Uniqueness in the City: Urban Morphology and Location Privacy

Hancheng Cao[1,2], Jie Feng[2], Yong Li[2], Vassilis Kostakos[3]

Stanford Univ.[1], Tsinghua Univ.[2], Univ. of Melbourne[3]

Oct. 11th, 2018

http://hanchengcao.me/

http://fi.ee.tsinghua.edu.cn/~liyong/

https://people.eng.unimelb.edu.au/vkostakos/

In this paper, we investigate a deceivingly simple question…

By revealing the types of nearby Points-of-Interests (POIs), do we reveal our actual location?

"Hey Mike! Near me there are 2 restaurants, 3 schools and a hospital. Guess where I am? "

"That's easy! you must be at …"

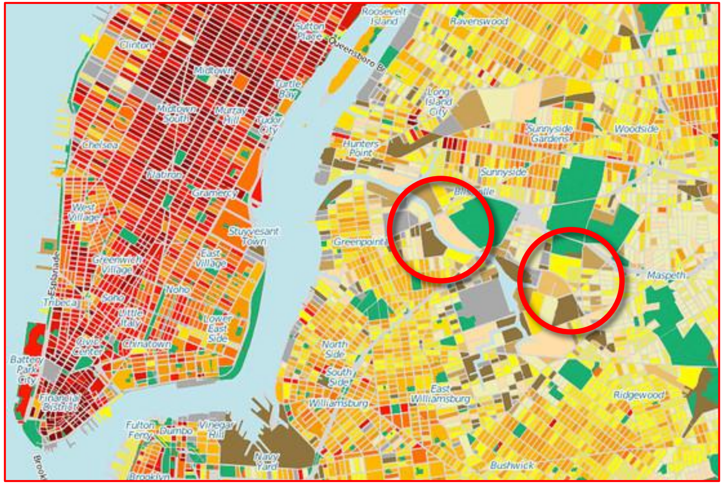
Motivation 1

Better understand location privacy and provide insights for designing privacy-aware location-based services

- Ubiquity of Location-based services: recommendation system, social media (e.g., Twitter, Facebook), appstore, etc.
- Status Quo: directly sharing user GPS information lead to privacy leak. Methods proposed using nearby POI types instead for recommendation (e.g. app/website promotion) to ensure user privacy (e.g., Yu et al.).
- Our work: first work to reveal that merely using nearby POI types is **NOT SAFE** enough!

Motivation 2

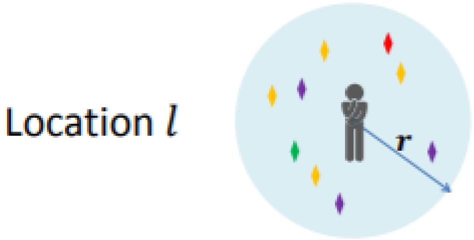
Better understand urban morphology in cities around the globe



- Are there two areas in a city with the same POI compositions?
- Our work: **first work** to empirically show that POI composition is highly unique in cities around the world.

Problem Formulation

- We model geographic area as circle, as in location-based service.
- We measure the level of **location privacy** for location l through **location uniqueness**, i.e. how many other locations show the same POI composition as l?



Location Re-identification: We count the frequency of all POI types within a given radius r around a particular location l, which gives us a **POI type distribution** vector P ($P = [n_{p1}, n_{p2}, \dots n_{pm}]$), where n_{pi} represents the frequency of POI type p_i within radius r around l. Then, we try to **re-identify** this location through P from a location pool L and result in a number of possible locations L_C, known as candidate location.

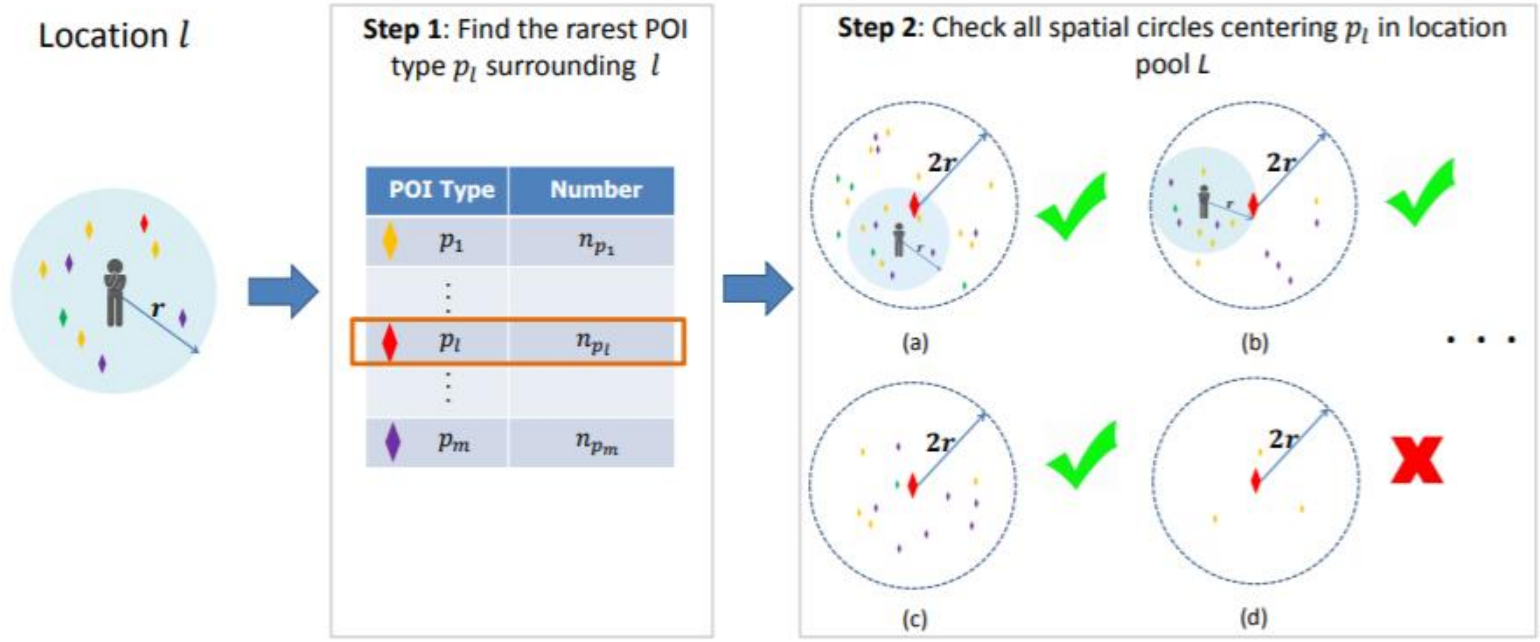
Measure of Location Privacy

- **Number of Candidate Locations:** The greater the number of candidate locations, the lower level of location uniqueness, thus higher level of location privacy

- **Privacy Index:** defined as number of candidate locations divided by the total number of locations in the location pool L

Framework

- Accurate location re-identification requires intractable computation, we therefore relax our demand and focus on providing lower bound on location uniqueness.



Experiment Datasets

POI data from OpenStreetMap (https://www.openstreetmap.org/) extracted by Mapzen (https://mapzen.com/products/)

- OpenStreetMap: easily accessible, wide popularity, global coverage, same format as other map service

  (POI name || POI type || GPS Location)

- We selected five representative cities worldwide: New York, Melbourne, Vancouver, Zurich and Shanghai.

Experiment Datasets

- **We selected five representative cities worldwide: New York, Melbourne, Vancouver, Zurich and Shanghai.**

	Country	Population($million$)	Urban area(km^2)	Num of POIs	Num of POI types	Studied area (km^2)
New York	U.S.	8.538	1213.37	26202	125	118*89
Melbourne	Australia	3.848	9992.5	17735	145	151*144
Vancouver	Canada	0.647	114.97	5267	74	22*16
Zurich	Switzerland	0.391	87.88	22000	147	41*39
Shanghai	China	24.2	6341	9618	94	395*365

- New York: global metropolis
- Melbourne: Oceanian metropolis
- Vancouver: middle-size north American city
- Zurich: middle-size European city
- Shanghai: Asian metropolis

Experiment Datasets

- **We selected five representative cities worldwide: New York, Melbourne, Vancouver, Zurich and Shanghai.**

City	Top 10 most popular POI types
New York	bicycle parking, restaurant, school, place of worship, cafe, fast food, bench, bank, fire station, drinking water, post box
Melbourne	restaurant, bench, cafe, fast food, toilets, post box, parking, drinking water, bicycle parking, waste basket, telephone
Vancouver	bench, restaurant, bicycle parking, cafe, fast food, post box, waste basket, bank, toilets, bicycle rental, drinking water
Zurich	bench, restaurant, drinking water, waste basket, vending machine, post box, parking, parking entrance, recycling, bicycle parking, fast food
Shanghai	bicycle rental, restaurant, bank, cafe, toilets, fast food, parking, community centre, fuel, school, bench

Experiment Setups

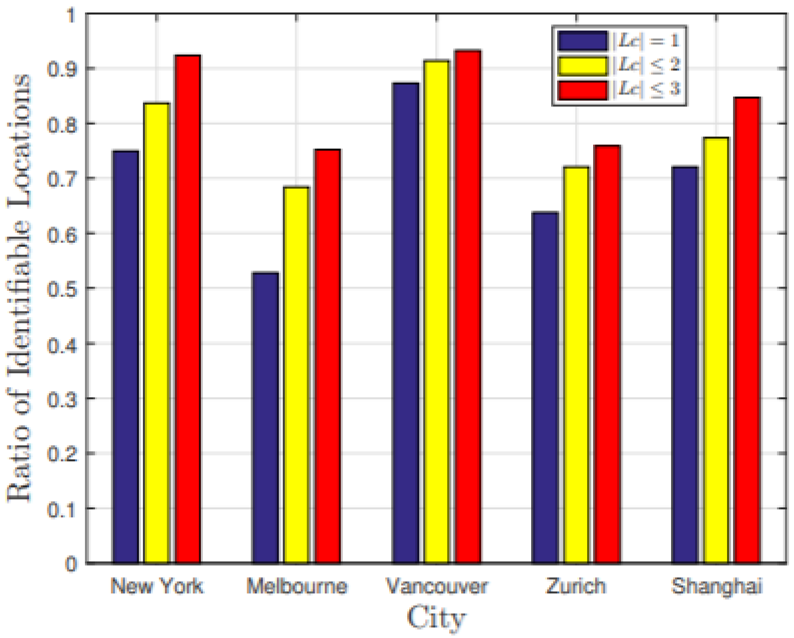
- To best capture the city structure, for each city, we uniformly sampled 1,200,000 unique locations and calculated POI type statistics using a varying radius of 0.1km, 0.25km, 0.5km, 1km, 2km and 4km to represent different spatial granularity under various application scenarios.

- We then filtered out location with low POI density (threshold: $50/\pi$ km^{-2}).

- We finally ran the proposed location re-identification method.

Results

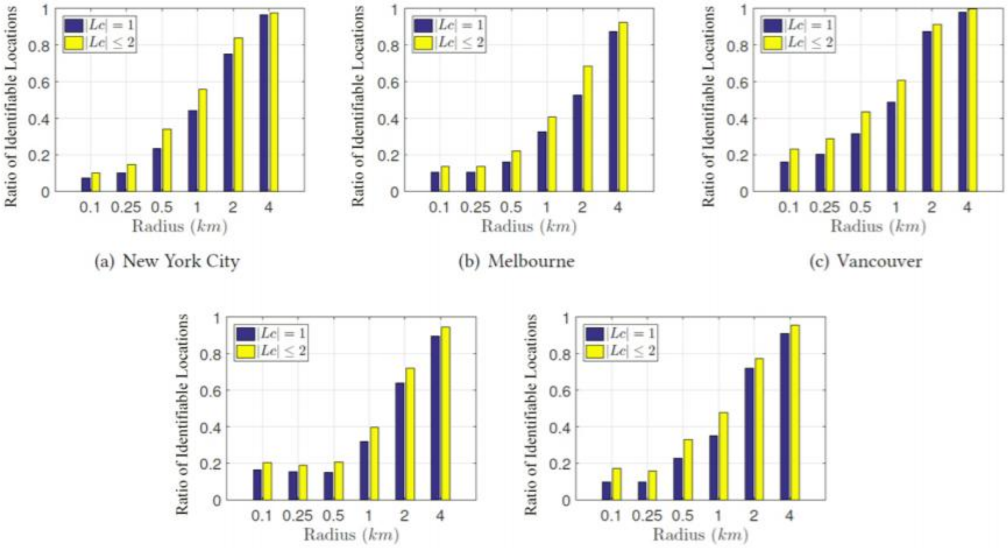
1. Location Uniqueness



- Surprisingly high level of location uniqueness in all five cities!
- When r=2km, 75%, 53%, 87%, 64% and 72% of randomly selected locations can be uniquely identified in New York, Melbourne, Vancouver, Zurich and Shanghai

Results

2. Location Privacy vs. Spatial Granularity



(a) New York City (b) Melbourne (c) Vancouver

(d) Zurich (e) Shanghai

Fig. 3. Percentage of locations which can be identified within one ($|L_c| = 1$) or two ($|L_c| = 2$) possible regions in the city with respect to radius.

- Greater radius leads to greater location uniqueness, and in turn lower level of location privacy

Results

2. Location Privacy vs. Spatial Granularity

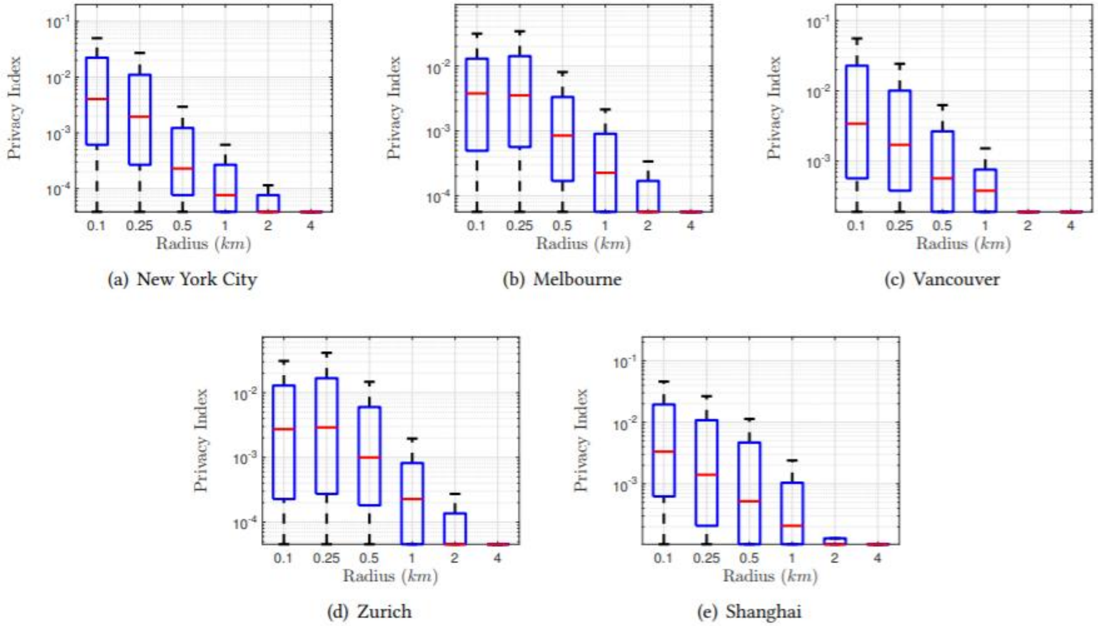


Fig. 4. Privacy index (location uniqueness) with respect to Radius.

- Greater radius leads to greater location uniqueness, and in turn lower level of location privacy

Results

3. Location Privacy vs. POI Density

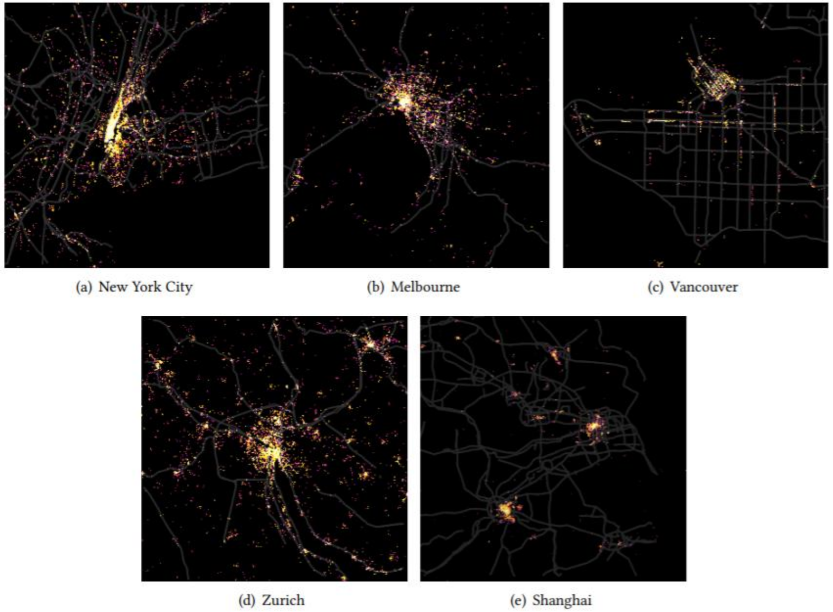


Fig. 6. Spatial distribution of POI density. Brighter color means higher density.

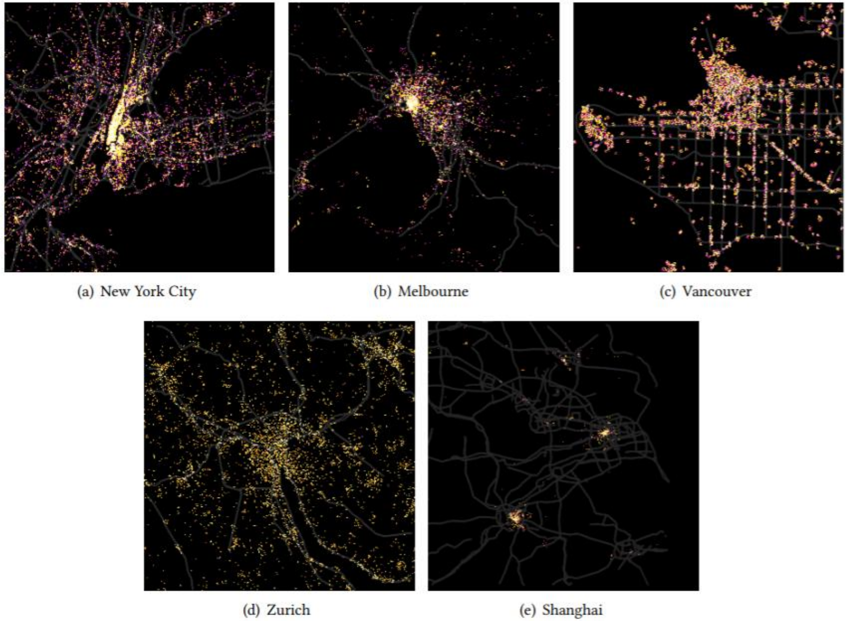
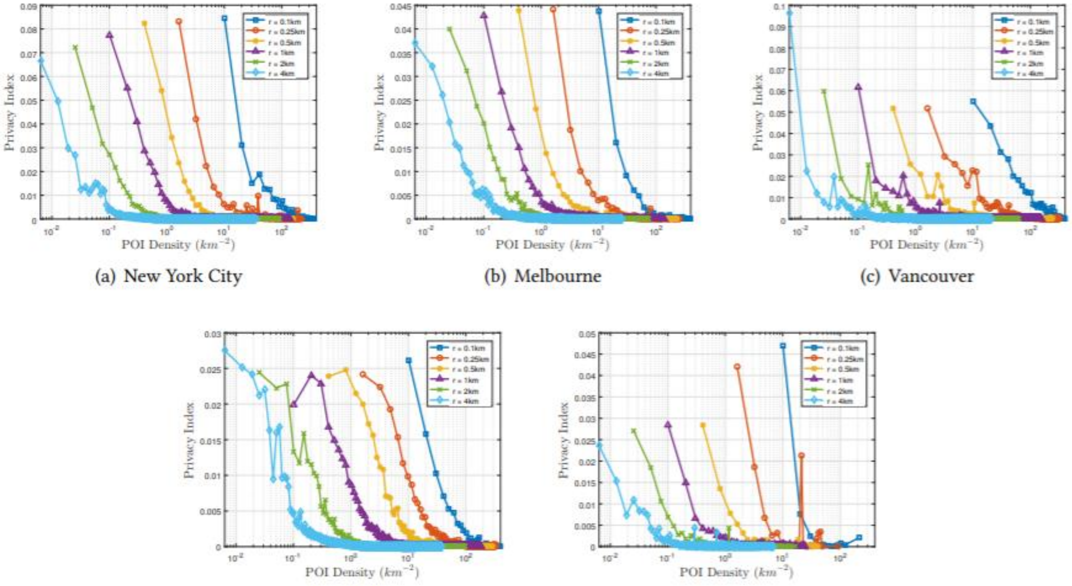


Fig. 7. Spatial distribution of location uniqueness. Brighter color means higher uniqueness.

Results

3. Location Privacy vs. POI Density



(a) New York City (b) Melbourne (c) Vancouver

(d) Zurich (e) Shanghai

Fig. 8. Relationship between privacy index (location uniqueness) and POI density.

- Greater POI density leads to greater location uniqueness, and in turn lower level of location privacy

Results

1. Location Privacy vs. POI Popularity

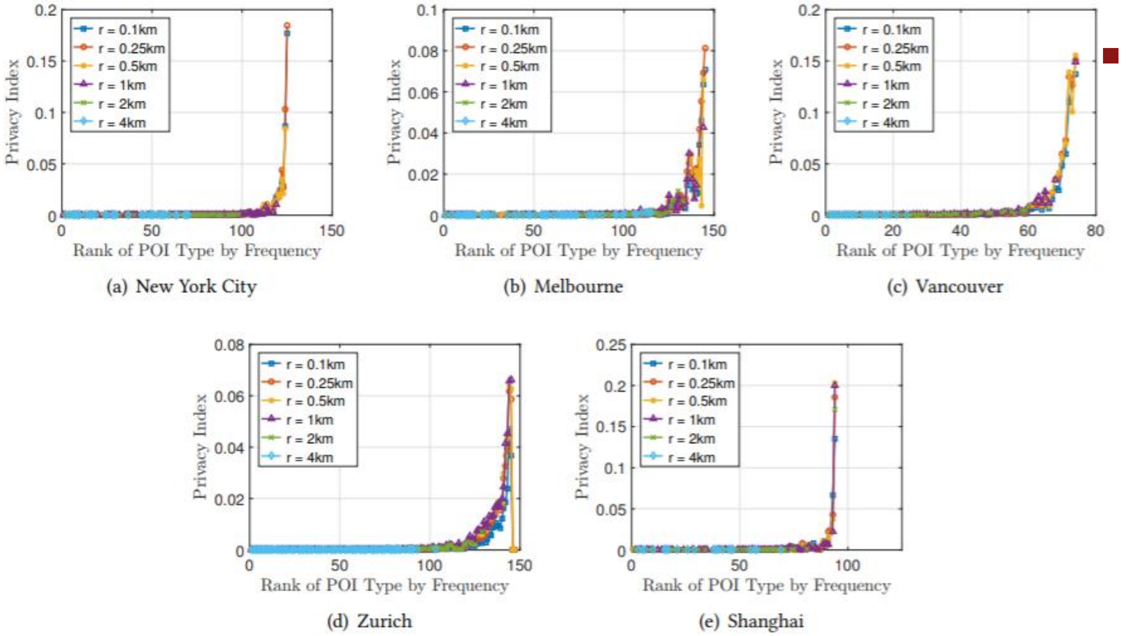


Fig. 9. Relationship between privacy index (location uniqueness) and POI Uniqueness/Ranking. POIs of rank 1 are the most popular in their city.

- Rare POI type leads to greater location uniqueness, and in turn lower level of location privacy

Results

1. Location Privacy vs. Distance to City Center

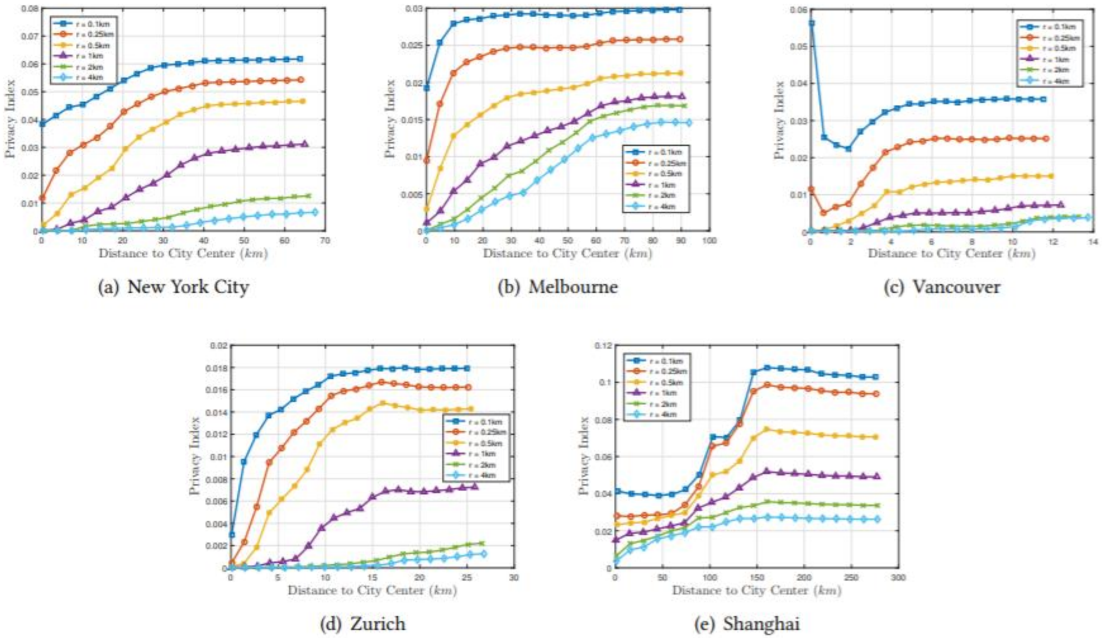


Fig. 10. Relationship between privacy index (location uniqueness) and distance to city center.

- The nearer to the city center, the greater level of location uniqueness, and the lower level of location privacy

Conclusions & Discussions

- POI composition in city is **highly unique**, thus revealing nearby POI types poses threat on user location privacy.

- Location has higher level of uniqueness (lower level of location privacy) with **larger radius, rarer POI types, higher POI density, and if nearer to city center**. Results are consistent over cities of different characteristics, which point to fundamental characteristics of urban morphology.

- POIs as context: immediate surroundings provide “generic” personalization while larger surroundings enable **richer personalization experience**

- Future work: generalize the findings to other map services; improve location re-identification algorithm; design privacy-aware systems with good personalization experience

Thanks! Questions?

Hancheng Cao[1,2], Jie Feng[2], Yong Li[2], Vassilis Kostakos[3]

Stanford Univ.[1], Tsinghua Univ.[2], Univ. of Melbourne[3]

Oct. 11th, 2018

http://hanchengcao.me/

http://fi.ee.tsinghua.edu.cn/~liyong/

https://people.eng.unimelb.edu.au/vkostakos/