GeoAI for Human Mobility Analytics and Privacy Protection

Song Gao

Geospatial Data Science Lab

University of Wisconsin-Madison

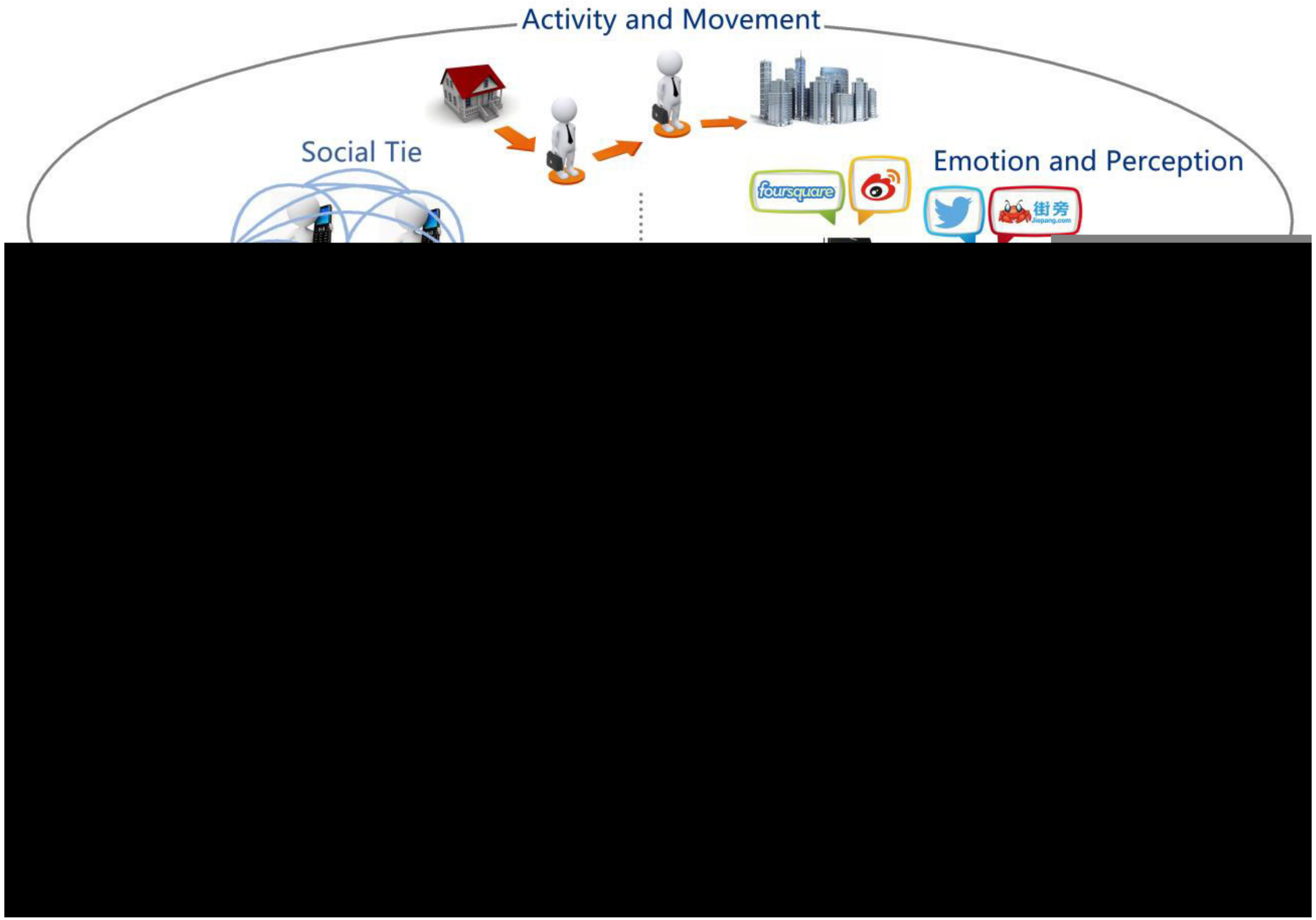
Email: song.gao@wisc.edu

GeoDS Lab: https://geods.geography.wisc.edu/

- A multidisciplinary team across geography, statistics, and computer sciences, etc.
- Research on theory, methods, and applications of geospatial data science and GeoAI.
- Advance GIScience to support geospatial industry innovation and solve real-world challenges.

Data-Driven Human Mobility Research

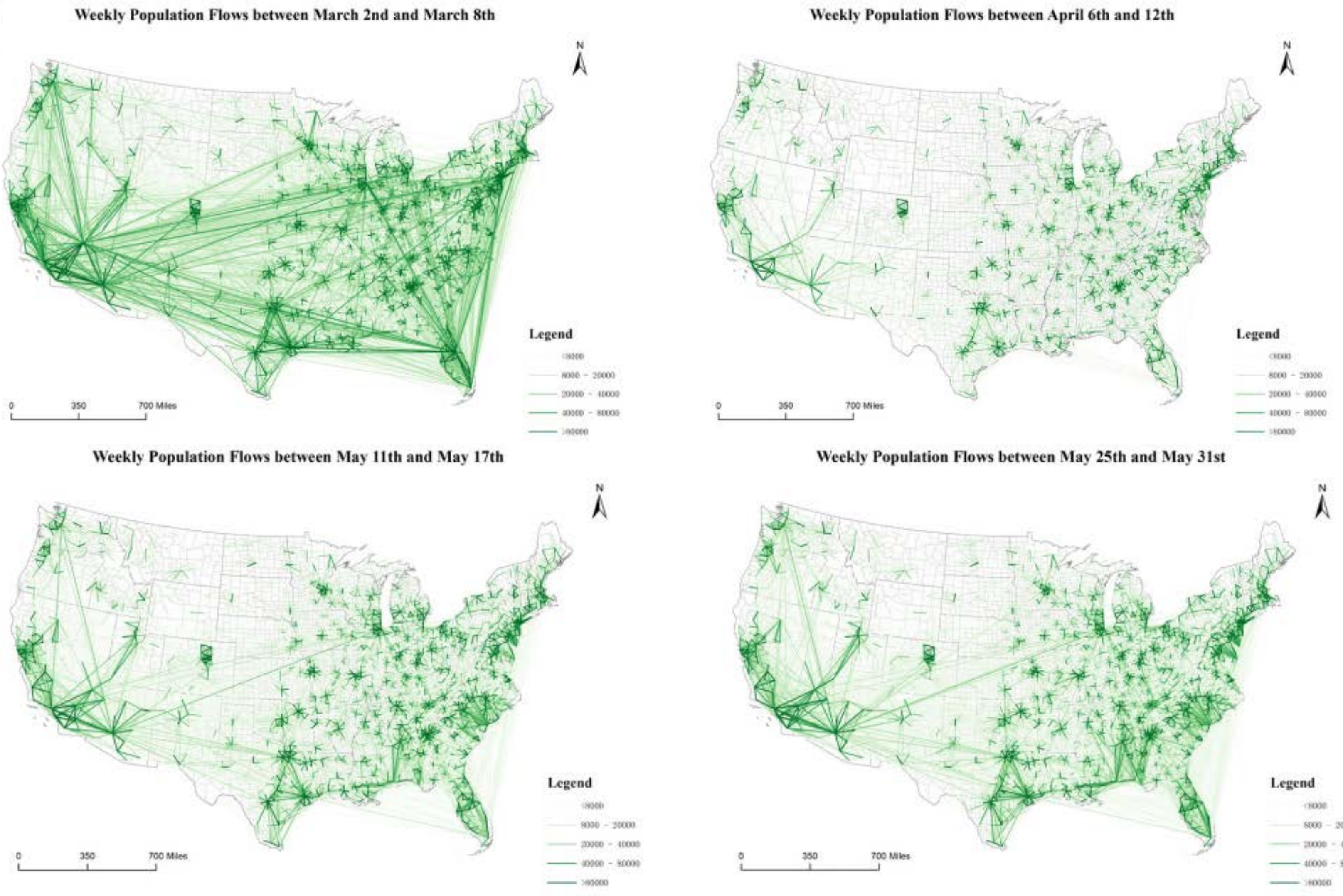
- Trip Surveys
- GPS trajectories
- Call detail records (CDRs)
- Social media
- Location-based apps



[redacted]

Liu, Y et al.. (2015). Social sensing: A new approach to understanding our socioeconomic environments. ***Annals of AAG***. 105(3). 512-530.

Multiscale Dynamic O-D Flow Open Data



Geospatial Data Science Lab UW-Madison

SAFEGRAPH

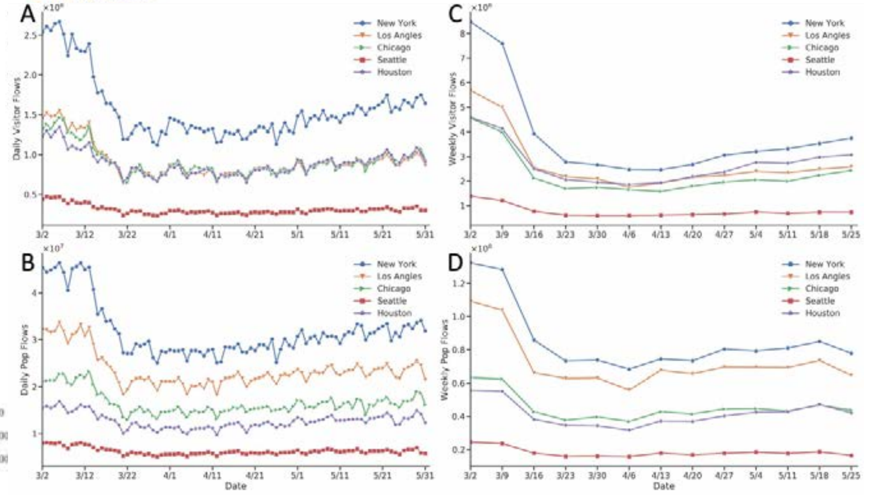
NSF

Multiscale Dynamic Human Mobility Flow Dataset in the U.S. during the COVID-19 Epidemic

GeoDS Lab, Department of Geography, University of Wisconsin-Madison.

Website · View Demo

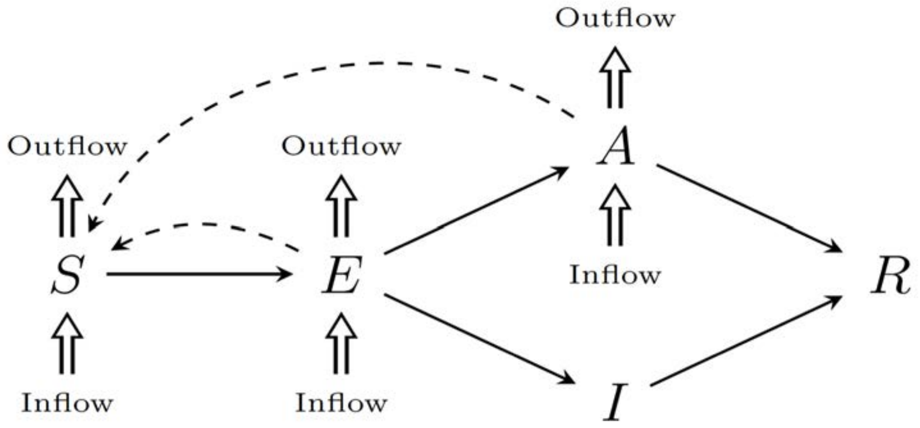
Table of Contents



Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., & Kruse, J. (2020). Multiscale Dynamic Human Mobility Flow Dataset in the US during the COVID-19 Epidemic. ***Scientific Data.*** *7, 390. https://www.nature.com/articles/s41597-020-00734-5*

https://github.com/GeoDS/COVID19USFlows

State-Specific Geospatial Modeling of COVID-19 Spread

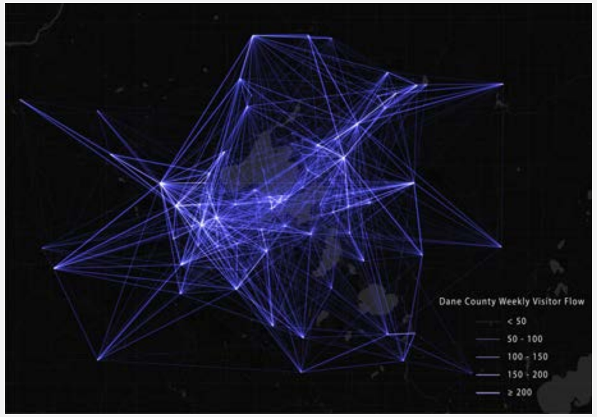


$$
\begin{cases}
\dfrac{dS_i}{dt} = -\dfrac{b_i S_i (A_i + \gamma E_i)}{P_i} + \boxed{\sum_{j \neq i} \alpha_t n_{ij} \dfrac{S_j}{P_j} - \sum_{j \neq i} \alpha_t n_{ji} \dfrac{S_i}{P_i}} \\
\dfrac{dE_i}{dt} = \dfrac{b_i S_i (A_i + \gamma E_i)}{P_i} - \dfrac{E_i}{D_e} + \boxed{\sum_{j \neq i} \alpha_t n_{ij} \dfrac{E_j}{P_j} - \sum_{j \neq i} \alpha_t n_{ji} \dfrac{E_i}{P_i}} \\
\dfrac{dI_i}{dt} = r_i \dfrac{E_i}{D_e} - c_I \dfrac{I_i}{D_c} - (1 - c_I) \dfrac{I_i}{D_l} \\
\dfrac{dA_i}{dt} = (1 - r_i) \dfrac{E_i}{D_e} - c_A \dfrac{A_i}{D_c} - (1 - c_A) \dfrac{A_i}{D_l} + \boxed{\sum_{j \neq i} \alpha_t n_{ij} \dfrac{A_j}{P_j} - \sum_{j \neq i} \alpha_t n_{ji} \dfrac{A_i}{P_i}} \\
\dfrac{dR_i}{dt} = c_I \dfrac{I_i}{D_c} + (1 - c_I) \dfrac{I_i}{D_l} + c_A \dfrac{A_i}{D_c} + (1 - c_A) \dfrac{A_i}{D_l}
\end{cases}
$$

A new mobility-network-based SEIR compartmental model

- Chen, S., Li, Q., Gao, S., Kang, Y., & Shi, X. (2020). State-specific projection of COVID-19 infection in the United States and evaluation of three major control measures. ***Scientific Reports***. 10, 22429. https://www.nature.com/articles/s41598-020-80044-3

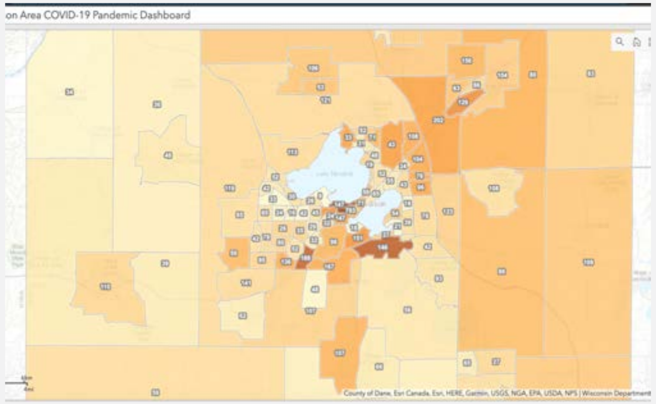
Intra-County Modeling of COVID-19 Spread

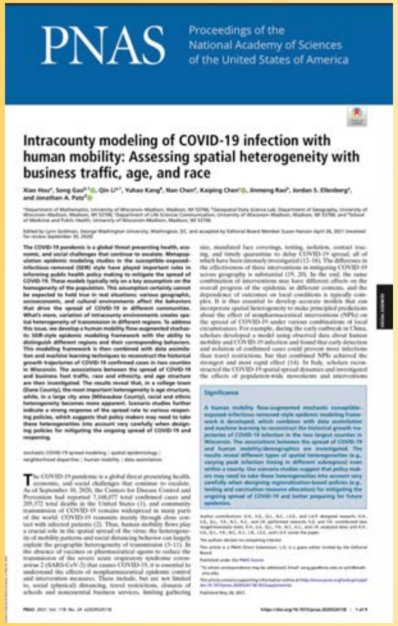


Mobility-based Modeling

Spatial Clustering

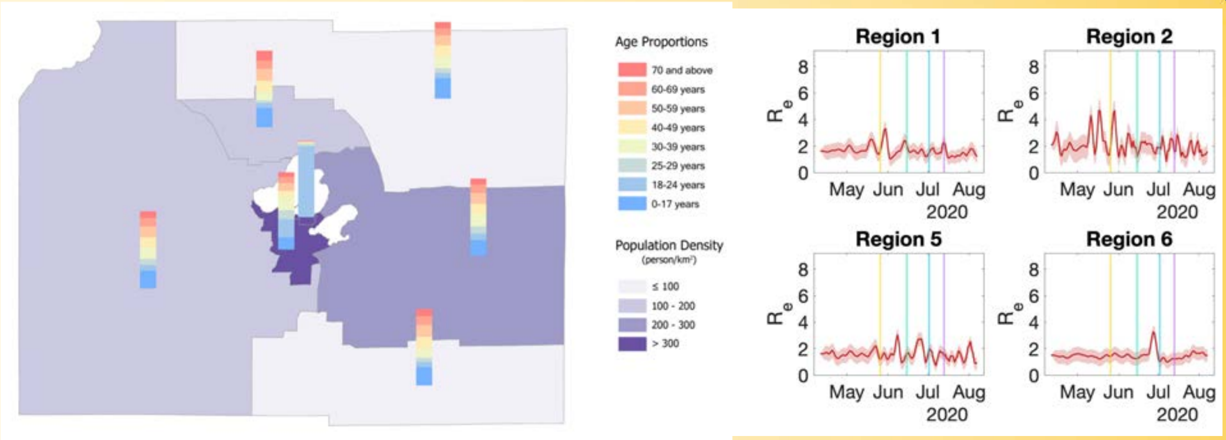
SEIR model




PNAS Proceedings of the National Academy of Sciences of the United States of America

Intracounty modeling of COVID-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race

Divide a county into multiple smaller regions according to traffic flows.



Hou, X., Gao, S., Li, Q., Kang, Y., Chen, N., Chen, K., Rao, J., Ellenberg, J. & Patz, J. A. (2021). Intra-county modeling of COVID-19 infection with human mobility: assessing spatial heterogeneity with business traffic, age and race. ***Proceedings of the National Academy of Sciences,*** 2021, 118 (24) e2020524118

GeoAI Meetings

- **ACM SIGSPATIAL GeoAI Workshops (GeoAI'2017, 2018, 2019, 2021)**
- **AAG GeoAI and Deep Learning Symposium (AAG 2018, 2019, 2020)**
- **CPGIS & HPSCIL@CUG GeoAI Webinars (2020, 2021)**

Topics: deep learning, intelligent transportation, urban informatics, cartography, business intelligence, public health, digital agriculture, disaster management, GIS tool development, etc.


International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond

Krzysztof Janowicz, Song Gao, Grant McKenzie, Yingjie Hu & Budhendra Bhaduri

To cite this article: Krzysztof Janowicz, Song Gao, Grant McKenzie, Yingjie Hu & Budhendra Bhaduri (2019): GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2019.1684500

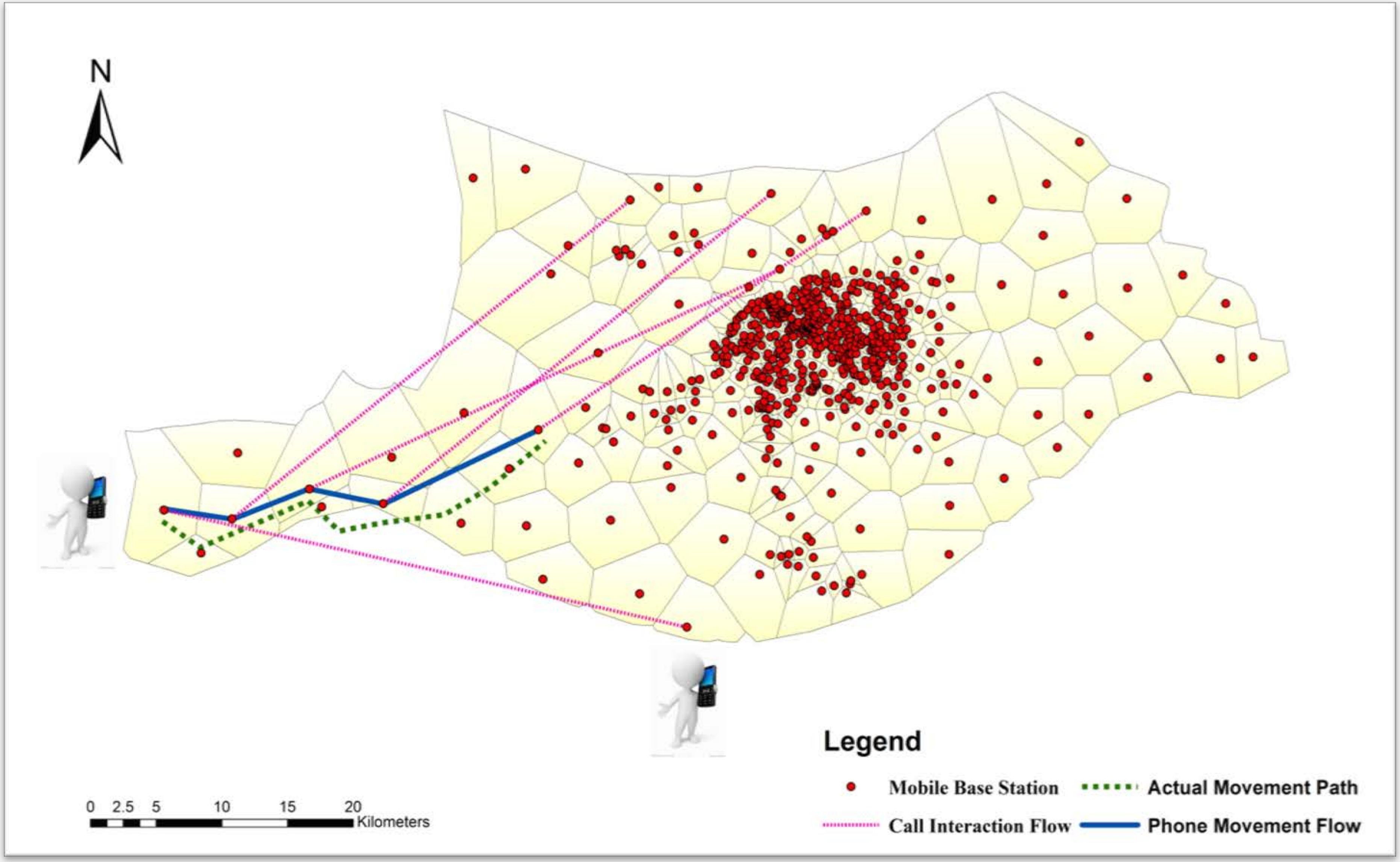
To link to this article: https://doi.org/10.1080/13658816.2019.1684500

AMERICAN ASSOCIATION of GEOGRAPHERS

ANNUAL MEETING • April 7-11, 2021 • Seattle, WA

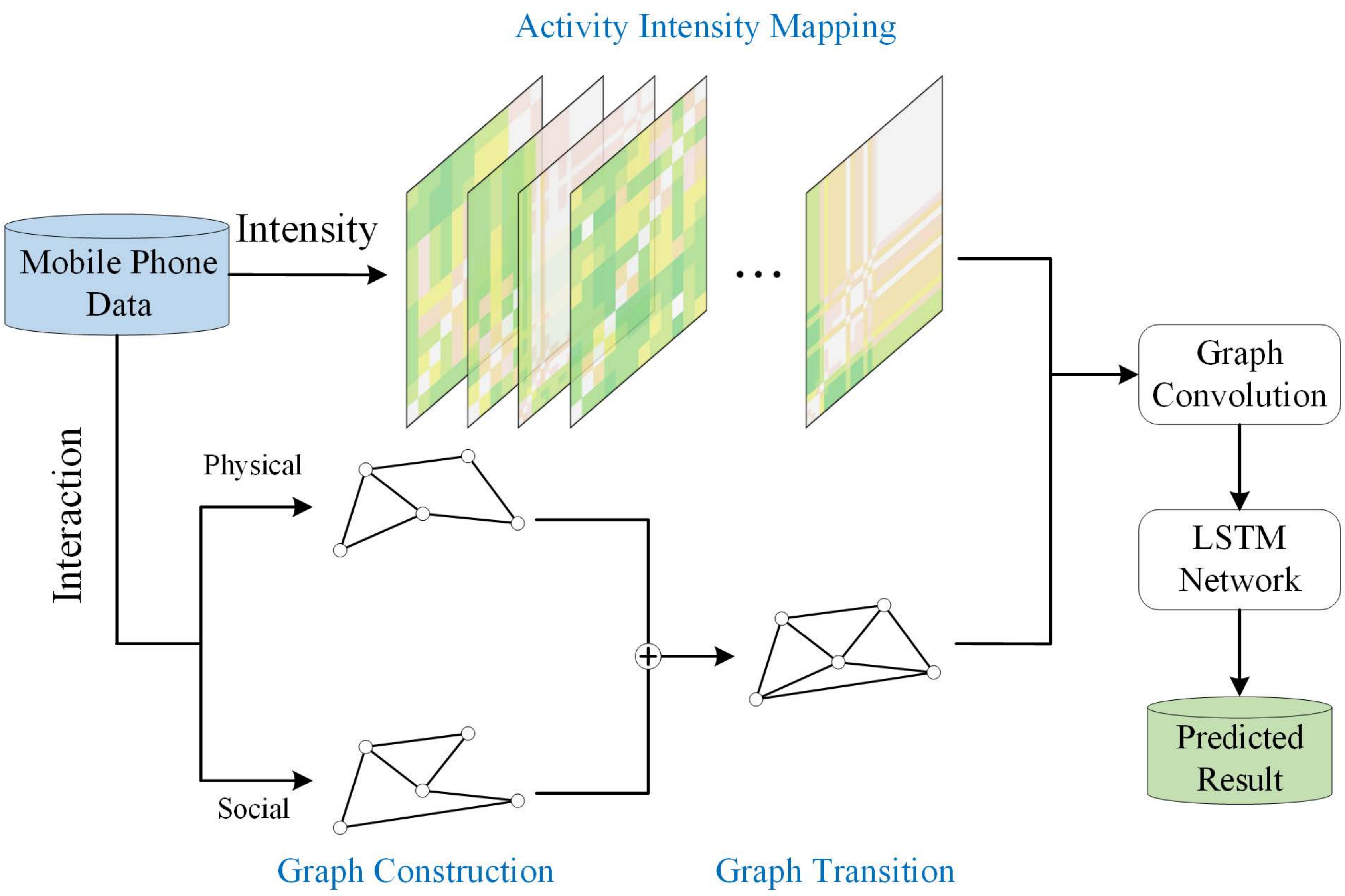
Prediction of human activity intensity using the interactions in physical and social spaces through graph convolutional networks

Phone Call Detail Records (CDRs)



Gao, S., Liu, Y., Wang, Y., & Ma, X. (2013). Discovering spatial interaction communities from mobile phone data. *Transactions in GIS,* 17(3), 463-481.

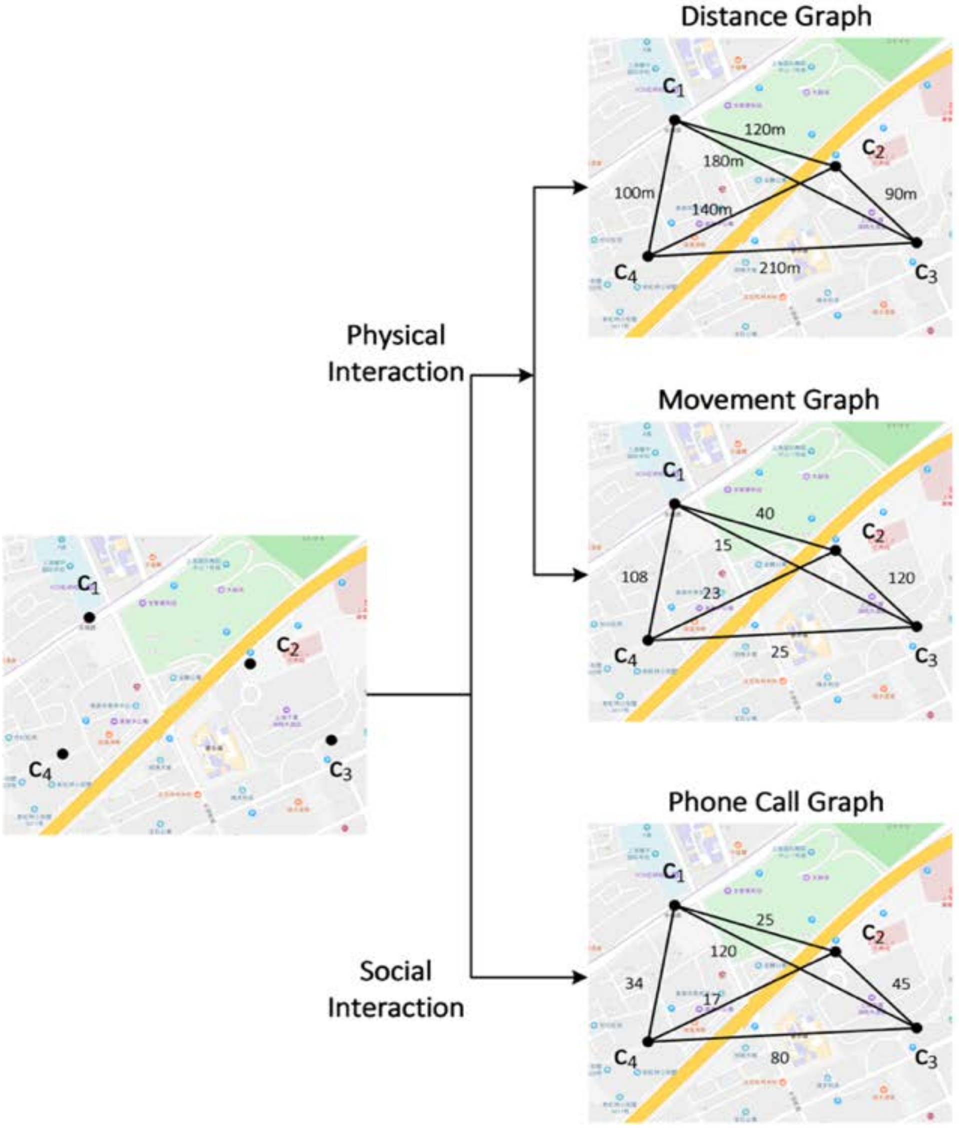
Problem Statement



Li, M., Gao, S., Lu, F., Liu, K., Zhang, H., & Tu, W. (2021). Prediction of human activity intensity using the interactions in physical and social spaces through graph convolutional networks. *International Journal of Geographical Information Science*, 1-28.

Multiple Types of Spatial Interactions

- Geographical distance
- Movement interaction
- Social network interaction



$$G_D=\begin{bmatrix} 0 & 120 & 180 & 100 \\ 120 & 0 & 90 & 140 \\ 180 & 90 & 0 & 210 \\ 100 & 140 & 210 & 0 \end{bmatrix} \quad (a)$$

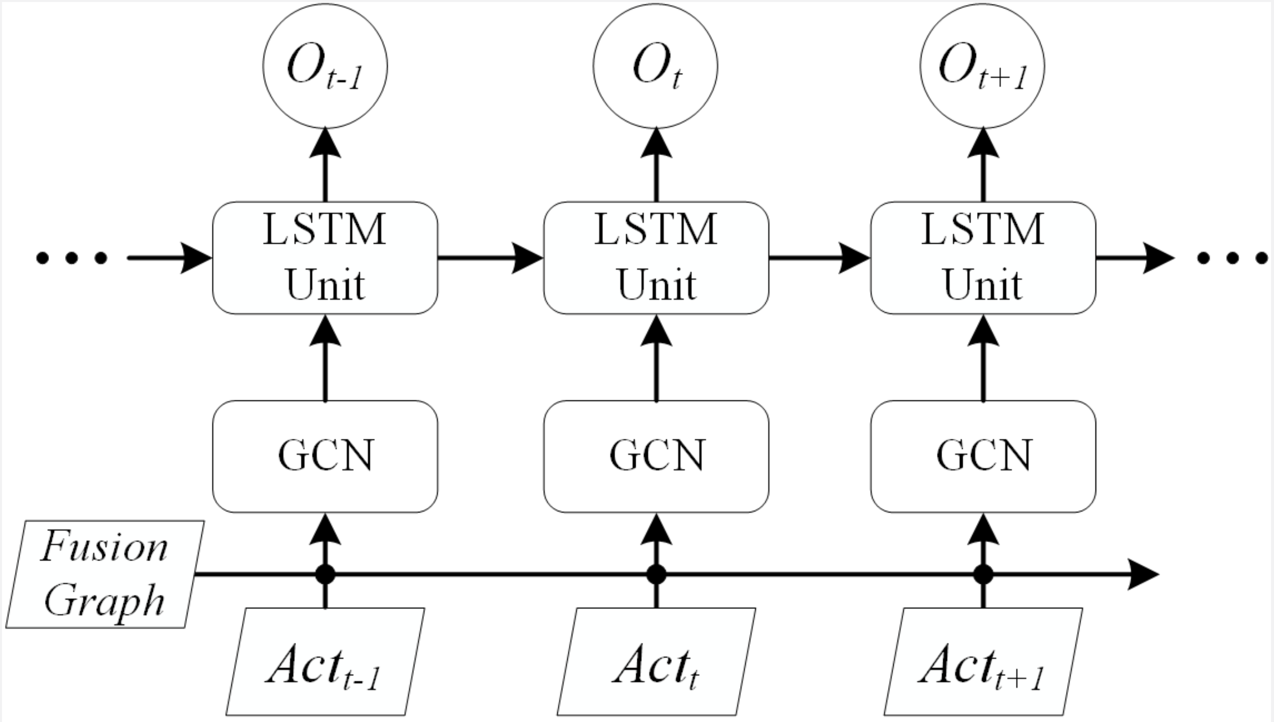
$$G_M=\begin{bmatrix} 0 & 40 & 15 & 108 \\ 40 & 0 & 120 & 23 \\ 15 & 120 & 0 & 25 \\ 108 & 23 & 25 & 0 \end{bmatrix} \quad (b)$$

$$G_S=\begin{bmatrix} 0 & 25 & 120 & 34 \\ 25 & 0 & 45 & 17 \\ 120 & 45 & 0 & 80 \\ 34 & 17 & 80 & 0 \end{bmatrix} \quad (c)$$

The proposed AIP-IPS model

AIP-IPS: Activity Intensity Prediction using the Interactions in Physical and Social spaces.

- Graph convolutional network (GCN) layer: modeling the impacts of spatial interactions
- Long short-term memor (LSTM) layer: modeling the temporal tendency and variations of human activity intensity



Li, M., Gao, S., Lu, F., Liu, K., Zhang, H., & Tu, W. (2021). Prediction of human activity intensity using the interactions in physical and social spaces through graph convolutional networks. *International Journal of Geographical Information Science*, 1-28.

The proposed AIP-IPS model

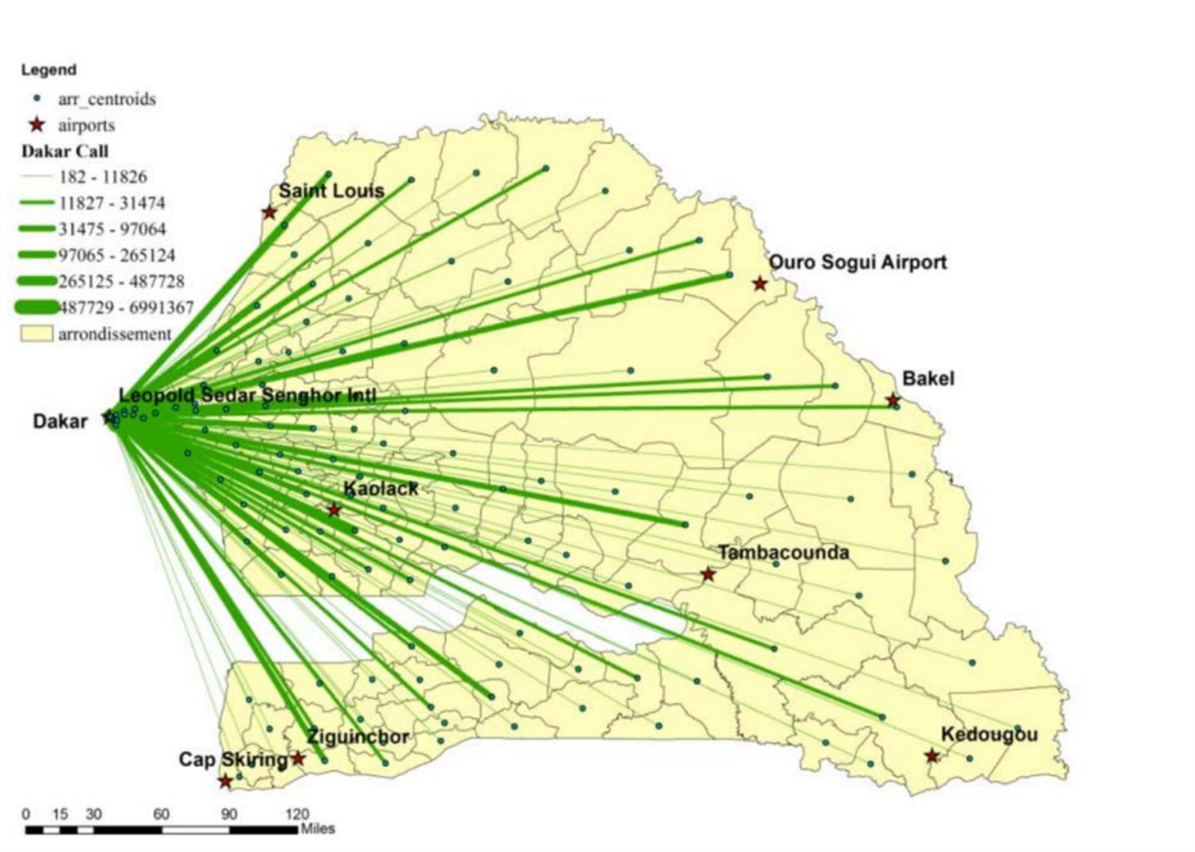
Parameter	Value	Parameter	Value
Distance threshold	30000	Distance graph weight	0.3
Movement threshold	20	Movement graph weight	0.1
Social threshold	20	Social graph weight	0.6
LSTM hidden layers	1	Time sequence length	6
GCN hidden layers	1	Num of hidden units	64
Graph convolution kernel size	1	Time step	6

Case Study

About 9 million users in Senegal through the "**Data for Development (D4D)**" Senegal challenge (De Montjoye et al., 2014).

Dataset 1: a tower-to-tower call frequency data for all towers on an hourly basis of the entire population

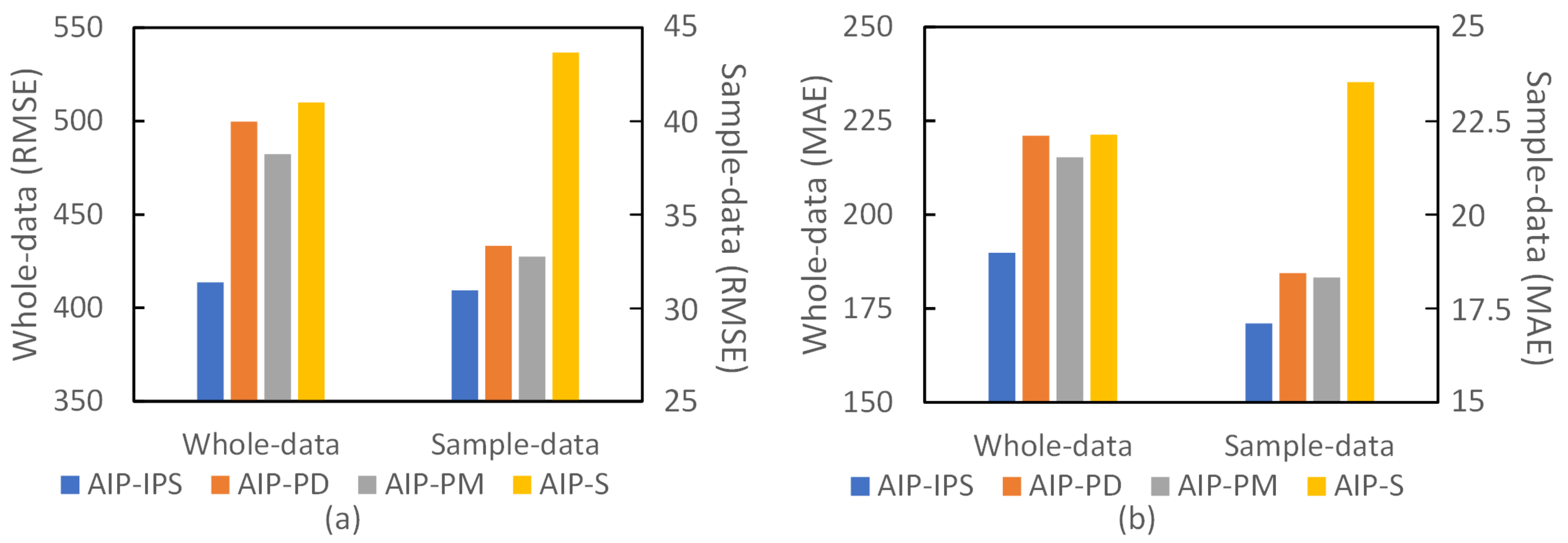
Dataset 2: a cell tower-level CDR data at the individual level for about 10% of mobile phone users



Results: Graph Fusion Performance

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(A'_{c_{i,t}} - A_{c_{i,t}})^2}{N*T}}$$

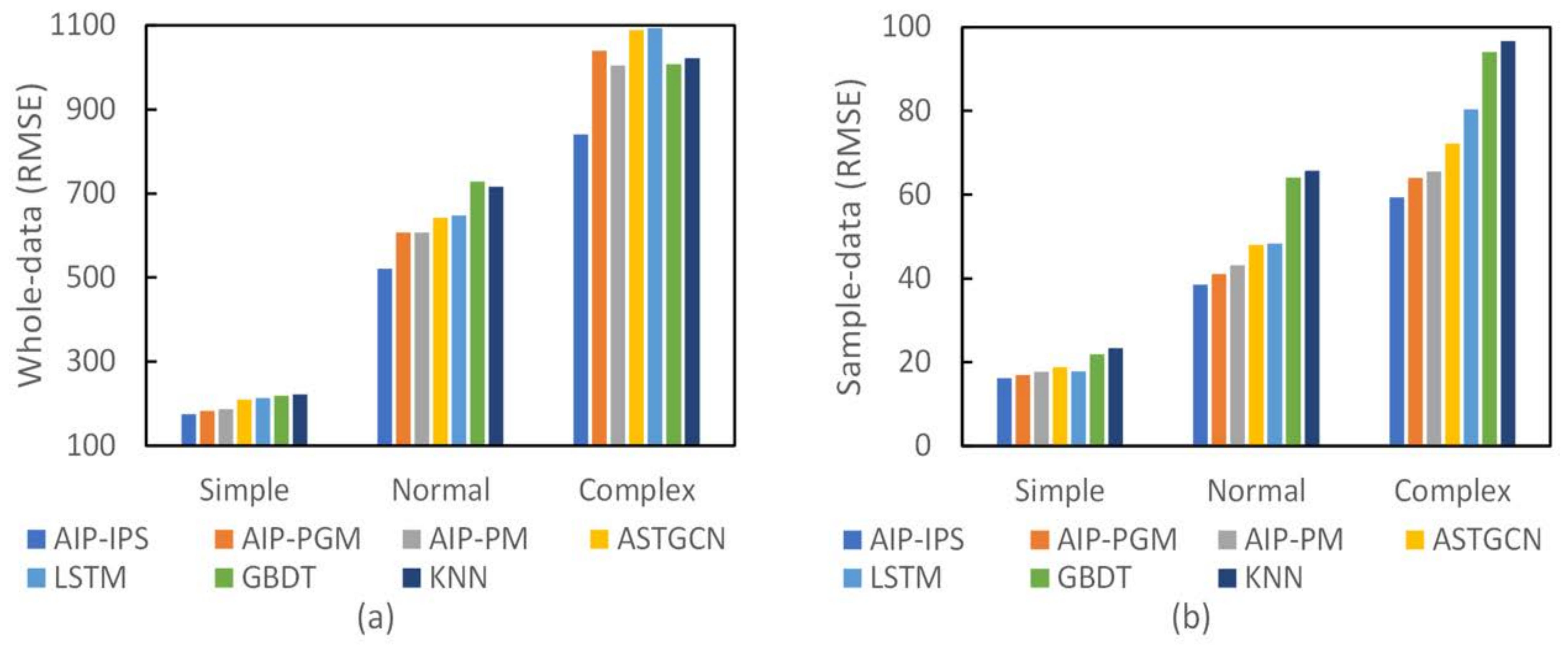
$$MAE = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left|A'_{c_{i,t}} - A_{c_{i,t}}\right|}{N*T}$$



Results: Comparison with Other Models

Category	Method	Whole-data		Sample-data	
		RMSE	MAE	RMSE	MAE
Deep Learning	AIP-IPS	**413.62**	**189.88**	**30.93**	**17.11**
	AIP-PM	482.22	215.27	34.36	18.32
	AIP-PGM	488.41	219.68	33.01	17.87
	ASTGCN	518.56	251.02	37.79	20.56
	LSTM	522.10	271.19	39.24	19.57
Classical Machine Learning	KNN	536.71	192.80	50.62	19.18
	GBDT	538.51	203.71	49.12	19.10
Time series	ARIMA	1529.88	823.03	120.21	67.53

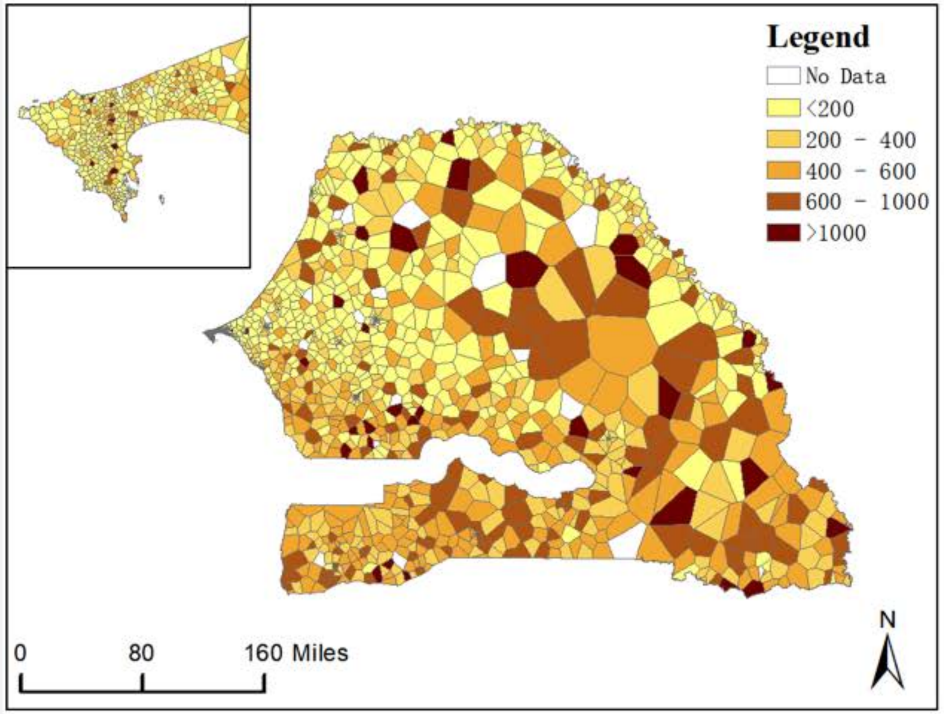
Results: Impact of Entropy (Regularity)



Summary

- **The importance of integrated physical and social interactions on activity intensity estimation**

- **A deep learning prediction model using fused GCN and LSTM layers**



Li, M., Gao, S., Lu, F., Liu, K., Zhang, H., & Tu, W. (2021). Prediction of human activity intensity using the interactions in physical and social spaces through graph convolutional networks. *International Journal of Geographical Information Science*, 1-28.

Trajectory (Location) Privacy Protection

GeoPrivacy

- **Geoprivacy** refers to an individual's rights to prevent the disclosure of personal sensitive locations including but not limit to their home, workplace, daily activities, or travel trips.

Gao, S., Rao, J., Liu, X., Kang, Y., Huang, Q., & App, J. (2019). Exploring the effectiveness of geomasking techniques for protecting the geoprivacy of Twitter users. Journal of Spatial Information Science, 2019(19), 105-129.

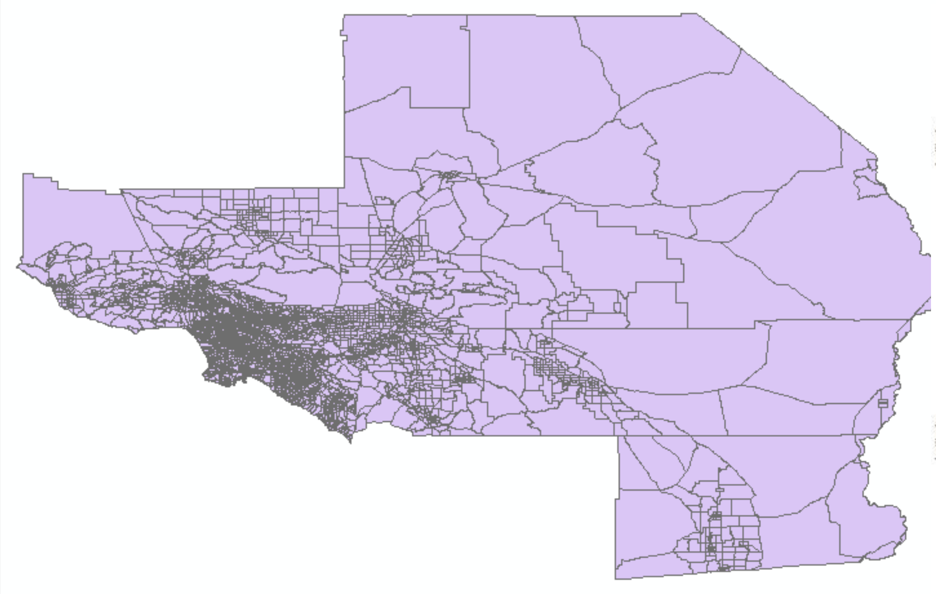
Location Privacy Protection Methods

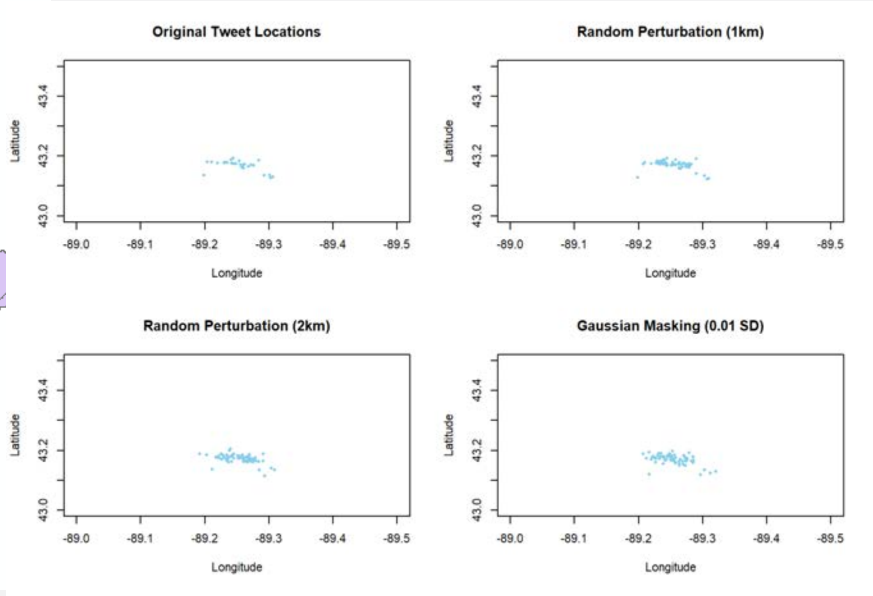
Aggregation

- Aggregating individual location into anonymized large-group characteristics (e.g. city blocks, census tracts)
- User privacy becomes a k-anonymity problem
- Reduces the spatial resolution of analysis and reduce the effectiveness of the analysis

Geomasking

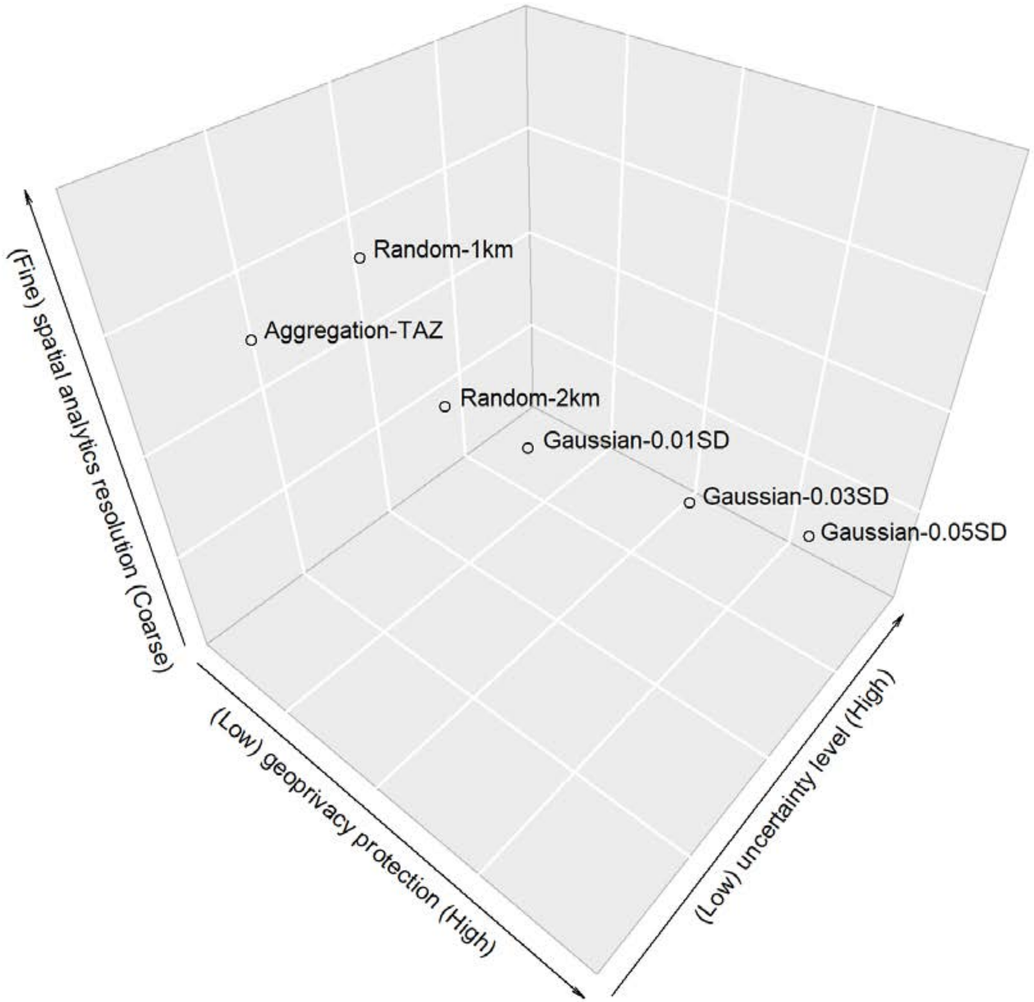
- Random Perturbation
- Donut Method
- Gaussian Method
- Voronoi Masking





Method Evaluation Framework

- **Spatial Analytics Resolution**
- **Uncertainty Level**
- **Geoprivacy Protection Capability**



Gao, S., Rao, J., Liu, X., Kang, Y., Huang, Q., & App, J. (2019). Exploring the effectiveness of geomasking techniques for protecting the geoprivacy of Twitter users. Journal of Spatial Information Science, 2019(19), 105-129.

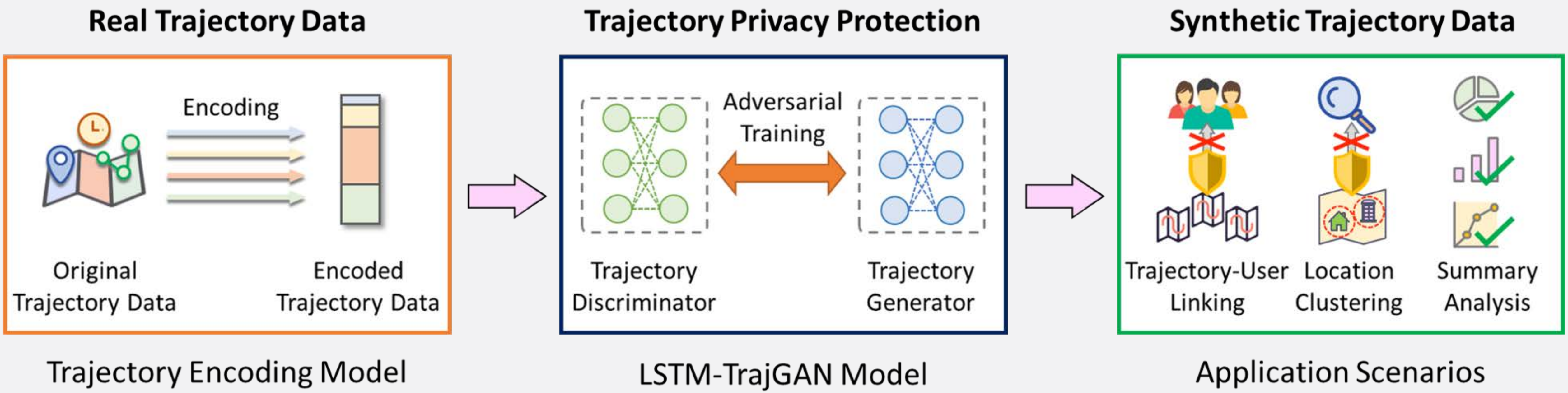
Research Questions

- **RQ1: How effective is our proposed deep learning model in protecting the trajectory creators from being re-identified? (i.e., privacy protection effectiveness)**

- **RQ2: Can the synthetic trajectories preserve the semantic features (spatial-temporal-thematic characteristics) compared to real trajectories? (i.e., utility)**

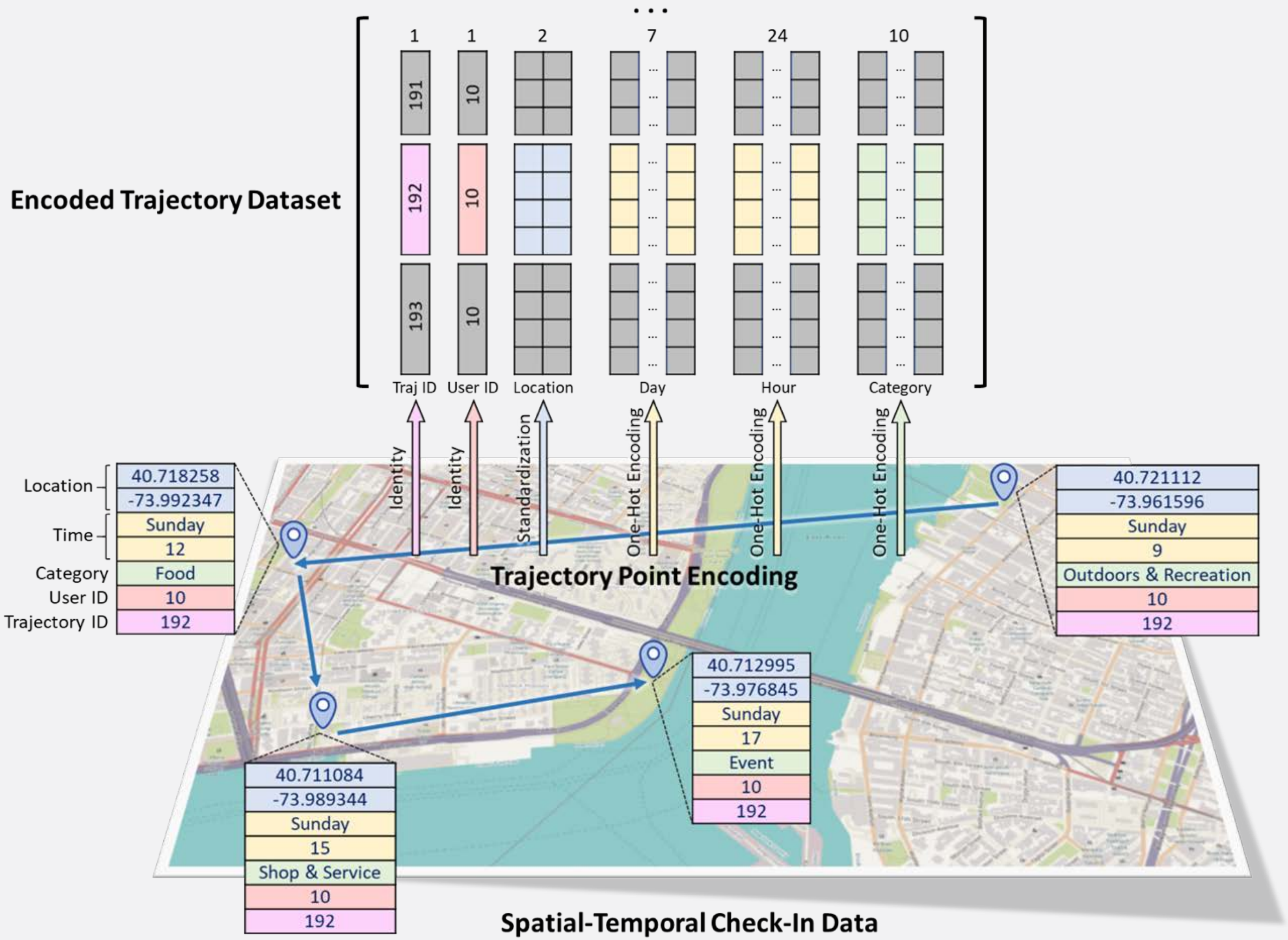
Rao, J., Gao, S., Kang, Y., & Huang, Q. (2020). LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection. *In the Proceedings of the 11th International Conference on Geographic Information Science (GIScience 2021), No. 12; pp. 12:1–12:17.* DOI: 10.4230/LIPIcs.GIScience.2021.12

Overview of the LSTM-TrajGAN Model

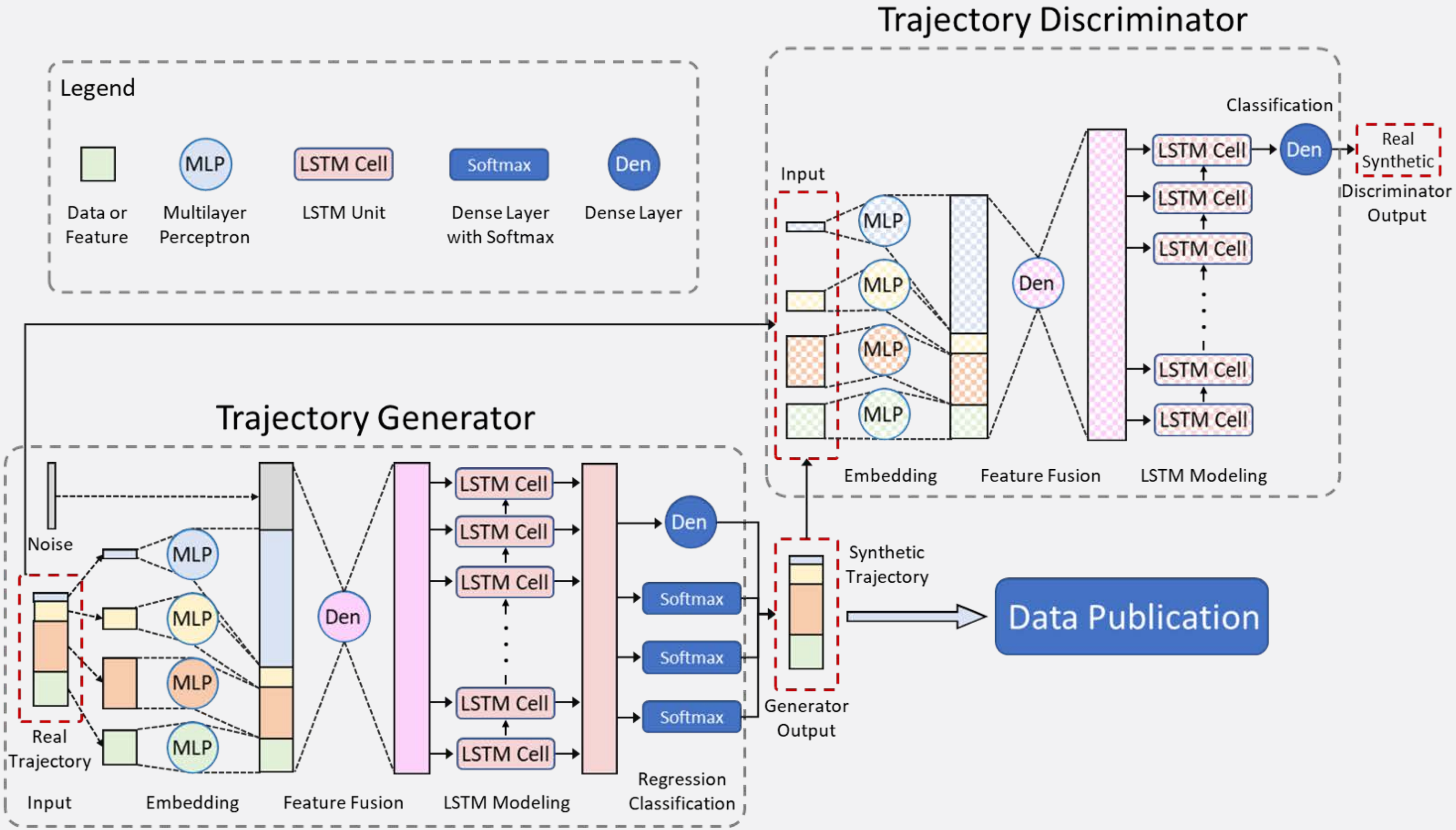


Rao, J., Gao, S., Kang, Y., & Huang, Q. (2020). LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection. *In the Proceedings of the 11th International Conference on Geographic Information Science (GIScience 2021), No. 12; pp. 12:1–12:17.* DOI: 10.4230/LIPIcs.GIScience.2021.12

Trajectory Encoding Model

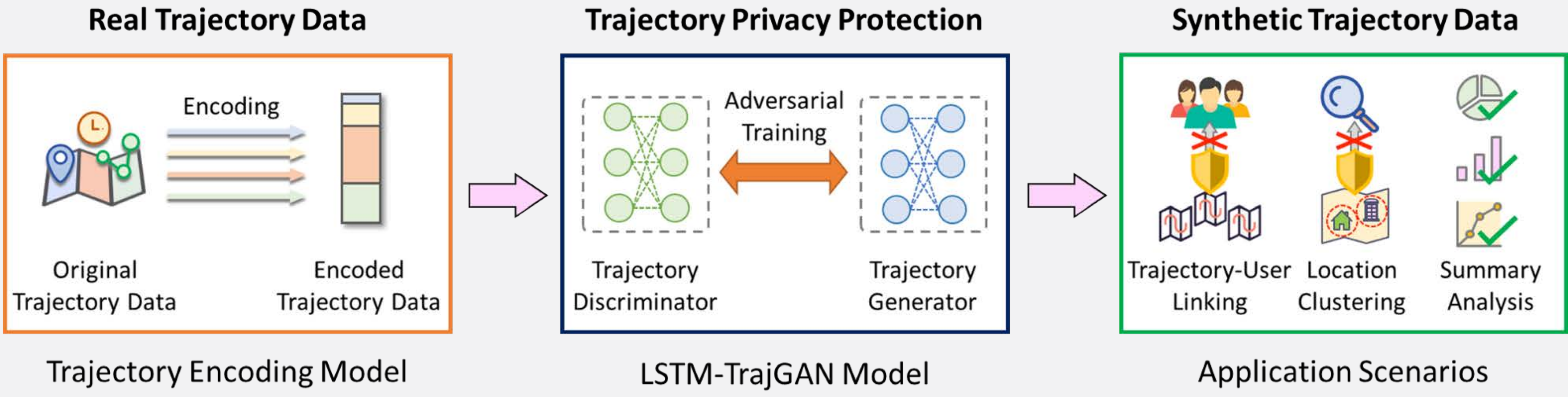


LSTM-TrajGAN Model



Rao, J., Gao, S., Kang, Y., & Huang, Q. (2020). LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection. *In the Proceedings of the 11th International Conference on Geographic Information Science (GIScience 2021), No. 12; pp. 12:1–12:17.* DOI: 10.4230/LIPIcs.GIScience.2021.12

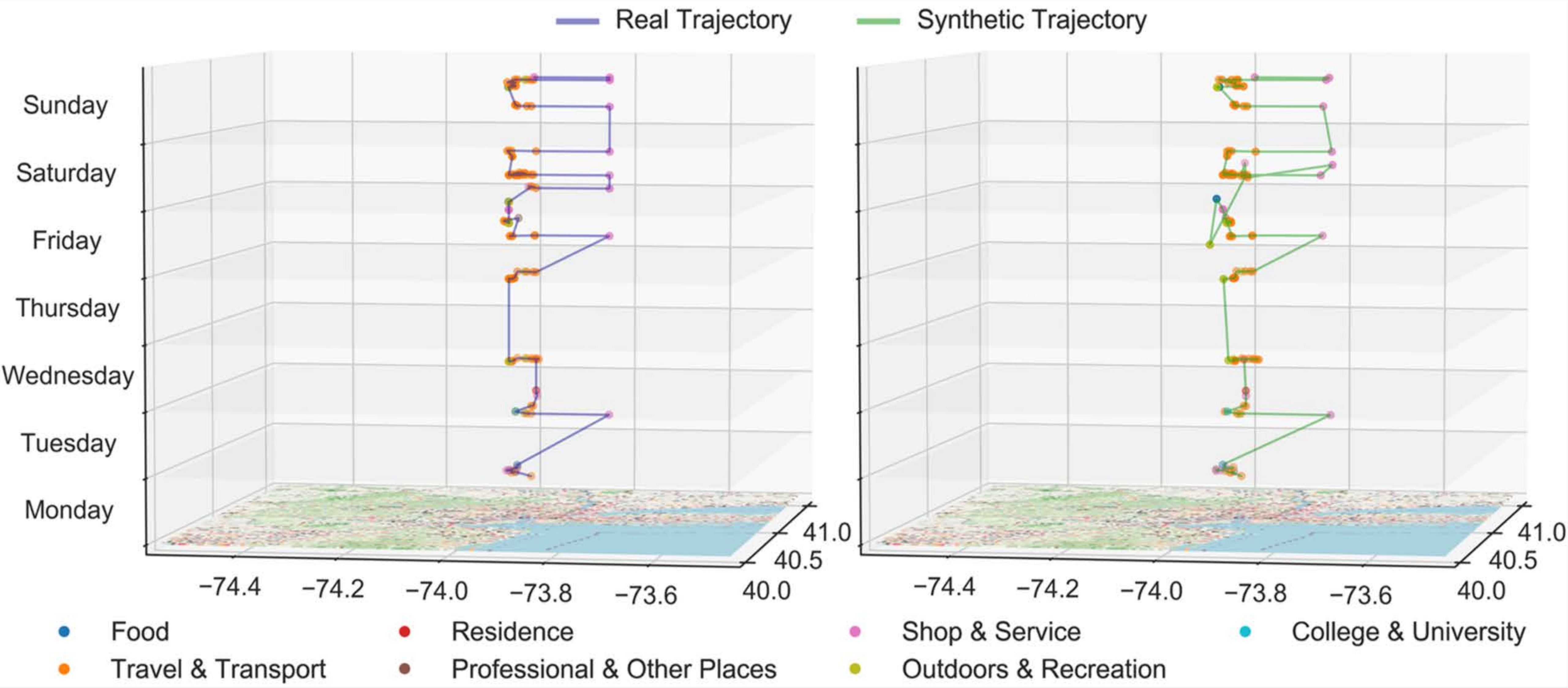
LSTM-TrajGAN Loss Function



$$TrajLoss(y^r, y^p, t^r, t^s) = \alpha L_{BCE}(y^r, y^p) + \beta L_s(t^r, t^s) + \gamma L_t(t^r, t^s) + c L_c(t^r, t^s)$$

Rao, J., Gao, S., Kang, Y., & Huang, Q. (2020). LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection. *In the Proceedings of the 11th International Conference on Geographic Information Science (GIScience 2021), No. 12; pp. 12:1–12:17.* DOI: 10.4230/LIPIcs.GIScience.2021.12

Trajectory Generation



Rao, J., Gao, S., Kang, Y., & Huang, Q. (2020). LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection. *In the Proceedings of the 11th International Conference on Geographic Information Science (GIScience 2021), No. 12; pp. 12:1–12:17.* DOI: 10.4230/LIPIcs.GIScience.2021.12

Trajectory Generation Results

The privacy protection effectiveness of different privacy protection methods on the Trajectory User Linking task.

Method	ACC@1	ACC@5	Macro-F1	Macro-P	Macro-R
Original	0.938	0.976	0.925	0.937	0.927
RP (Spatial Only)	0.777	0.934	0.758	0.806	0.764
RP (Spatial-Temporal)	0.668	0.888	0.640	0.711	0.654
Gaussian (Spatial Only)	0.561	0.832	0.522	0.573	0.537
Gaussian (Spatial-Temporal)	0.486	0.766	0.431	0.488	0.470
LSTM-TrajGAN	**0.459**	**0.722**	**0.381**	**0.429**	**0.428**

Note: RP stands for Random Perturbation; Gaussian stands for Gaussian Geomasking

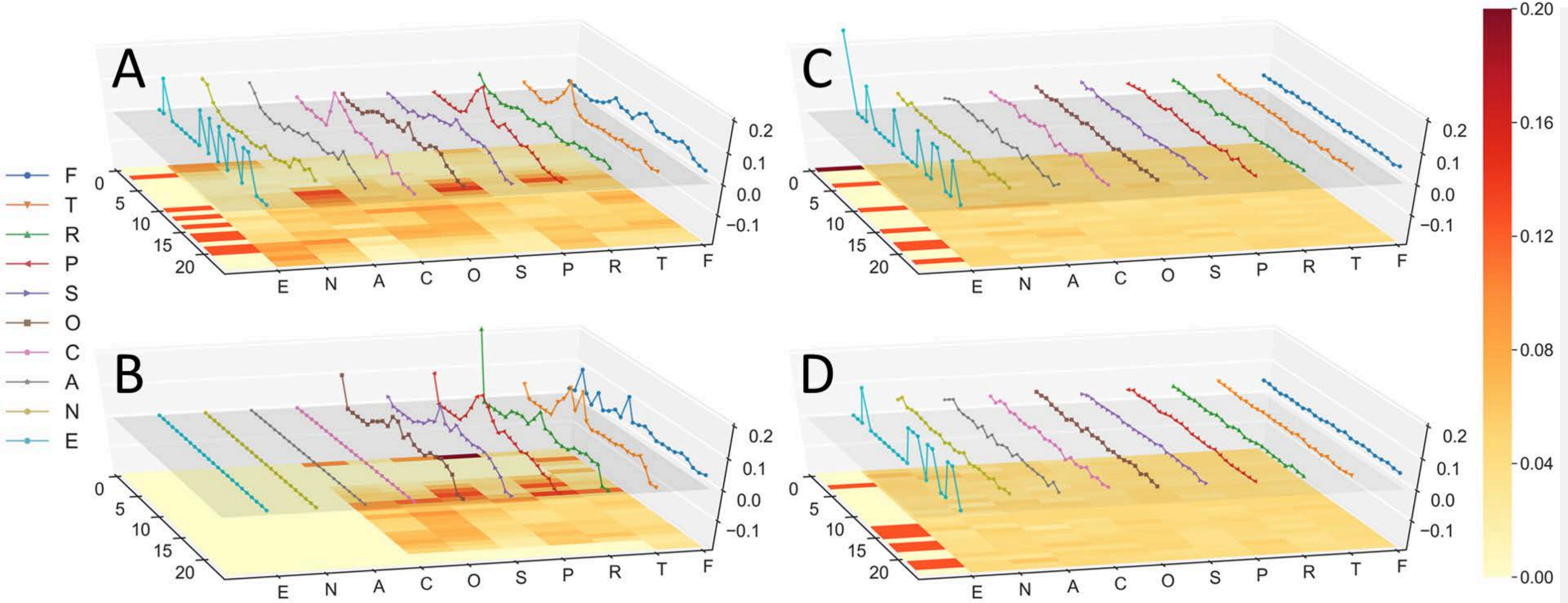
Trajectory Generation Results

Spatial characteristics evaluation based on Hausdorff Distance and Jaccard Index:

Method	Hausdorff Distance				Jaccard Index			
	Min	Max	Std	Mean	Min	Max	Std	Mean
RP	0.001	0.006	0.001	0.004	0.000	0.977	0.194	0.763
Gaussian	0.001	0.034	0.005	0.014	0.000	0.933	0.231	0.478
LSTM-TrajGAN	0.001	0.046	0.006	0.012	0.000	0.951	0.234	0.582

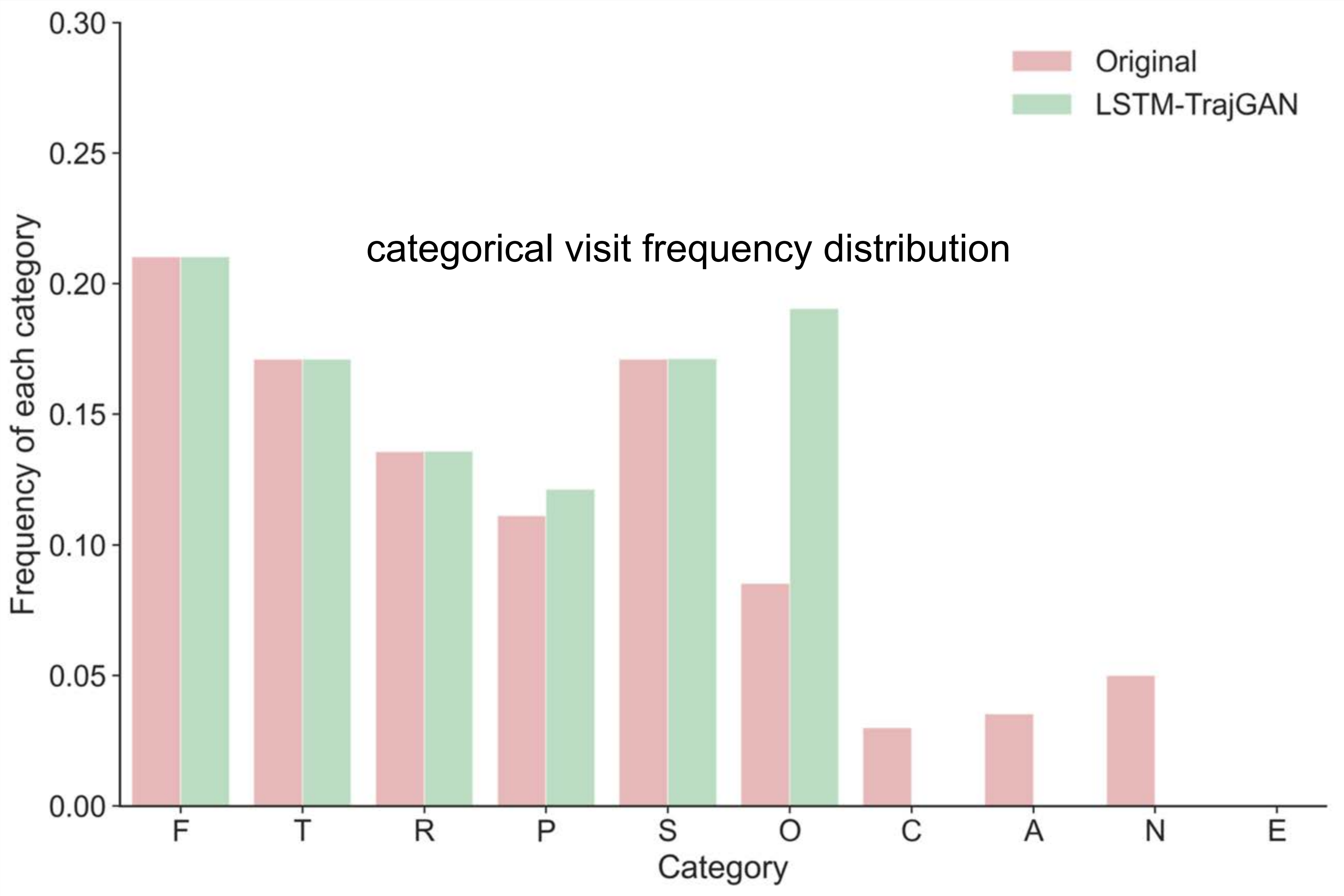
Note: RP stands for Random Perturbation; Gaussian stands for Gaussian Geomasking

Trajectory Generation Results



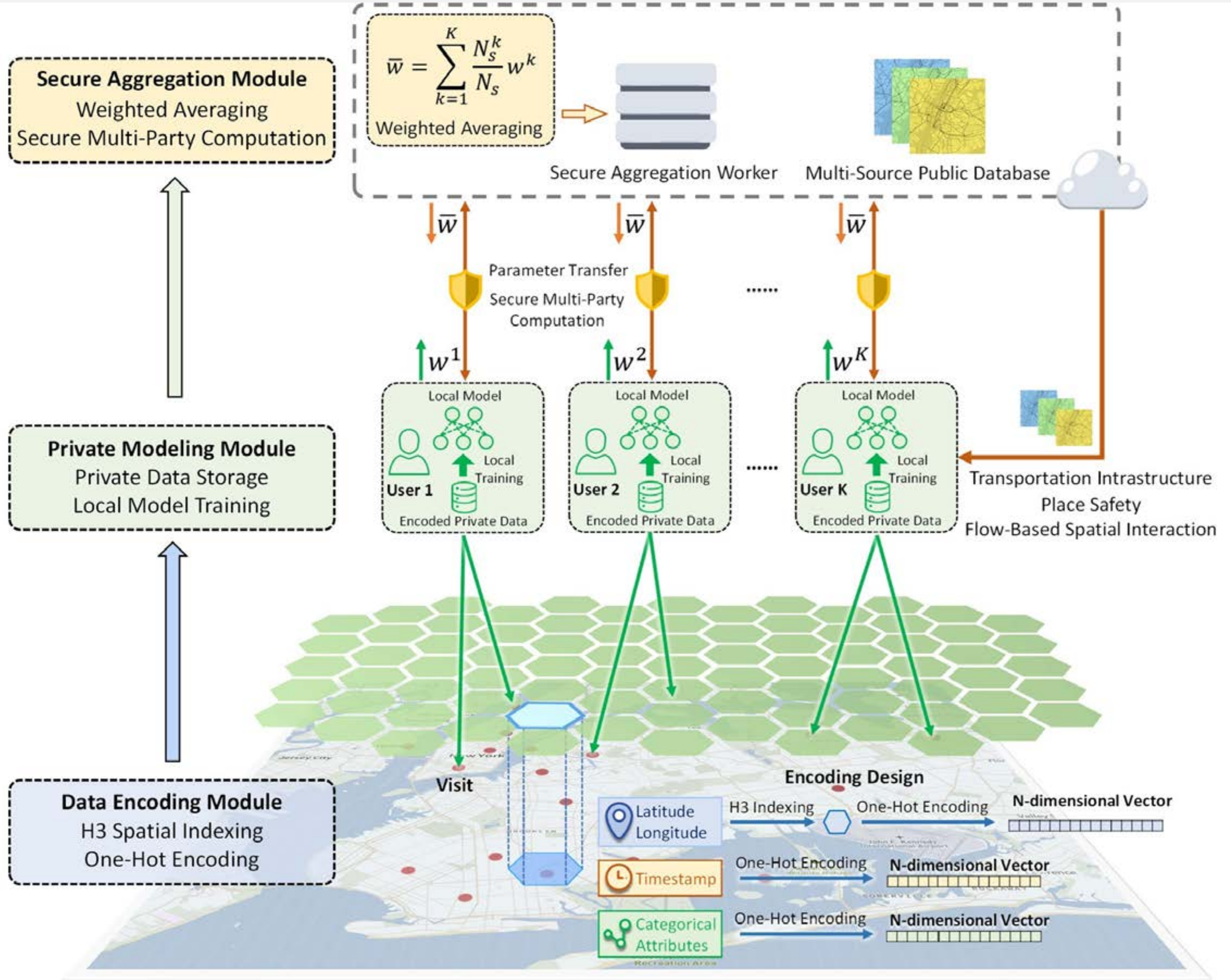
The hourly temporal visit probability distribution for each POI category by (A) Original data, (B) LSTM-TrajGAN, (C) Random Perturbation (within 24 hours), and (D) Gaussian Geomasking data.

Trajectory Generation Results



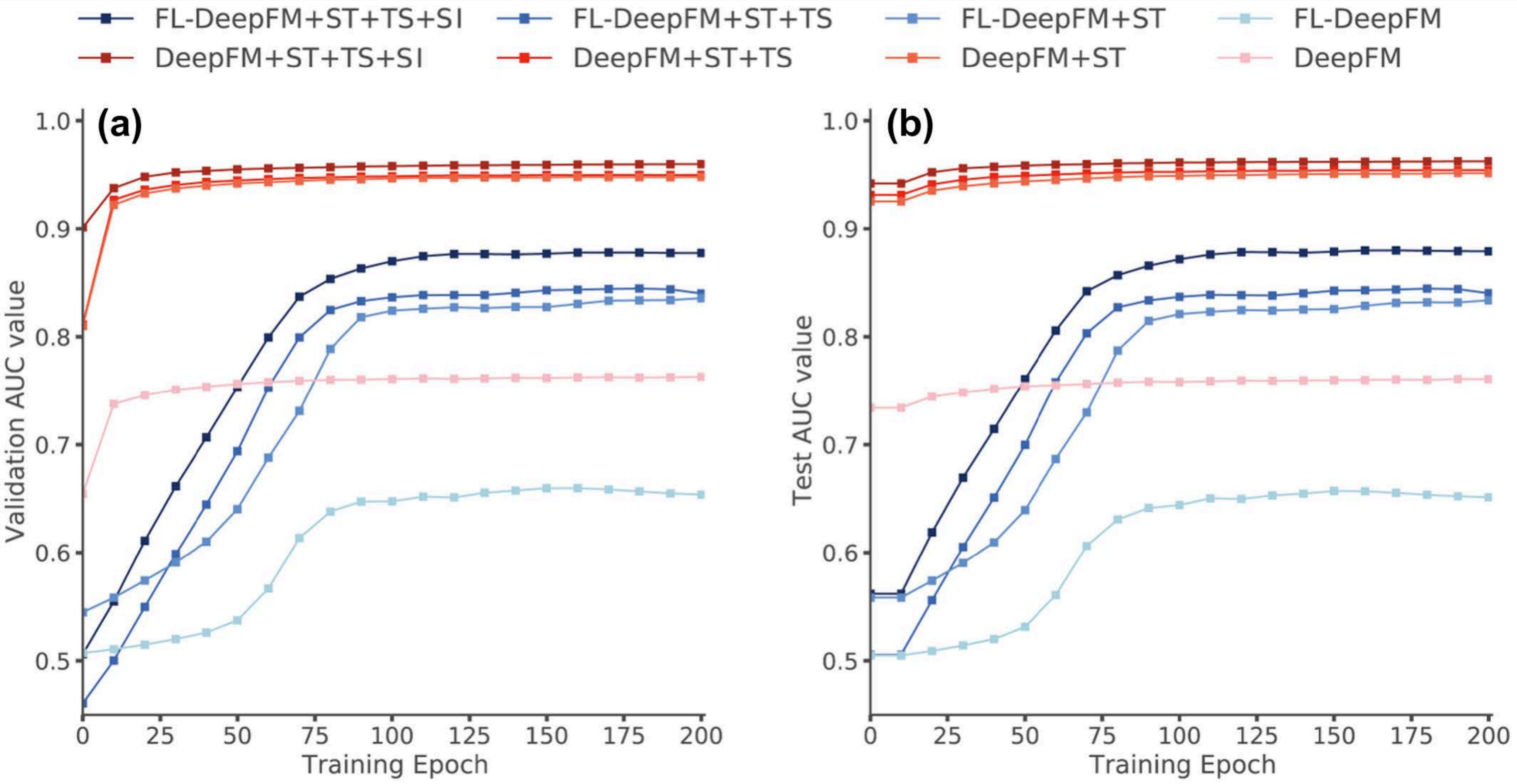
Rao, J., Gao, S., Kang, Y., & Huang, Q. (2020). LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection. *In the Proceedings of the 11th International Conference on Geographic Information Science (GIScience 2021), No. 12; pp. 12:1–12:17.* DOI: 10.4230/LIPIcs.GIScience.2021.12

Federated Learning-Based Framework



Rao, J., Gao, S., Li, M., & Huang, Q. (2021) A privacy-preserving framework for location recommendation using decentralized collaborative machine learning. Transactions in GIS, 25(3), 1153-1175.

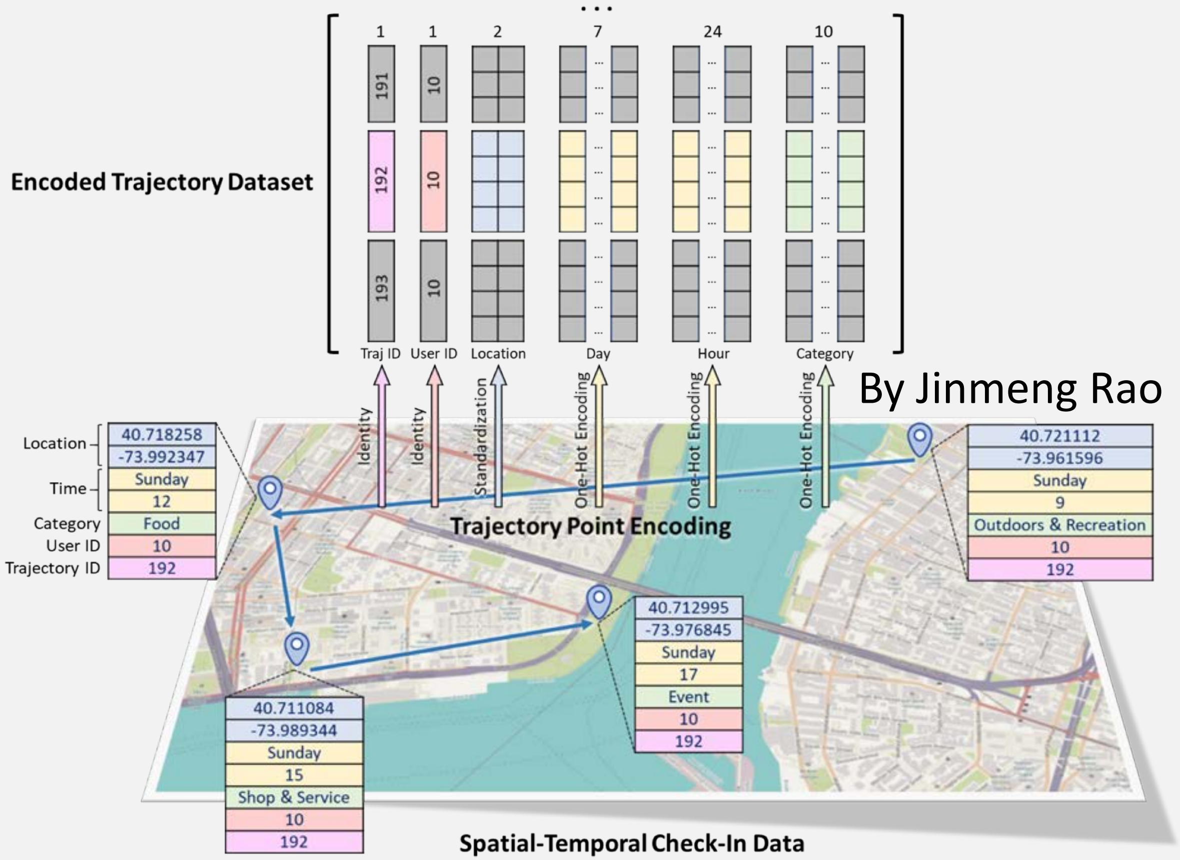
Federated Learning-Based Framework



Rao, J., Gao, S., Li, M., & Huang, Q. (2021) A privacy-preserving framework for location recommendation using decentralized collaborative machine learning. Transactions in GIS, 25(3), 1153-1175.

Challenges

- Structured vs. unstructured inputs
- Design an appropriate loss function
- Effective learning strategies
- Labelled data samples



By Jinmeng Rao

Remarks

- Technology and data-driven advancement in GeoAI
- AI <--> GIScience (Spatially explicit AI methods)
- Bias, Ethics, and Privacy Issues in GeoAI
- Benchmark Data and Tool Sharing for Replicability and Reproducibility in GeoAI research

GeoAI Special Issues

GeoInformatica

An International Journal on Advances of Computer Science for Geographic Information Systems

https://www.springer.com/journal/10707/updates/17573556

Call for papers: Geospatial artificial intelligence

Geoinformatica welcomes submissions to our new special issue on "Geospatial artificial intelligence".

Geospatial artificial intelligence (GeoAI) provides novel approaches for addressing a variety of problems in both our natural environment and human society. Novel artificial intelligence (AI) techniques are transforming a range of fields from computer vision and natural language processing to autonomous driving. With the advancements of software and hardware technologies, deep learning algorithms, scalable computation platforms, and availability of high-resolution geospatial data are enabling the fast-growing field of GeoAI. This special issue (SI) aims to bring together the latest research on this topic. Particularly, this SI encourages the development and sharing of open GeoAI tools that can be reused for GIScience & GeoInformatics research and education. Potential topics of interest include, but are not limited to:

Transactions in GIS

https://onlinelibrary.wiley.com/page/journal/14679671/homepage/featured-collections#spatialdomain

Open Calls for Papers

Call for Papers: Special Issue on Explainable Artificial Intelligence in the Spatial Domain (X-GeoAI)
Submission Deadline: November 30, 2021

Special Issue on Virtual Geographic Environment Construction Based on Ubiquitous Geographic Information
Deadline: October 20, 2021

Call for Papers: Special Issue on Explainable Artificial Intelligence in the Spatial Domain (X-GeoAI)

Development in artificial intelligence (AI) research in combination with the large-scale availability of data has led to a great leap towards a new era of data intensive science. Within Geographic Information Science (GIScience), GeoAI encompasses research directions that employ AI methods to better analyze and predict complex geographic phenomena. Detection of terrain features or predictive analytics in crime activity, natural disasters or health are some examples that indicate the great potential of combining AI with geographic information. Machine learning is a core element of GeoAI applications; it is used to extract generalizable patterns that describe a phenomenon in question, in the form of mathematical models. These models excel in efficiency, scalability and accuracy, however, their complex nature impedes their descriptive capabilities greatly affecting explainability and human comprehension. Thus, a clear reasoning path from data to

Thank you!

Geospatial Data Science Lab
UW-Madison

https://geods.geography.wisc.edu

Twitter: @gissong