Privacy-Preserving Online Task Assignment in Spatial Crowdsourcing with Untrusted Server

Hien To[1], **Cyrus Shahabi**[2], Li Xiong[3]

[1] Amazon Mechanical Turk

[2] University of Southern California

[3] Emory University

[1] This work has been done while the author was a PhD student at University of Southern California

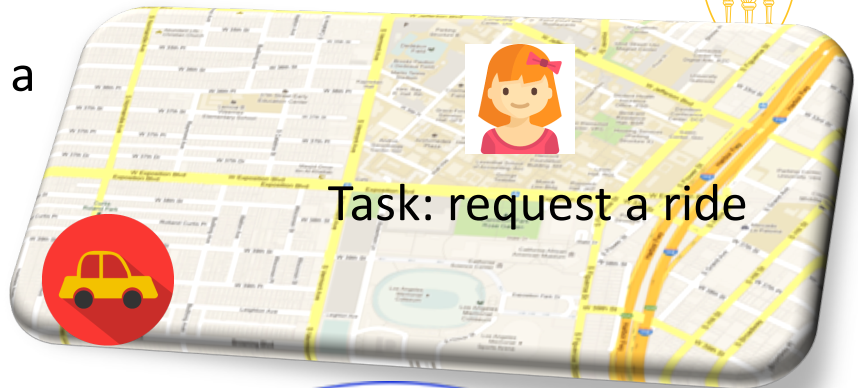
Outlines

- Introduction & Motivation
- Related work
- Background
- Proposed Approach
- Evaluation
- Conclusions

Spatial Crowdsourcing (SC)

Crowdsourcing: outsourcing a set of tasks to a set of workers amazon mechanical turk™ Artificial Artificial Intelligence

Spatial crowdsourcing (SC): requires workers to ***physically*** travel to task's location



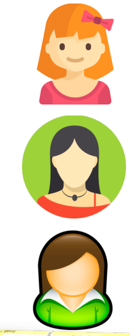
Ubiquity of mobile users — **6.5 billion** mobile subscriptions, 93.5% of the world population [*]

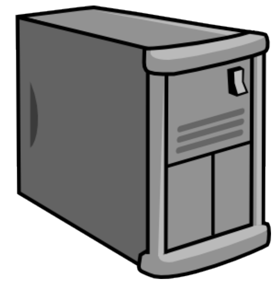
Technology advances on mobiles — Smartphone's **sensors**. e.g., video cameras

Network bandwidth improvements — From 2.5G (384Kbps) to 3G (14.7Mbps) and recently **4G** (100 Mbps)

Task Assignment in SC

Requesters (e.g., request a ride)





Server (e.g., Uber)

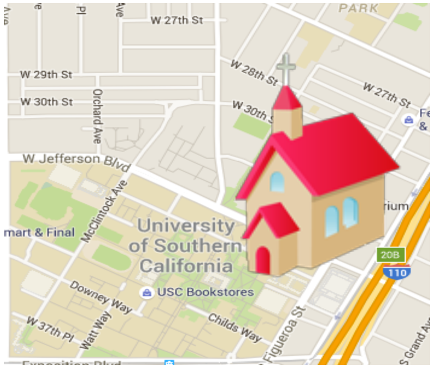
Workers (e.g., drivers)

Server chooses best workers for a task based on task-worker proximity *e.g., [Kazemi'12, Pournajaf '14, To'17]*

Server knows locations of workers and tasks ☹

Risks of Location Leaks

Location leaks sensitive information, e.g., religious view, health status



Attacks based on locations:

PRIVACY ROAD KILL 4/26/16 2:40 PM

If you use Waze, hackers can stalk you

'God View': Uber Allegedly Stalked Users

"Uber treated guests to Creepy Stalker View, showing them the whereabouts and movements of 30 Uber users in New York in real time."

Forbes

Location Privacy

Anonymity based (e.g., cloaking)

- Pseudonymity *[Pfitzmann et al. 2010]*
- K-anonymity/Cloaking *[Sweeney'02]*

Encryption-Based

- Private information retrieval *[Ghinita et al. SIGMOD 2008]*
- Space transformation *[Khoshgozaran & Shahabi SSTD 2007]*

Perturbation (e.g., differential privacy)

- Geo-indistinguishability *[Andrés et al CCS 2013]*
- δ-location set-based differential privacy *[Xiao & Xiong CCS 2015]*

Apple and Google adapted **differential privacy** to discover usage patterns from a large number of users

- Google Chrome web browser [1]
- Apple QuickType/Emoji [2] suggestions.

[1] Erlingsson et. al. *Rappor: Randomized aggregatable privacy-preserving ordinal response. ACM SIGSAC 2014.*
[2] Learning with Privacy at Scale.
https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html

Related work

Privacy-Preserving Task Assignment

Papers	*Privacy Techniques*			*Protection*		*Trusted Server*	
	Cloak	*Encrypt*	*Perturb*	*Worker*	*Task*	*Yes*	*No*
[Pournajaf et al. 2014]	x			x		x	
[Sun et al. 2017]	x			x		x	
[Pham et al. 2017]	x			x	x	x	
[Hu et al. 2015]	x			x		x	
[Shen et al. 2016]		x		x			x
[Liu et al. 2017]		x		x	x		x
[To et al. 2014]			x	x	☹	x	☹
[Gong et al. 2015]			x	x	☹	x	☹
[Zhang et al. 2015]			x	x	☹	x	☹
[To et al. 2016]			x	x	☹	x	☹

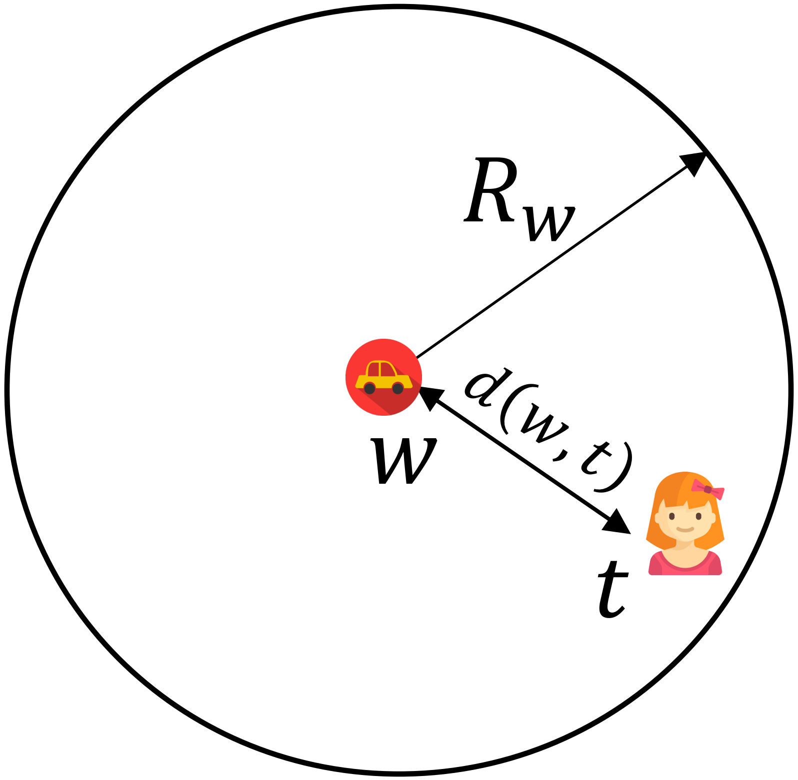
Existing work that use perturbation technique protect worker location only and assume trusted server ☹

Outlines

- Introduction & Motivation
- Related work
- Background
- Proposed Approach
- Evaluation
- Conclusions

Notations

Notation	Description
w, t	Actual locations of a worker, a task
w', t'	Perturbed locations
R_w	Reachable distance of worker w
$d(w, t)$	Euclidean distance between w and t



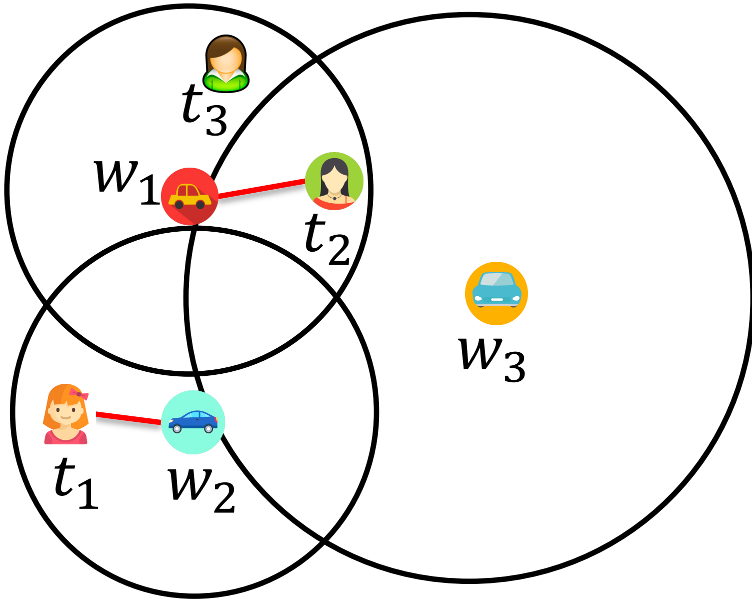
Task t is ***reachable*** from worker w if $d(w, t) \leq R_w$

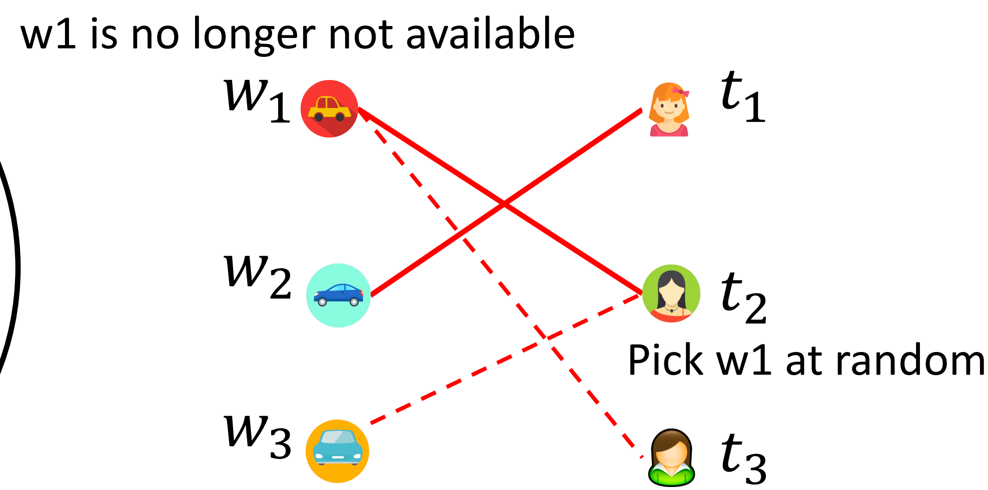
d can be non-Euclidean & R_w can be complex shapes like polygon

Background

Online Task Assignment

Worker set is known, each task arrives one-by-one





Assign as many tasks as possible to workers

Ranking algorithm[*] is optimal, competitive ratio 0.63

- Permutes workers and assigns a **random rank** to them
- Each task is matched to a reachable worker of the highest rank

USC Viterbi School of Engineering *Integrated Media Systems Center*

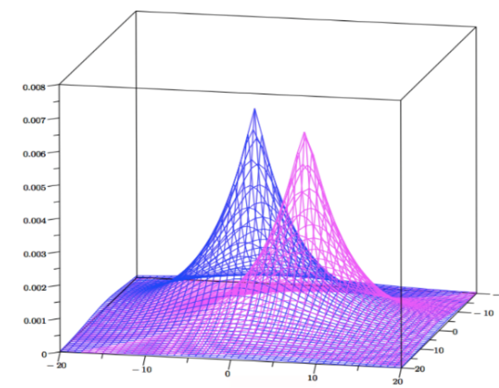
[*] Karp et al. *An optimal algorithm for on-line bipartite matching*, STC'90

Background

(ϵ, r) Geo-indistinguishability[*]

<u>The goal:</u> An adversary cannot distinguish locations which are at most r distance away

<u>Approach:</u> Any two locations at distance at most r produce "similar" observations (bounded by ϵ),



More formally:

Mechanism A satisfies (ϵ, r)-Geo-I iff for all x, y such that $d(x, y) \leq r$:

$$d_p\big(A(x), A(y)\big) \leq \epsilon d(x, y) \leq \epsilon r$$

- $d(x, y)$: Euclidean distance between x, y
- $d_p(,)$: multiplicative distance between two distributions

[*] Andrés et al. *Geo-indistinguishability: differential privacy for location-based systems*, CCS'13

Background

(ϵ, r) Geo-indistinguishability[*]

it is sufficient to achieve (ϵ, r)-Geo-I by generating random point z (from actual point x ∈ X) according to planar Laplace distribution.

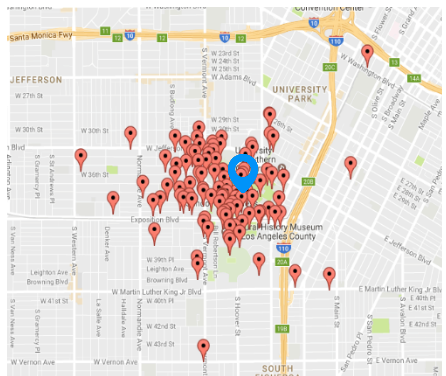
r (in meters) is the radius within which privacy is guaranteed

ϵ tunes how much privacy, smaller ϵ means higher privacy

achieve privacy by injecting planar Laplace noise

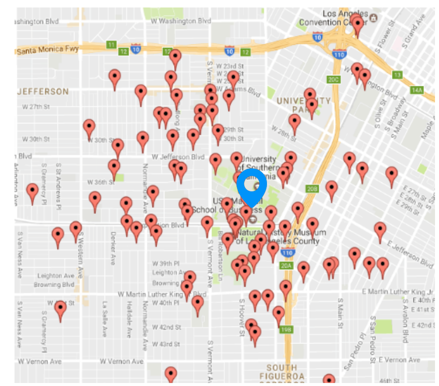
True locations

Perturbed locations



$\epsilon = \log(6)$

r=1 km



Better privacy: $\epsilon = \log(2)$

r=1 km

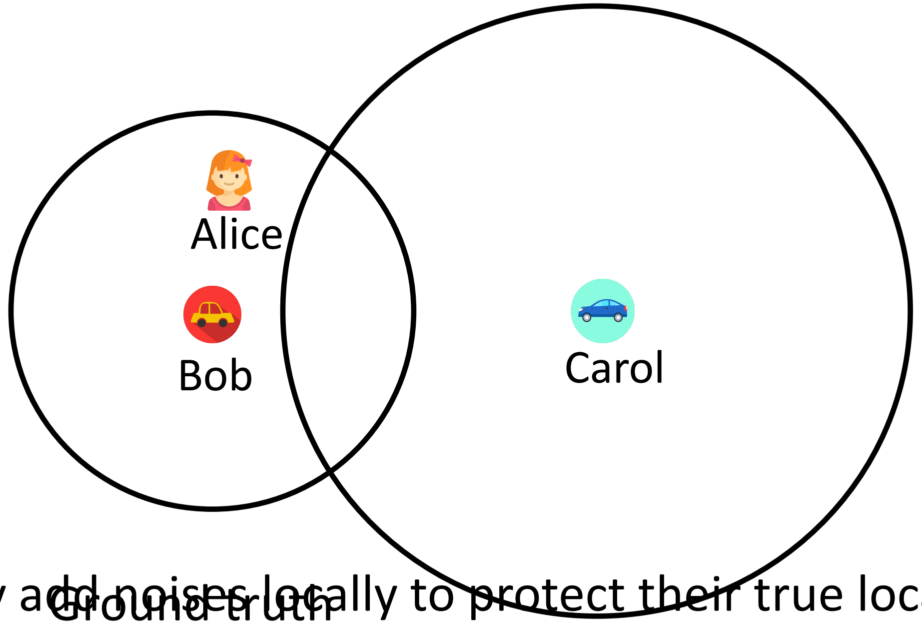
USC Viterbi
School of Engineering
Integrated Media Systems Center

[*] Andrés et al. *Geo-indistinguishability: differential privacy for location-based systems*, CCS'13

Challenges with Perturbed Locations

Background

Reachable worker-task pair is observed as unreachable, and vice versa



They add noises locally to protect their true locations

Ground truth

Alice is not assigned to Bob (not reachable) ☹

Alice's location is disclosed to Carol *unnecessarily* ☹

Outlines

- Introduction & Motivation
- Related work
- Background
- Proposed Approach
- Evaluation
- Conclusions

Three-Phase Framework

Alice requests for a ride

Finds ***candidate drivers*** *for* Alice: Bob Carol Dave

Server does not know *anyone's* location (works in perturbed space for both riders and drivers)

Sends perturbed locations of drivers to Alice

System Overhead

Finds the **most likely reachable driver:** Bob

Alice does not know *any driver's* location (works in perturbed space for drivers but knows her own location)

Reveals her location to Bob

Location Disclosure

Bob checks if Alice is reachable

Reachable → accepts (happy case)

Not reachable → rejects

Repeat until either task is assigned or no candidate worker left

Baseline Approach

System Overhead: size of the worker candidate set, captures communication and computational overhead
Location Disclosure (false hit): privacy leak occurs when Alice estimates an unreachable worker as reachable & reveals her location
Utility: number of assigned tasks
Worker Travel Cost: captures travel cost or assignment quality

"Oblivious" algorithm

- assumes perturbed locations as actual ones
- Direct adaptation of Ranking algorithm[*] to our framework
  - Consider both **random rank** and **distance-based rank**

Core idea:

- to use underlying distributions of noisy locations to estimate real locations

USC Viterbi School of Engineering Integrated Media Systems Center

[*] Karp et al. *An optimal algorithm for on-line bipartite matching*, STC'90

Worker-Task Reachability

Compute the **reachability probability** of a worker-task pair given their observed distance

$: \Pr(d(w,t) \leq R_w \mid d(w', t'))$

$: \Pr(d(w,t) \leq R_w \mid d(w', t))$

I. Analytical approach, based on estimating the reachability probability

- Derive PDF of $d(w,t)$, given w', t'

  Subsequently, the reachability probability can be computed efficiently

- Planar Laplace distribution is difficult to analyze so we approximate it by bivariate normal distribution (BND)

II. Empirical approach, based on synthetic or historical data

Proposed Approach

Bivariate Normal Distribution (BND)

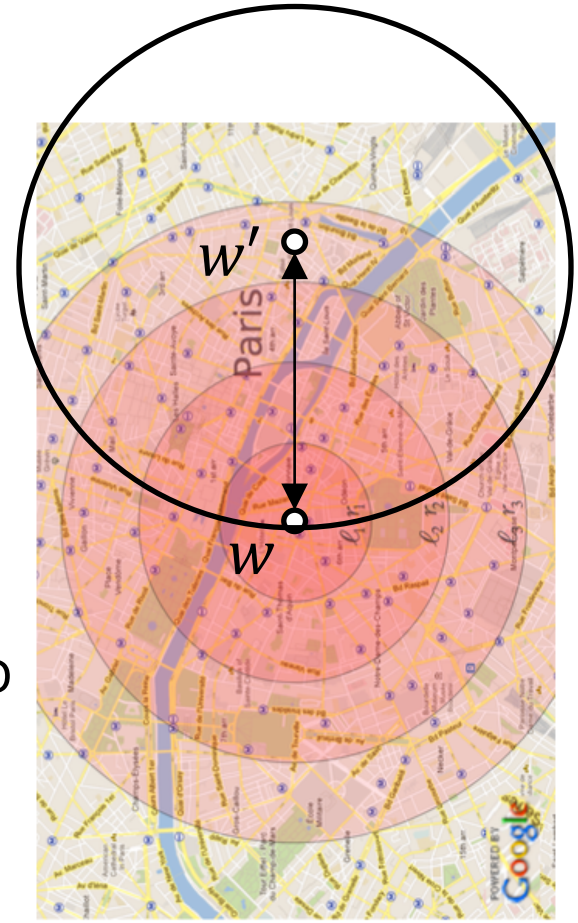
(ϵ, r)-Geo-Indistinguishability uses planar Laplace distribution (PLD) to inject noise

- PLD is difficult to analyze

Approximate PLD by a circular BND with same mean (w_x, w_y) & covariance matrix $\begin{bmatrix} \frac{2r^2}{\epsilon^2} & \\ & \frac{2r^2}{\epsilon^2} \end{bmatrix}$

- BND is made up of two random variables x and y; both normally distributed
- PLD is symmetric to its center → approximated BND should be symmetric to the same center

w' is known → w follows circular BND centering at w': circular $BND(w', \Sigma)$



USC Viterbi School of Engineering Integrated Media Systems Center

[*] Andrés et al. *Geo-indistinguishability: differential privacy for location-based systems*, CCS'13

Proposed Approach

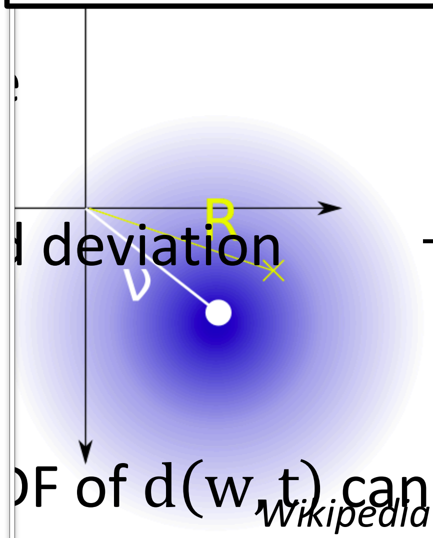
derives PDF of $d(w,t)$

Given true location of Alice t and perturbed location of Bob w'

estimates PDF of $d(w,t)$

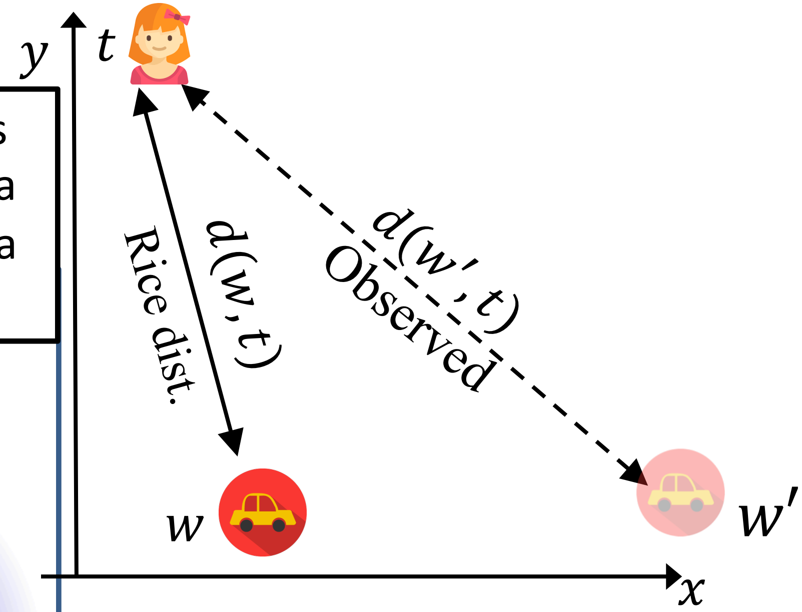
In the 2D plane, pick a fixed point at distance ν from the origin. Generate a distribution of 2D points centered around that point, where the x and y coordinates are chosen independently from a gaussian distribution with standard deviation σ (blue region). If R is the distance from these points to the origin, then R has a Rice distribution.

Rice distribution is the magnitude of a circular BND with a non-zero mean



l deviation

F of d(w,t) can be found in the paper



[*] Stüber. *Principles of mobile communication*, volume 2. Springer, 2001

Probability-based Solution

The key idea is to use the probabilistic model (either the analytical or the empirical approach), for quantifying reachability between a worker and a task.

finds *candidate drivers* N_j based on *reachability threshold* α

$$N_j = \{w_i : \Pr(reachability(w'_i, t'_j)) \geq \alpha\}$$

The smaller α, the higher the overhead, but less chance of missing a reachable worker

reveals her location to highly *likely reachable drivers*

$$Rank_{w_i} = \Pr(reachability(w'_i, t_j))$$

Heuristic:

can reduces disclosure of her location based on *reachability threshold* β $(\beta > \alpha)$

e.g., if $Rank_{w_i} < \beta$, cancel this task

Outlines

- Introduction & Motivation
- Related work
- Background
- Proposed Approach
- Evaluation
- Conclusions

Experimental Evaluation

- GPS-equipped taxis dataset [1]
  - Workers' locations are the most recent drop-off locations
  - Tasks' locations at the pick-up locations
  - 500 tasks and 500 workers were randomly sampled

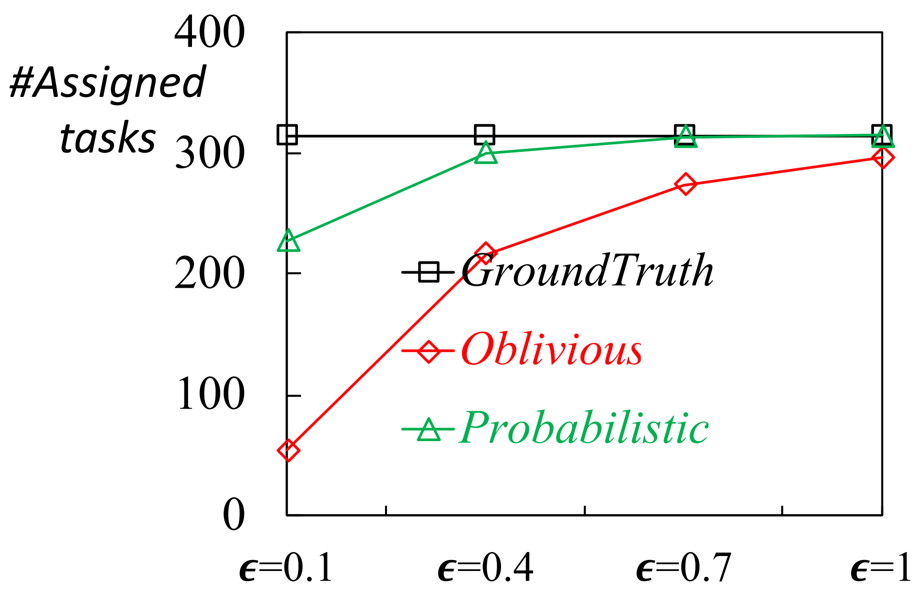
	#Passengers	#Drivers	Area
T-Drive	100,000+	9,019	Beijing City

- Performance metrics
  - **Utility**: number of assigned tasks
  - **Worker Travel Cost**: captures travel cost or assignment quality
  - **System Overhead**: size of the worker candidate set, captures communication and computational overhead
  - **Location Disclosure** (false hit): privacy leak occurs when requester estimates an unreachable worker as reachable

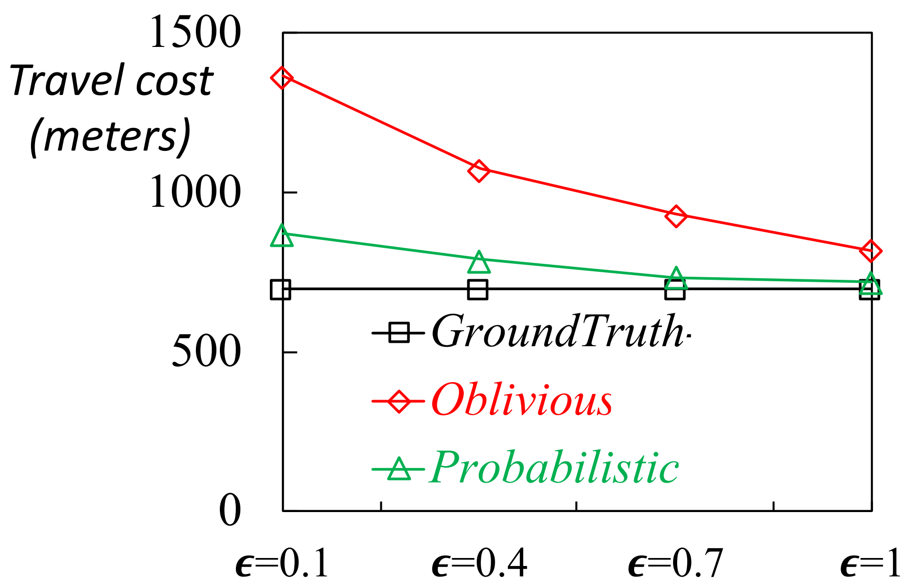
[1] Yuan et al. *T-drive: driving directions based on taxi trajectories*. SIGSPATIAL 2010

Utility and Travel Cost

GroundTruth	Has access to exact locations (distance-based rank)
Oblivious	Assumes perturbed locations as actual ones (distance-based rank)
Probabilistic	Estimates worker-task reachability (probability-based rank)



Probabilistic obtains much **higher utility** than *Oblivious* (by 300%)



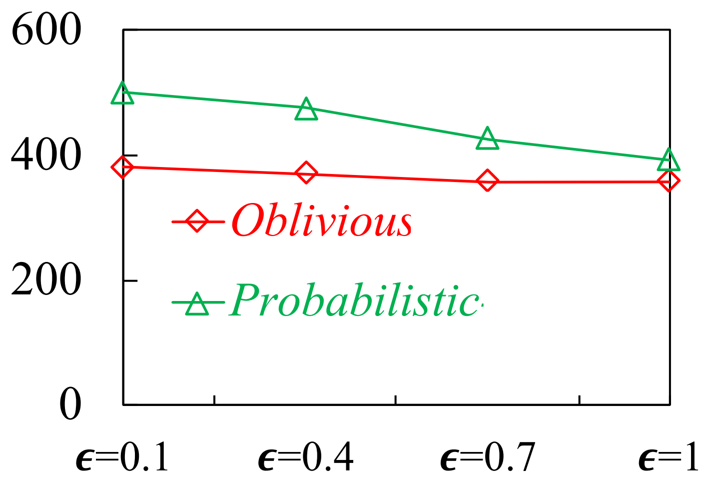
Probabilistic obtains significantly **lower travel cost** than *Oblivious* (by 30%)

Evaluation

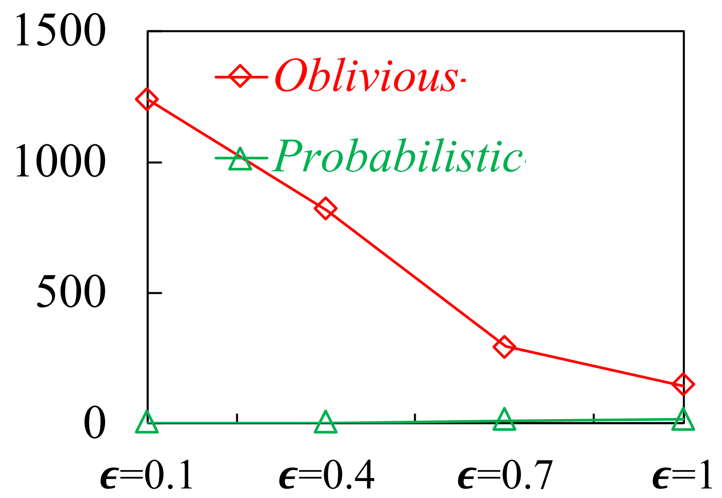
System Overhead and Privacy Leak

Oblivious	Assumes perturbed locations as actual ones (distance-based rank)
Probabilistic	Estimates worker-task reachability (probability-based rank)

#workers (overhead)



#false hits (disclosure)



Although the overhead of *Probabilistic* is slightly higher than *Oblivious's*, *Probabilistic* has **much smaller false hits**

Average **#false hits** before a task can be assigned: 23 workers vs 1.05 workers

USC Viterbi School of Engineering *Integrated Media Systems Center*

Conclusions and Future Work

- Protected locations of both workers and tasks
  - Introduced privacy-aware framework with untrusted server
  - Proposed models for quantifying worker-task pair reachability
  - Proposed algorithms, heuristics for effective online tasking
- Confirmed the cost of privacy is practical
  - Low cost and low overhead without compromising utility
- Future directions
  - Consider malicious adversaries: requesters send fake tasks to estimate workers’ locations, server colludes with workers (driverless cars)
  - Consider protection for dynamic workers and task: workers’ traces and task locations of individual requesters can follow a specific pattern
  - Consider tasks that may require redundant assignment: taking pictures of a particular location, reporting how crowded a restaurant is

Unintended Consequences of Disclosing Location Data

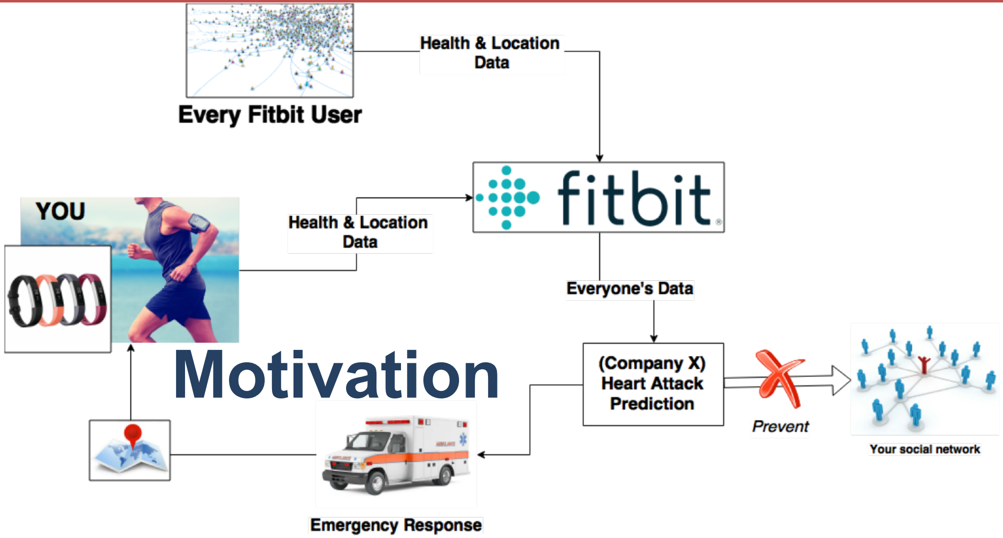
Cyrus Shahabi, Ph.D.
Professor of Computer Science, Electrical Engineering & Spatial Sciences
Chair, Department of Computer Science
Director, Integrated Media Systems Center (IMSC)
Viterbi School of Engineering
University of Southern California
Los Angeles, CA 900890781
shahabi@usc.edu

Outline

- Motivation: Geo-social Privacy
- Prior Work: Inferring Social Behaviors
- Current Efforts: Protecting against social inferences
  - But allow location disclosure
- Open Problem: Protecting against location disclosure
  - But allow social inferences

Motivation

Location Data is necessary for service but social connectivity is sensitive.



Enable LBS to provide recommendation, advertisement, and other services.

Outline

Motivation: Geo-social Privacy

Prior Work: Inferring Social Behaviors

Current Efforts: Protecting against social inferences

- But allow location disclosure

Open Problem: Protecting against location disclosure

- But allow social inferences

Privacy Twist

Inferring Social Relationships

- Privacy attack

walk2friends: Inferring Social Links from Mobility Profiles [CCS, Nov ‘17] Backes M, Humbert M, Pang J, Zhang Y.

walk2friends: Inferring Social Links from Mobility Profiles [CCS, Nov '17] Backes M, Humbert M, Pang J, Zhang Y.

- Can we do better in very dense datasets ?

- Feature learning method – Unsupervised
  - As opposed to EBM's supervised linear regression.
  - Claims to exploit followship in addition to EBM's co-occurrence

- Inspired by Deep Learning in NLP – word2vec
  - Skip-gram Model
    (Tomas Mikolov et. al., at Google Research, 2013)

A glance at the Skip-Gram Model

Goal: Given a specific word in a sentence, tell us the probability for every word in our vocabulary of being the "nearby word" to the one we chose.

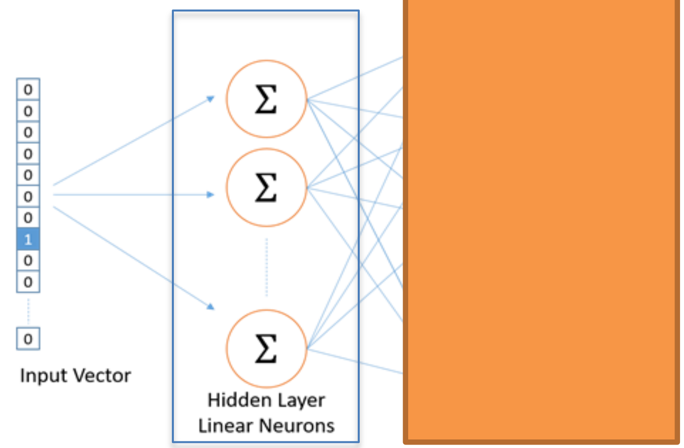
Corpus training (NN)

The quick brown fox jumps over the lazy dog.

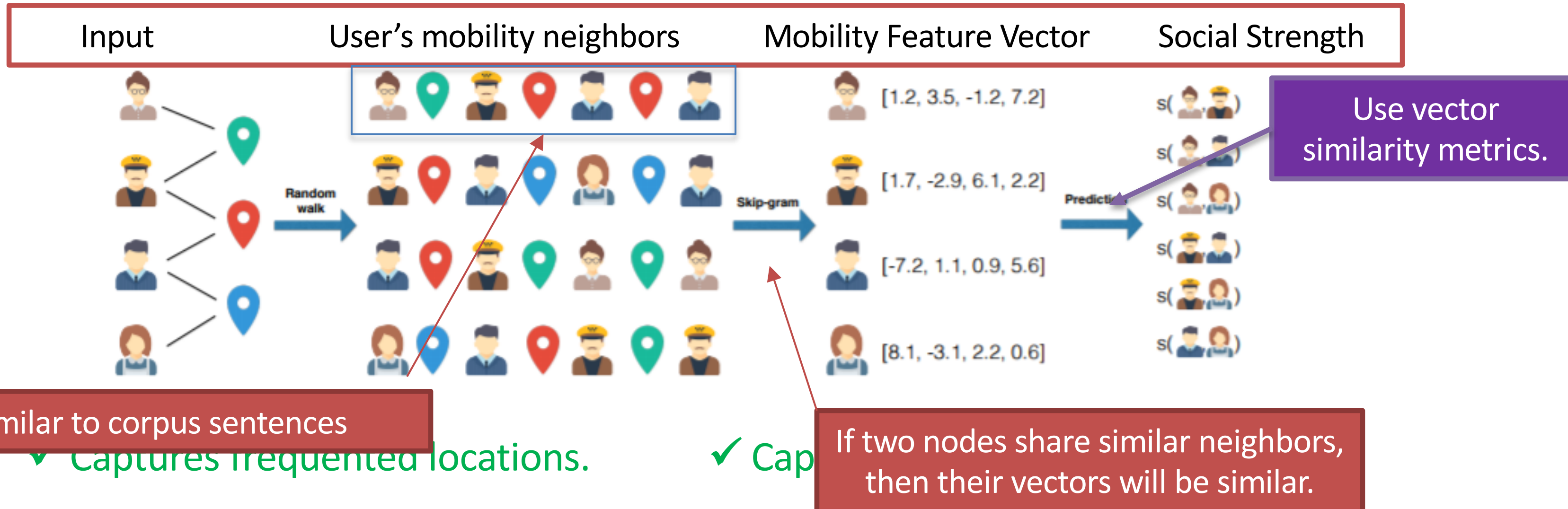
→ (fox, quick)
(fox, brown)
(fox, jumps)
(fox, over)

Keep weights of Hidden Layer i.e. word to vector!

Discard Output Layer!



walk2friends: Extending to locations based networks.



✓ Captures frequented locations. ✓ Cap

✓ Performs~10-15% percent better than EBM on relatively dense datasets.

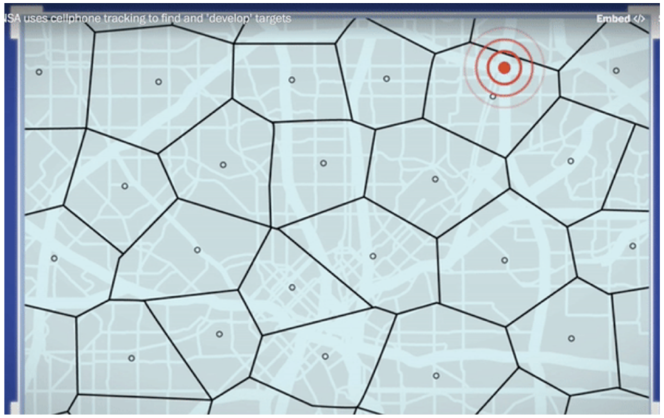
✗ 3-5% worse on sparse datasets.

Outline

- Motivation: Geo-social Privacy
- Prior Work: Inferring Social Behaviors
- Current Efforts: Protecting against social inferences
  - But allow location disclosure
- Open Problem: Protecting against location disclosure
  - But allow social inferences

Co-Location Privacy Risks

1. NSA PRISM (began 2007):
   Mass surveillance of location data from Google, FB, Microsoft.



[Source: Washington Post]

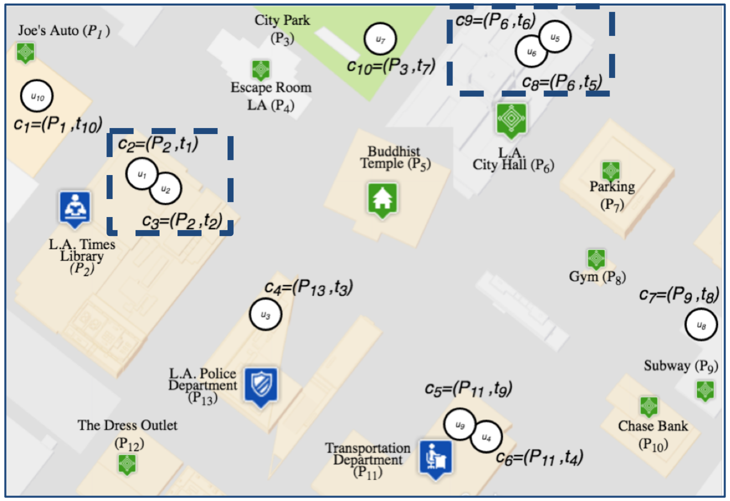
2. NSA's Co-Traveler program (exposed 2013):
   Identifies unknown associates of a known target.

3. Domestic prosecution facilitated by co-location information as evidence of wrongdoing. [United States v. Jones, 132 S.Ct. 945 (2012)]

Target Co-locations

The building blocks for social inference techniques.

Co-Location: Two people at *roughly* the same geographic locale at roughly the same time.



We quantify 'roughly' based on parameters Δ_s and Δ_t .

In running example,

- Assume buildings are points

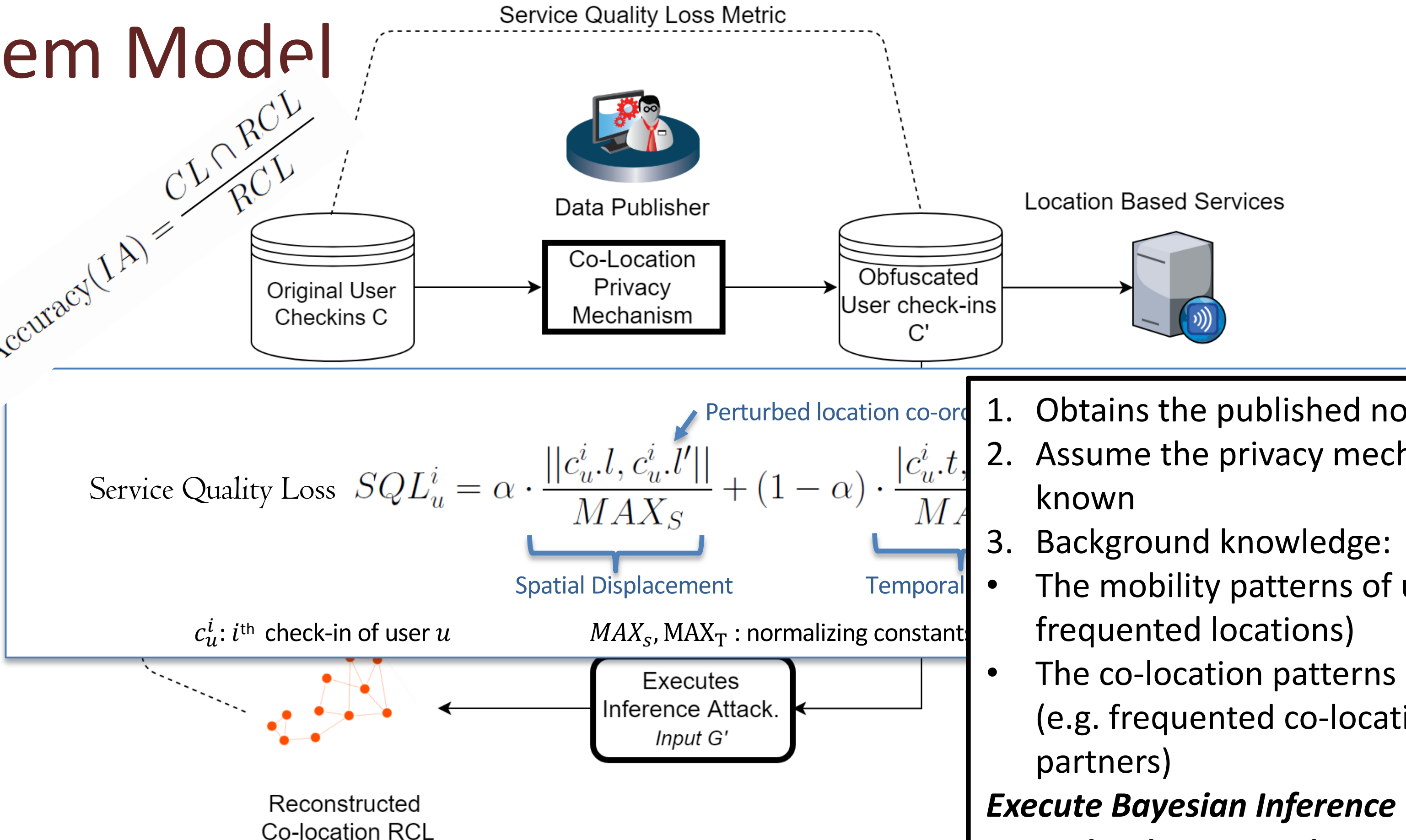
$\Delta_s = SameBuilding, \Delta_t = 1t$

Co-Locations: $(u_1, u_2), (u_5, u_6)$

Δ_s and Δ_t are application specific.

System Model

$$\text{Inference Accuracy}(IA) = \frac{CL \cap RCL}{RCL}$$



1. Obtains the published noisy data
2. Assume the privacy mechanism is known
3. Background knowledge:

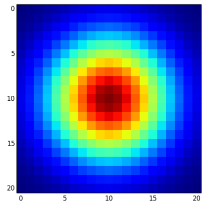
- The mobility patterns of users. (e.g. frequented locations)
- The co-location patterns of users. (e.g. frequented co-locating partners)

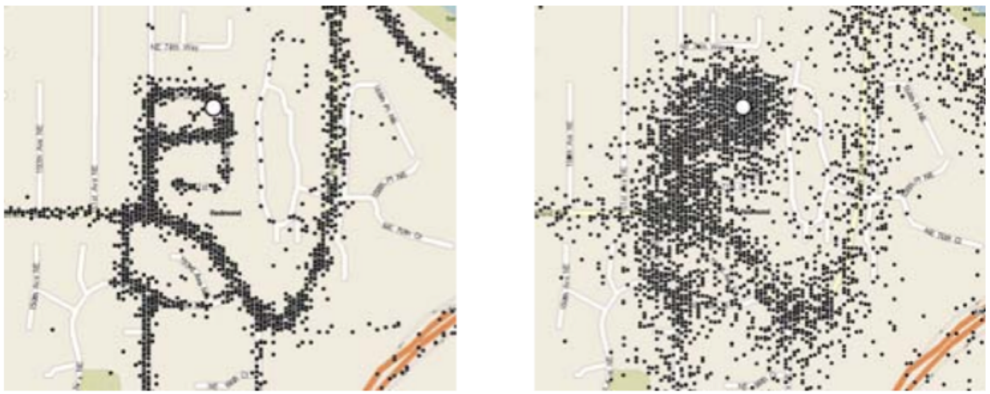
Execute Bayesian Inference to reconstruct as accurate as possible representation of the original co-locations.

Method 1: Gaussian Perturbation (Naïve)

Popular method in statistical data privacy and location privacy.

Method: 1. For every co-location.
2. Translate coordinates with 2d-gaussian noise.
3. Translate timestamp with 1d-gaussian noise





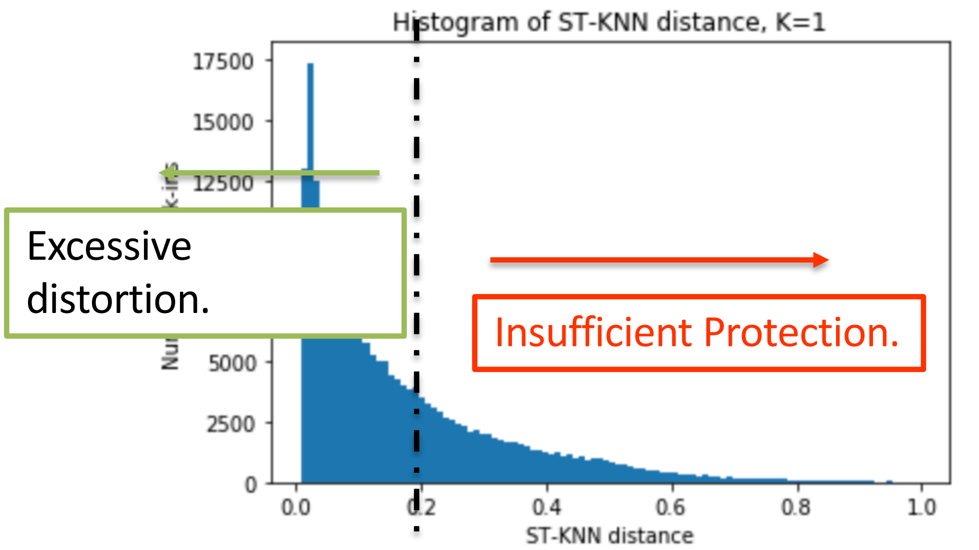
(a) Original GPS data (b) Additive Gaussian noise

Krumm, [PerCom'07]

Shortcomings of Gaussian Perturbation

1. Skewed nature of the distribution of the closest neighbor: many have NN very close, and some have NN very far.

2. Any fixed magnitude of noise will leave co-locations with
   - Low Privacy: Under-protected in sparse areas
   - Low Utility: Over-protected In dense areas inhibiting quality of LBSs.



$ST_{dist} = 0.1 = (100m, 40min)$

Method 2: Adaptive Perturbation

Use the presence of spatio-temporal nearest neighbors as an estimate for density.

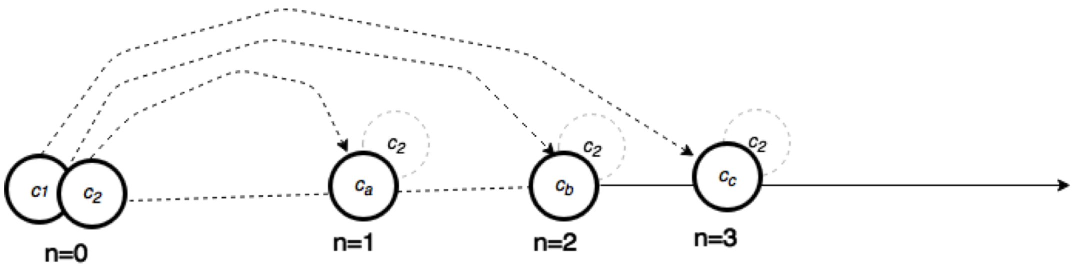
Method: 1. For every check-in in a co-location pair

2. Chose a point p uniformly over the set of

(i) the k nearest neighbors,

(ii) together with the current location.

3. Move to p.



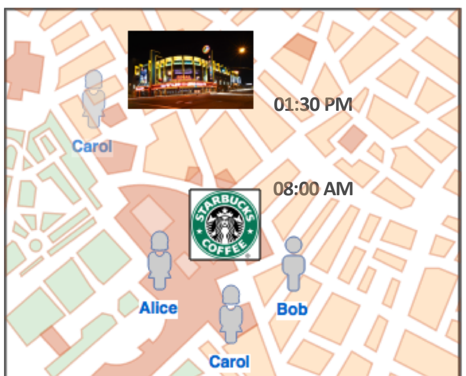
Move c2 to any of 'b=4' positions at random

$^*ST_{dist}(c, c')$ = sum of normalized spatial and temporal distances

Method 3: Co-Location Masking

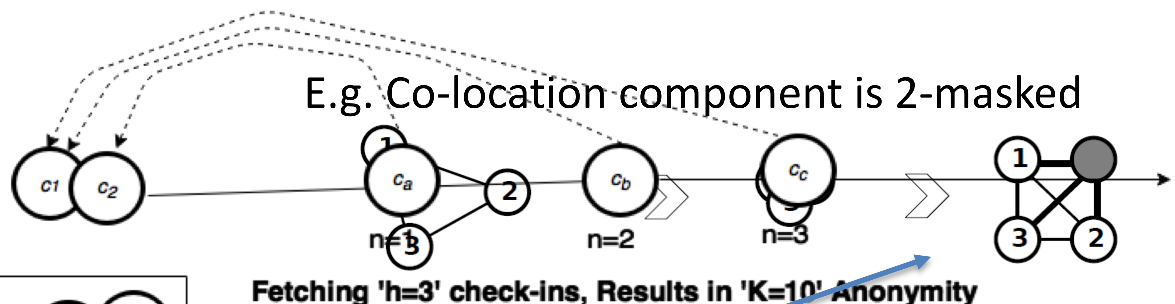
Definition: A co-location is b-masked if it is spatio-temporally indistinguishable to $b-1$ other co-locations.

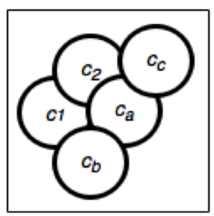
Method: For every co-location pair
Move an "h" number of closest check-ins to form a group.



The co-Location between Alice-Bob is now 3-masked.

E.g. Co-location component is 2-masked



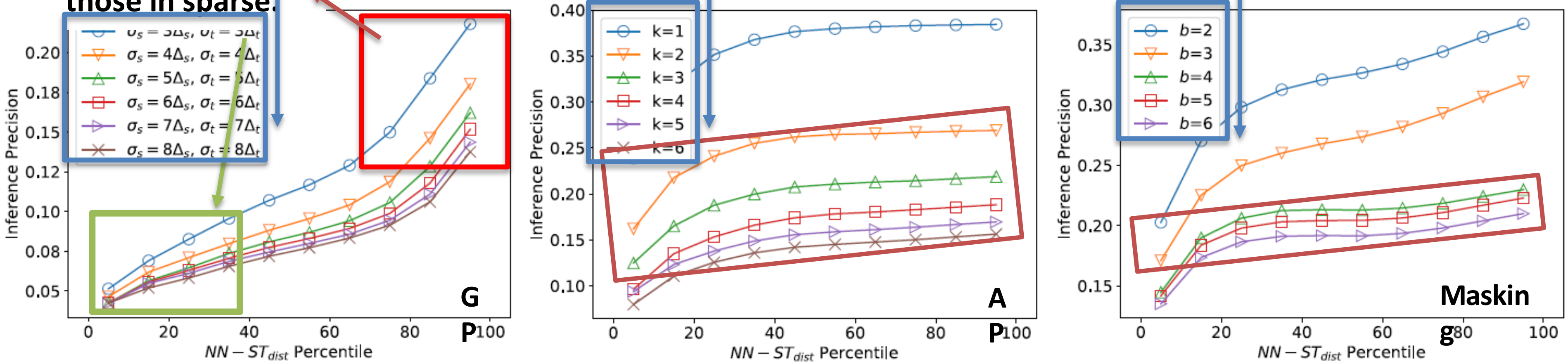


On seeing any co-location the adversary can only tell it's truthfulness with a chance of $3/6$ $(i.e.\ 1/b)$.

Attack Accuracy on Privacy Mechanisms

Over-protect dense at the exp those in sparse.

Increasing level of distortion.



Dense

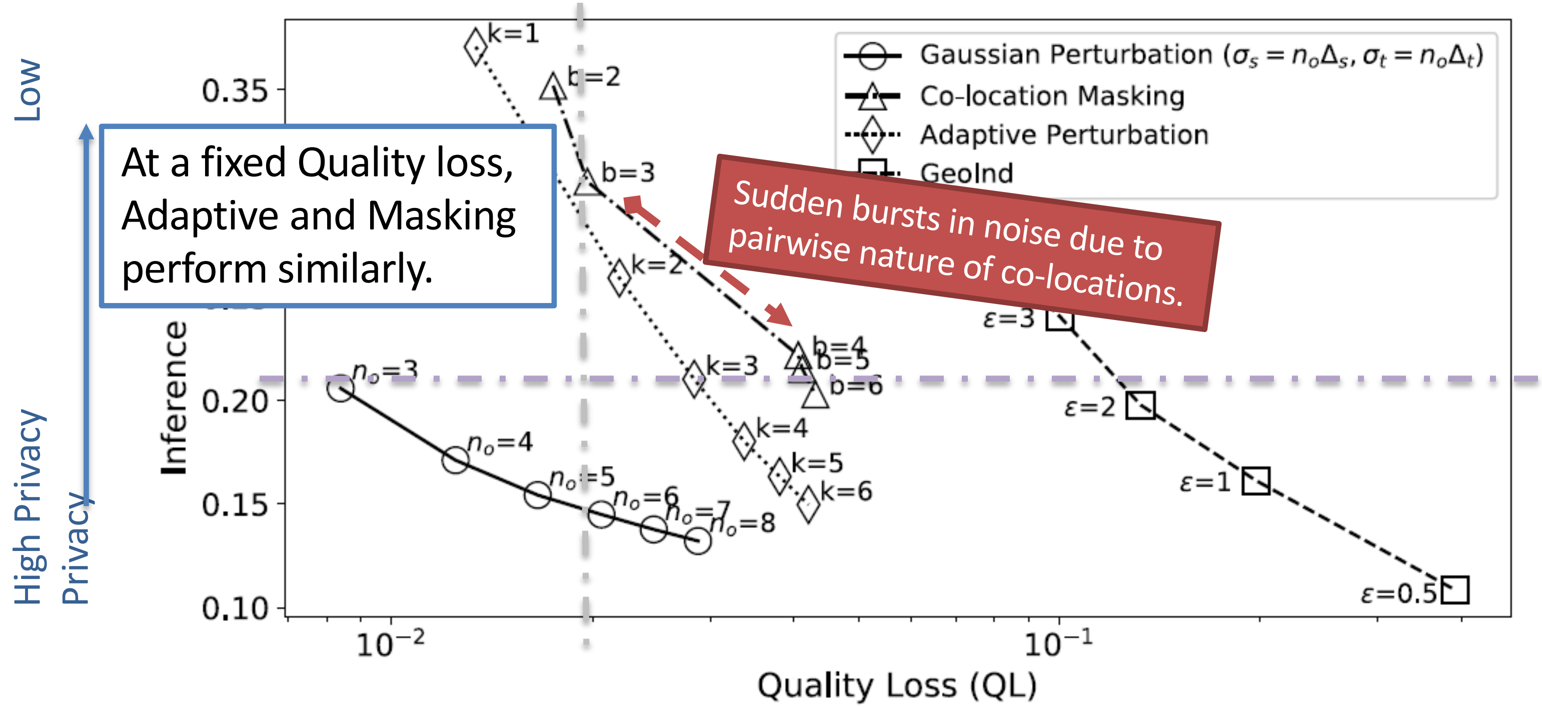
Sparse

Gaussian Perturbation exposes a significant portion of the population to highly accurate inferences.

Adaptive Perturbation and Masking provide consistent protection (i.e. with low variance) against an adversary.

Masking guarantees privacy according to definition.

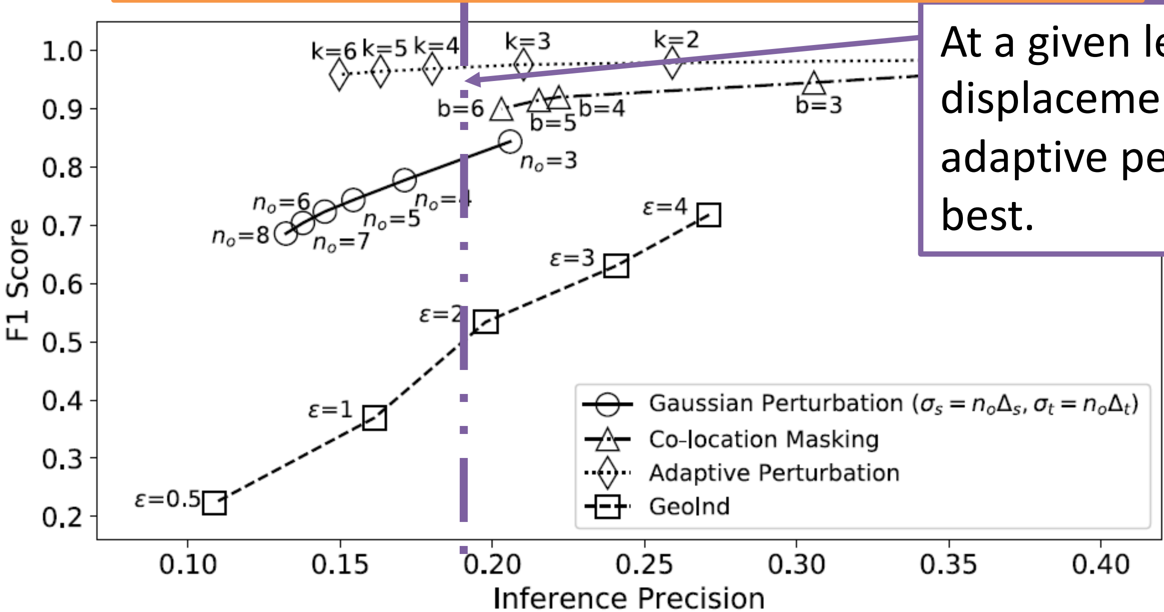
Analysis of Quality Loss



- ❖ Gaussian offers better average privacy but completely exposes those in sparse areas.
- ❖ Location privacy methods such as $\epsilon-GeoInd$ obliterate data utility.
- ❖ Co-location masking offers limited flexibility in calibrating noise.

Impact on Range Queries

Spatial range queries emulate real-world workload.

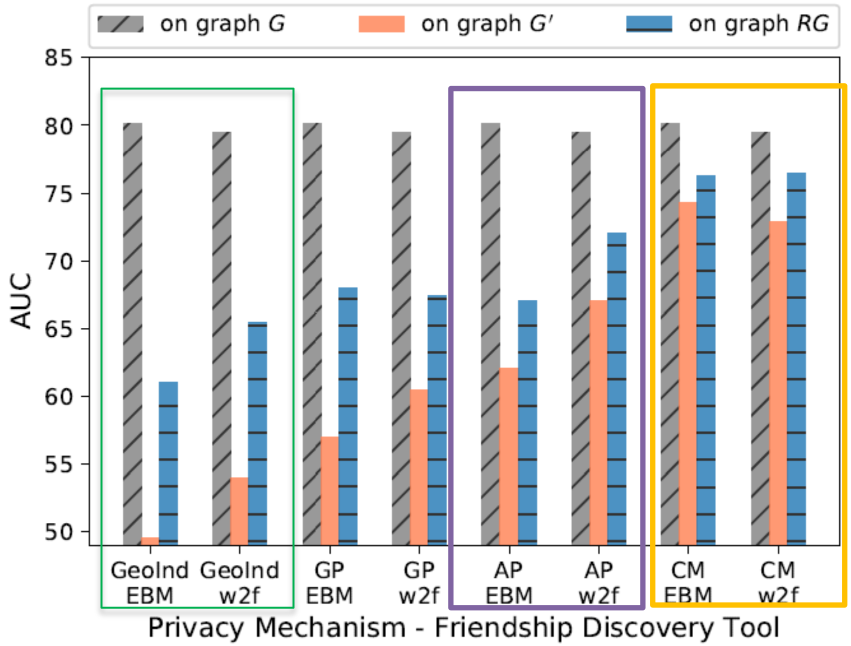


At a given level of total displacement, adaptive performs best.

❖ Adaptive Perturbation distorts to the NNs, hence is ideal for location-based advertising.

Evaluation of Friendship Discovery

Area Under the ROC Curve (AUC) ranges from [0.5, 1], where 0.5 is equivalent to random guessing, and 1 is perfect guessing.



the original graph G

obfuscated to G ‘

reconstructed graph RG

GeoInd is not effective in protecting against friendship discovery due to spatial-only noise.

W2f is less affected due to the random-walk processin building mobility features being more resilient to the nature of AP.

Masking leaves the underlying co-locations unperturbed. In longitudinal dataset, repeat co-location exposures reveal the friendship correlation.

Outline

Motivation: Geo-social Privacy

Prior Work: Inferring Social Behaviors

Current Efforts: Protecting against social inferences

- But allow location disclosure

Open Problem: Protecting against location disclosure

- But allow social inferences

Two Sides of the Coin

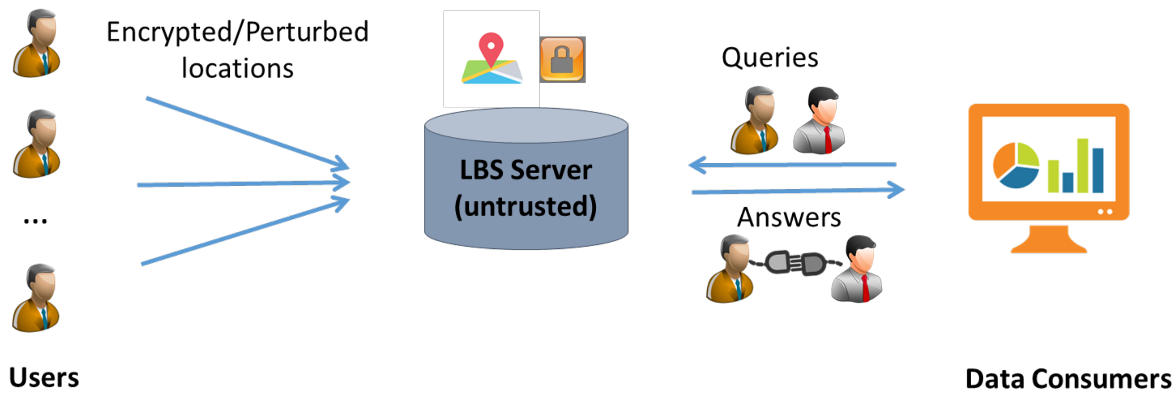
Protecting against location disclosure
**** But allow for Social Inference***

Privacy-Preserving Social Inferecne

Criminology
- identify the new or unknown members of a criminal gang or a terrorist cell

Epidemiology
- spread of diseases through human contacts

Policy
- induce local influence in electing a tribal representative



Backup

Challenges

1. How to quantify the protection against social inferences?

2. A privacy mechanism may result in

insufficient protection

OR

over-protection

at the cost of utility if only social inferences need to be protected.

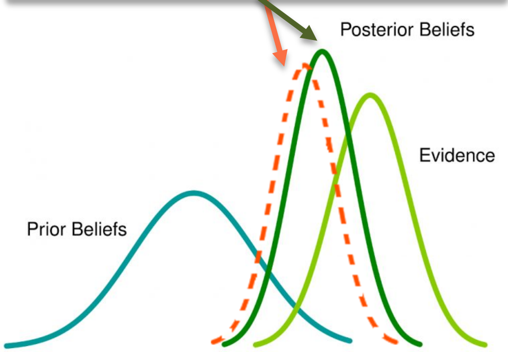
3. How to account for the background knowledge of a potential adversary ?

Modelling the Adversary

Objective: A conservative estimate of co-location privacy of users after adding noise.

1. After the adversary obtains the published noisy data. **(Evidence)**
2. Assume the privacy mechanism is known to the adversary. **(Evidence)**
3. Supply the adversary with background knowledge on **(Prior)**
   - The mobility patterns of users. (e.g. frequented locations)
   - The co-location patterns of users. (e.g. frequented co-locating partners)
4. Execute Bayesian Inference to reconstruct as accurate as possible representation of the original graph and co-locations. **(Posterior)**

The inferred posterior may still not be the true co-location distribution.



Inference Attack

1. Disciplined in the Bayesian technique of reasoning about privacy.
   i. Obtain the posterior distribution over all possible co-locations of a user's check-in.
   ii. Move the check-in to its most probable co-location.

2. Privacy is defined as the error in the adversary's inference attack.

$$\text{Inference Accuracy}(IA) = \frac{CL \cap RCL}{RCL}$$

CL: Original set of co-locations. RCL: Reconstructed set of co-locations.

3. Utility of the privacy mechanism = the total noise added to the original data.
   For a single check-in :

$$\text{Service Quality Loss } SQL_u^i = \alpha \cdot \underbrace{\frac{||c_u^i.l, \overbrace{c_u^i.l'}^{\text{Perturbed location co-ordinate}}||}{MAX_S}}_{\text{Spatial Displacement}} + (1-\alpha) \cdot \underbrace{\frac{|c_u^i.t, \overbrace{c_u^i.t'}^{\text{Perturbed timestamp}}|}{MAX_T}}_{\text{Temporal Distortion}}$$

c_u^i: i^{th} check-in of user u $MAX_s, \mathrm{MAX_T}$: normalizing constants. α : weighting factor

References

- [KDD’03] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '03).
- [PNAS’10] David J. Crandall, Lars Backstrom, Dan Cosley, Siddharth Suri, Daniel Huttenlocher, and Jon Kleinberg. “Inferring social ties from geographic coincidences.” PNAS 2010
- [KDD’10] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. 2010. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '10).
- [Ubi’10] Justin Cranshaw, Eran Toch, Jason Hong, Aniket Kittur, Norman Sadeh.“Bridging the Gap Between Physical Location and Online Social Networks.” Ubicom 2010
- [VLDB’12] H. Shirani-Mehr, F. Banaei-Kashani, and C. Shahabi. “Efficient reachability query evaluation in large spatiotemporal contact datasets.” Proc. VLDB Endow, 5(9), May 2012
- [SIGMOD’13] Pham, Huy, Cyrus Shahabi, and Yan Liu. "Ebm: an entropy-based model to infer social strength from spatiotemporal data." Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data. ACM, 2013
- [Nature’13] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel. “Unique in the Crowd: The privacy bounds of human mobility.” Scientic Reports, 3, Mar.2013
- [ICDE-Bulletin’15] Cyrus Shahabi, Huy Pham: Inferring Real-World Relationships from Spatiotemporal Data. IEEE Data Eng. Bull. 38(2): 14-26 (2015)
- [CNS’15] B. Wang, M. Li, H. Wang, and H. Li. “Circular range search on encrypted spatial data.” In IEEE CNS, 2015
- [CCS’15] Y. Xiao and L. Xiong. “Protecting locations with differential privacy under temporal correlations.” In CCS, 2015
- [PerCom’07] Krumm, John. "Inference attacks on location tracks." *Pervasive computing* (2007): 127-143.
- [ACM TOPS’17] Argyros, George, et al. "Evaluating the Privacy Guarantees of Location Proximity Services." *ACM Transactions on Privacy and Security (TOPS)* 19.4 (2017): 12.
- [RSLB 2003] K. T. Eames and M. J. Keeling, “Contact tracing and disease control,” *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 270, no. 1533, pp. 2565–2571, 2003.

Reference (other)

- Chen et al., "Scalable influence maximization for prevalent viral marketing in large-scale social networks", ACM SIGKDD, 2010
- Goyal et al., "A data-based approach to social influence maximization", VLDB, 2011
- Kempe et al., “Maximizing the spread of influence through a social network”, ACM SIGKDD, 2003
- Leskovec et al., “Cost-effective outbreak detection in networks”, ACM SIGKDD, 2007
- Saito et al., "Prediction of information diffusion probabilities for independent cascade model" Knowledge-Based Intelligent Information and Engineering Systems, Springer, 2008.
- Goyal et al., “Learning influence probabilities in social networks” ACM WSDM, 2010
- Cho et al., “Friendship and mobility: user movement in location-based social networks”, ACM SIGKDD, 2011
- Zhang et al., “Understanding spatial homophily: the case of peer influence and social selection”, WWW, 2014
- Cranshaw et al., “Bridging the gap between physical location and online social networks”, ACM UbiComp, 2010
- Liben-Nowell et al., “The link-prediction problem for social networks”, Journal of the ASIST, 2007
- Pham et al., "Ebm: an entropy-based model to infer social strength from spatiotemporal data", ACM SIGMOD, 2013