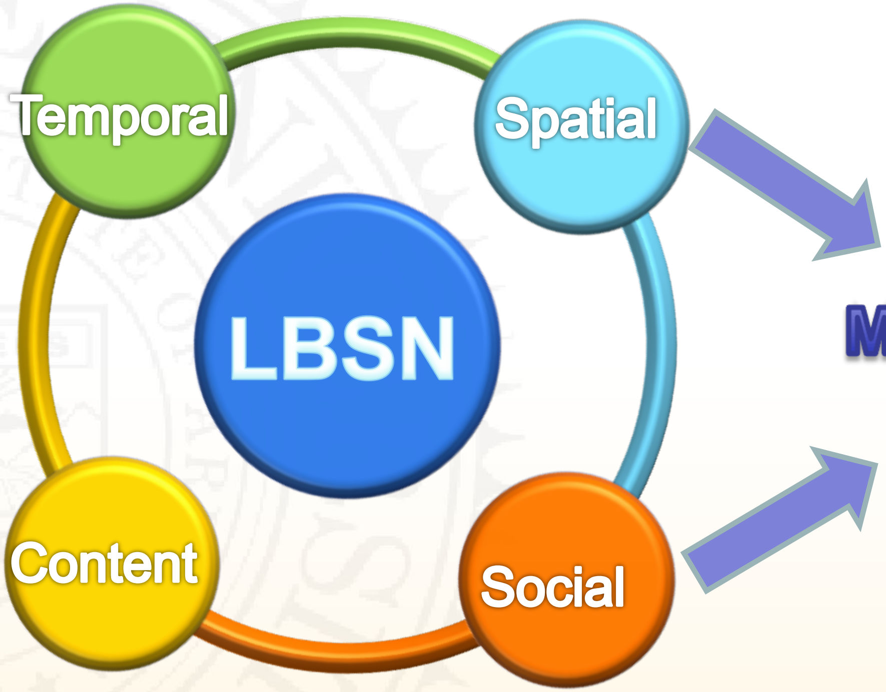
Methods	**Top-1**	**Top-2**	**Top-3**
EsVm	17.88%	24.06%	27.86%
EsSm	16.20%	21.92%	25.43%
VsSm	16.49%	22.28%	25.92%
RsSm	14.93%	20.30%	23.70%
RsVm	15.23%	20.85%	24.50%
gSCorr	**19.21%**	**25.19%**	**28.69%**

Experiments

➢ **Effect of Different Geo-Social Circles**

Methods	Top-1	Top-2	Top-3
	6.51%	8.31%	9.32%
	3.65%	4.75%	5.34%
	18.37%	24.10%	27.34%
	18.62%	24.44%	27.79%
	19.01%	24.95%	28.35%
	8.33%	10.79%	12.23%
	19.21%	**25.19%**	**28.69%**

Mobile Location Prediction in Spatio-Temporal Context



Mobile Location Prediction in Spatio-Temporal Context

DMML

Arizona State University

Problem Statement

The probability of checking in at location *l* given the check-in time at t and latest check-in

$$p(v_i = l \mid t_i = t, v_{i-1} = l_k)$$

$$= \boxed{p(t_i = t \mid v_i = l)}\,\boxed{p(v_i = l \mid v_{i-1} = l_k)}$$

Temporal Constraint

The probability of the i-th visit happening at time t, observing that the i-th visit location is l.

Spatial Prior

The probability of next visit at location l given the current visit at l_k

Historical Model

Temporal Constraint

Temporal Constraint:

$$p(t_i = t \mid v_i = l)$$

$$= p(h_i = h, d_i = d \mid v_i = l)$$

$$= p(h_i = h \mid v_i = l)\, p(d_i = d \mid v_i = l)$$

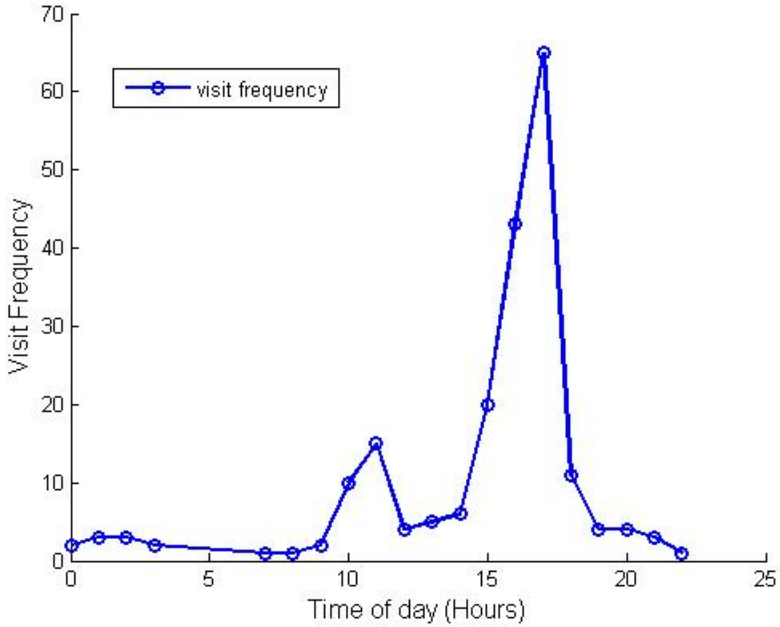
Hourly Constraint **Daily Constraint**

h: Hour of the day, i.e., 10:00am, 3:00pm
d: Day of the week, i.e., Monday, Sunday

Temporal Constraint

Compute $p(h_i = h \mid v_i = l)$ and $p(d_i = d \mid v_i = l)$

- Distribution of a user's visits at a specific location in 24 hours. (user id: 013; place id: 3)



$$p(h_i = h \mid v_i = l) = N_l(h \mid \mu_h, \sigma_h^2)$$

$$p(H \mid v_i = l) = \prod_{i=1}^{N_l} N_l(h_i \mid \mu_h, \sigma_h^2)$$

$$(h_i \in H, |H| = N_l)$$

Maximizing Likelihood $\Rightarrow \begin{cases} \mu_h \\ \sigma_h^2 \end{cases}$

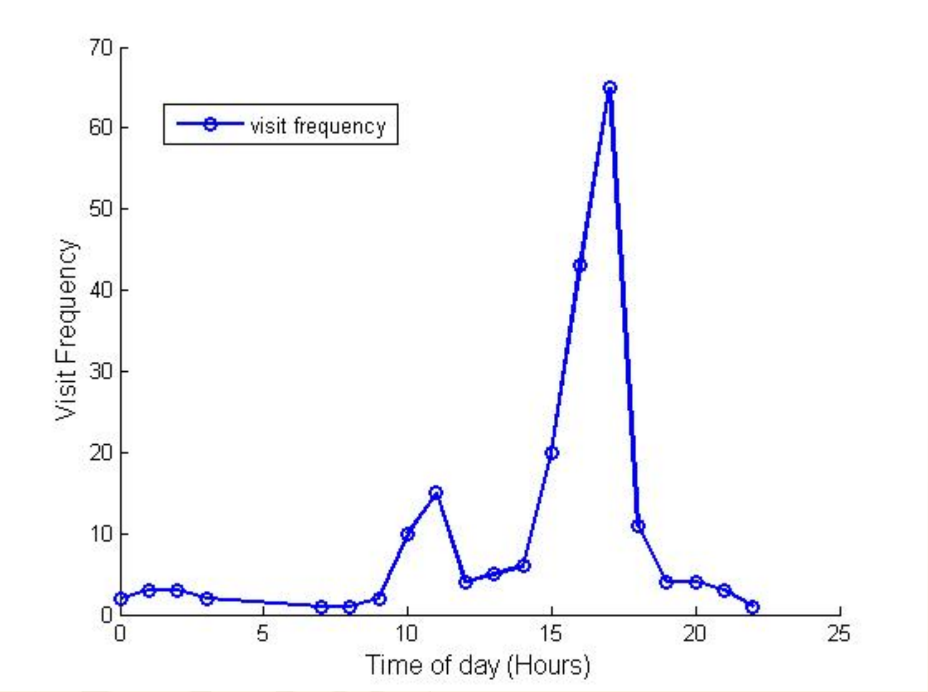
DMML

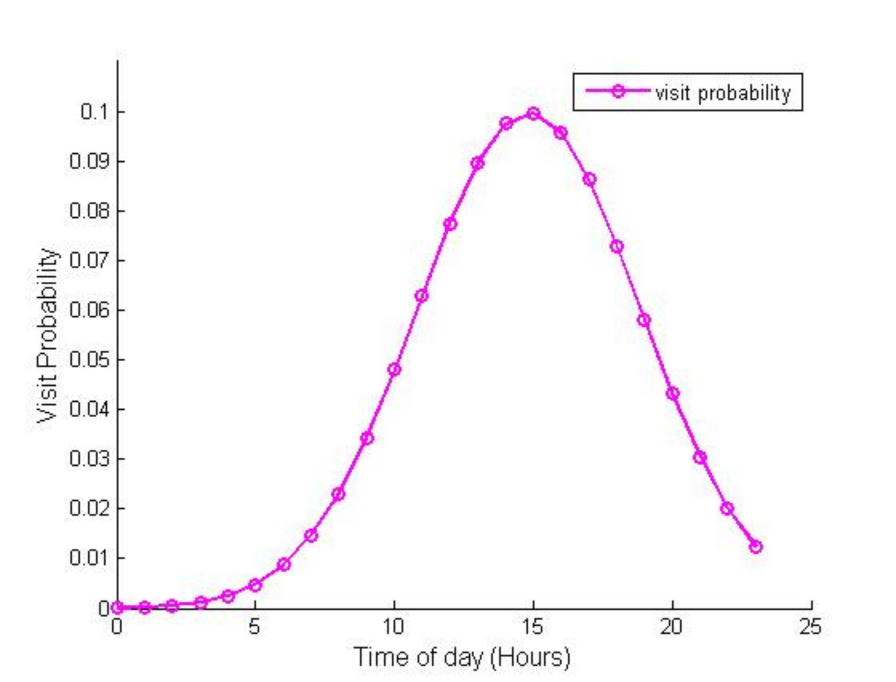
ARIZONA STATE UNIVERSITY

Temporal Constraint

Curve Fitting:

[user id: 013; place id: 3]





Location Prediction

Probability of visiting location l at time t with the latest visit at l_k

$$p(v_i = l \mid t_i = t, v_{i-1} = l_k)$$

$$= p(v_i = l \mid v_{i-1} = l_k)\, p(h_i = h \mid v_i = l)\, p(d_i = d \mid v_i = l)$$

$$= p(v_i = l \mid v_{i-1} = l_k)\, N_l(h \mid \mu_h, \sigma_h^2)\, N_l(d \mid \mu_d, \sigma_d^2)$$

HPY Prior Gaussian Gaussian

HPY Prior Hour-Day Model (HPHD)

Experiments – Together is Better

- Results

Table 1: Location Prediction Results

	Models	Correct No.	Accuracy
Spatial-based	MFV	1148	0.3402
	OMM	1466	0.4345
	FMM	1583	0.4692
	HP	1610	0.4772
Temporal-based	MFH	1462	0.4333
	MFD	1156	0.3426
	MFHD	1538	0.4558
Spatio-temporal	HPH	1680	0.4979
	HPD	1583	0.4692
	HPHD	**1705**	**0.5053**

Rank 3rd among 21 participated teams in Nokia Mobile Competition

Some of Our Recent Findings

- Social-Historical Ties on Location-Based Social Networks (ICWSM'2012)
  - Are two types of ties equally important?
- Geo-Social Correlation (CIKM'2012)
  - Handling the Cold Start Problem
- Mobile Location Prediction in Spatio-Temporal Context in Next Location Prediction in 2012 *Nokia Mobile Data Challenge Workshop*, 3rd Prize
  - Together is better

Acknowledgments: The projects are, in part, sponsored by ONR grants.

THANK YOU