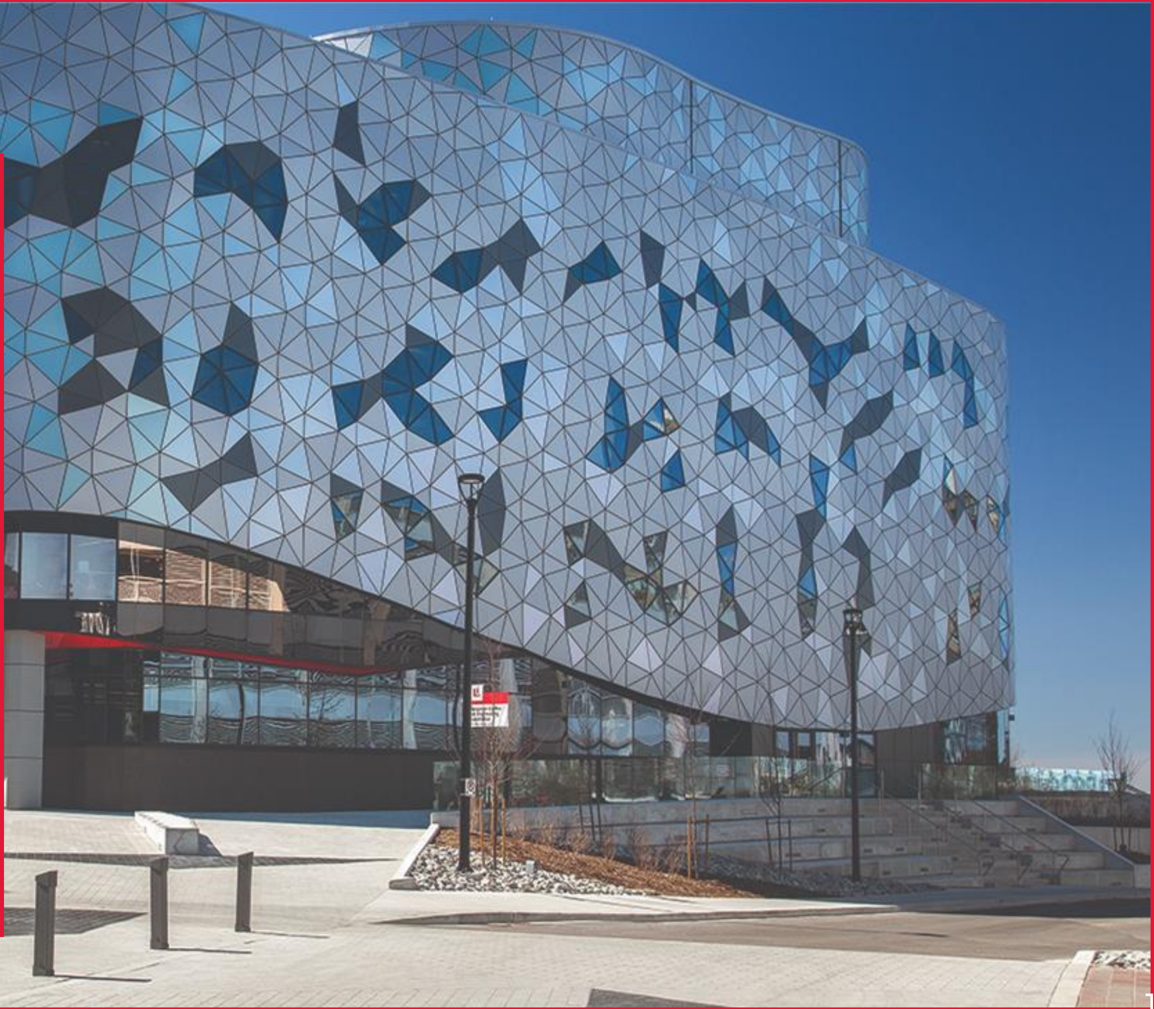
AI-Driven Mobility Data Analytics

Autonomous Vehicles and Connected Cities Symposium 2025

Ontario Transportation Council (OTC)



Manos Papagelis

Thu, Oct 30, 2025

YORK U

Introduction

Manos Papagelis / Data Mining Lab @ YorkU

Academic Background
- Associate Prof., EECS, York University
- Data Mining Lab, EECS, York University

Education
- PhD, Computer Science, University of Toronto
- BSc/MSc, Computer Science, University of Crete, Greece

Research Interests
- Data Mining / Graph Mining / Machine Learning
- Data Systems / Big Data Analytics / Knowledge Discovery
- Natural Language Processing

Move Mobility Innovation Centre

https://lassonde.yorku.ca/move/

Research Areas: mobility analytics, connected mobility, autonomous mobility, sustainable mobility and augmented/virtual reality mobility

Email: papaggel@gmail.com
Office: LAS 3050

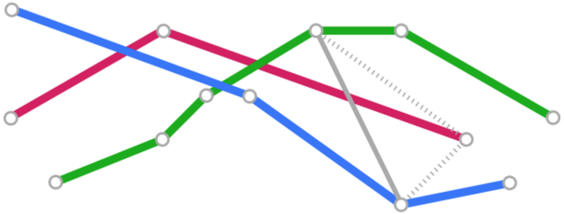
Background & Motivation

Mobility/Trajectory Data

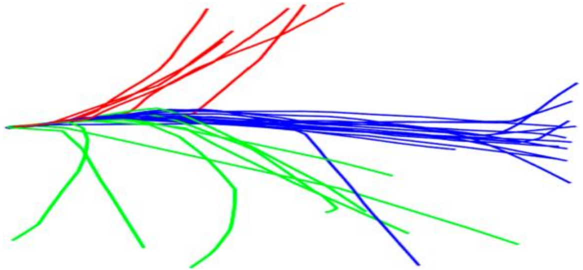
Trajectory: A Sequence of (Spatiotemporal) Points

Vast Amounts of Trajectory/Mobility Data

Mobility Data Analytics / Trajectory Data Mining



trajectory similarity



trajectory clustering

trajectory anomaly detection
trajectory network mining
trajectory classification
...

challenging computational problems

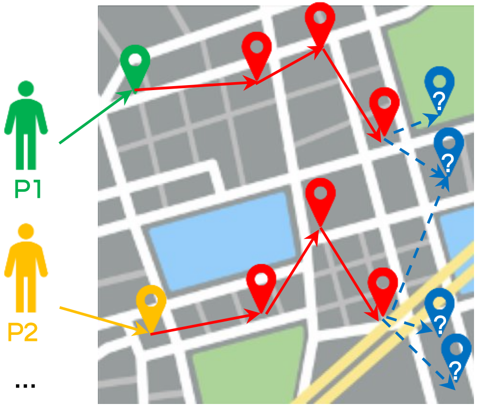
Mobility Data Analytics in the Age of Big Data and AI

a symbiotic relationship that presents a new strategy for addressing complex problems in mobility data analytics

Image source: This image was created with the assistance of DALL·E 3

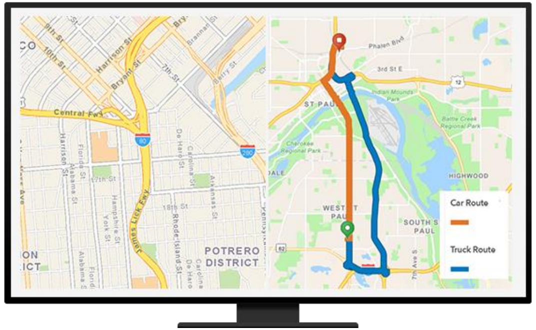
Plethora of Applications

ridesharing



trip/POI (point-of-interest) recommendation

traffic analysis



route planning and optimization

Our Journey on Mobility Data Analytics

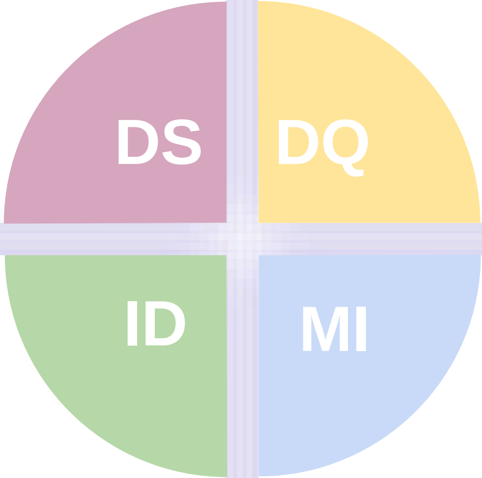
- Trajectory dataset and resources [ACM SIGSPATIAL '23]
- Trajectory prediction [ACM TSAS, '25]
- Trajectory/mobility network prediction [ACM TKDD, Submitted]
- Trajectory classification [IEEE MDM '23]
- Trajectory unlearning [ACM TSAS, Submitted]
- Trajectory simplification [ACM SIGSPATIAL '23]
- Trajectory similarity [Submitted]
- Trajectory network analysis [Big Data Research, IEEE MDM '20, GeoInformatica, IEEE BigData '18, 2 x IEEE MDM '18]
- Mobility + epidemics [ACM SIGSPATIAL/SpatialEpi '24, ACM SIGSPATIAL/SpatialEpi '23, IEEE MDM '22]
- Transportation optimization [ACM SIGSPATIAL '22, ACM SIGSPATIAL '22]

Higher-order Mobility Flow Data

Challenges of Working with Trajectory Data

Data Sparsity
Irregular or incomplete geolocation sampling

Data Quality
Outliers, noisy, low accuracy



80% of the data is generated by 20% of the users

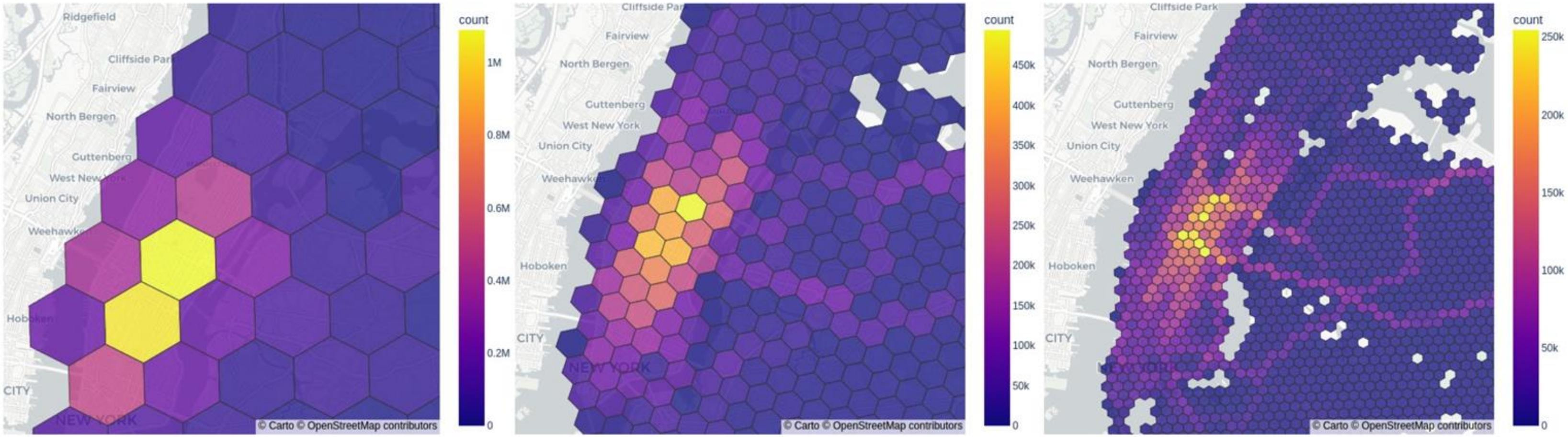
Imbalanced Data

Not compatible with well-known ML models

Model Incompatibility

Map Tessellation

Low resolution → High resolution

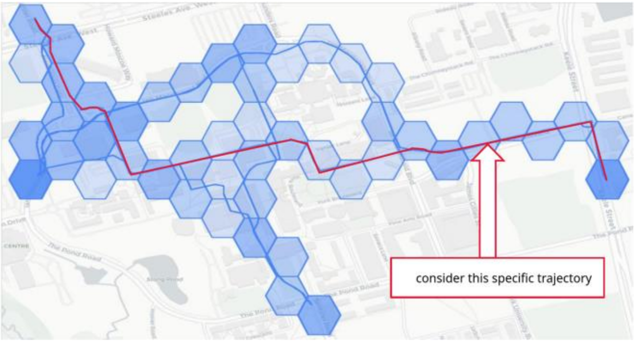


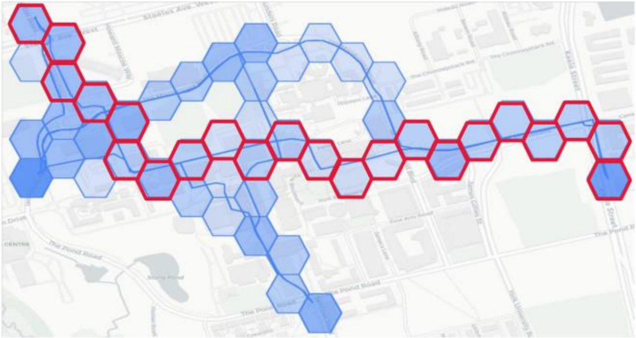
Why hexagons?

- More circular that fully tessellates the space
- Same distance to all adjacent neighbours

YORK U

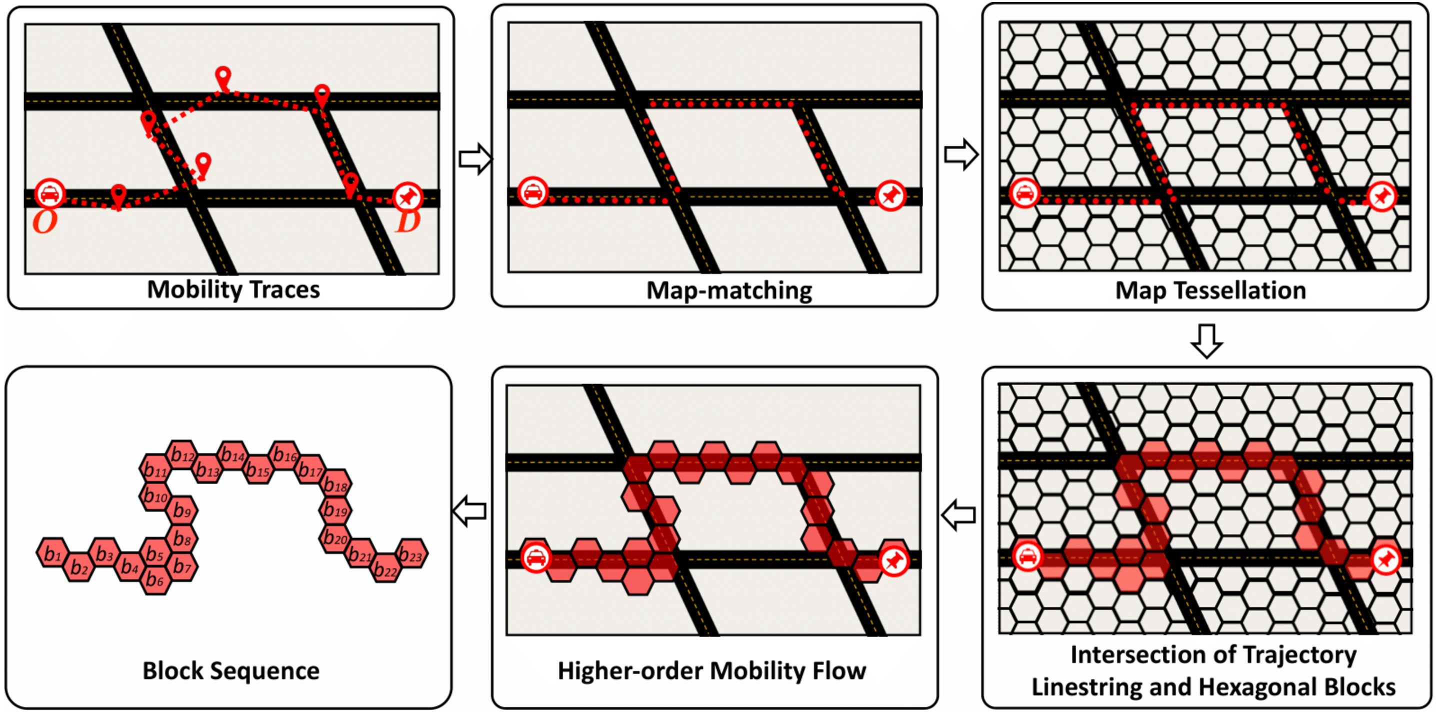
Trajectories: Sequences of Hexagons





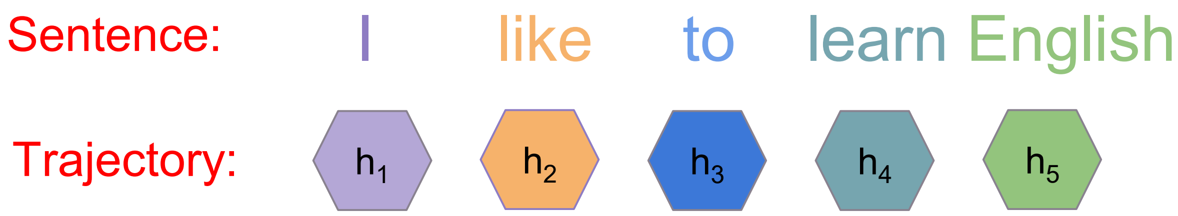
Trajectory: h_1, h_2, h_3 ... h_{20}, h_{21}, h_{22}

Overview of the Transformation Pipeline



Treat Trajectories as Language Statements

Treat Trajectories as Language Statements



Analogy:

Natural Language → tokens, sentence

Mobility data → hexagon ids, trajectory

Advantages

- Reduced data sparsity
- More compatible with established ML architectures (e.g., sequence models, LLMs)

(Dis)advantages

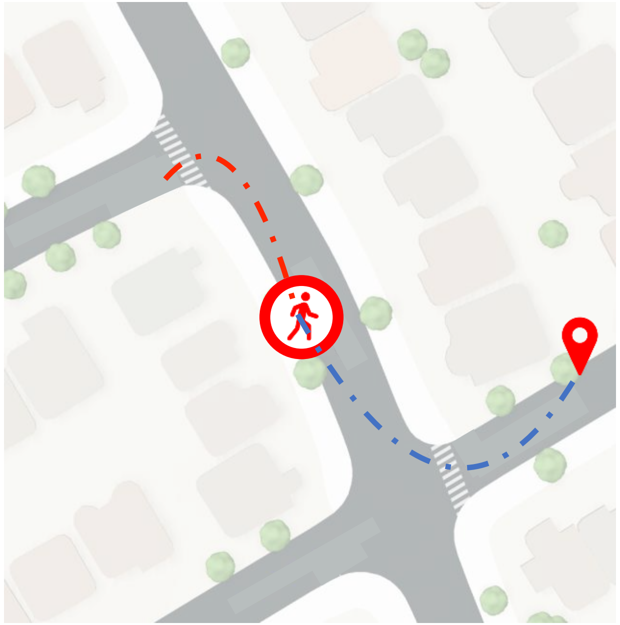
- Requires training models from scratch on a “new language”

YORK U

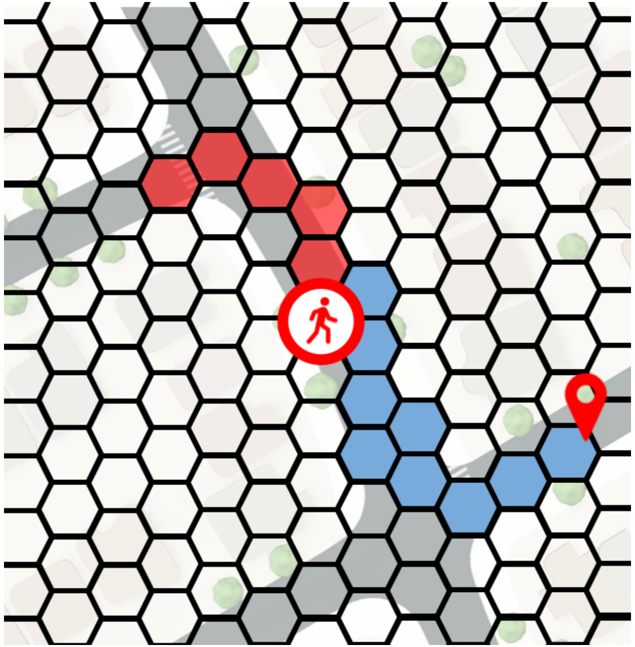
Trajectory Prediction

Predict the Next-k Trajectory Steps Problem

Problem of Interest: Trajectory Prediction



predict future trajectory



predict the next k hexagons
(model as a sequence prediction problem)

Problem Definition

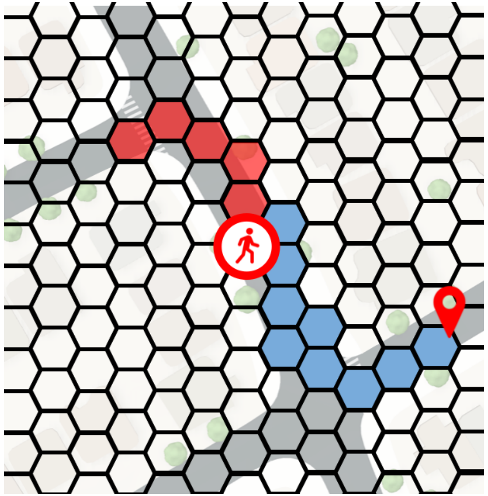
Let

- an observation area
- a set of objects and their history trajectories
- an observation period

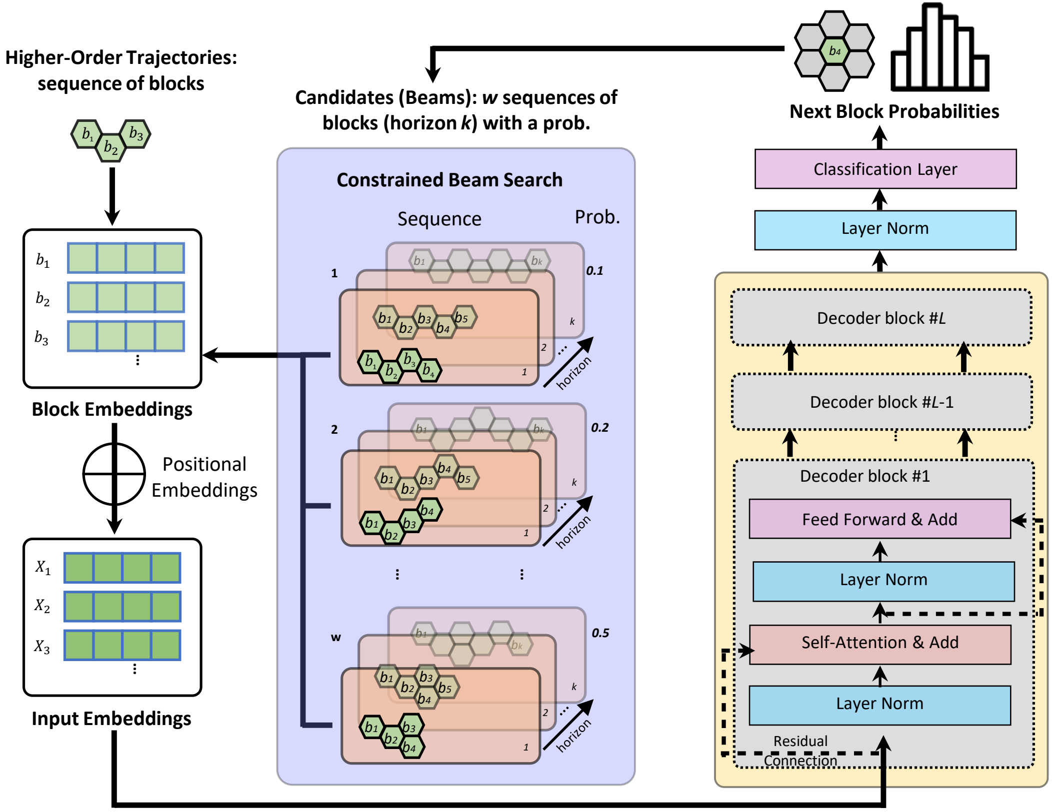
Input: Given

- a moving object n
- a partial trajectory = <p1, p2,...,pt>
- a prediction horizon k > 0

Output: We want to
predict the **next k hexagons** of the input partial trajectory



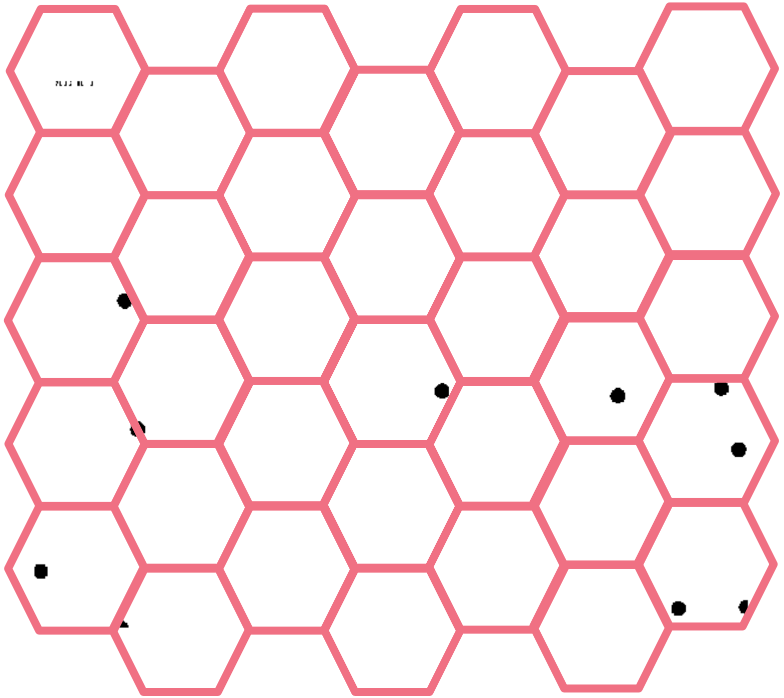
TrajLearn Overview



TrajLearn outperforms the SOTA & baselines by up to ~60%

Mobility Network Prediction

Trajectories of moving objects



Which objects will interact in the future?

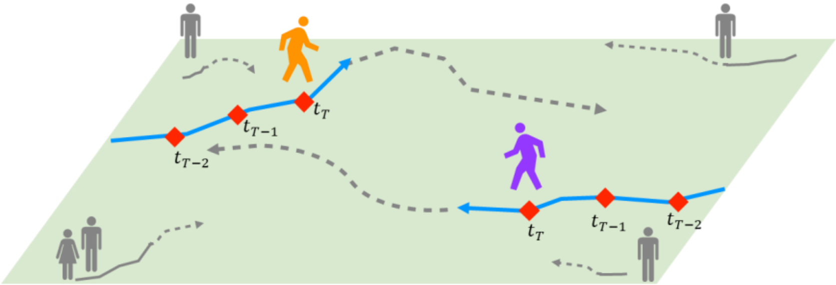
Interaction (x, y):

Both x and y are at hexagon h at time t

Mobility Network Prediction

Let

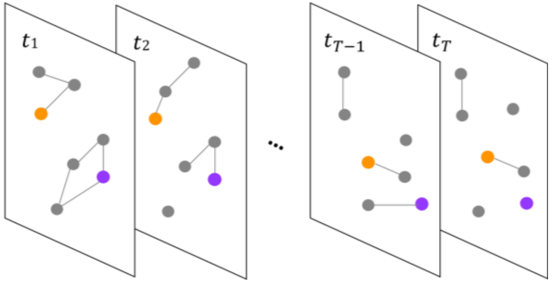
- an observation area $\mathcal{M}$
- a set of objects $\mathcal{N}$ and their history trajectories $\mathcal{T}^l$
- an observation period $[0, \mathcal{W}]$



Input: Given

- a prediction horizon $k > 0$

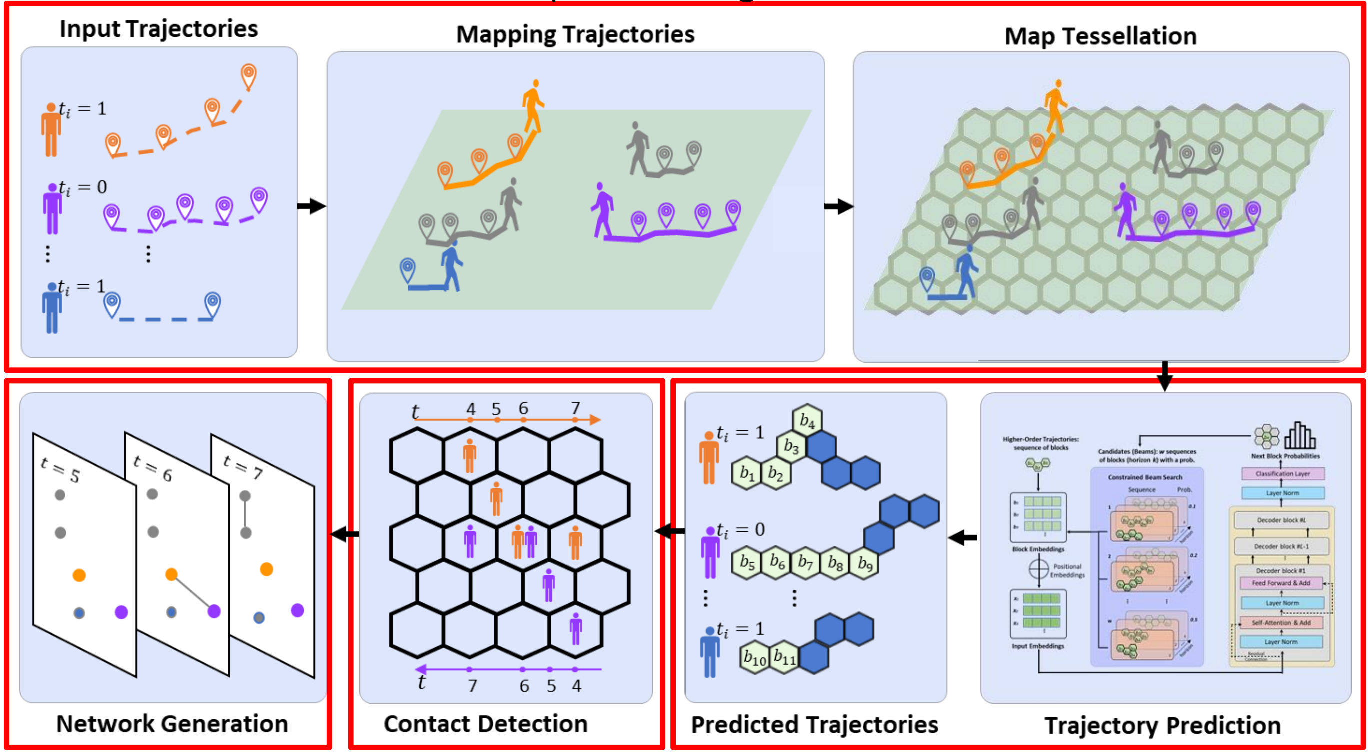
Output: We want to
predict the next k future proximity networks $< G_{i_{w+1}}, G_{i_{w+2}}, \dots, G_{i_{w+k}} >$



YORK U

MobiNetForecast Overview

Preprocessing



Module 3 Module 2 Module 1

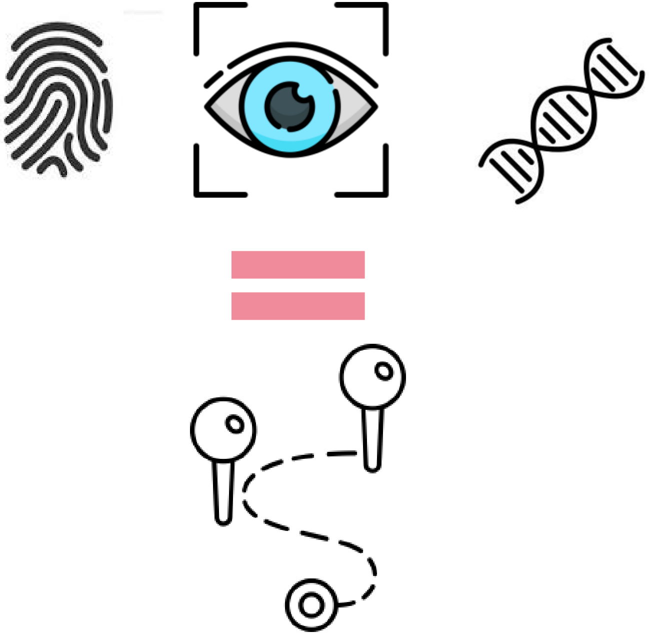
MobiNetForecast outperforms SOTA temporal link prediction methods by **up to two orders of magnitude** in F1-score

MobiNetForecast outperforms SOTA sequence models by **15–20%** on key metrics

Trajectory Classification

The Trajectory-User Linking Problem

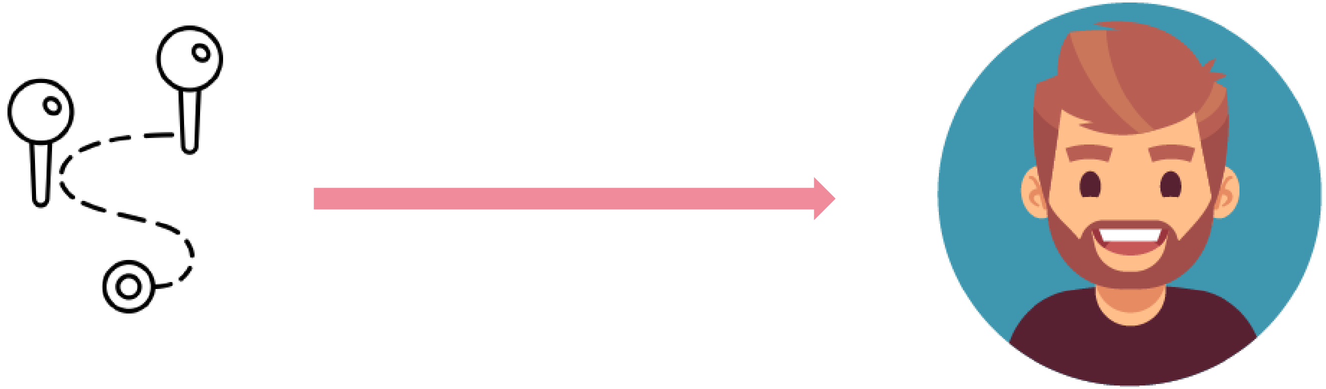
YORK U



can trajectories

help to identify a person?

Trajectory-user Linking (TUL)



trajectory-user linking **aims at linking** anonymous trajectories to users who generate them

Problem Definition

Given:

$\mathcal{U} = \{u_1, u_2, u_3, .., u_c\}$ – users

$\mathcal{T} = \{Tr_1, Tr_2, \ldots, Tr_n\}$ – unlinked trajectories

TUL is defined as **a multiclass classification problem**

$$\min_{f \in \mathcal{F}} \mathbb{E}[\mathcal{L}(f(Tr_i), ui)] \ over \ \mathcal{F}$$

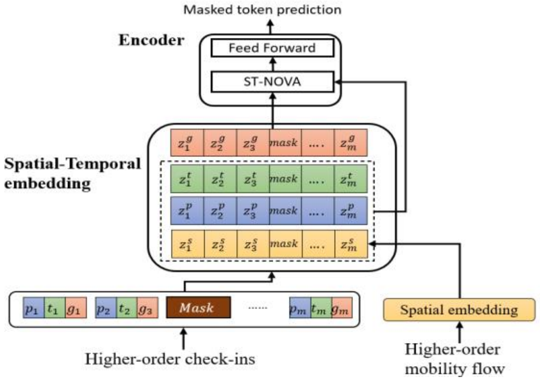
where $\mathcal{F}$ is the set of all classifiers in the hypothesis space

$\mathcal{L}(\cdot)$ is the loss between the predicted label $f(Tr_i) \in \mathcal{U}$ and the true label $u_i \in \mathcal{U}$

YORK U

TULHOR: A Transformer-based Trajectory-user Linking Model

Higher-order mobility flow data generation



TULHOR: A spatiotemporal model that deals with sparsity and low data quality of the TUL problem

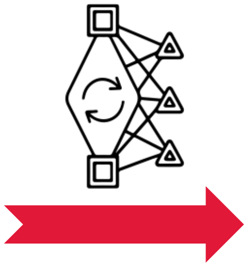
TULHOR outperforms baselines by up to 8%

Trajectory Unlearning

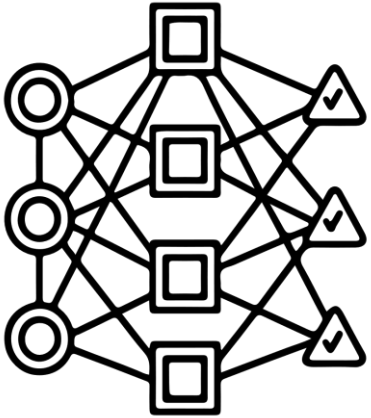
Machine Unlearning of Mobility Data

Deep Learning Approach

Large Historical Mobility Data



Training Algorithm



Trained Neural Model

What if some training data needs to be deleted?

Data Privacy Legislation

GDPR

- Right to be Forgotten
- Right to restriction of processing

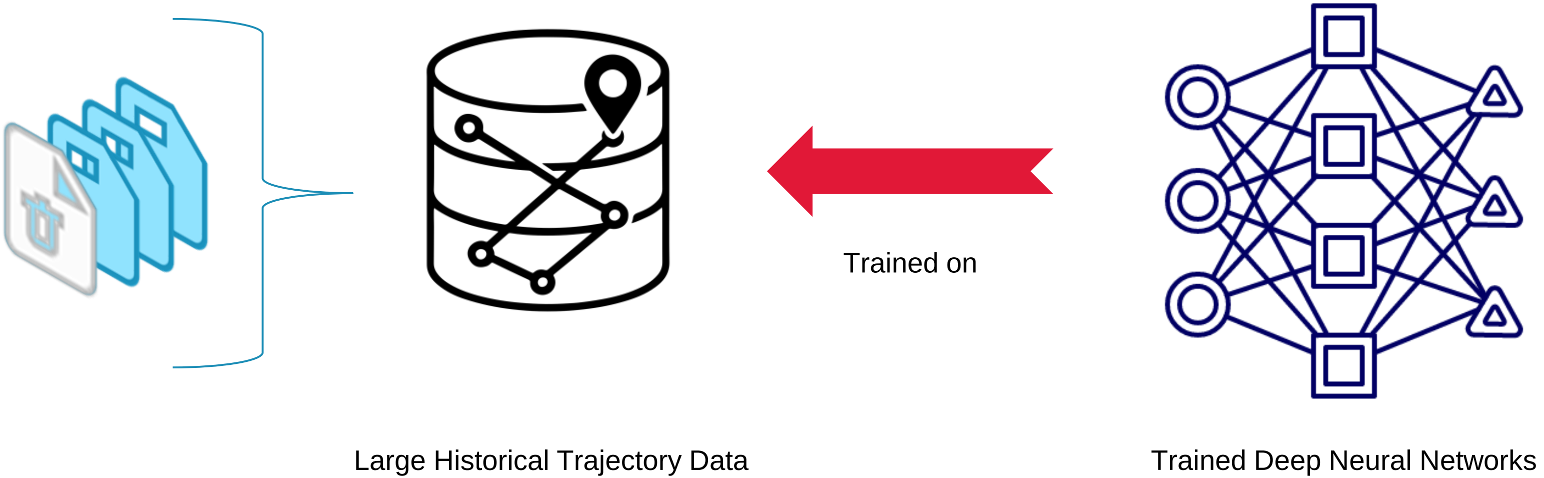
CCPA

- Right to Opt-Out
- Right to Delete

PIPEDA

- New one: CPPA
- Personal Information Protection

Machine Unlearning (MU) for Trajectory Data

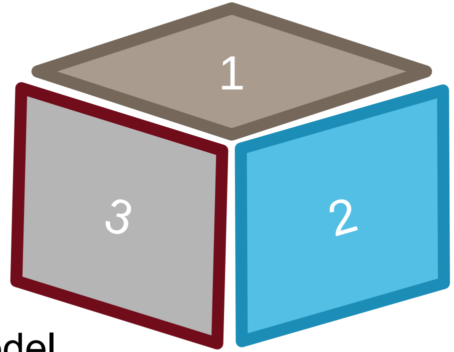


Goal: Eliminate specific trajectory influence **without retraining**

YORK U

The Three Main Goals of MU

Computation Efficiency
Faster than retraining



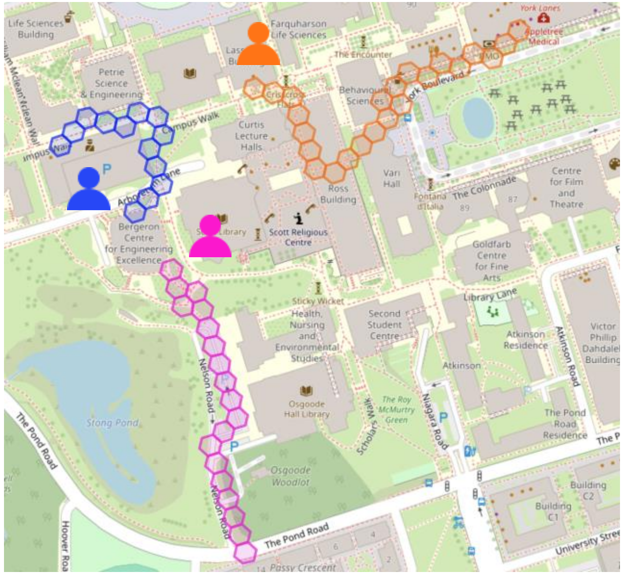
Unlearning Efficacy
Truly remove the unlearning-targeted information

Preserved Model Utility
Maintain utility of “unlearned” model

[Liu et al., CVPR ‘24]

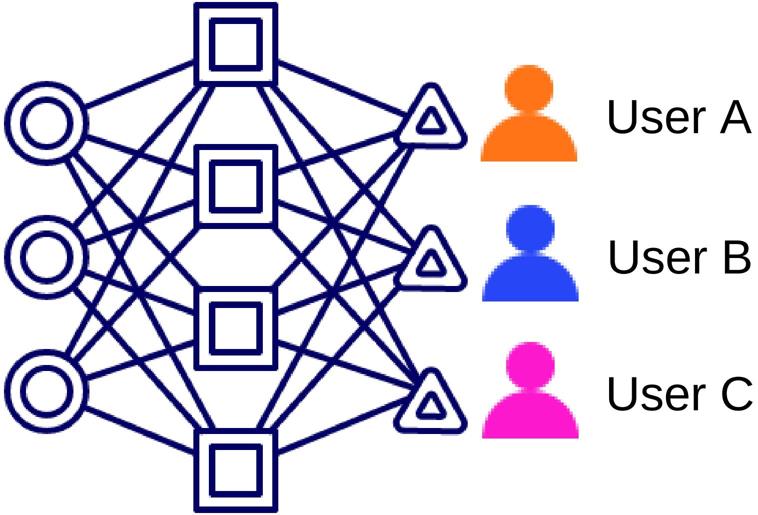
YORK U

Trained Model: Trajectory Classification



Users' Trajectory

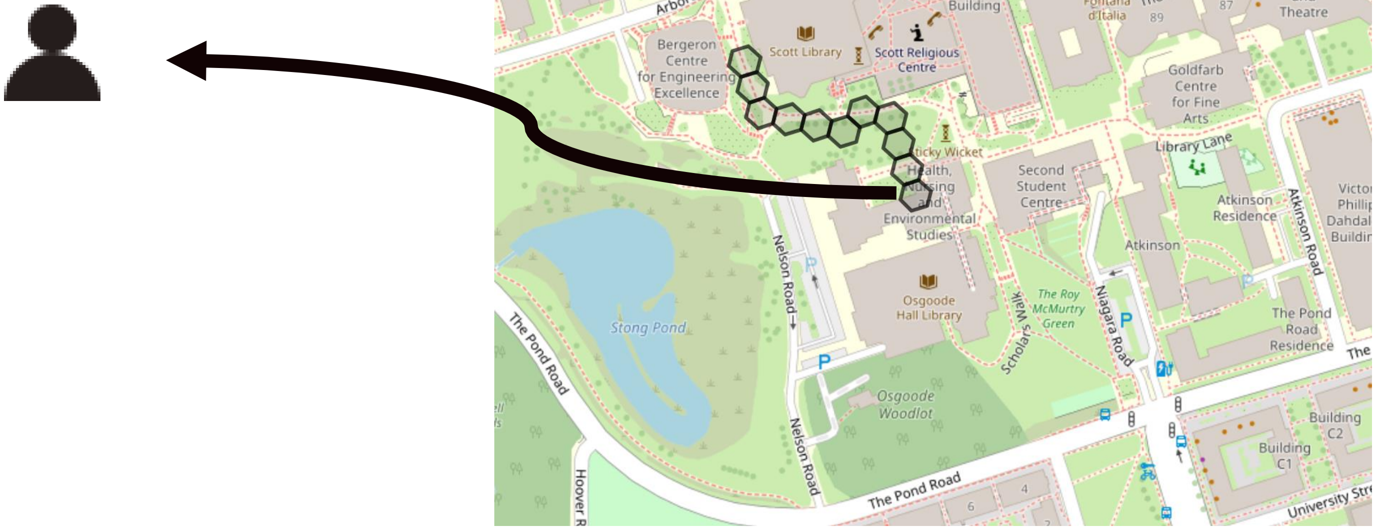
Training Algorithm



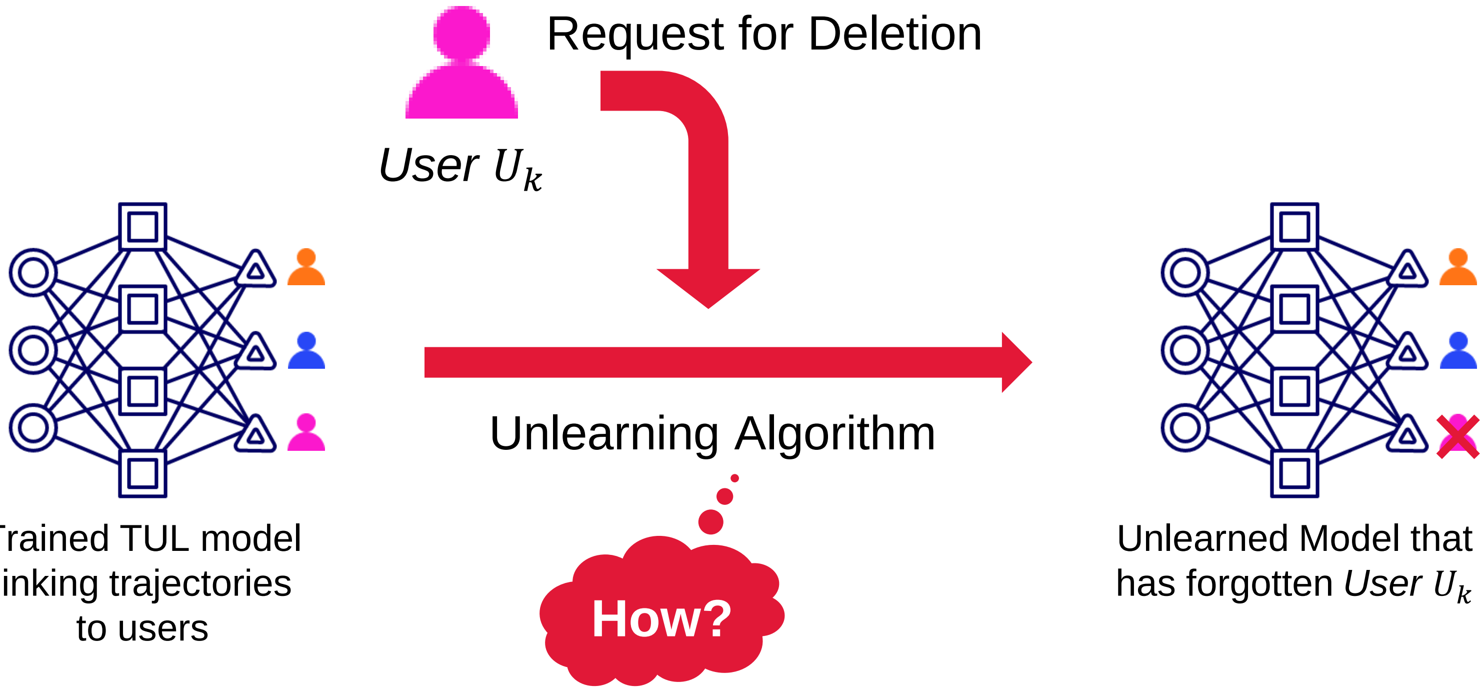
Classifier

Trajectory User Linking (TUL)

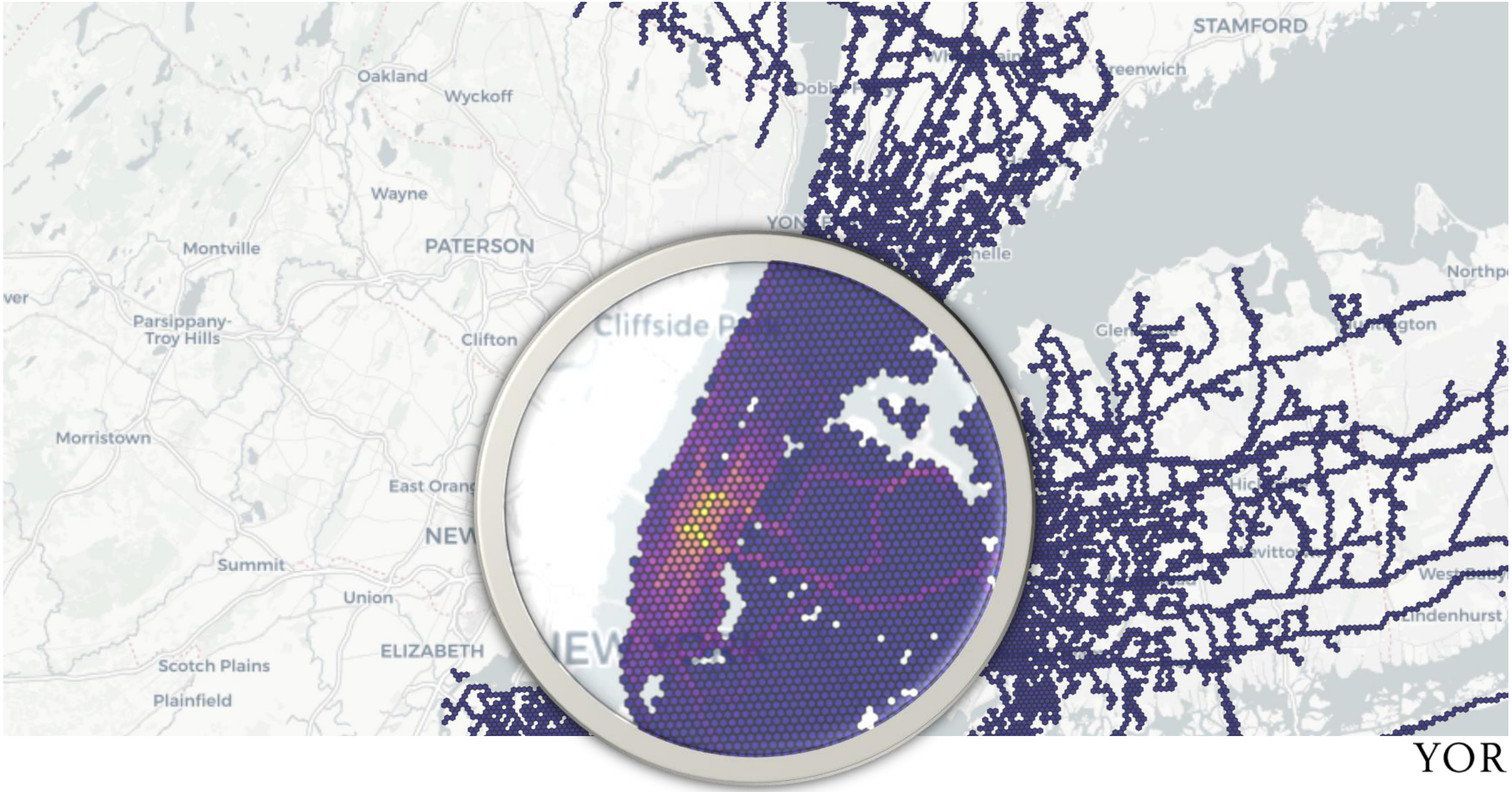
{A, B, C} ?



Unlearning in TUL

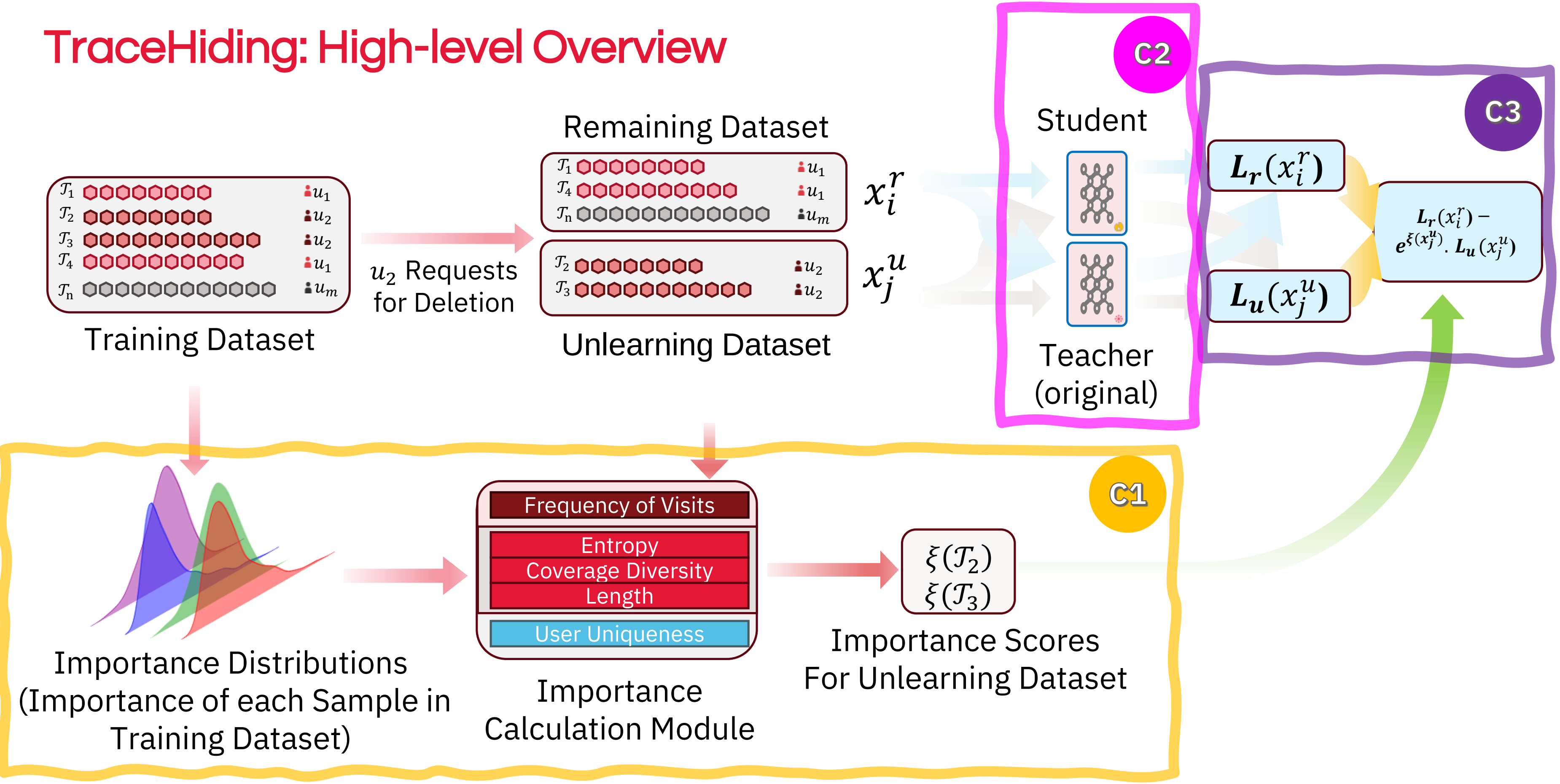


Our Key Idea: Not All Samples Are Equally Important



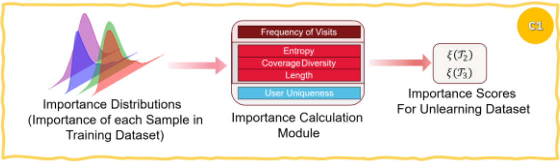
YORK U

TraceHiding: High-level Overview

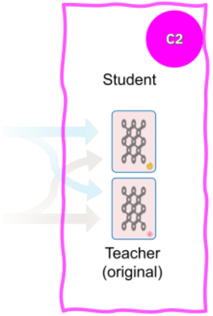


YORK U

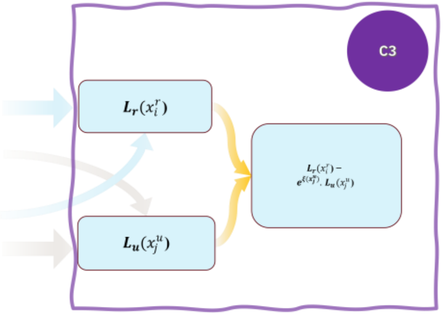
TraceHiding: Components



C1) Data-driven Importance Score



C2) Teacher-Student Model



C3) Loss Function

TraceHiding

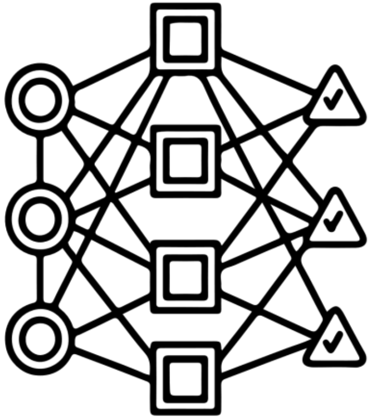
- achieves superior **unlearning accuracy**
- competitive **membership inference attack** (MIA) resilience
- up to **40× speedup** over retraining with minimal test accuracy loss

Conclusions

Concluding Remarks

Deep learning models for **mobility data analytics**:

- **TrajLearn**: Trajectory Prediction
- **MobiNetForecast**: Mobility Network Prediction
- **TULHOR**: Trajectory-User Linking
- **TraceHiding**: Machine Unlearning for Mobility Data



Reproducibility

- See paper details for open-source **datasets**, **source code**, and **models**

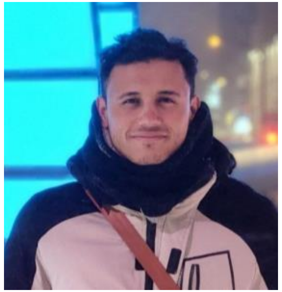
Privacy Considerations & Ethical Guidelines

- Need to balance the use of mobility data with privacy/ethical considerations

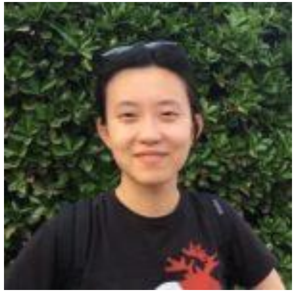
Credits

Amirhossein Nadiri

Mahmoud Alsaeed



Ali Faraji



Jing Li

Nina Yanin

Ghadeer Abuoda

TrajLearn: Trajectory Prediction Learning using Deep Generative Models. A. Nadiri, J. Ling, A. Faraji, G. Abuoda, M. Papagelis. **ACM TSAS, 11 (3), 2025.**

Mobility Network Forecasting: A Trajectory-based Contact Prediction Approach. A. Nadiri, J. Ling, G. Abuoda, M. Papagelis. **ACM TKDD, submitted.**

Trajectory-User Linking using Higher-order Mobility Flow Representations. M. Alsaeed, A. Agrawal, M. Papagelis. **IEEE MDM 2023.**

TraceHiding: An algorithmic framework for machine unlearning in mobility data. A. Faraji, M. Papagelis. **ACM TSAS, submitted**.

Point2Hex: Higher-order Mobility Flow Data and Resources. A. Faraji, J. Ling, G. Alix, M. Alsaeed, N. Yanin, A. Nadiri, M. Papagelis. **ACM SIGSPATIAL 2023**.

Questions

Thank you!