Your Trajectory Privacy Can Be Breached Even If You Walk in Groups

Kaixin Sui, Youjian Zhao, Dapeng Liu, Minghua Ma, Lei Xu, Li Zimu, Dan Pei

清華大學

Tsinghua University

Trajectory in a Large-scale Wi-Fi Network

Wi-Fi access point

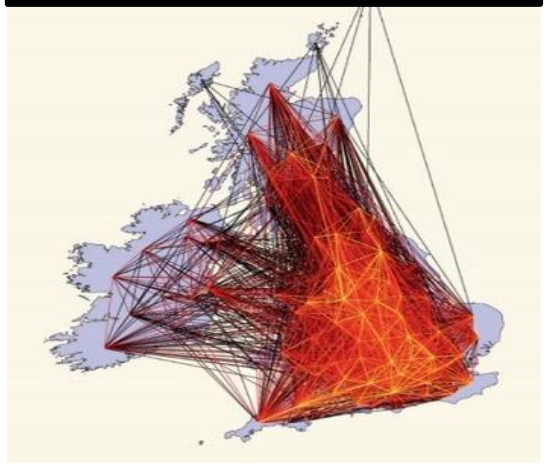
Tsinghua Campus

Trajectory in a Large-scale Wi-Fi Network

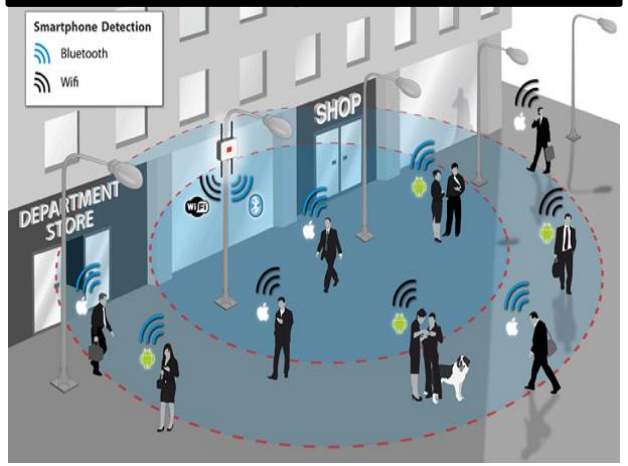
Trajectory in a Large-scale Wi-Fi Network

Value of Trajectory datasets

Mobility pattern mining

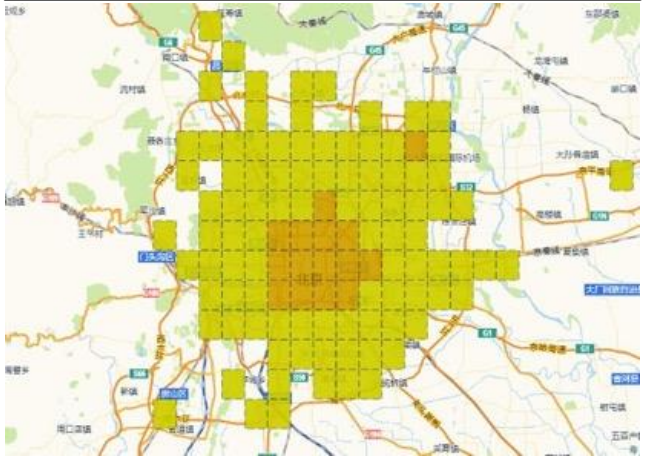


Proximity marketing



Location-based service

Transportation control



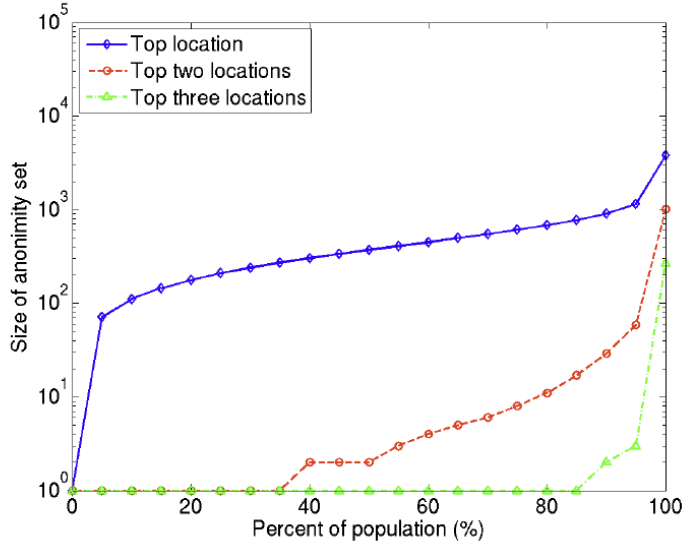
Privacy Issue of Trajectory datasets

Mobile Trajectories are **highly unique**

→ **Potential re-identification**

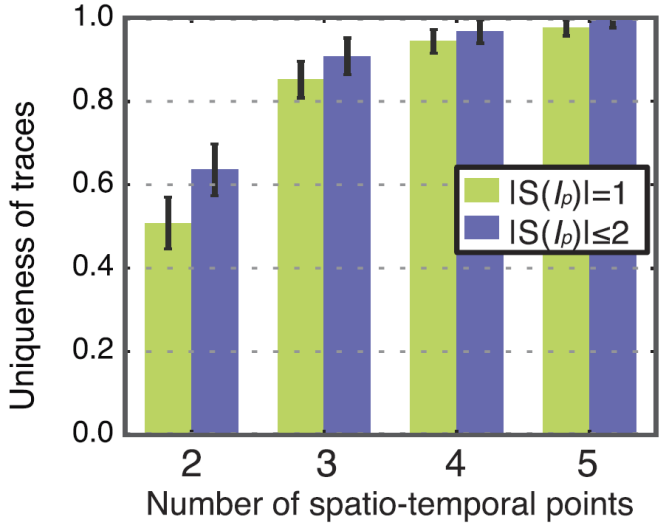
→ **Privacy risk**

MobiCom 2011



Three top locations pinpoint 50% users

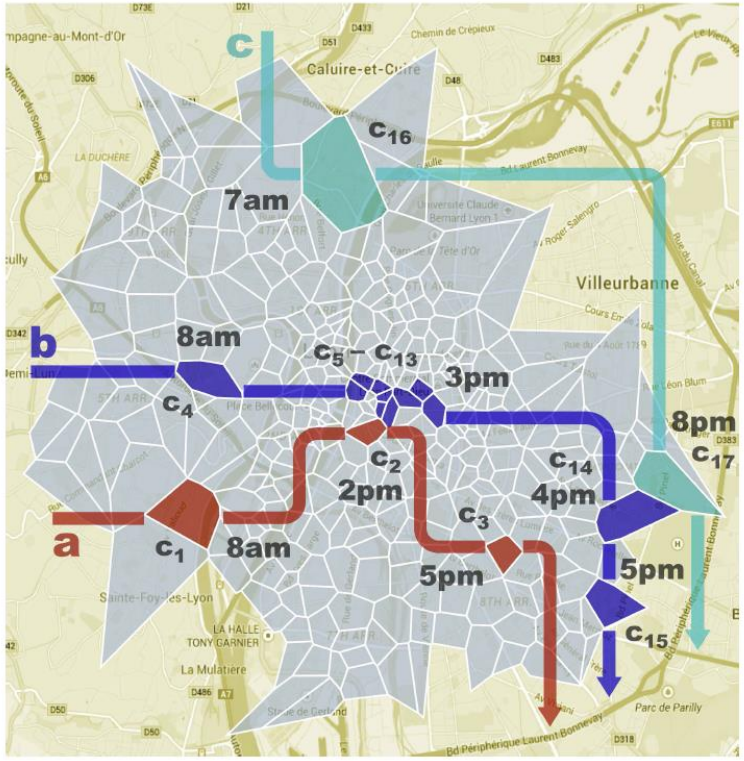
Nature Scientific Reports 2013



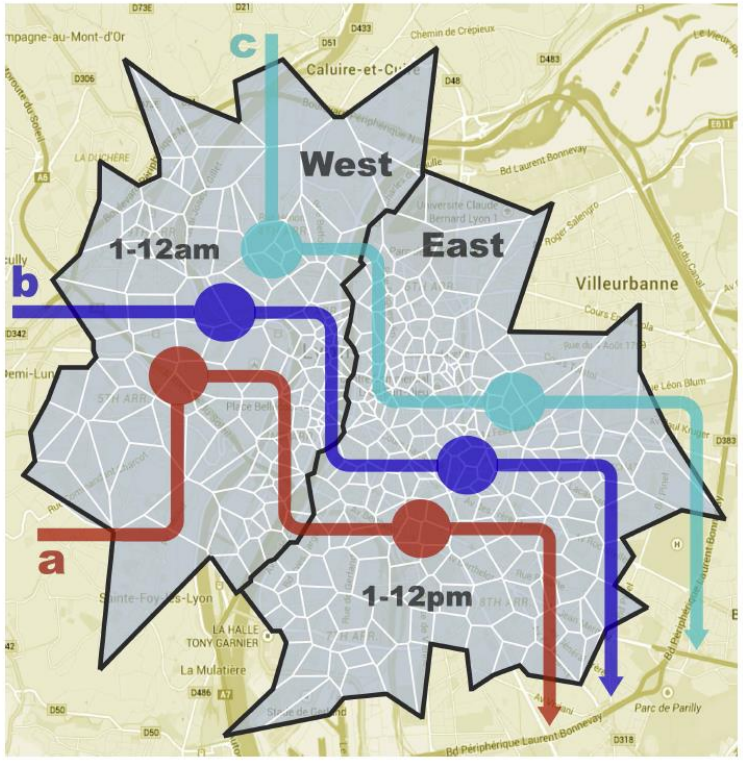
Five random points pinpoint 95% users

Counter-measure: K-anonymity

- **Key idea: each user should be indistinguishable from at least k − 1 others (hidden in an anonymity set no smaller than k)**

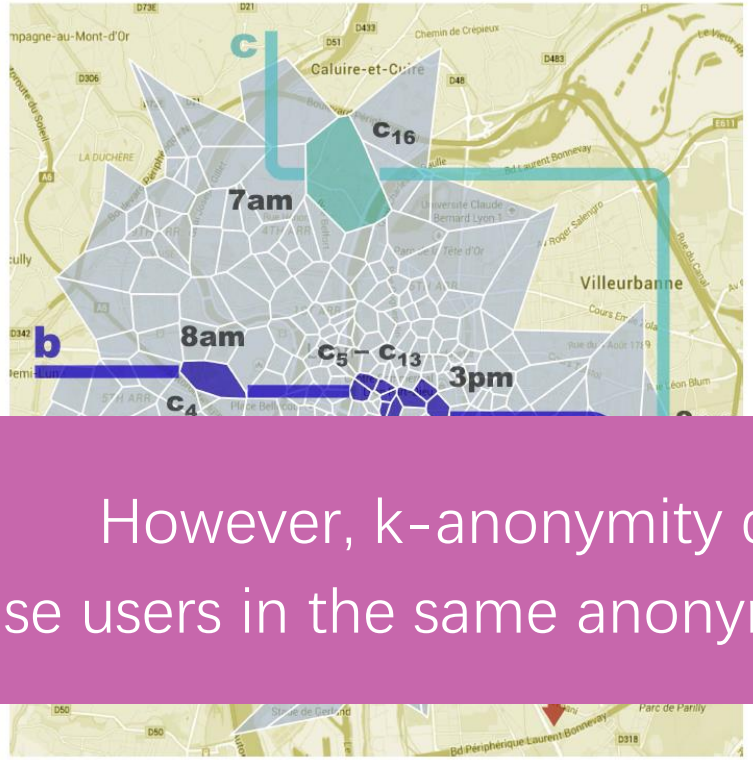


e.g.
3-anonymity

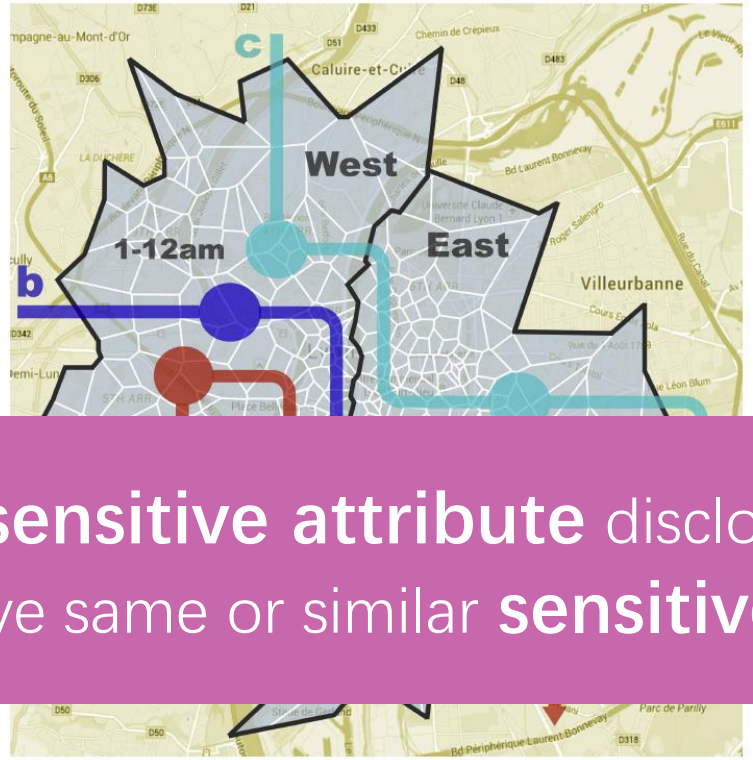


Counter-measure: K-anonymity

- **Key idea: each user should be indistinguishable from at least k − 1 others (hidden in an anonymity set no smaller than k)**



e.g.
3-anonymity



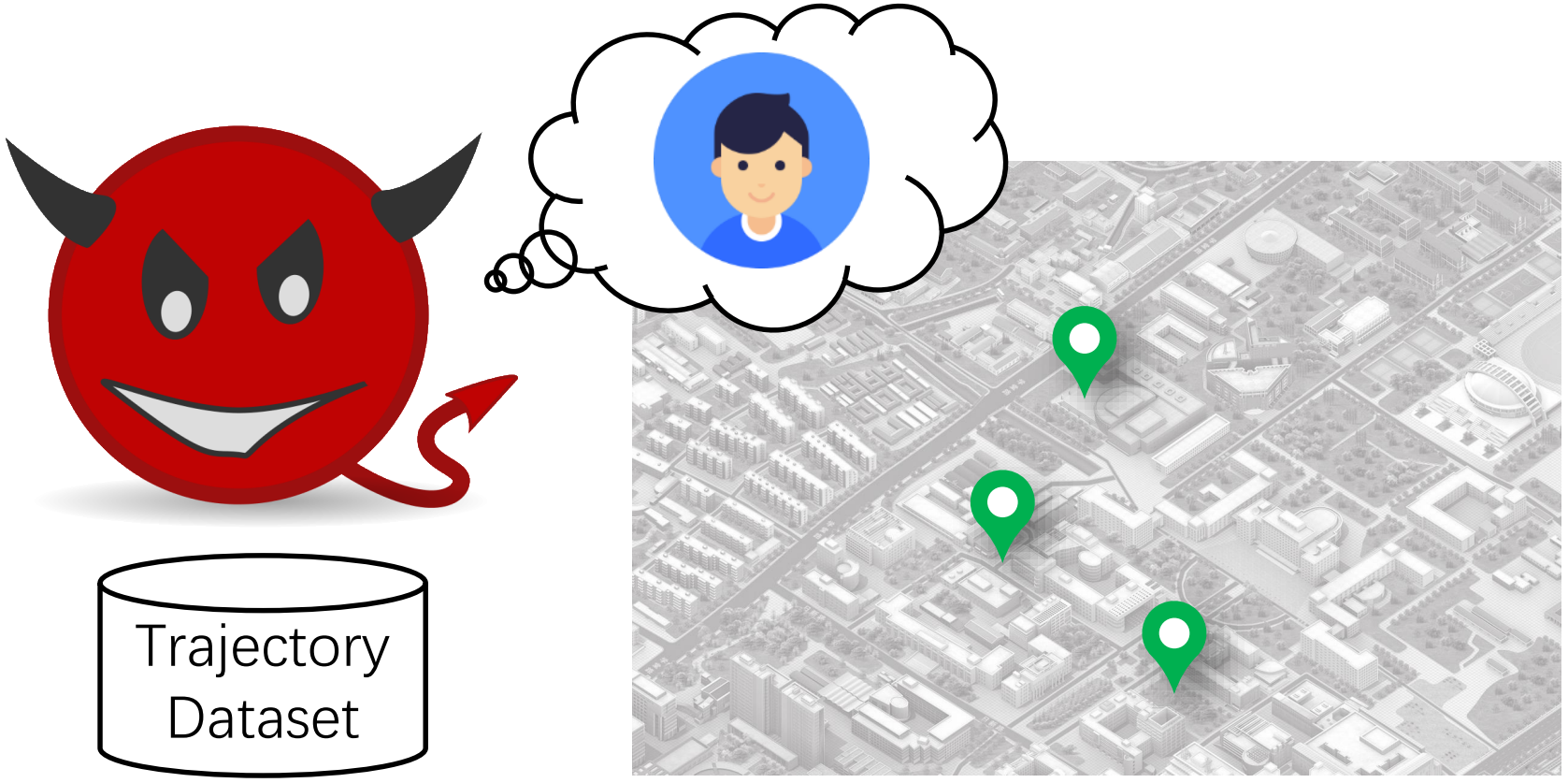
However, k-anonymity cannot prevent **sensitive attribute** disclosure
Because users in the same anonymity set may have same or similar **sensitive attributes**

Example

The adversary

knows his **three spatiotemporal samples** e.g. from social networks

and wants to know his **top two most visited locations** e.g. home and work place



Example

The adversary

finds **four** people with the same **three spatiotemporal samples**

Trajectory Dataset

Example

The adversary

cannot identify the target (thanks to 4-anonymity!)

Example

The adversary

cannot identify the target (thanks to 4-anonymity!)

But, what if these people's top two locations are the same ?

Trajectory Dataset

Problem

low diversity of the sensitive attributes

We present a large-scale measurement study to answer:

- What is the diversity of the trajectory dataset?
- What is the relationship between k and the diversity?
- Does a larger k help improve the diversity?

Wi-Fi Based Trajectory dataset from Tsinghua University

$4km^2$

150,000
Mobile devices

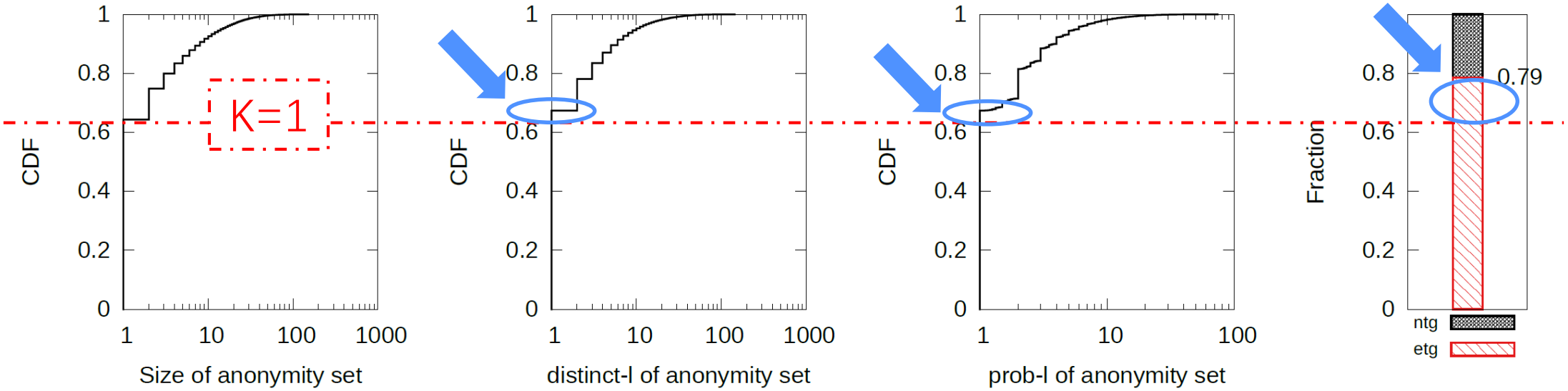
2,670
Access points

111
Buildings

4
Weeks

Observation #1

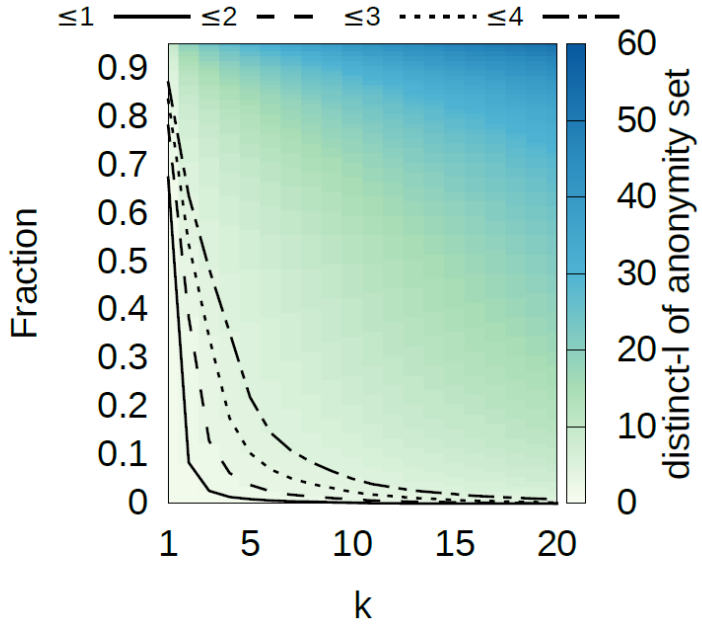
The risk of low diversity is high, even for K>1

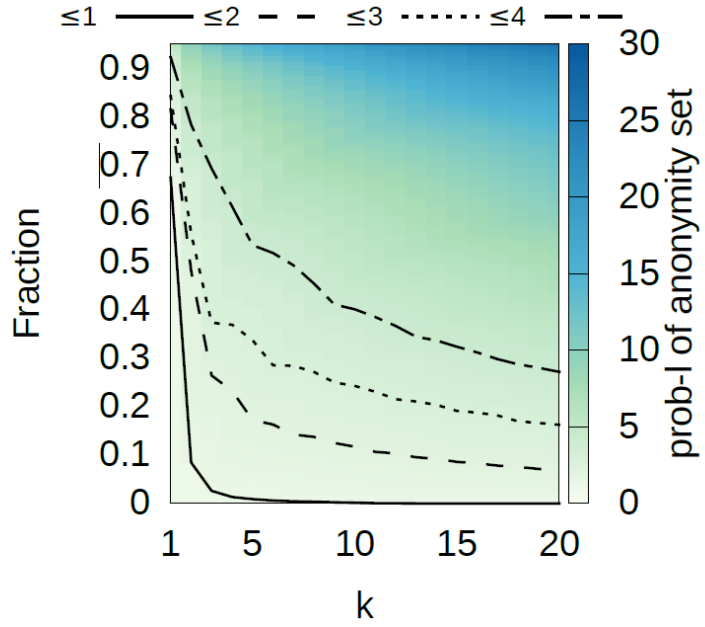


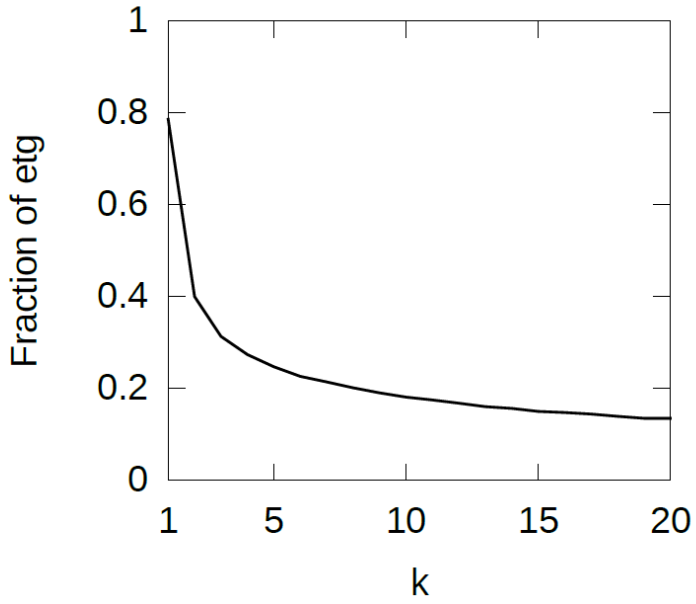
Three metrics that measure the diversity

Observation #2

Larger k helps improve the diversity, but the effect degrades exponentially



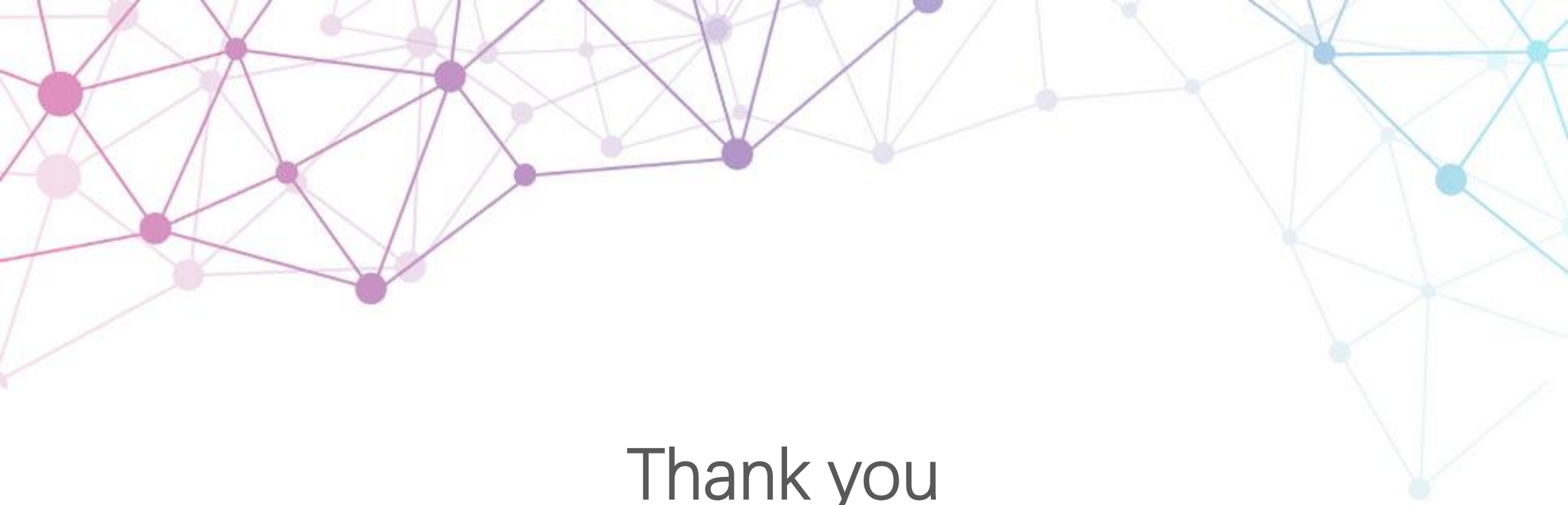




See more results in the paper

Conclusion

- A large-scale trajectory measurement shows that
  - The risk of low diversity is high for the trajectory dataset
  - K-anonymity cannot help solve the problem effectively
  - Calls for diversity-oriented solutions beyond k-anonymity (future work)



Thank you

清華大學

Tsinghua University