PENNSTATE

1855

PGT: Measuring Mobility Relationship using Personal, Global and Temporal Factors

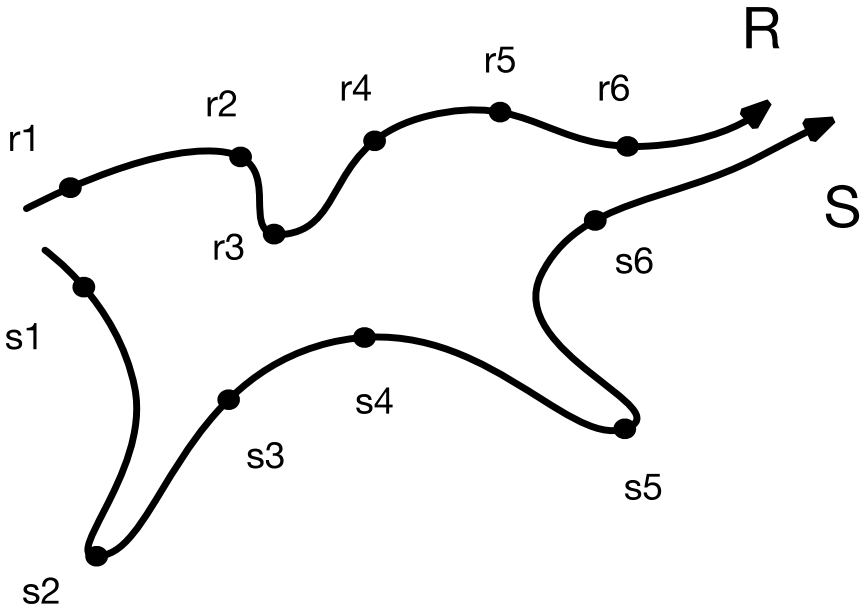
Hongjian Wang, Zhenhui Li, Wang-Chien Lee
Penn State University

ICDM 2014 Shenzhen

Measure the mobility relationship strength

- Given trajectories of two users, measure their relationship strength

ID	Location	Time-stamp
R	40.812, -77.856	2014-11-22 13:00:00
R	40.770, -77.855	2014-11-22 13:30:40
R	40.774, -73.975	2014-12-27 10:00:00
…	…	…



- Application
  - Recommendation
  - Crime investigation

Baseline Method -- Meeting Frequency

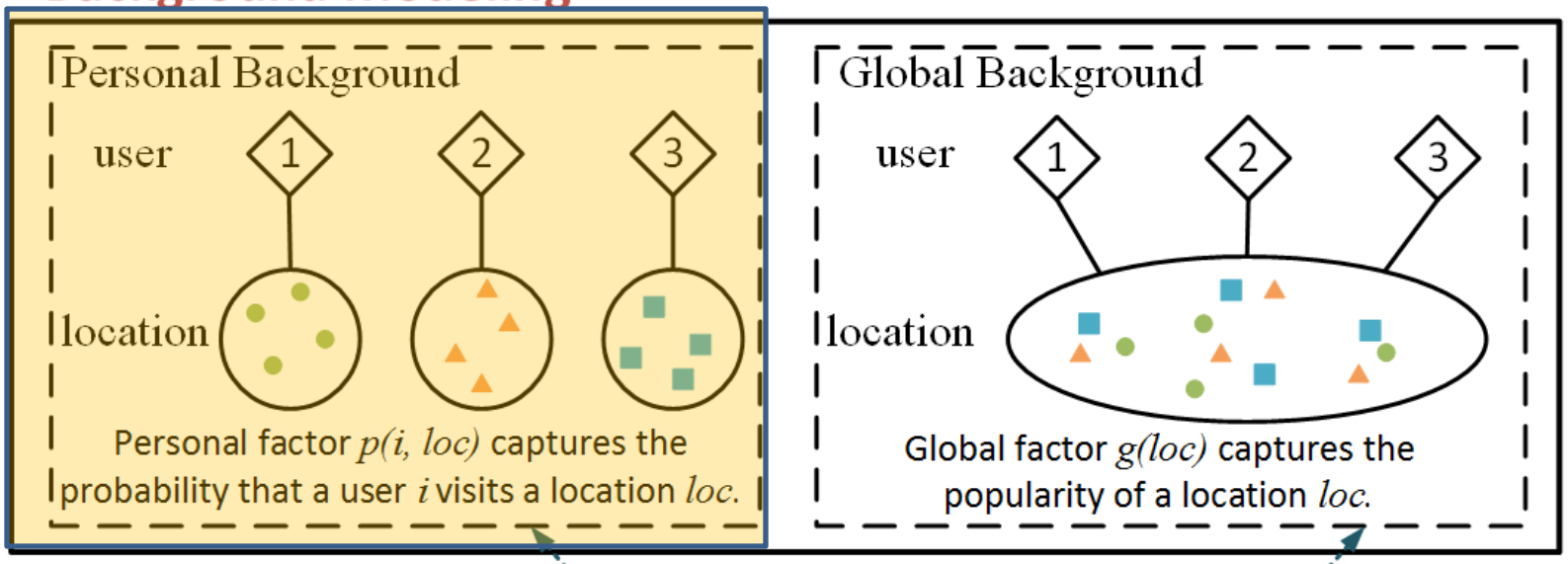
the more frequently you co-locate with another person,

less frequently

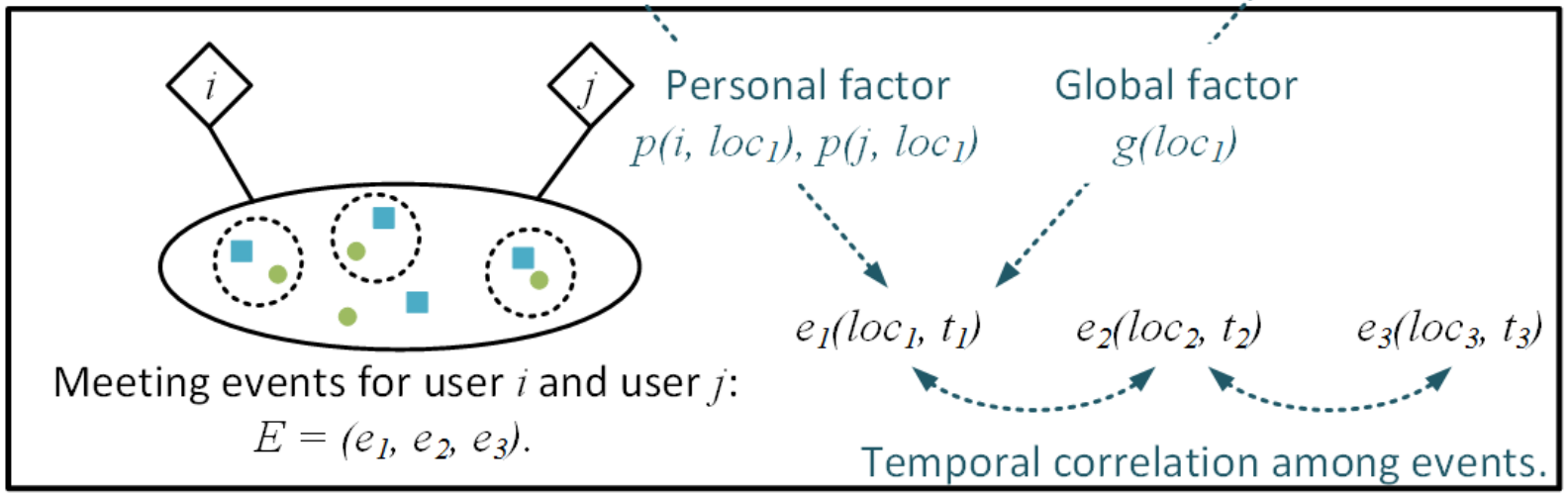
the stronger the mobility relationship is.

weaker

Background Modeling



Mobility Relationship Mining



Personal Background is important

Shanghai has a lower probability to be visited.
Co-location in Shanghai is less likely, but it happens.

Co-location event in Shanghai should carry higher weight.

Personal Background Formulation

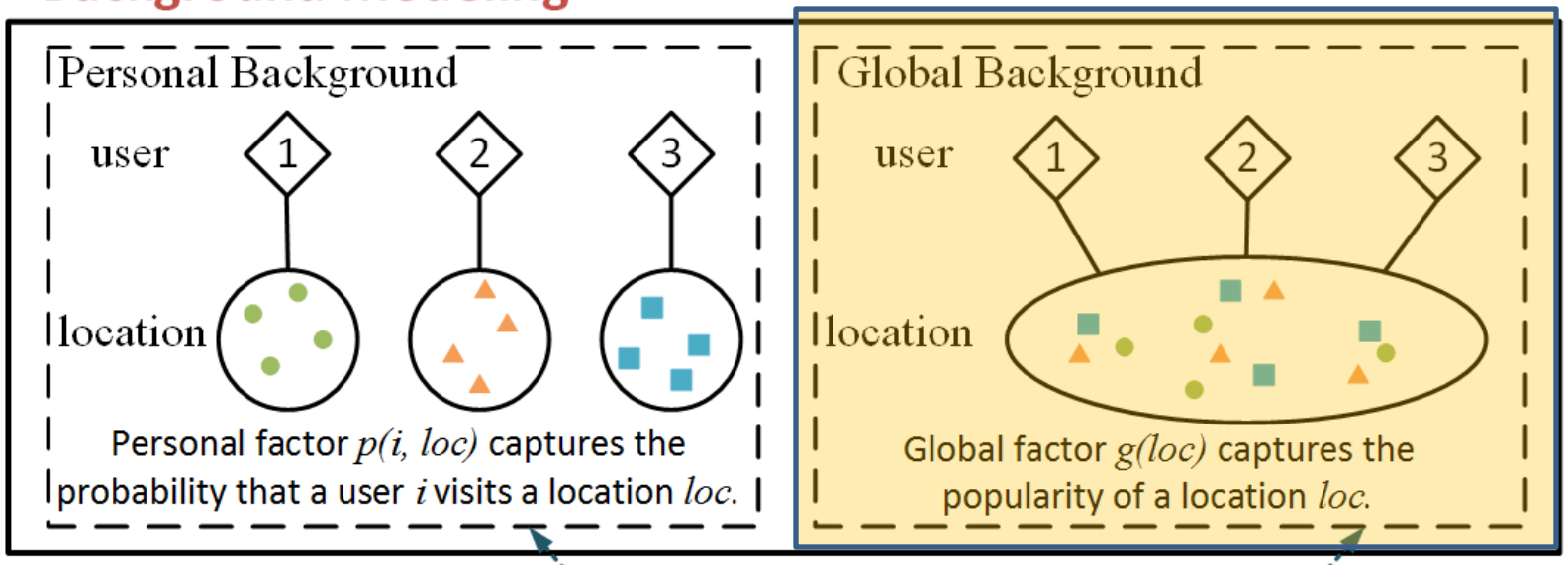
- For given user i, the probability of visiting location loc_k is

$$\rho(i, loc_k) = \sum_{loc_q^i \in S_i} e^{-c \cdot \frac{dist\left(loc_k, loc_k^i\right)}{|S_i|}}$$

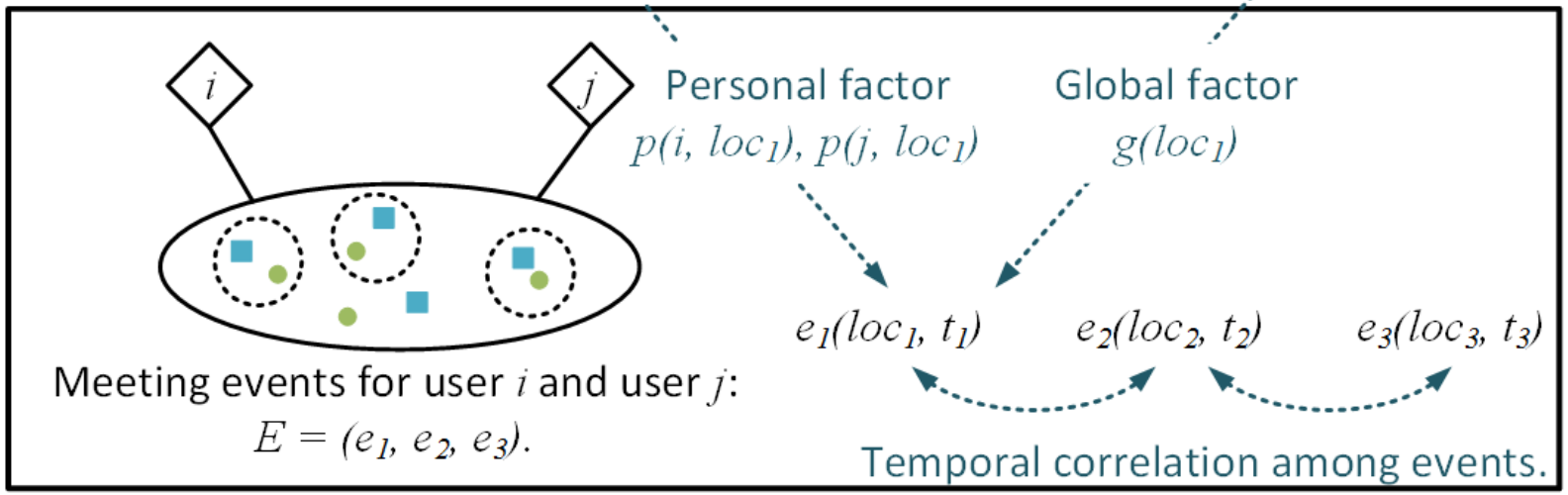
The visited location is far from others, the probability is low.

Judge whether visited location is close to others.

Background Modeling

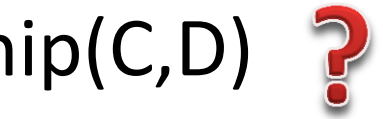


Mobility Relationship Mining



Global Background Matters

- A and B meet in downtown for 10 times.
- C and D meet in D's house for 10 times.

Relationship(A,B) = Relationship(C,D) 

Global Background Formulation

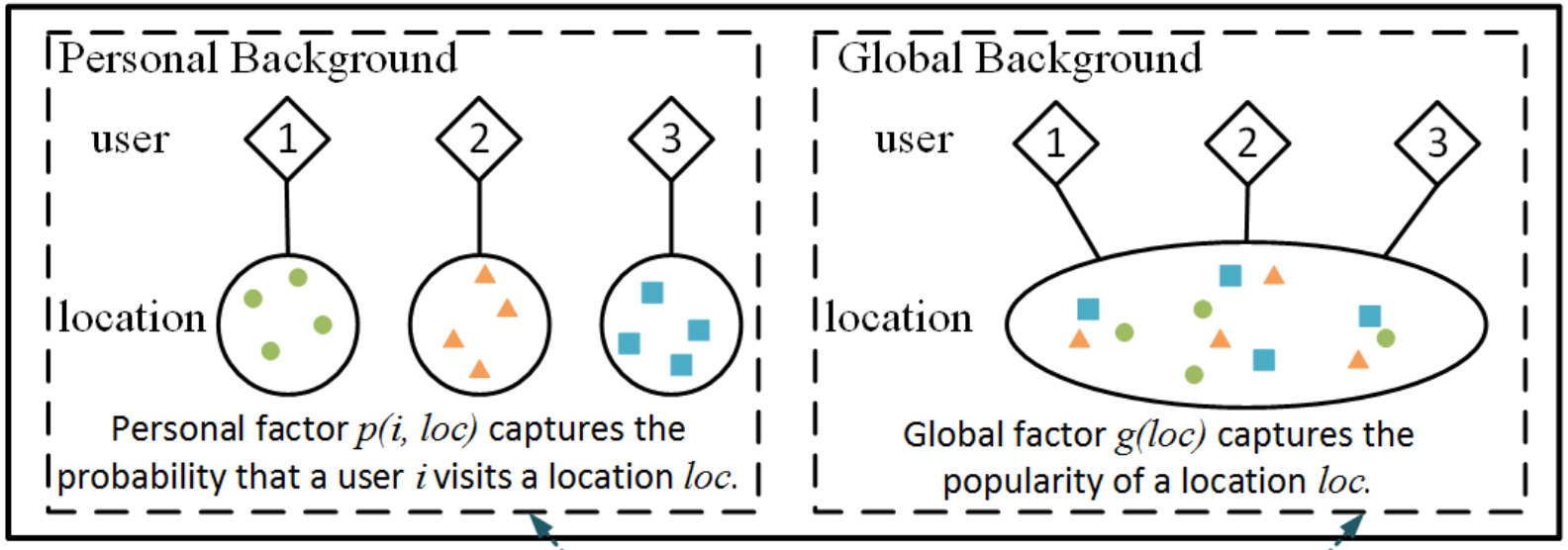
$$P(i, loc_k) = \frac{|S_i(loc_k)|}{\sum_i |S_i(loc_k)|}$$

At loc_k, the probability of observing different use *i*.

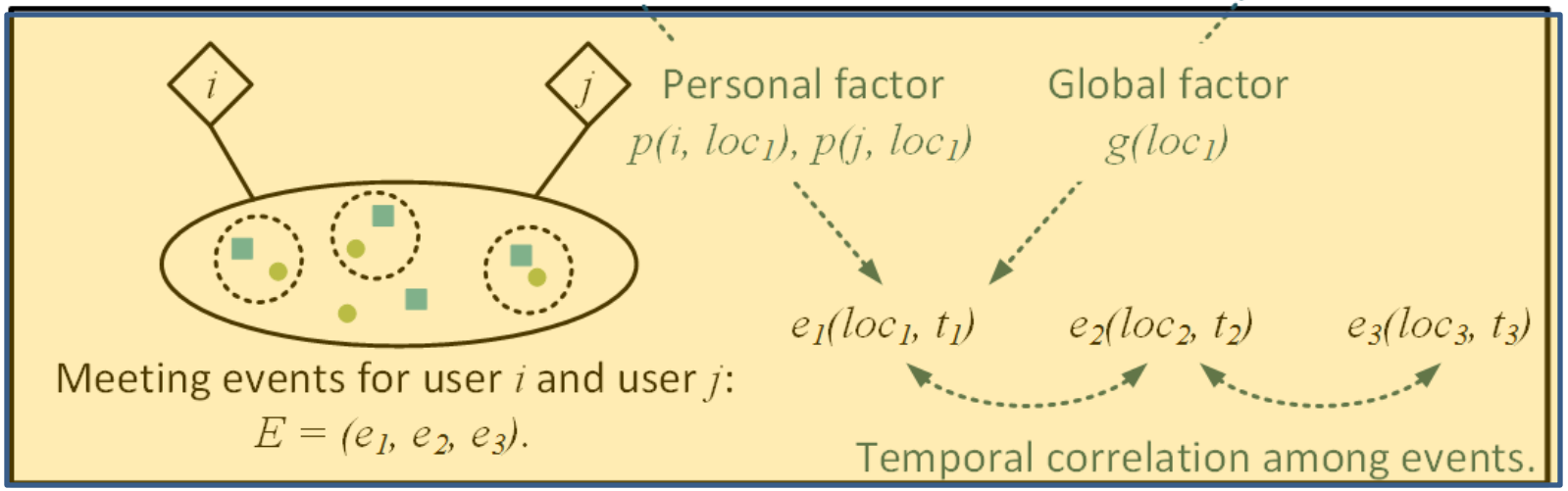
$$g(loc_k) = -\sum_{i:P(i,loc_k)\neq 0} P(i, loc_k) \cdot \log P(i, loc_k)$$

Entropy of loc_k.
Less users visited -> lower entropy -> more private location

Background Modeling



Mobility Relationship Mining



Temporal Correlation Between Events

03-26 10:00	03-26 11:20	03-26 14:30	03-26 15:36	03-26 15:37

Continuous meeting events → probably one-time trip?

03-01 10:00	04-23 09:20	05-01 11:30	06-21 10:46	06-26 08:37

Sporadic meeting events → a stronger relationship indication

Related Work

- Co-location frequency as measure (without considering background):
  - ***Kalnis et al. SSTD, 2005***
  - ***Jeung et al. VLDB, 2008***
  - ***Li et al. VLDB, 2010***
  - ***Cranshaw et al. Ubicomp, 2010.***
  - ***Zheng et al. ICDE, 2013***
- Global factors: ***Pham et al. SIGMOD, 2013.***
- Personal factors: ***None***
- Temporal factors: ***None***

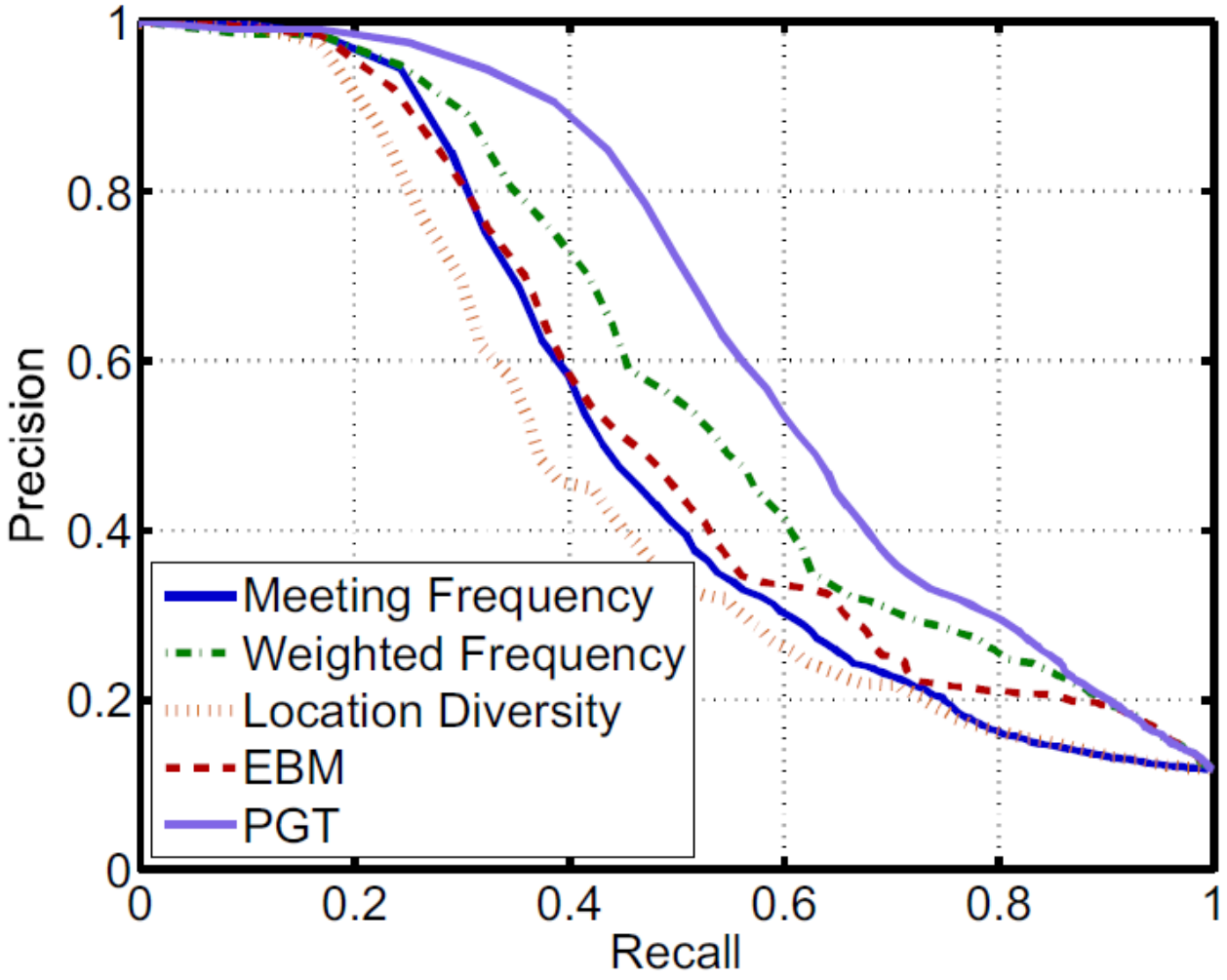
Experiments

- Datasets – two location-based social networks check-in data*
  - Gowalla (Feb, 2009 – Oct, 2010)
  - Brightkite (Apr, 2008 – Oct, 2010)

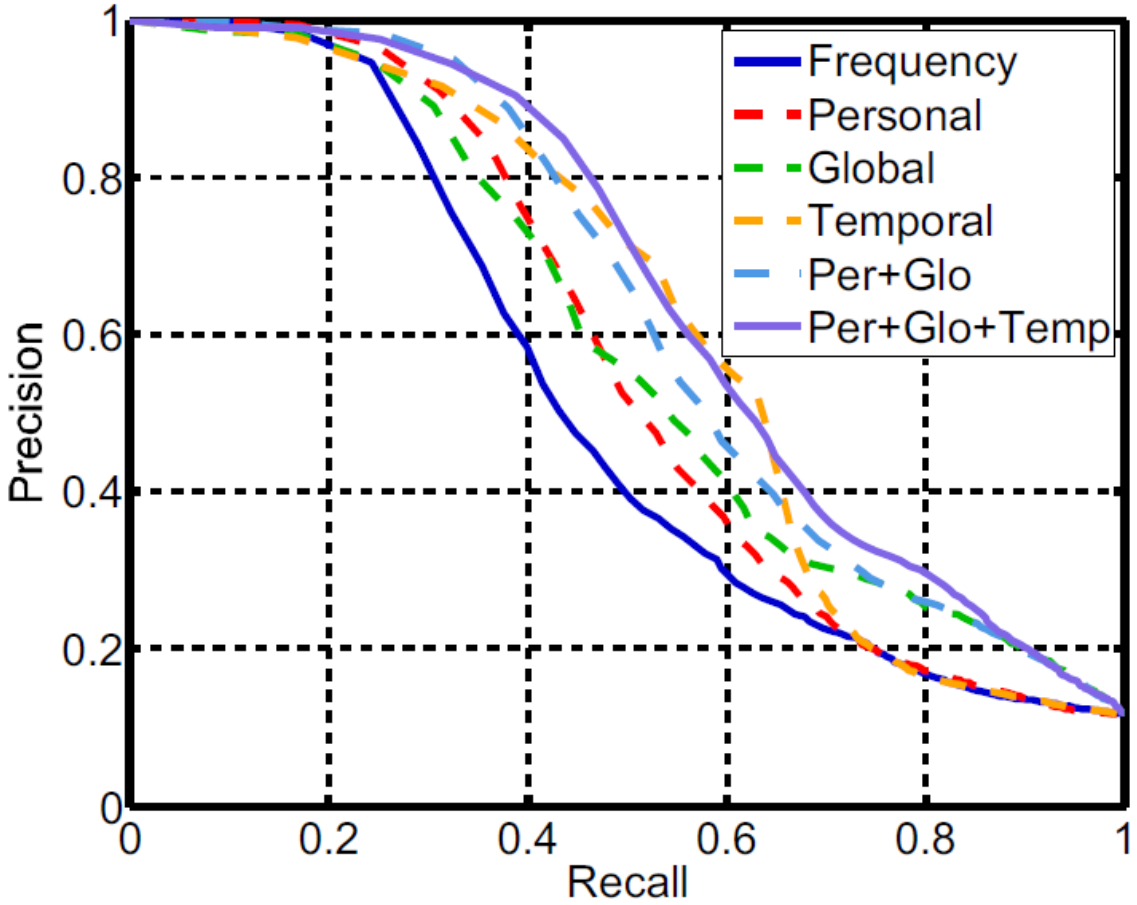
	Gowalla	Brightkite
No. of users	107,092	58,228
No. of friend pairs	950,327	214,078
No. of check-ins	6,442,890	4,491,143
Average check-ins per user	60	78

**** E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in Proc. KDD, 2011.***

Experiments: Compare with the State of the Art on Gowalla

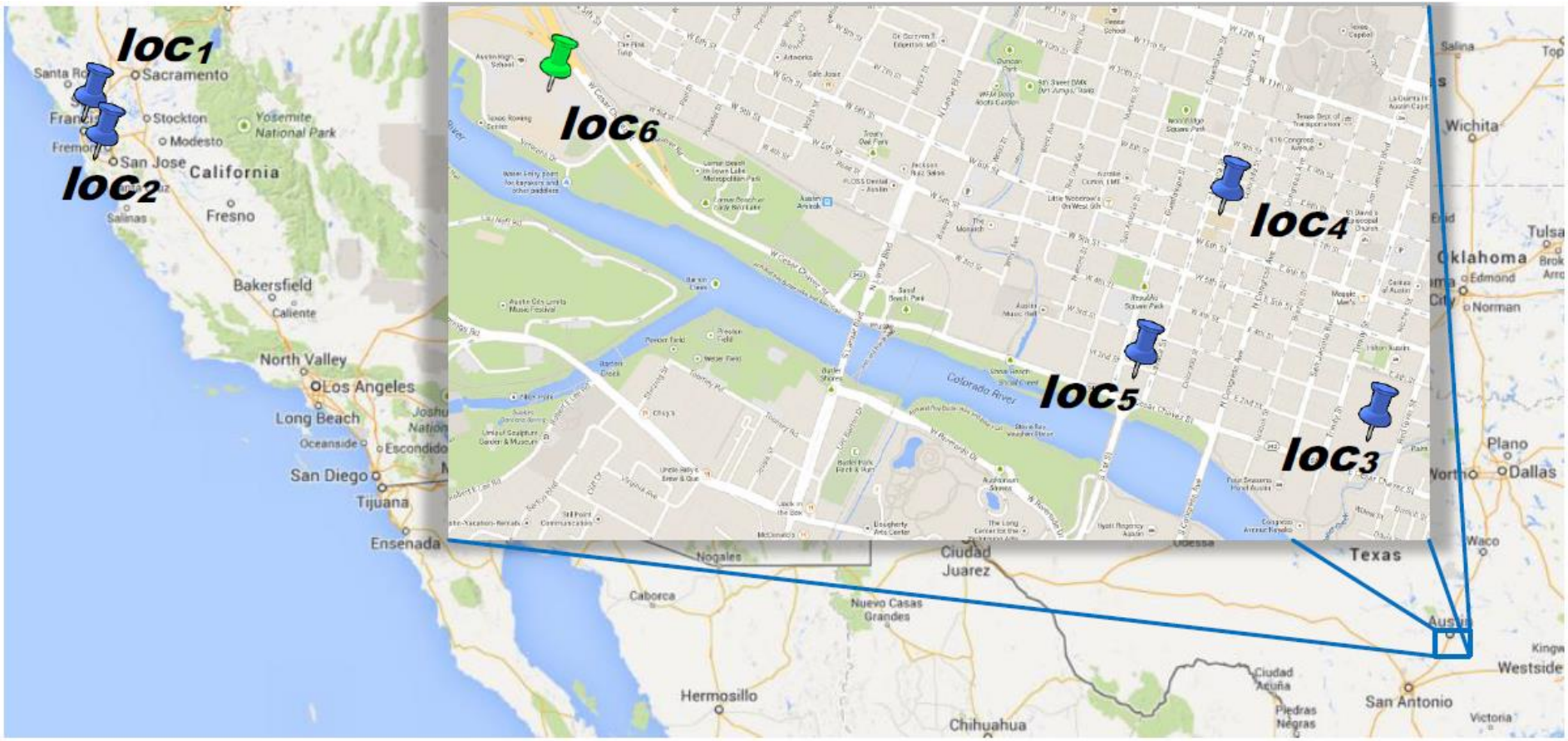


Experiments: Compare Various Factors



The precision-recall curves on top 5000 users from Gowalla.

Case Study: Personal Factor works



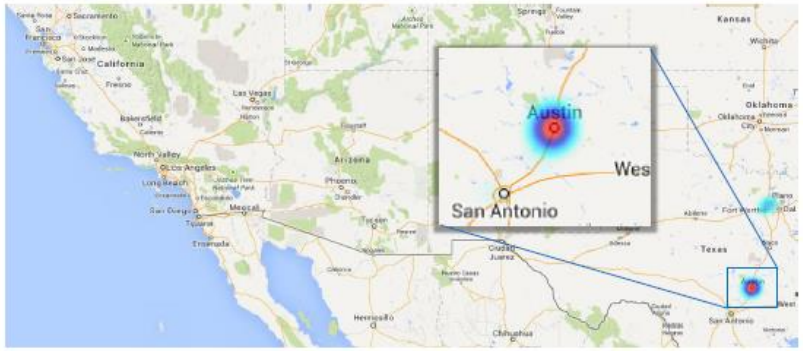
Both pairs meet 5 times in total. Blue Pair are friends. Green not.

Personal Profile of the Four Users

(a) User #267

(b) User #510

Friend



(c) User #350

(d) User #6138

Not

Results using Different Measures

User Pair	Friends / Not	Frequency	Personal Factor
#267, #510	Yes	5	22.03
#350, #6138	No	5	9.72

First Pair is more likely to be friends.

Summary

- We propose a unified framework to measure the strength of relationship based on two users' mobility.
- Our model is simple and deterministic, which considers:
  - Personal probability visiting a location
  - Location popularity from general public
  - Temporal correlation among co-locations

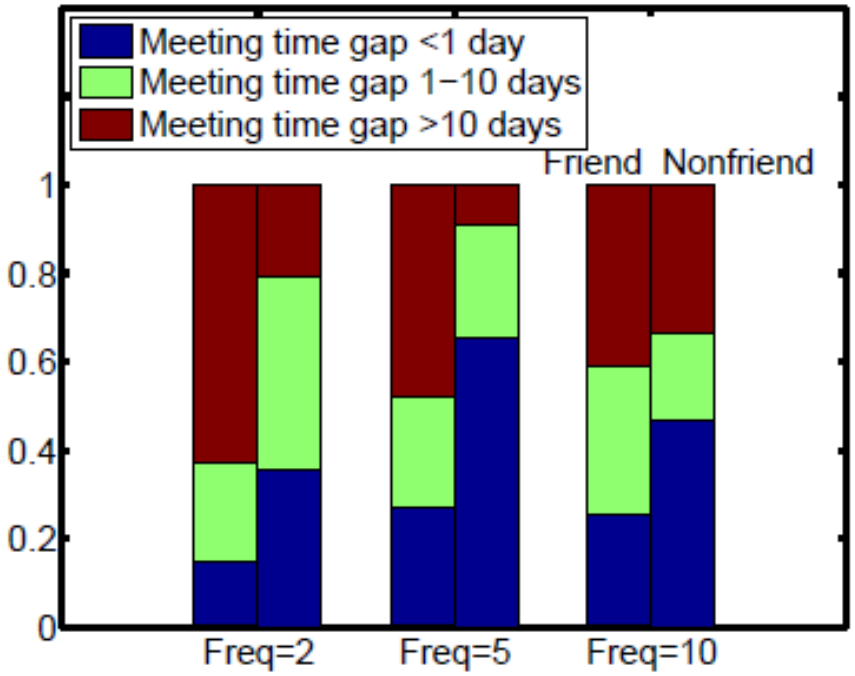
Future work

- Extend this work from identifying pairwise relationships to discovering common interest groups.
- Further combine the context at each location, such as the activity at that location.

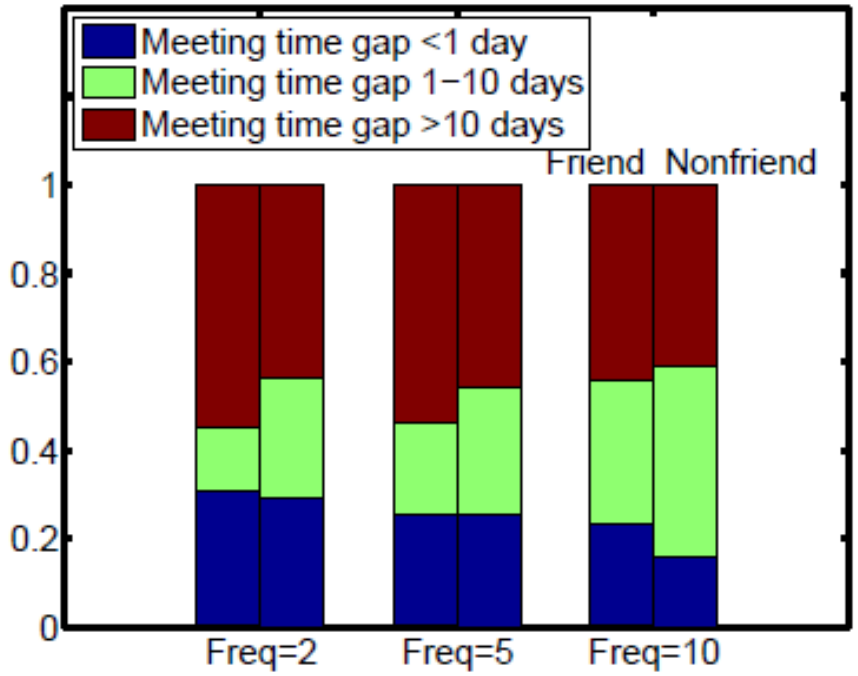
Dataset Properties

- The Gowalla users tend to check-in at featured spots, and recommend places and trips for others.
- The Brightkite users tend to check-in with acquaintance to maintain personal social circle.
- As a result, check-ins in Gowalla are mostly made on popular places.

Datasets Have Different Properties



(a) Gowalla



(b) Brightkite

The distribution of time gaps between consecutive meeting events for three representative groups (meeting frequency = 2; 5; 10).

Social Relation From Geospatial Data

- Diversity of co-locations

Table 1: Example of Diversities

Co-occurrence Vector	Shannon Entropy	D_{ij} Value	Diversity	Likelihood of Coincidences	Prob. of a Friendship
$C_{12} = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0)$	1.609	5.000	High	Low	High
$C_{23} = (1, 2, 1, 1, 0, 0, 0, 0, 0, 0)$	1.332	3.789	Medium	Medium	Medium
$C_{13} = (0, 0, 4, 0, 0, 0, 0, 0, 0, 0)$	0.000	1.000	Low	High	Low

High diversity -> high probability of friendship

H. Pham, C. Shahabi, and Y. Liu, "Ebm: An entropy-based model to infer social strength from spatiotemporal data," in Proc. SIGMOD, 2013.