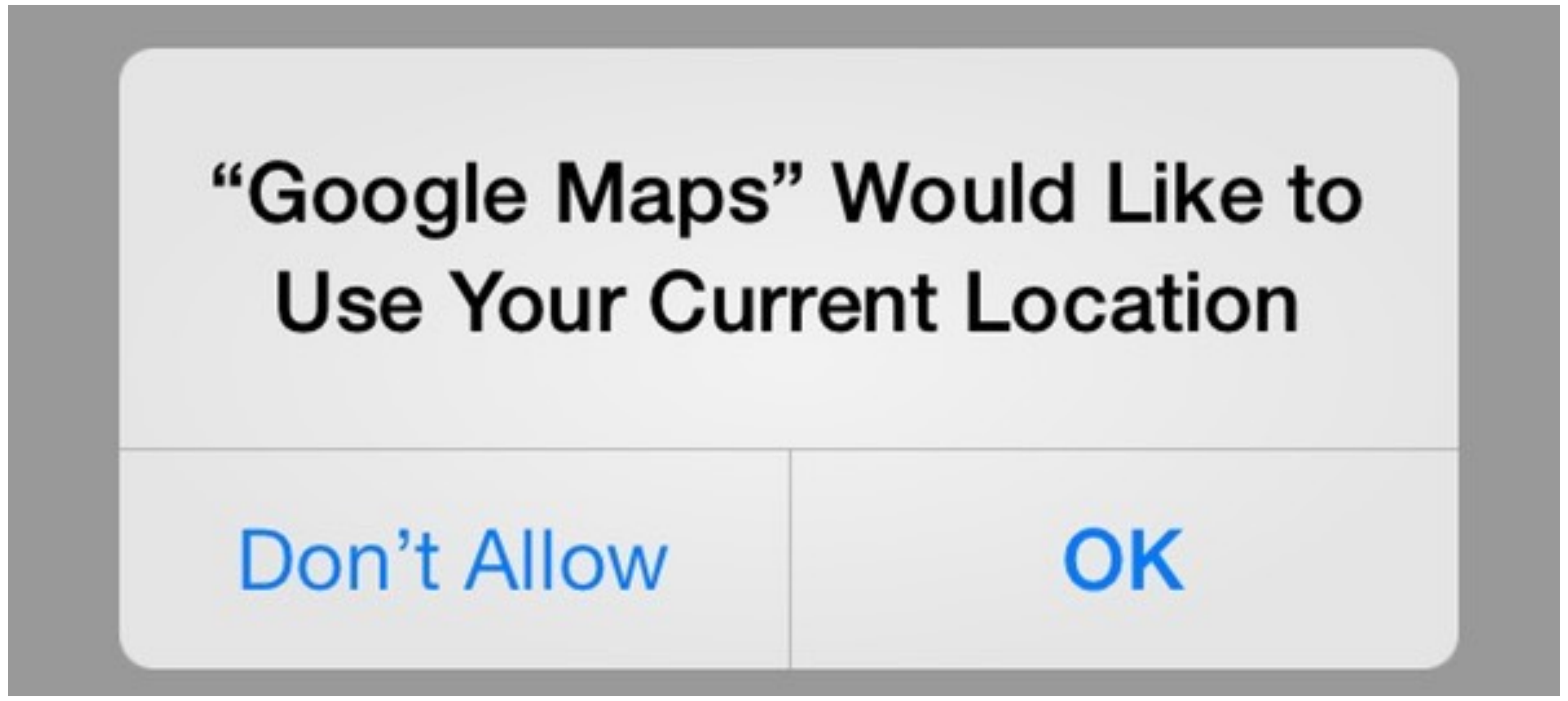
walk2friends: Inferring Social Links from Mobility Profiles

Yang Zhang

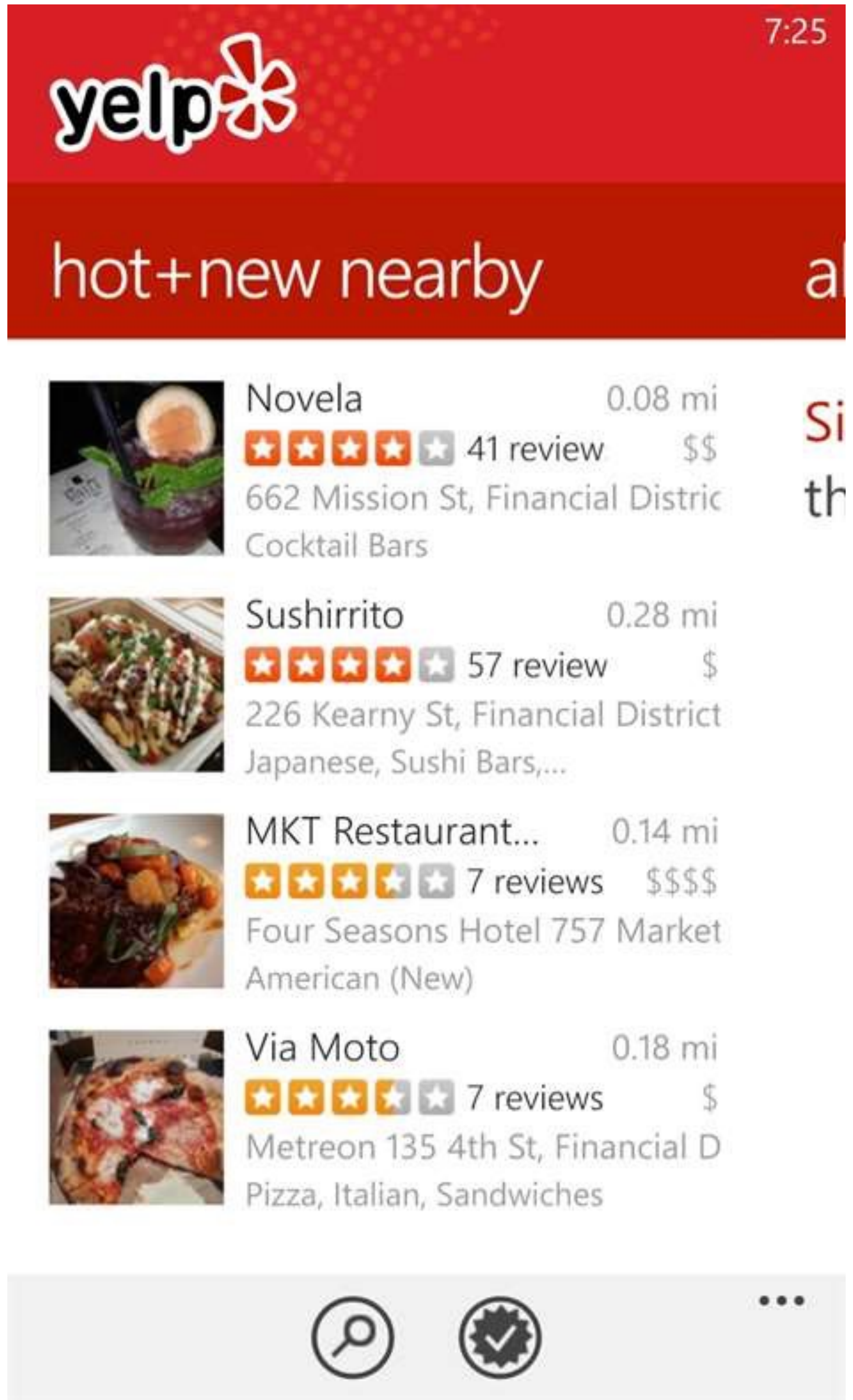
joint work with Michael Backes, Mathias Humbert, and Jun Pang

CISPA

Center for IT-Security, Privacy and Accountability


"Google Maps" Would Like to Use Your Current Location
Don't Allow
OK


Tell your friends you're here
facebook
facebook.com/your-Page-here


7:25
yelp
hot+new nearby
Novela 0.08 mi
41 review $$
662 Mission St, Financial Distric
Cocktail Bars
Sushirrito 0.28 mi
57 review $
226 Kearny St, Financial District
Japanese, Sushi Bars,...
MKT Restaurant... 0.14 mi
7 reviews $$$$
Four Seasons Hotel 757 Market
American (New)
Via Moto 0.18 mi
7 reviews $
Metreon 135 4th St, Financial D
Pizza, Italian, Sandwiches

Location Privacy

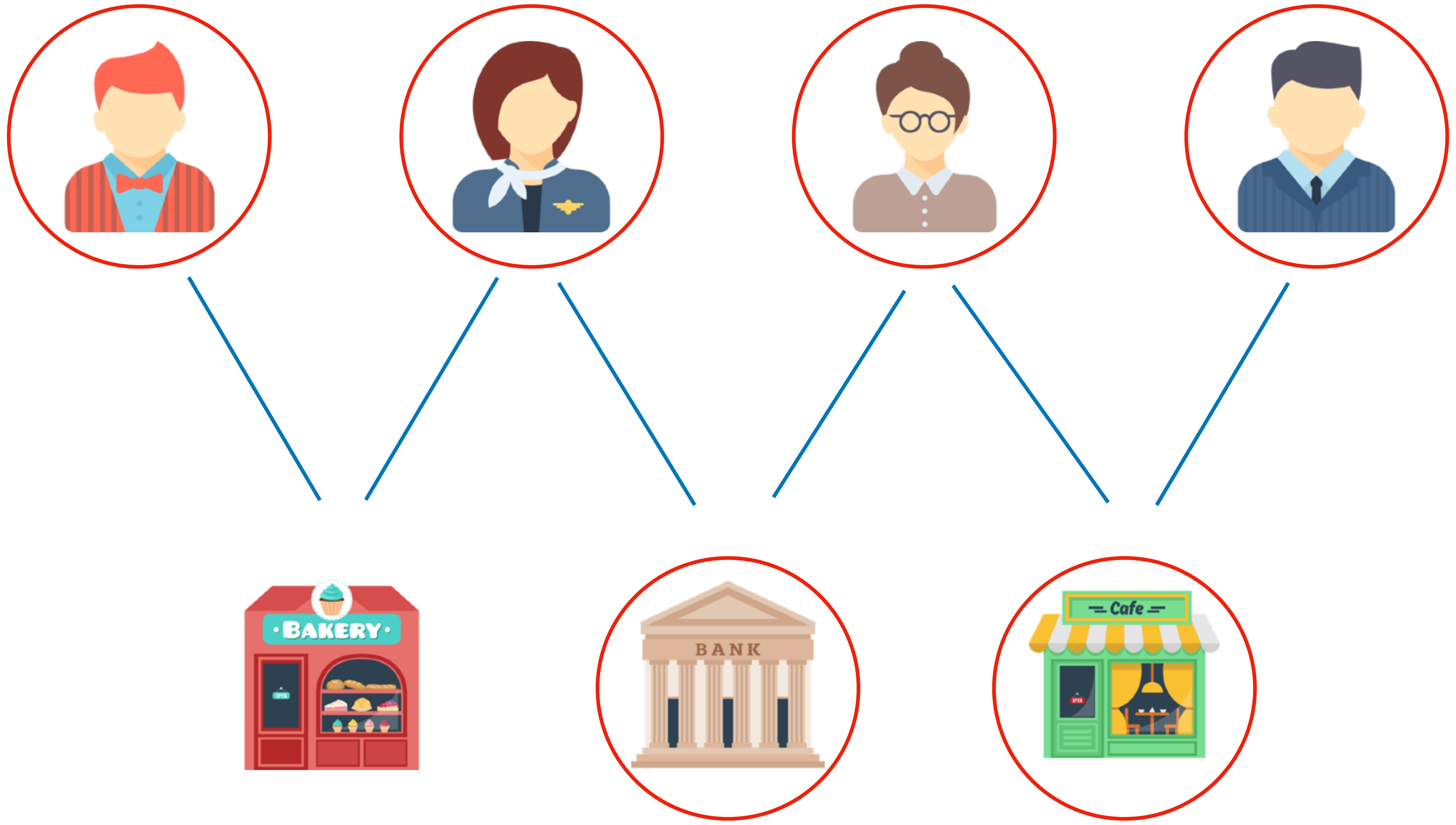
- 4 spatial-temporal points can identify 95% of the individuals
- Mobility traces can be effectively de-anonymized
- You are where you go
  - Demographics
  - Social relations

Social Relation Privacy

- Social relations can be sensitive, e.g., office romance
- 17.2% -> 56.2% (Facebook users in New York)
- NSA’s co-traveler program

Predict whether two users are friends based on the locations they have visited

- Solution 1: common locations two users have visited
  - Almost all data mining approaches take this way
    - Location entropy
  - Can’t work when two users share no common locations



- Solution 2: mobility profiles/features
  - Summarize each user's mobility profiles
  - Friends share similar mobility profiles than strangers
  - Feature engineering
    - Tedious efforts and domain expert knowledge
    - Time consuming

Every Single Time!!!

Representation Learning

- Learning features (representation/deep learning)
  - Follow a general object (unsupervised)
- Graph representation learning (graph embedding)
  - Preserve each user’s neighbors in a social network
- Mobility feature learning

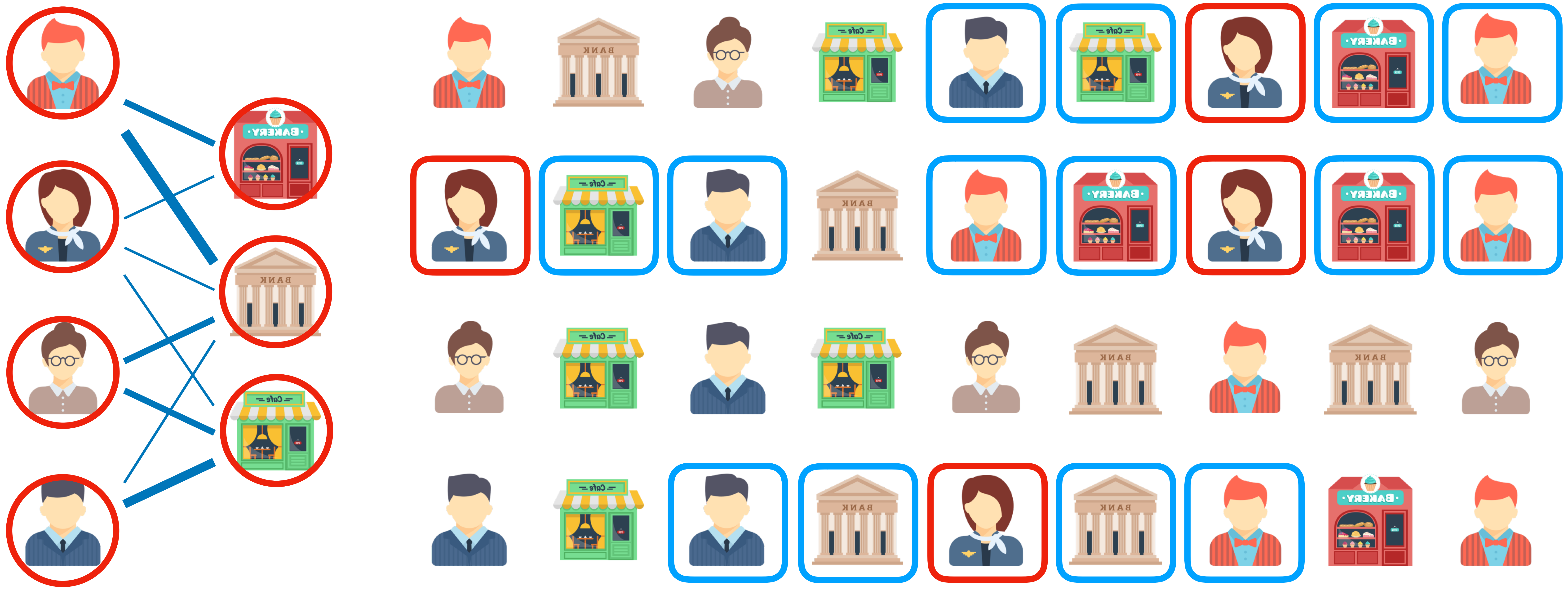
Assumption: A user's mobility neighbors can reflect his mobility profile/features

- Define each user's mobility neighbors
- Learn mobility features/profiles
- Infer two users' social relation

Mobility Neighbors

- A user's mobility neighbors include
  - Locations a user has visited
  - Others who have visited similar locations and their locations
- Breadth first search
  - Not considering the visiting frequencies
- Random walk sampling

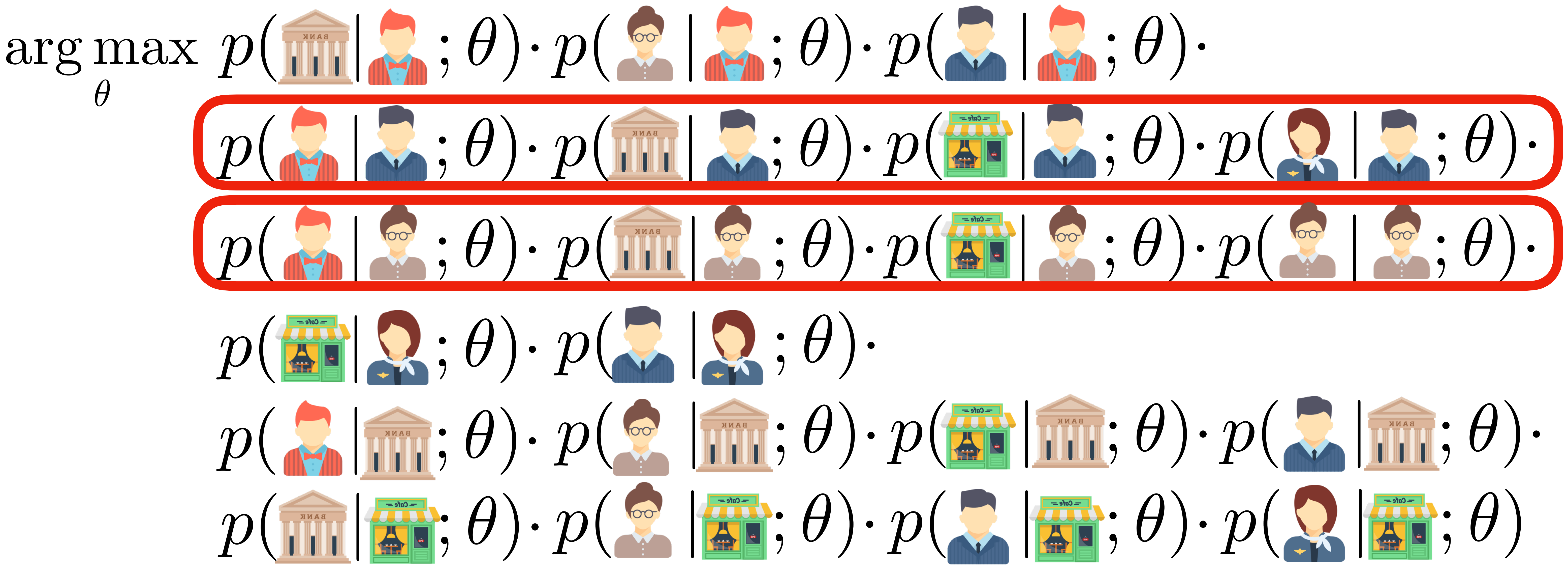
Mobility Neighbors



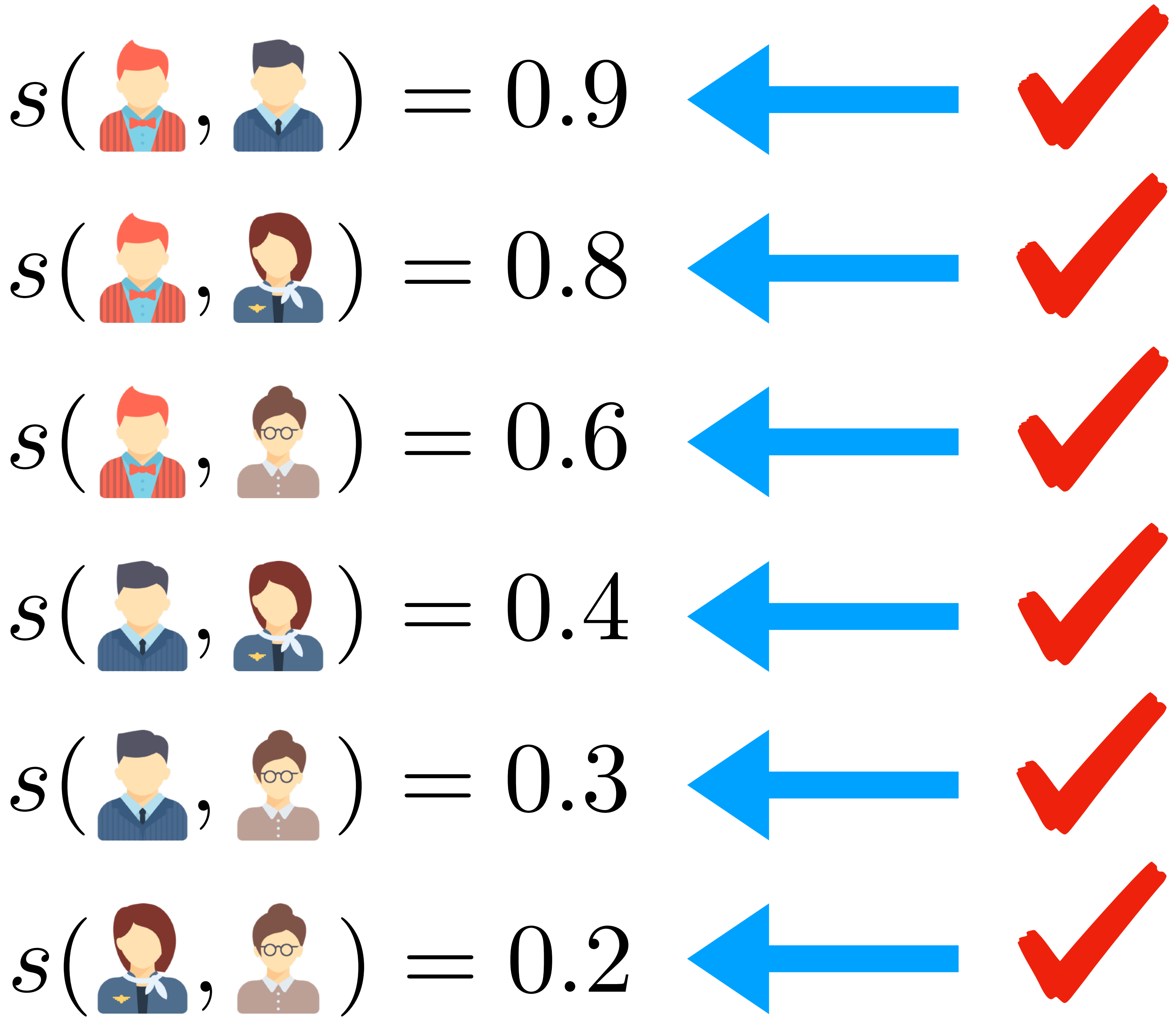
CISPA Center for IT-Security, Privacy and Accountability

Feature Learning

- Learn a function: $\theta : \mathcal{U} \rightarrow \mathbb{R}^d$
- Each node to predict it's neighbors
- $p(\cdot \mid \cdot ; \theta)$ Softmax



Social Relation Inference



- Cosine similarity
- Unsupervised
- Predict any social relation

CISPA
Center for IT-Security, Privacy and Accountability

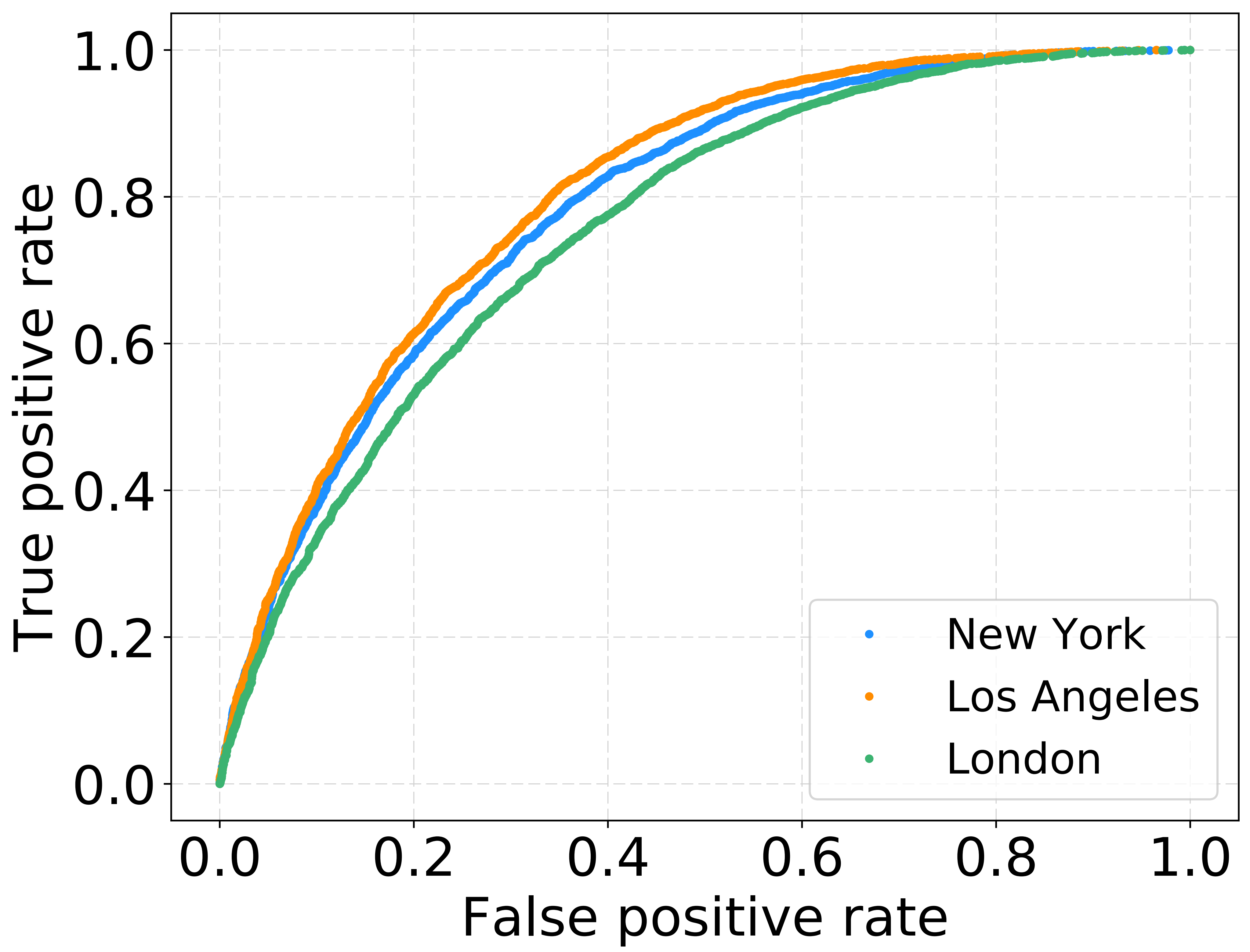
Evaluation: dataset

- Instagram users' check-ins
    - New York, Los Angeles and London
- Foursquare (location semantics)
- Social relations (two users follow each other)

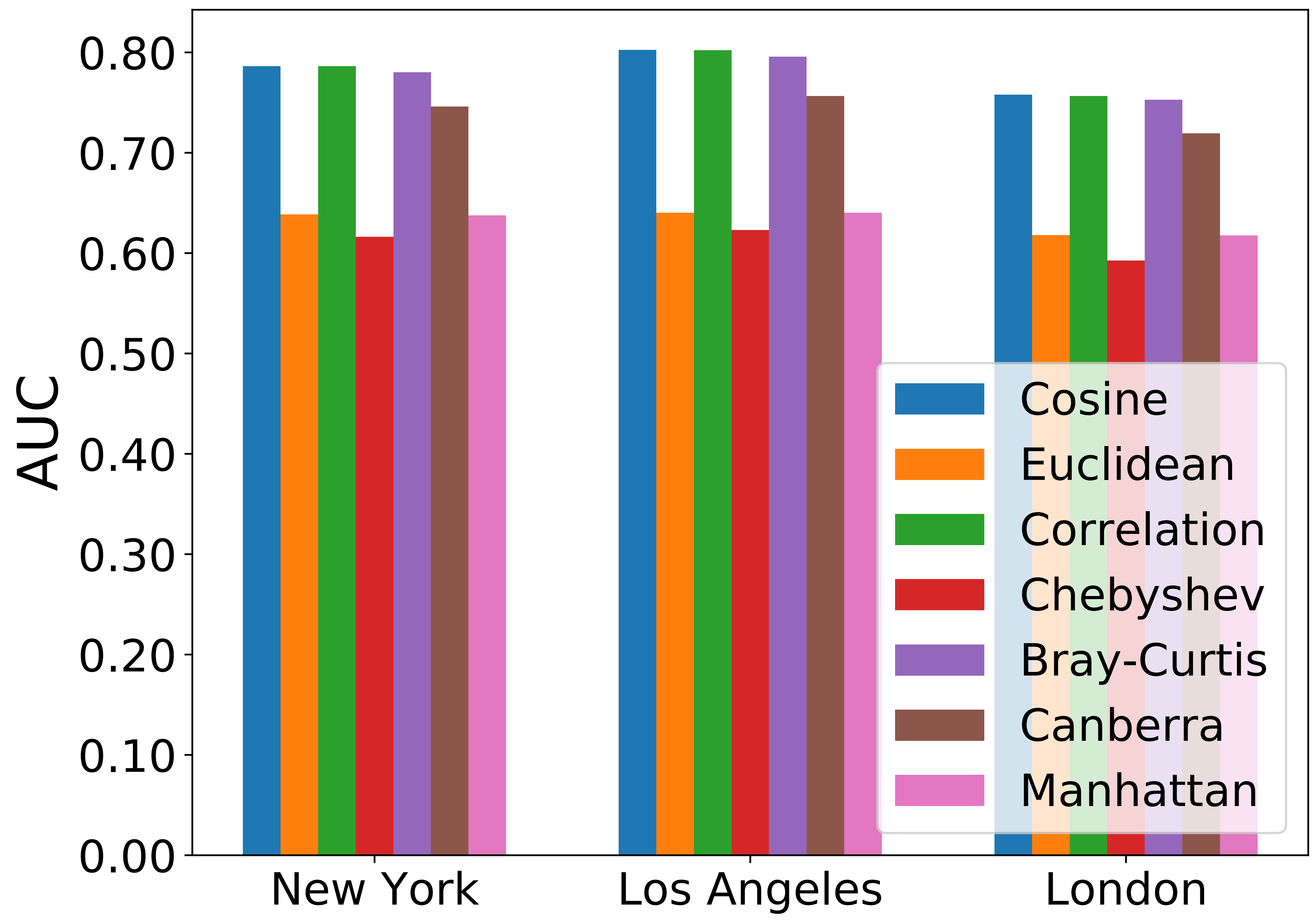
	New York	Los Angeles	London
No. check-ins	1,843,187	1,301,991	500,776
No. locations	25,868	22,260	10,693
No. users	44,371	30,679	13,187
No. social links	193,995	129,004	25,413

Center for IT-Security, Privacy and Accountability

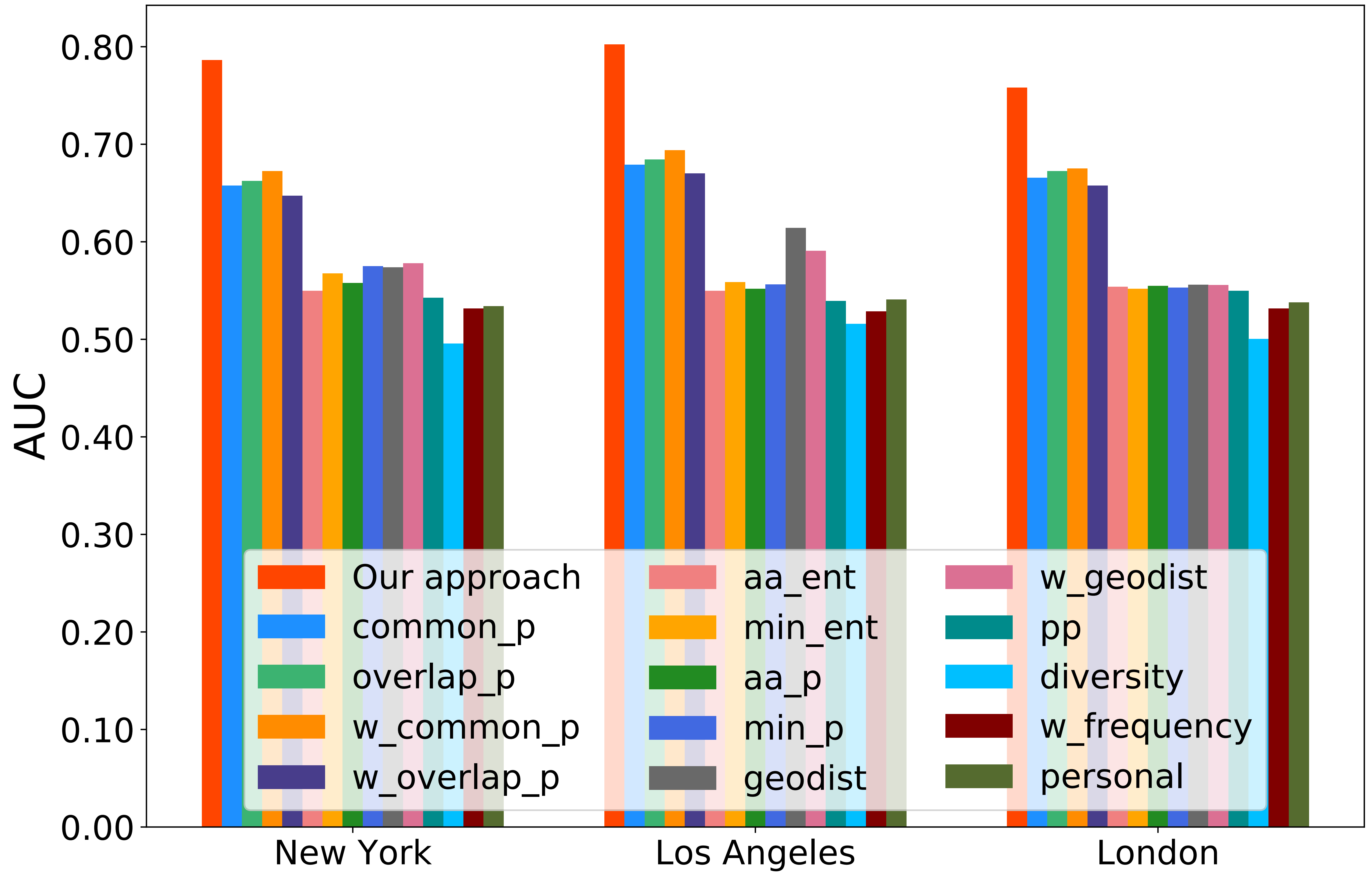
Evaluation: ROC curve



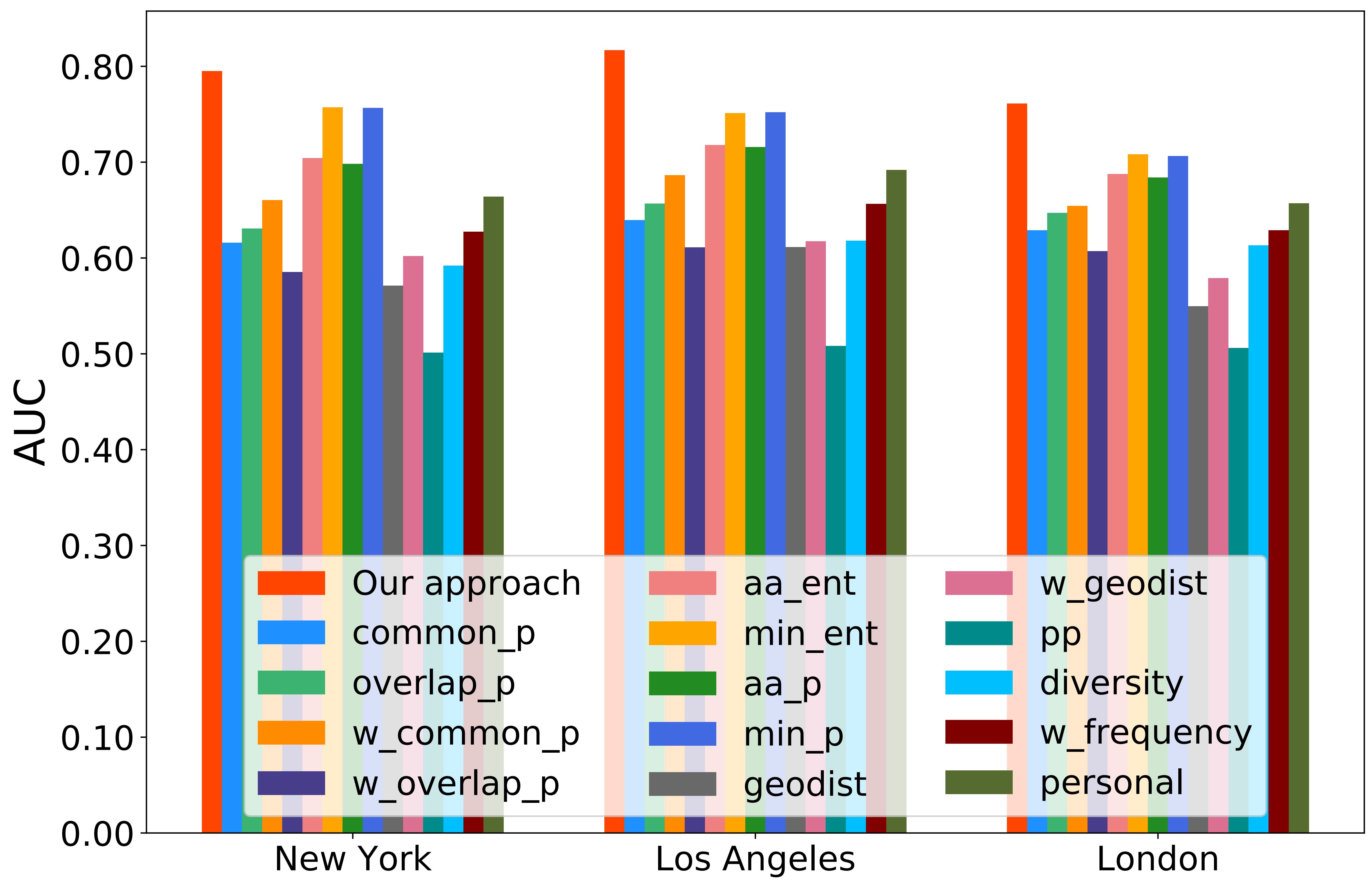
Evaluation: distance metric



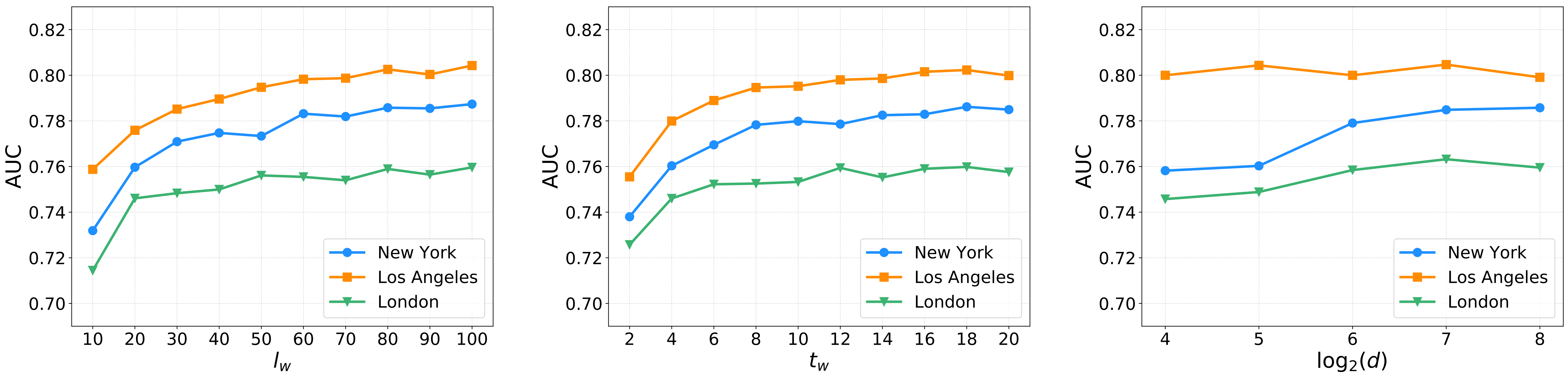
Evaluation: baseline models



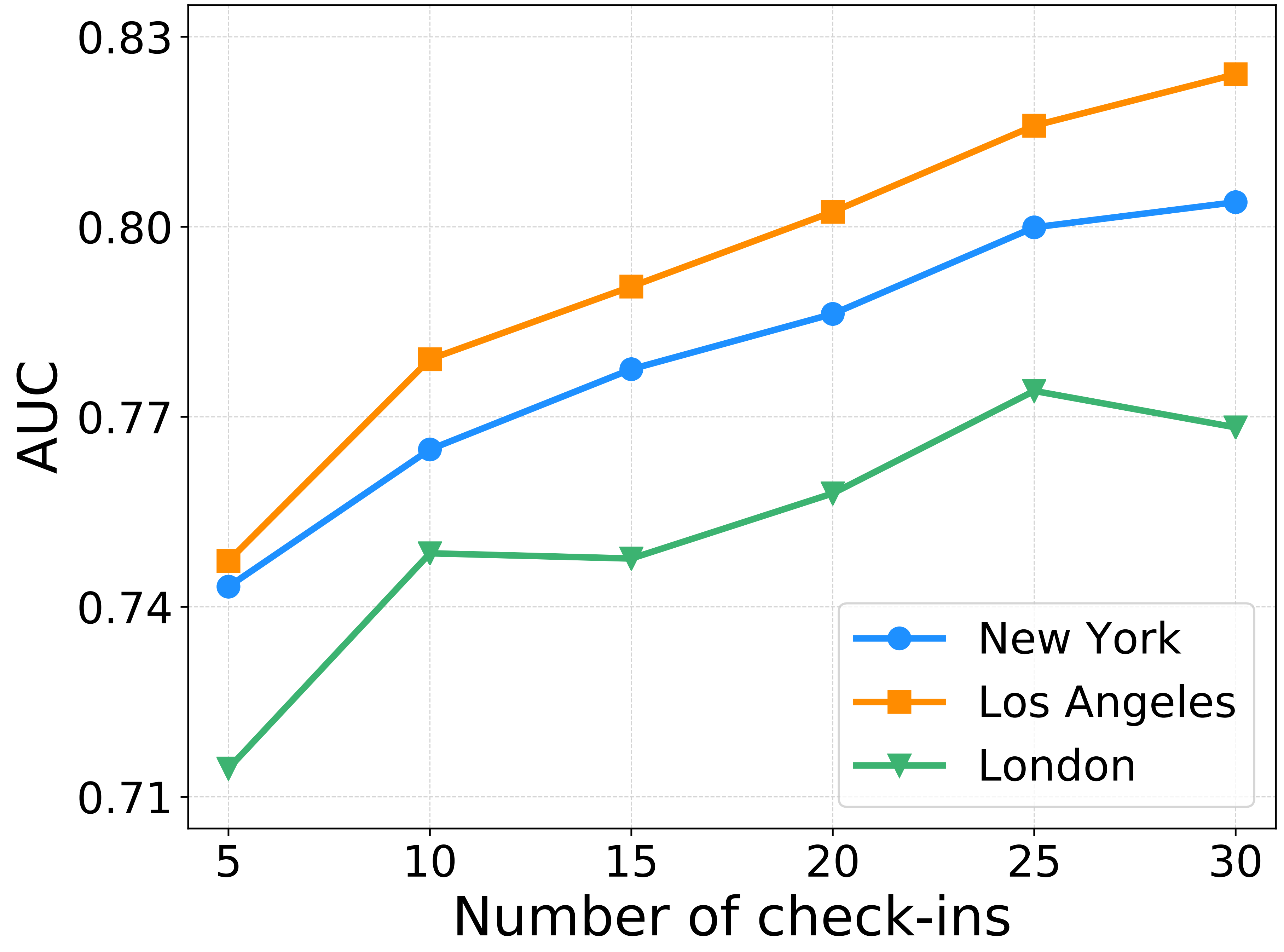
Evaluation: baseline models



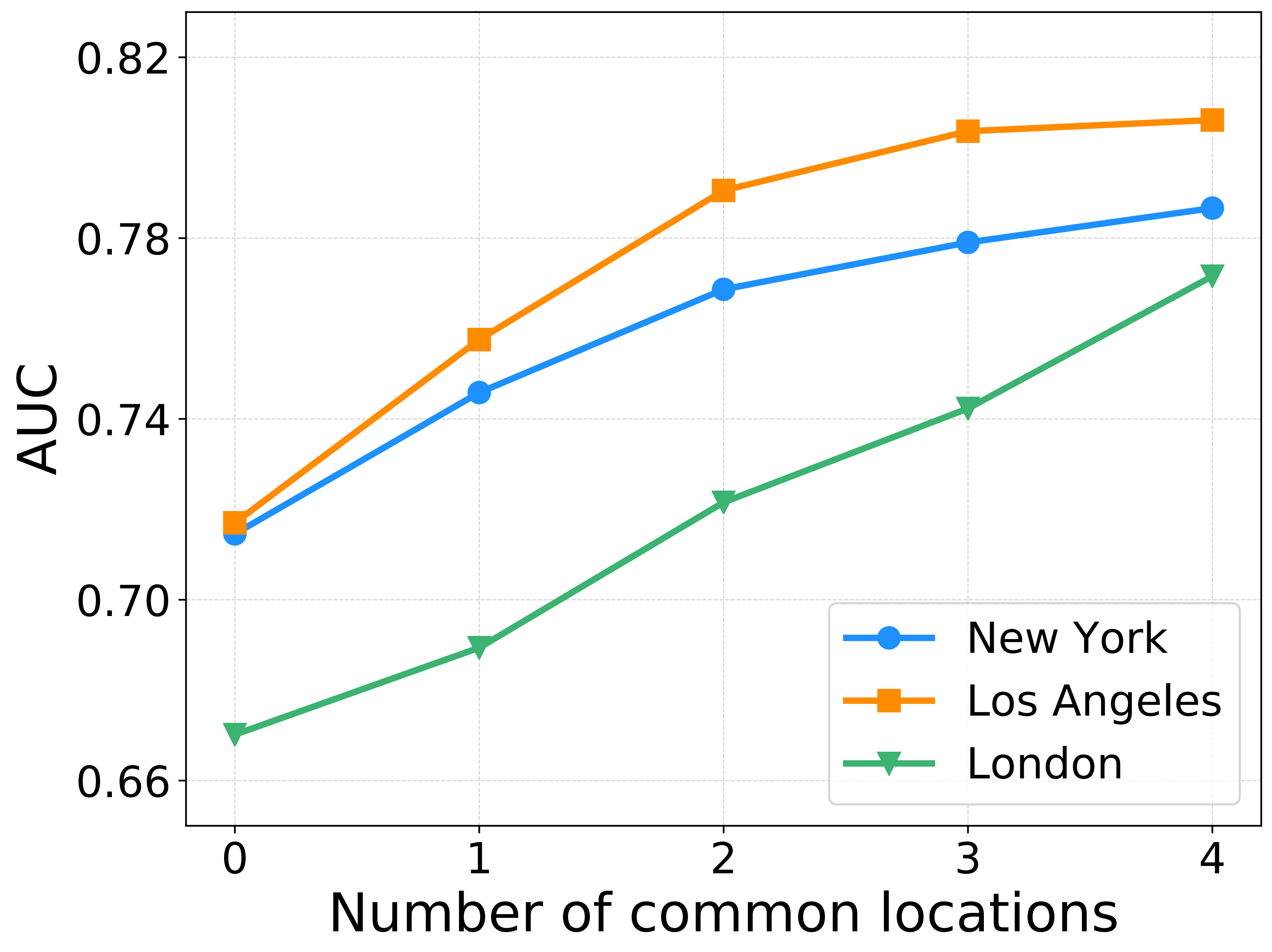
Evaluation: hyperparameters



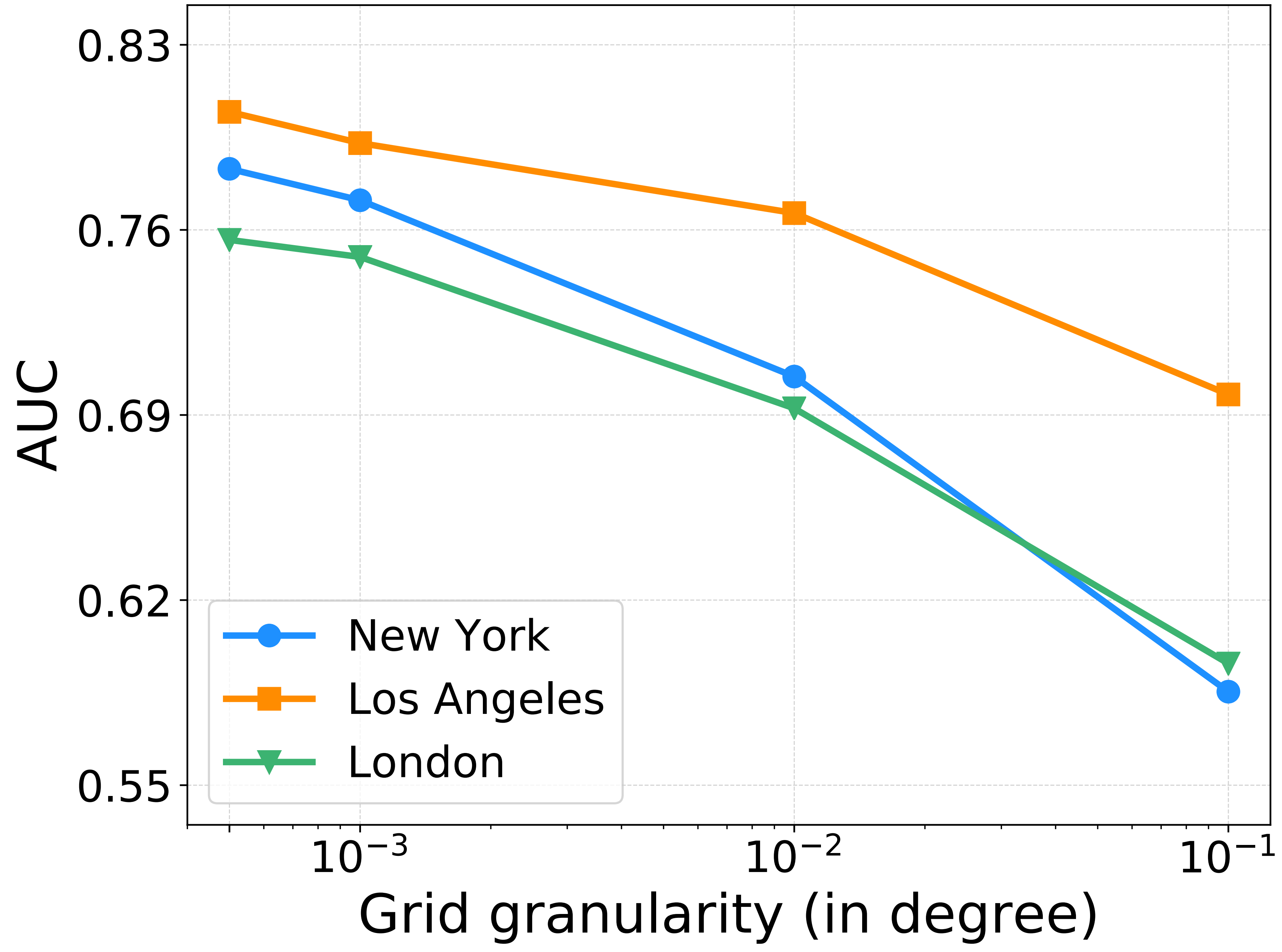
Evaluation: check-in numbers



Evaluation: common locations



Evaluation: geo-coordinates



Defense Mechanisms

- Hiding
    - Delete certain proportion of check-ins
- Replacement
    - Random walk to replace locations

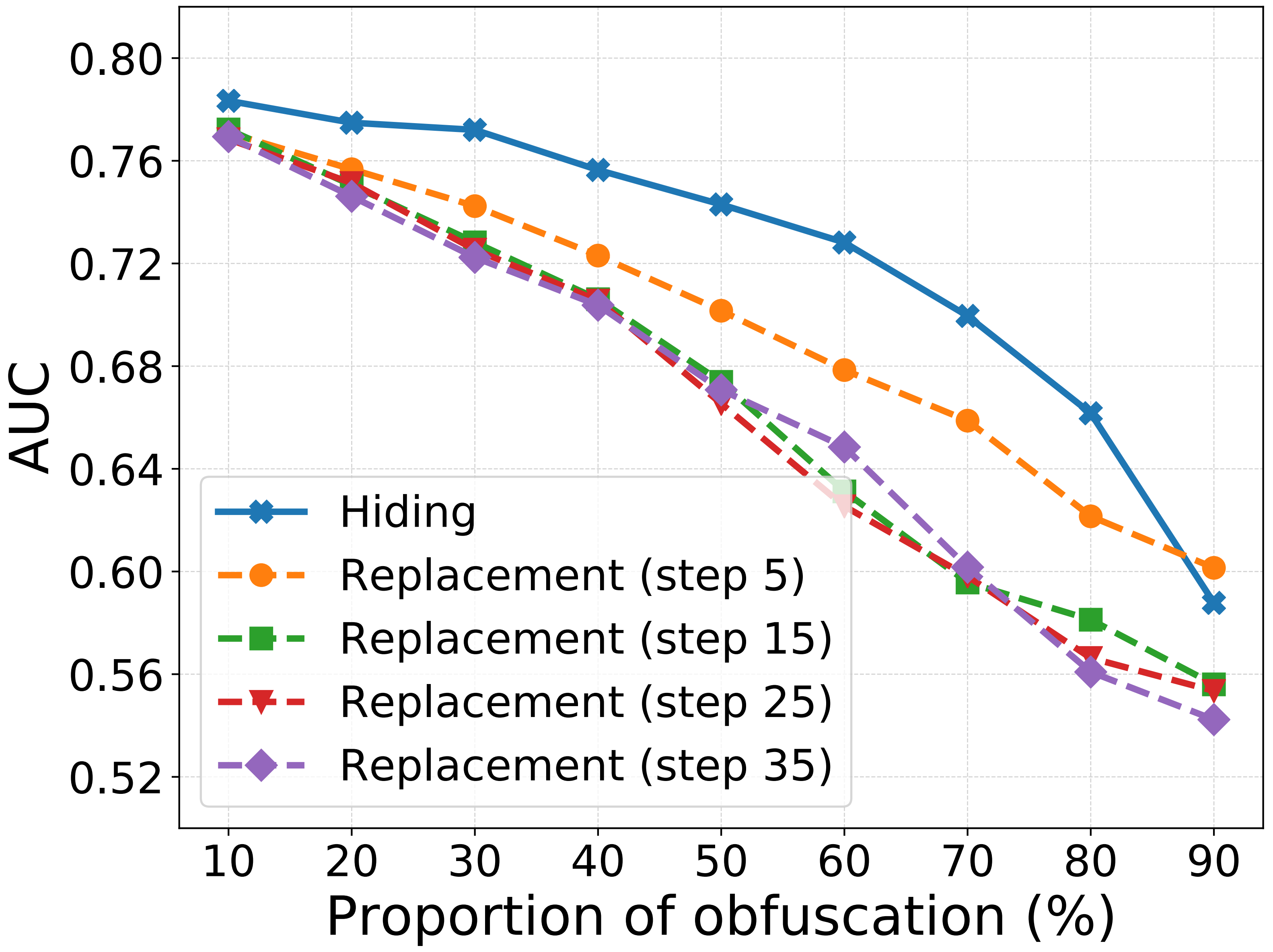
Defense Mechanisms

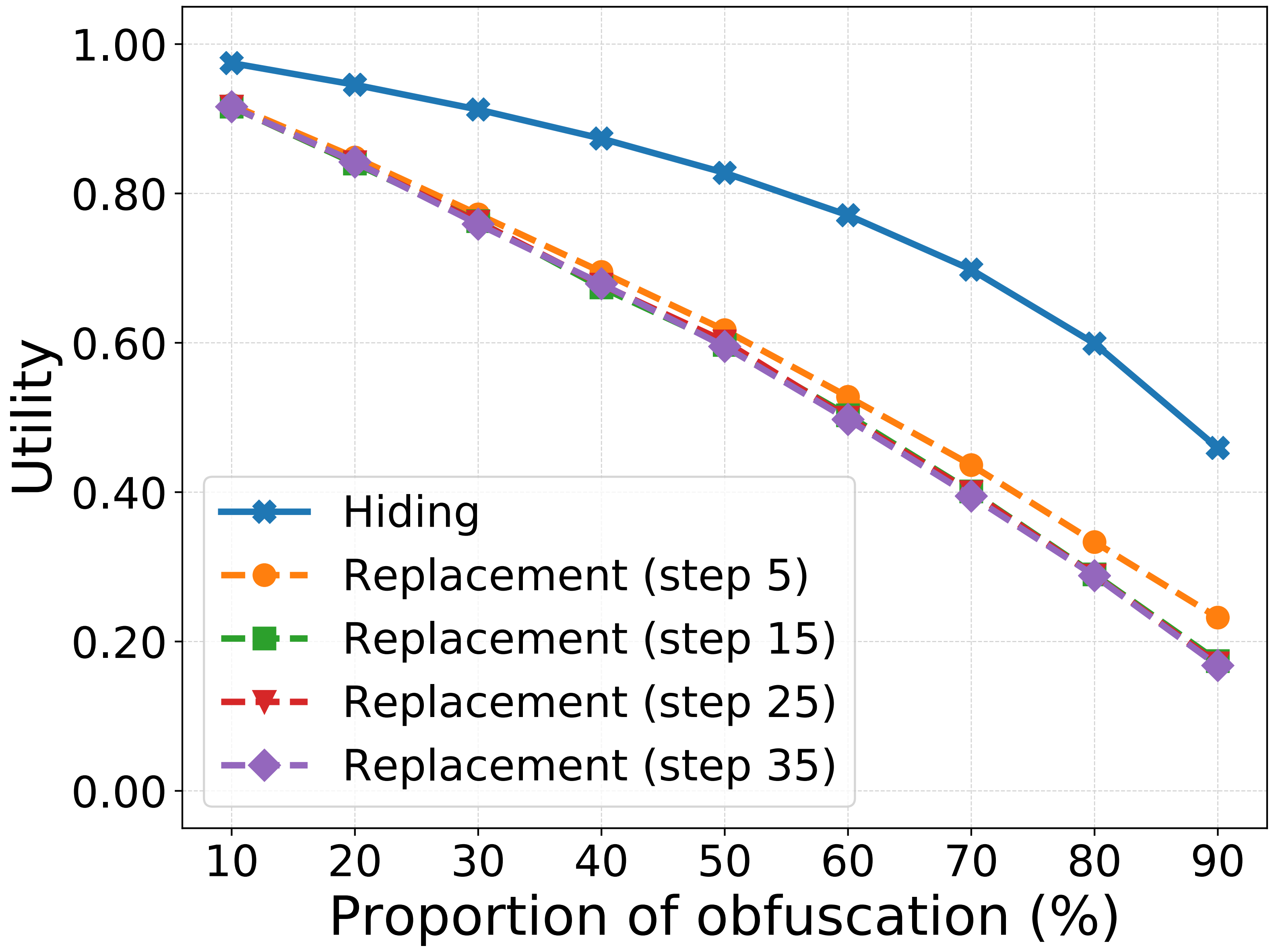
- Generalization
  - Geo-coordinate and location semantics
    - MoMA -> art (40.76N, -73.97W)
  - Recover location first
    - art (40.76N, -73.97W) -> MoMA or Tom Otterness Frog?

Utility Metric

- Each user's check-in distribution
    - Both original and obfuscated
- Jensen-Shannon divergence
- Average over all users

Defense Evaluation

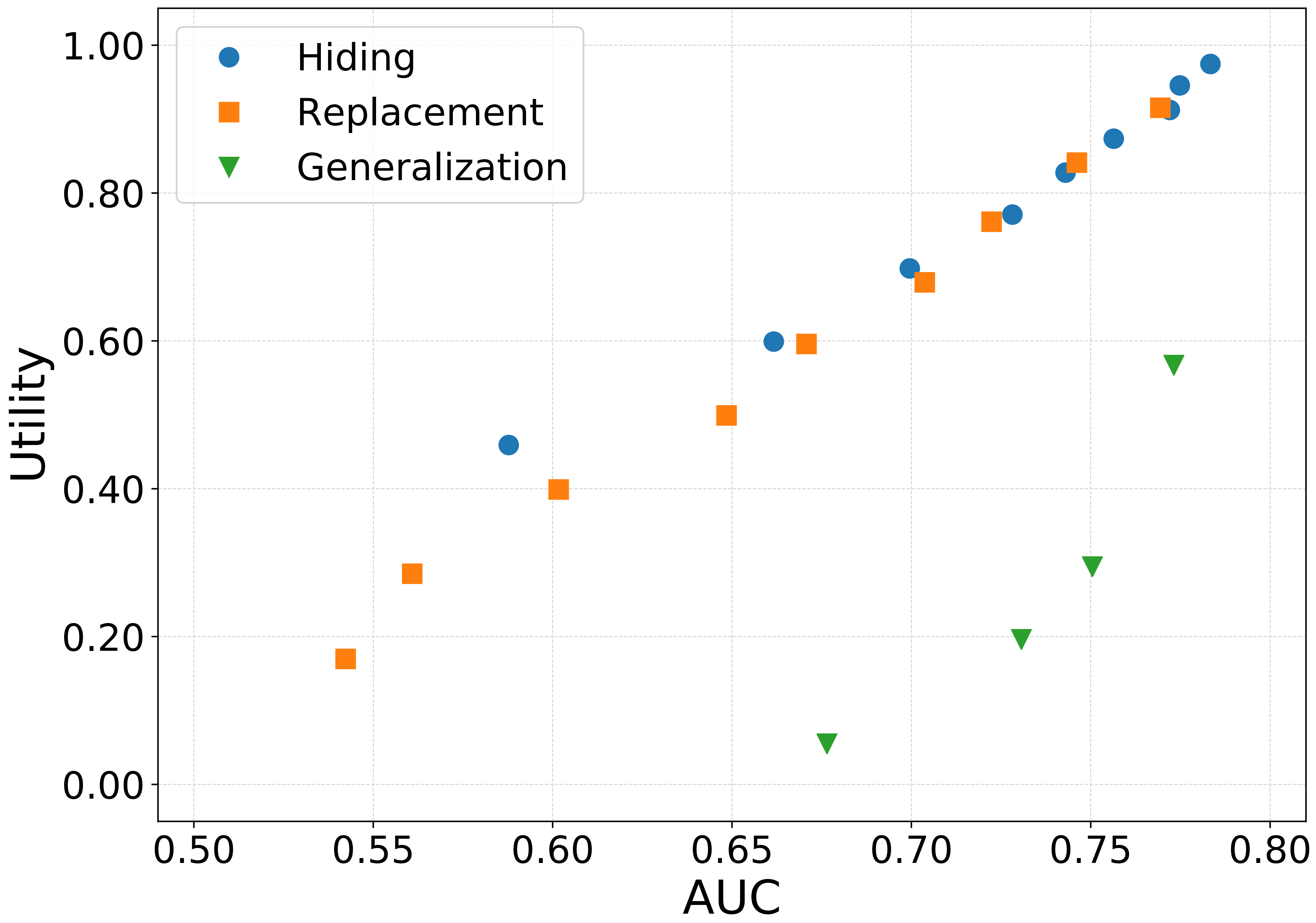




Defense Evaluation

	AUC		Utility		Recovery rate	
	ls	*hs*	*ls*	*hs*	*ls*	*hs*
lg	0.77	0.75	0.57	0.30	52%	23%
hg	0.73	0.67	0.20	0.06	14%	2%

Defense Evaluation



yang.zhang@cispa.saarland
@yangzhangalmo

Conclusion

Center for IT-Security, Privacy and Accountability

- A new social relation inference attack with mobility profiles
  - Learning user profiles
  - Unsupervised and predict any social relations
- Three general defense mechanisms
  - Replacement and hiding outperform generalization