Balancing Public Health Need and Privacy Concerns

GEORGETOWN UNIVERSITY McCourt School of Public Policy | MASSIVE DATA INSTITUTE | TECH & SOCIETY

MDI Scholar: Yundi "Wendy" Shi, MS in Data Science for Public Policy, SP2025; MDI Scholar

Research Team: Giulia Pullano, Ph.D, Postdoctoral Fellow; Shweta Bansal, Ph.D.; Kobbi Nissim, Ph.D., Georgetown University

Introduction

Differential privacy is a widely adopted standard for protecting individual data while preserving analytical value. This study is among the first to implement both Central Differential Privacy (CDP) [1] and Local Differential Privacy (LDP) [2] in mobility data for disease modeling [3]. We focus on the role of human mobility in inter-county infectious disease spread, as captured through mathematical models.

Key Objective: The main goal of this research to examine the trade-off between privacy protection and model performance.

Figure 1: Daily Individual Mobility Data in San Diego

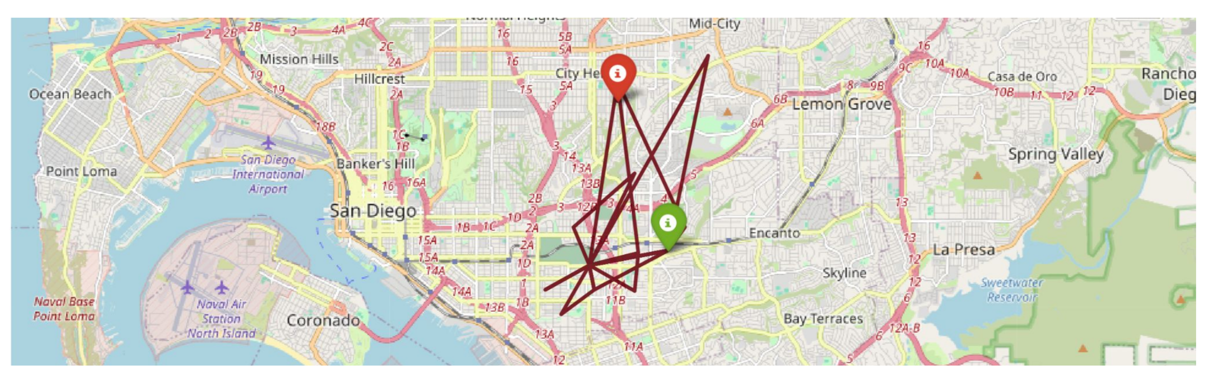
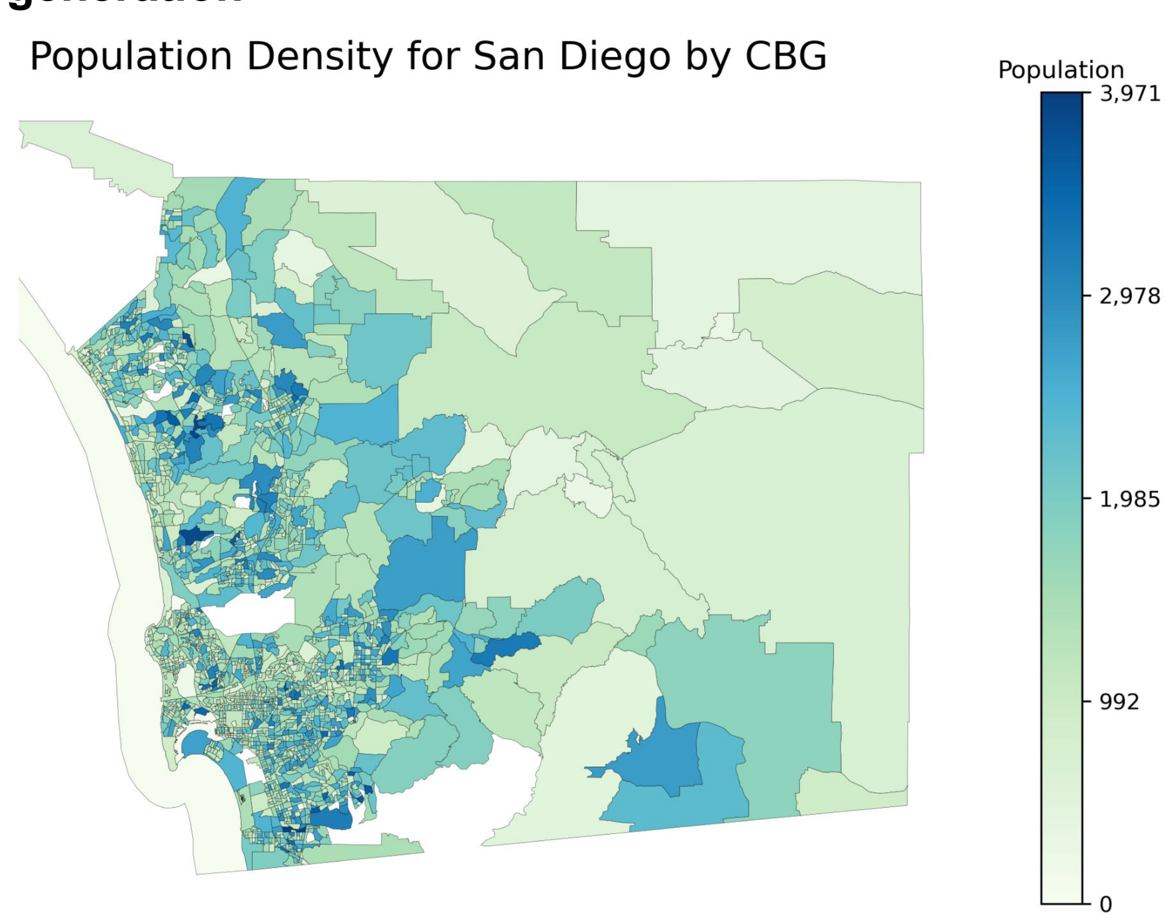


Figure 2: Shapefile & pop density for trajectory generation



Methodology

- Generating 10,000 synthetic mobility trajectories for each of three states—NY, CA, and WS—using the `scikit-mobility` Python package.
- We use the population density and shape-file by Census Block Group (CBG) to guide mobility data generation.
- Differential privacy was applied at two levels:
  - **State-level**: Aggregated mobility matrices
  - **Individual-level**: Daily CBG visit counts
- Apply noise to daily averaged visit counts to the `OpenDP` library. Convert and aggregate matrix mobility matrices as inputs to an disease model simulating COVID-19 spread from March 15, 2020.

Figure 3: Logged outflow weight and proportion of links being zero for LDP (top) and CDP (bottom)

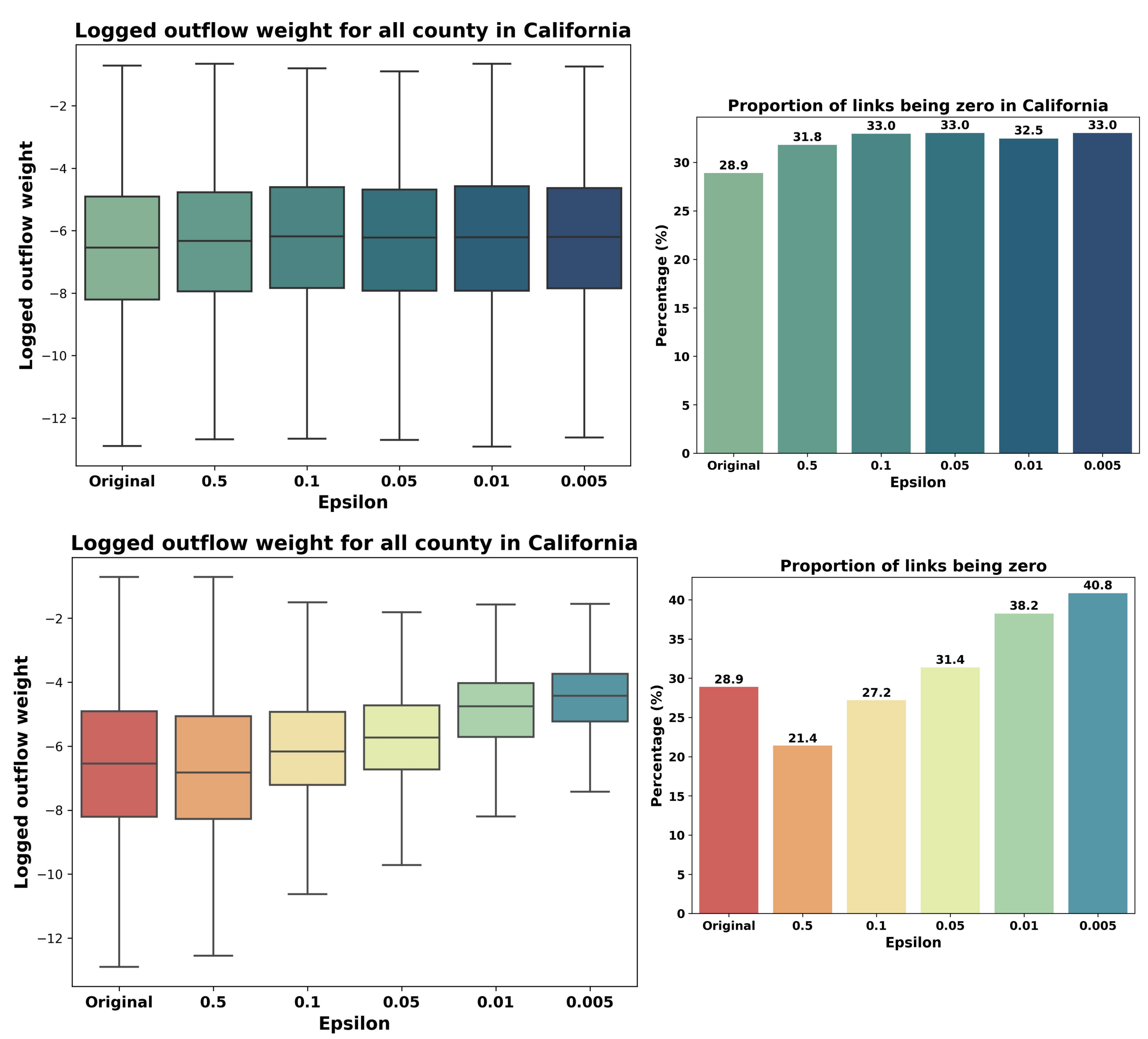


Figure 4: Arrival time (left) and peak time (right) with LDP

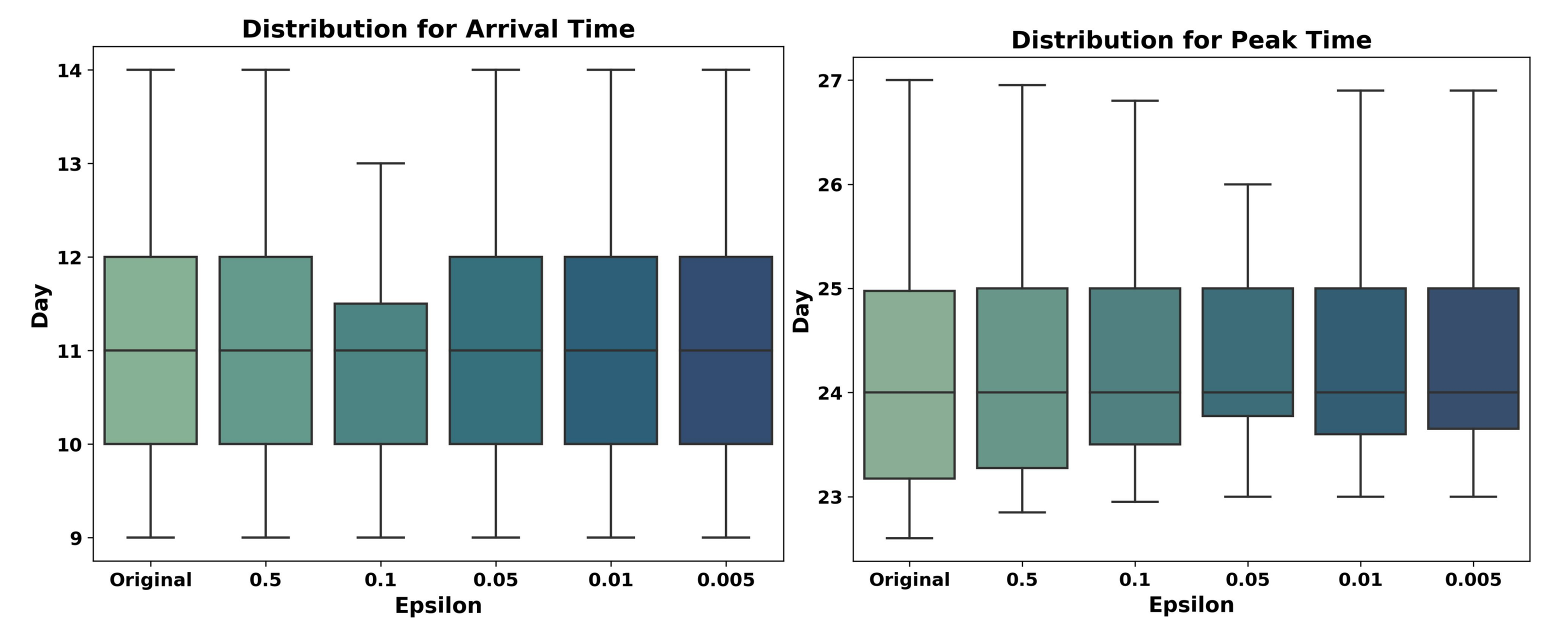
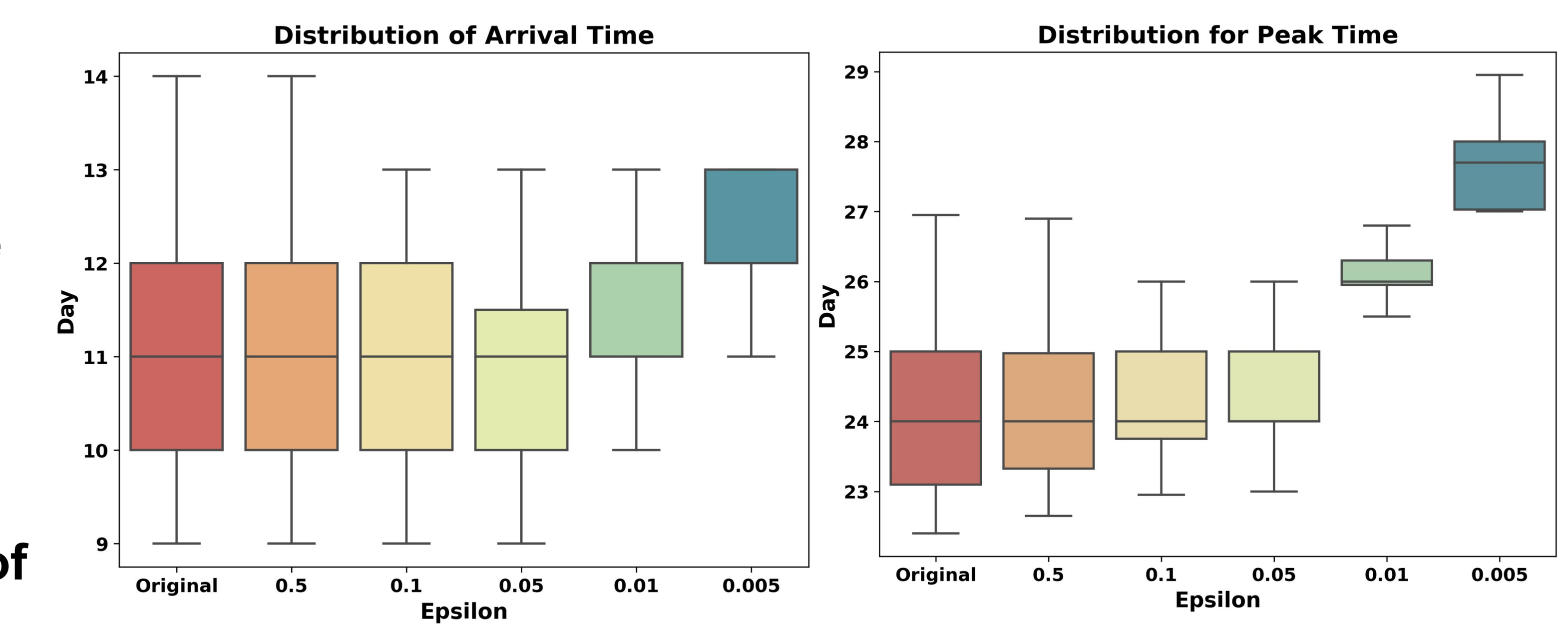


Figure 5: The distribution of arrival time (left) and peak time (right) with CDP



Results

Figure 3 shows that adding noise at different intensities increases outflow weights and reduces the number of near-zero connections. This suggests that while inter-county travel probabilities rise, the noise-adjusted mobility matrix concentrates flows toward a few counties while severing connections with others.

Figures 4 and 5 demonstrate that when noise is applied at the individual CBG level before aggregating to counties, the timing of arrival and peak infection for diseases like COVID-19 remains stable across different privacy levels. However, when noise is applied directly to pre-aggregated county mobility data, both arrival and peak times become more sensitive to the noise magnitude. The results suggest that for aggregated data, an epsilon value of 0.5 offers the best balance between privacy and SEIR model performance.

Limitation

Access to real-world individual-level mobility data remains a key challenge, we have not applied LDP to actual trajectory data. Our analysis currently includes NY, CA, and WA States. Future work will explore states with more variation in metro size and urban–rural patterns to assess whether arrival and peak infection trends hold.

Conclusion

Our findings show that applying Central Differential Privacy (CDP) at the county level leads to significant errors in predicting arrival and peak infection times, while Local Differential Privacy (LDP) at the individual level preserves temporal accuracy. Public health agencies should prioritize LDP-based data partnerships when model precision is critical, while recognizing the tradeoff between privacy and performance in CDP applications.

References

[1] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3* (pp. 265-284). Springer Berlin Heidelberg.

[2] Xiong, X., Liu, S., Li, D., Cai, Z., & Niu, X. (2020). A comprehensive survey on local differential privacy. *Security and Communication Networks*, *2020*(1), 8829523.

[3] Pullano, G., Alvarez-Zuzek, L. G., Colizza, V., & Bansal, S. (2025). Characterizing US Spatial Connectivity and Implications for Geographical Disease Dynamics and Metapopulation Modeling: Longitudinal Observational Study. *JMIR Public Health and Surveillance*, *11*, e64914.

Acknowledgment: This Project is funded by the Georgetown University ***Fritz Family Fellowship***