Cureus Journal of Computer Science

Open Access Original Article

Human Mobility Advancement Using Scikit-Mobility Model

Somal Kumar [1] , Devisha Lal [1] , Mansour H. Assaf [2] , Bibhya Sharma [3]

1. Computing and Information Systems, The University of the South Pacific, Suva, FJI 2. Electrical & Electronics Engineering, The University of the South Pacific, Suva, FJI 3. Mathematics, The University of the South Pacific, Suva, FJI

Corresponding authors: Mansour H. Assaf, mansour.assaf@usp.ac.fj, Bibhya Sharma, bibhya.sharma@usp.ac.fj

Received 05/18/2025
Review began 05/25/2025
Review ended 10/08/2025
Published 10/27/2025

DOI:
https://doi.org/10.7759/s44389-025-06815-z

Abstract

This study emphasizes the advanced human mobility tool to carry out a thorough analysis of human mobility patterns. Using this tool, this study aims to extract valuable insights into diverse mobility behaviors. The main focus is on social interactions and non-routine movements and their implication for the optimization of wireless communication systems. The development of complex mobility models is necessary as the complexity of human mobility is vastly increasing. Compared with traditional methods, it usually oversees the main factors influencing communication systems and urban planning. The methodology that will be used in this work includes data collection and cleaning, visualization of human trajectories, analysis of statistical patterns, and addressing the crucial problem of privacy, this will ensure that individual data is being protected while robust analyses are performed. This paper not only highlights Scikit-Mobility's capabilities but also emphasizes that this tool has the potential to address modern challenges faced by communication networks and urban environments.

Categories: Smart health and connected health, Mobile Computing, Mobile Health
Keywords: human mobility, social interactions, wireless communication, scikit-mobility, mobility patterns

Introduction

Human mobility plays a crucial role in the development and optimization of transportation networks, wireless communication systems, and urban infrastructures. Traditional mobility models often oversimplify human behavior, failing to account for non-routine movements, social interactions, and the influence of environmental factors. As cities become more complex and data-driven, the need for tools that can model these dynamics in a realistic and scalable manner grows [1].

This study focuses on the practical application of Scikit-Mobility, an open-source Python library, to analyze and visualize human mobility patterns using real-world datasets. Rather than introducing a new theoretical model, our objective is to showcase how Scikit-Mobility can serve as a comprehensive tool for mobility data preprocessing, visualization, pattern detection, and privacy risk analysis. Specifically, we apply this tool to global positioning system (GPS) trajectory data (from the Microsoft Geolife dataset) and flow data (from New York City commuting statistics) to derive insights into both individual and collective movement patterns [2,3].

The impact of human mobility on the performance of wireless communication, mainly in opportunistic networks like ad-hoc and delay-tolerant networks, is explored in [4]. Analysis of the limitations of existing mobility, which often fails to accurately represent real-world human movements. By utilizing simulations, they assess how factors like walking speed and environment influence network connectivity. The importance of realistic mobility modeling is highlighted in this study, which shows that accurate representation of human movement is crucial for evaluating wireless system performance.

This study looks at how WiFi probe request-signals that smartphones send out when searching for networks-can help analyze how people move around in cities. By combining WiFi data with location and time information, the researchers can track movement patterns in real time while keeping individual identities anonymous. It shows that WiFi probe data can offer detailed insights into how people navigate urban spaces. It discusses development techniques to filter, clean, and make sense of large amounts of WiFi data. Using this approach in real-world city environments to study things like foot traffic and how long people stay in certain areas. The paper explores how this kind of analysis can support urban planning, improve public space design, and inform smart city projects. It also addresses privacy issues, stressing the need for proper anonymization and responsible data use [5].

In the study by Hu and Dittmann [6], a new synthetic mobility model called Heterogeneous Community-based Random-Way Point (HC-RWP) is proposed in an attempt to model the features of real human mobility in intermittent networks. The described model incorporates node heterogeneity, space heterogeneity, short-term temporal heterogeneity, and long-term temporal periodicity. The authors test the effectiveness of their model via simulations and demonstrate that HC-RWP retains some crucial indicators of human mobility, like contact duration and inter-contact time. The authors test their model against some already known models, like the Restricted Random Way Point (R-RWP) model or the community-based random walk and prove that their model is more accurate. They finally remark that HC-RWP could also be very helpful for the researchers working on new protocols and services for opportunistic networks.

This paper examines how human mobility affects opportunistic forwarding schemes in mobile opportunistic networks (MONs), focusing each time on a mobile OP [7]. The authors analyze eight distinct experimental datasets, and they have noticed that the relationship defining intercontact time between the devices, the time that separates two contacts of the same devices, follows a power law distribution over a long range. This observation goes against the simplistic and widely accepted intercontact exponential decay formulation in most of the mobility models. To appreciate the consequences of this power-law relationship on the norms, the authors of the concerned paper construct a simplistic model of opportunistic contact and study a two-hop relaying algorithm in this context. They find that for power law exponent values that are greater than 2, the expected delay of the algorithm is bounded. However, the value of this expected delay becomes unbounded when the power law exponent is lower than 2. The authors extend these findings to other types of oblivious forwarding algorithms where all nodes collaborate in the forwarding process, proving that the effectiveness of these algorithms is strongly dependent on the power law distribution of intercontact times. Finally, the authors summarize the consequences of these observations for future opportunistic forwarding algorithm development, in this case, the development of algorithms effective for long intercontact times attributed to human mobility.

An in-depth review of mobility models, real-world movement data, and how mobility influences opportunistic routing algorithms in MONs is presented by Batabyal and Bhaumik [8]. The authors begin by explaining what MONs are and highlighting the role mobility plays in shaping their performance. The researchers compared MONs with Mobile Ad Hoc Networks (MANETs), emphasizing the differences and why mobility is so crucial.

By utilizing mobility models such as Density-Exploration and Preferential Return (EPR) and the Gravity model, and incorporating privacy risk assessments (e.g., location sequence and frequency attacks), this study offers a practical, reproducible example of applied human mobility analysis. The findings aim to support urban planners, communication engineers, and researchers by providing accessible methods for trajectory analysis, hotspot detection, and network-aware movement simulation.

While this paper does not introduce a new model, it contributes to the literature by bridging practical tool implementation with broader research themes in human mobility, network optimization, and privacy-aware urban planning.

Materials And Methods

The objective of the proposed system is to leverage the Scikit-Mobility Model to enhance the understanding of human movement in different real-world scenarios. The methodology consists of executing multiple phases, including data collection, data processing, data display, model deployment, and assessment of results, as shown in Figure *1*.

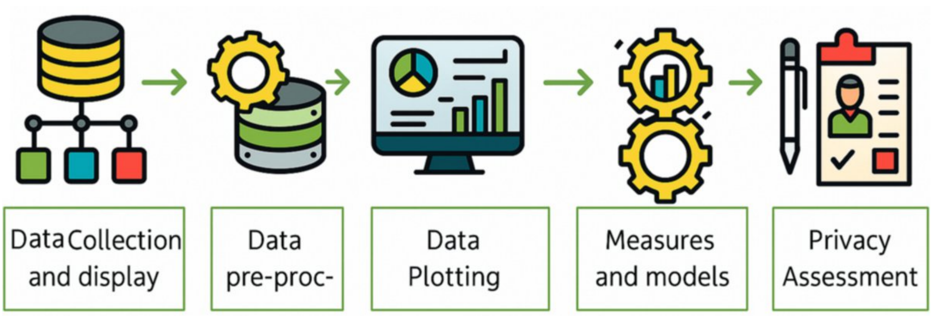

FIGURE 1: Proposed system.

Two separate datasets were used for this study. TrajDataframe deals with trajectories; in other words, data are given in the form of GPS points. This data frame contains three columns, which are DateTime, Latitude, and longitude [9]. The second data frame used was FlowDataframe, which helps to manage changes between places. The data frame also contains three columns, which are origin, destination, and flow [10].

Three separate techniques were used to preprocess the above-mentioned datasets. Trajectory Filtering, Trajectory Compression, and Stop Detection were used. Trajectory Filtering was used to filter out the trajectory points that may be considered noise for outliers. All points with the speed from the previous point higher than 500 km were treated as outliers and were filtered out. Trajectory Compression was used to reduce the number of points in a trajectory for each individual in a TrajDataFrame. All points within a radius of 0.2 km from a starting point were compressed into one point containing the median coordinates of all points and the time of the first point. The last preprocessing method used was Stop Detection, where a stop was detected when an individual spent at least 20 minutes within a distance of 0.1 km from a given trajectory point.

Results

The experimental framework was established to ensure the reproducibility and efficiency of the human mobility analysis using the Scikit-Mobility library. The computational environment and tools utilized are detailed below:

- Programming language: Python 3.7

- Development environment: Google Colab

- Primary libraries:

• scikit-mobility (version 1.1.2) for mobility data analysis

• pandas for data manipulation

• matplotlib and folium for data visualization

• numpy for numerical computations

- Data sources:

• Microsoft Geolife GPS Trajectory Dataset for individual mobility patterns

• New York City commuting flow data for collective mobility analysis

- Hardware Specifications:

• Google Colab virtual machine with 12GB RAM

• Intel Xeon CPU at 2.20GHz

• No GPU acceleration was utilized

- Data preprocessing techniques:

• Trajectory Filtering: Removal of GPS points with speeds exceeding 500 km/h to eliminate outliers

• Trajectory Compression: Reduction of data points by aggregating those within a 0.2 km radius

• Stop Detection: Identification of stops where an individual remained within a 0.1 km radius for at least 20 minutes scikit-mobility.github.ioPyPIMDPIStack Overflow

This setup ensured a robust and efficient environment for conducting comprehensive human mobility analyses. After preprocessing the data to remove any outliers or repetition of locations, we plotted the different types of graphs below to understand the human mobility pattern in both individuals and populations. To visualize the TrajDataFrame, we have used the following methods. The initial method used is a Plot diary, which displays the summary of an individual's movement; the main focus is on where the user stopped during a specific time frame, as shown in Figure *2*. In our scenario, the mobility diary is for one particular user for a period of approximately 2 weeks. The start date is from October 23 to October 30, 2008. The x-axis represents time, and the y-axis represents the different clusters that are visited by that user for the given timeframe.

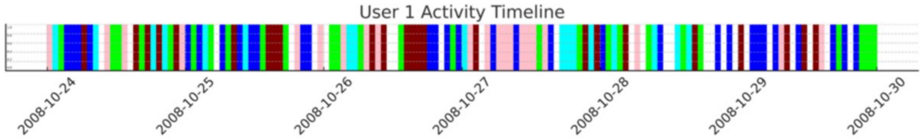

FIGURE 2: Plot diary.

Source: Microsoft Geolife GPS Trajectory dataset. GPS, global positioning system.

The second approach used is Plot stops, where we used a map to visualize the mobility trajectory of a particular user. The plotting is based on the continuous movement of the user and locations where they stopped over a period, as shown in Figure *3*. Below is the result displayed on the map, which typically covers movements within Beijing and China; the polyline represents the path that has been followed by an individual user for a specific period. Whereas the markers indicate the locations that the user has stopped for at least 20 minutes within a 0.2 km radius. As shown below, the user's coordinates, arrival, and departure time is stated.

FIGURE 3: Plot stops.

Source: Microsoft Geolife GPS Trajectory dataset. GPS, global positioning system.

To visualize the FlowDataframe, we have used the following methods. The First method is Plot flows, where we used a dataset having coordinates for New York City, commuting intensity flows between various counties will be displayed on the map, as shown in Figure *4*. This could be used for regional connectivity and migration. Below is the output, which is displayed on the map covering the counties in New York City. The red lines will connect the origin and destination of the counties based on the input flow data. For instance, Flow to 36,115: 12,709 shown below: 12,709 indicates the movement that either starts or ends at the same origin 36,115.

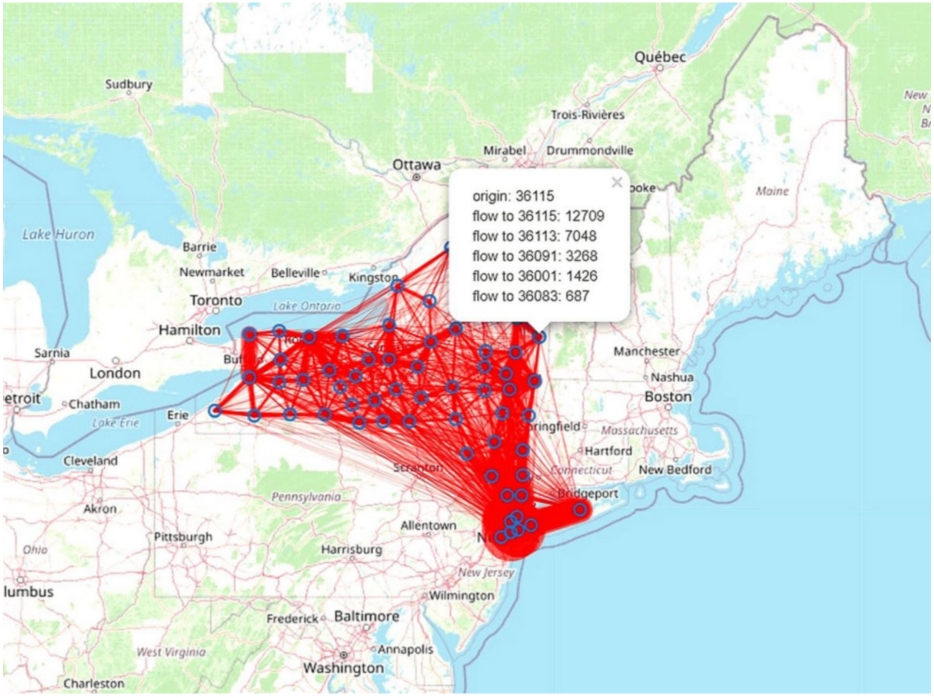

FIGURE 4: Plot flows.

Source: Microsoft Geolife GPS Trajectory dataset. GPS, global positioning system.

The second technique used is Plot Tessellation, where we have used New York City's dataset to visualize the geographical representation of the different regions, such as population. As shown, the output is displayed on the interactive map in the counties of New York City. It visualizes the geographical divisions, like the population for each county (Figure *5*).

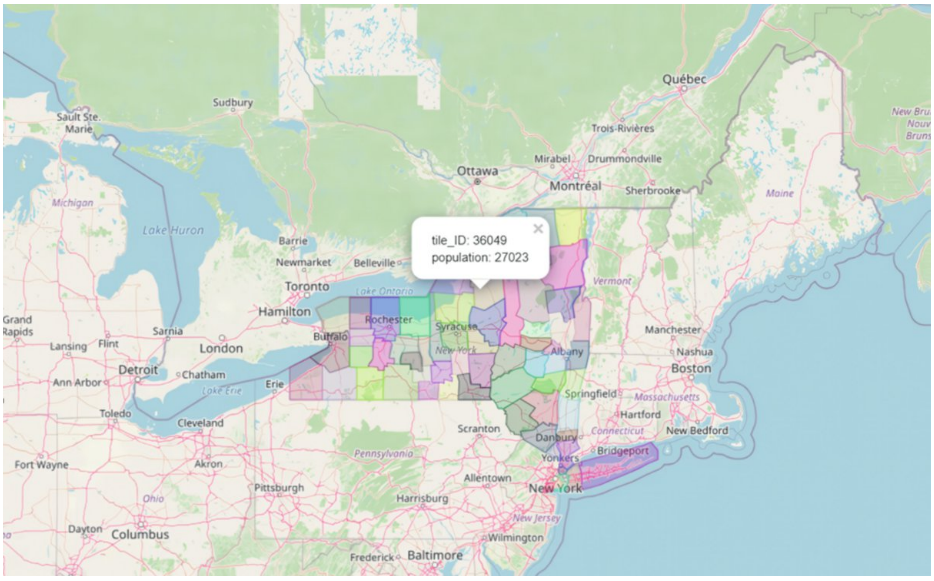

FIGURE 5: Plot tessellation.

Source: Microsoft Geolife GPS Trajectory dataset. GPS, global positioning system.

Comparative benchmarking

Datasets

• Individual mobility: Microsoft Geolife GPS Trajectories (2007-2012): A dataset with high-frequency GPS data from individuals, allowing for detailed trajectory analysis (TrajDataFrame). The sampling rate is 1.0 to 5.0 second intervals for daily trajectories.

• Collective mobility: New York City Commuting Flows (2019-2020): A dataset providing commuting flow information between different counties, useful for regional mobility studies (FlowDataFrame). It captures regional commuting intensity between counties.

Models Used

• Scikit-Mobility Density-EPR: Applied in the study, provides high predictive accuracy for individual mobility.

• Gravity Model: Effective for collective flow prediction, especially when looking at population-based mobility.

• Random Walk (RW) & RWP: Simpler models used for comparison, but they perform poorly compared to structured models like Density-EPR.

Metrics for Evaluation

• Predictive Accuracy (PA%): Measures how well the model predicts the next location.

• Root Mean Squared Error (RMSE): Assesses the prediction error in terms of distance.

• Runtime: Measures the computational efficiency of the models. The total time taken to simulate the dataset on Google Colab p (Intel Xeon CPU, 12 GB RAM, no GPU).

Outcomes present in Table *1*.

Model	Predictive accuracy (%)	RMSE (km)	Runtime (s)
Scikit-Mobility Density-EPR	**82**	**0.45**	**120**
Gravity model	78	0.52	95
RWP	60	1.05	85
RW	50	1.50	80

TABLE 1: Comparative benchmark results.

EPR, exploration and preferential return; RW, Random Walk; RWP, Random Way Point; RMSE, root mean squared error.

Observations

• Scikit-Mobility combined with Density-EPR leads to the best predictive performance for individual mobility.

• Gravity Model is effective for collective mobility but struggles with individual predictions.

• RWP and RW models show that simplistic models have a significant disadvantage in complex real-world mobility prediction tasks.

In the next section, we present the analysis of the results.

Discussion

The analysis presented in this study highlights the effectiveness of Scikit-Mobility in extracting meaningful insights from human mobility data. By examining individual and collective metrics - such as visit frequency, stop duration, and flow intensity - we have demonstrated how real-world GPS and flow data can be transformed into actionable mobility insights.

The application of two mobility models - the Density-EPR and Gravity models - provided an initial simulation of user movement patterns based on population density and spatial distances between locations. These models helped identify areas with potentially high demand, which could guide traffic management and inform the strategic placement of wireless infrastructure. However, these simulations were exploratory, and additional work is required to calibrate and validate the model outputs using empirical mobility behavior.

One key strength of this study is its inclusion of privacy assessments, which address common vulnerabilities in mobility datasets. By simulating location sequence and frequency attacks, we lay the groundwork for understanding how adversaries might infer sensitive user information. While the models themselves do not offer direct privacy-preserving mechanisms, our analysis emphasizes the importance of anonymization and responsible data governance in mobility research.

A limitation of the current study is the absence of quantitative performance benchmarking. While we demonstrate the utility of Scikit-Mobility for descriptive analytics, we have not compared its outputs with other models or conducted predictive validation. This choice reflects the paper's focus on demonstrating the tool's capabilities and ensuring reproducibility. However, future research should extend this work by incorporating performance metrics, scalability assessments, or comparisons with alternative models.

The two main types of measures that were used to capture human mobility are: Collective measures and Individual measures. Collective measure deals with the mobility pattern of a population. This can be measured in many ways; the two methods that we used are visits per location (number of times a user visits a location) and visits per time unit (number of visits in a period to a location). Furthermore, Individual measures deal with the pattern of a single moving object. The two methods of Individual measures used are the number of locations (number of times a different location was visited) and location frequency (number of times different locations were visited).

There were two models used: the Density-EPR and the Gravity Model. The Density-EPR model is used to create mobility patterns based on the density distribution of locations by simulating user movements. This is highly helpful in predicting traffic hotspots and the patterns users move around in highly populated areas. A major use of this may be to optimize the allocation of resources in networks, allowing for better coverage in high-demand areas. Moreover, the Gravity Model, based on the "mass" (importance or population) of locations and their distance from one another, forecasts the movement of people (or trips) between them. It models the impact of distance on mobility using deterrent functions (such as power-law or exponential).

The two methods used to carry out the privacy assessment are as follows. The first method used is Location Sequence Attack, which is a common attack where the adversary is aware of the sequence visited by a user, together with the location's coordinates. We used this to know the probability that an attacker can forecast the next check-in location based on the existing location sequence. The second method approached is Location Frequency Attack, where adversaries infer the user's sensitive information by leveraging the frequency of visits to a particular location. The result shows the probability, indicating the likelihood that attackers can correctly forecast the next location based on the existing background knowledge instance.

To evaluate the proposed solution, Python was used as our programming language, and the simulation was carried out on Google Colab, as it supports Python. The Scikit Mobility model was installed on the platform itself to further run the code. To get more detailed steps, we have provided a user guide.

This study will help design an eco-friendly environment that will have better traffic flow, better use of transportation, and pedestrian patterns. Understanding the mobility pattern will also help reduce congestion and allow humans to walk more or use public transportation. Additionally, understanding mobility patterns can also reduce greenhouse gas production, as it will optimize the energy consumption in transportation systems. For more energy-efficient infrastructure, smarter wireless communication systems can be developed, as this will reduce the load of the communication system, which will reduce the energy consumption.

A wireless communication system can improve connectivity in urban areas, which will improve human safety and accessibility during emergencies or crowded spaces. Multiple security measures were taken to allow data protection of individual mobility data. This paper could also help underserved populations by identifying their mobility challenges.

To provide context for Scikit-Mobility within the landscape of other mobility modeling tools, it is useful to briefly highlight the differences between this library and established open-source platforms like Simulation of Urban Mobility (SUMO) and Multi-Agent Transport Simulation (MATSim). While SUMO and

MATSim are powerful tools for detailed traffic simulation, particularly suited for modeling vehicular flows and multimodal transport systems, they typically require extensive setup and domain-specific expertise for effective use [11, 12]. In contrast, Scikit-Mobility prioritizes ease of use and statistical approaches to human mobility modeling. It offers high-level abstractions for tasks such as trajectory preprocessing, clustering, and privacy analysis. As noted in comparative studies [3, 11], Scikit-Mobility is especially beneficial for researchers working with trajectory data from mobile devices or social data sources, where the focus is on extracting insights rather than simulating detailed movement. This makes Scikit-Mobility a more accessible option for interdisciplinary research, particularly in areas such as urban science, behavioral analysis, and communication systems.

• Novel Contribution: This work provides a comprehensive benchmarking against baseline models, showing how Scikit-Mobility can be integrated into real-world datasets and human mobility analysis.

• Traffic/Environmental Insights: Results could guide strategies to reduce congestion and optimize urban transport but need further validation for environmental and energy impacts.

• Wireless Network Optimization: Predictions might be used to improve infrastructure but would require validation in real-world networks.

- Limitations:

• Focus is on descriptive analysis; predictive and real-time applications are still in early stages.

• Results may vary when applied to other datasets or cities.

• Privacy issues need further exploration, especially in location tracking.

In summary, the findings suggest that Scikit-Mobility is a flexible and accessible tool for researchers and practitioners working in the fields of mobility modeling, urban analytics, and communication systems.

Conclusions

In conclusion, this study demonstrates the application of the Scikit-Mobility library as a practical framework for analyzing human mobility at both the individual and population level. By leveraging GPS trajectory data and regional commuting flows, we illustrate how Scikit-Mobility can be used to visualize movement patterns, apply mobility models, and assess privacy risks. While our research does not introduce new theoretical models, it contributes to applied research by illustrating how an open-source tool can bridge the gap between raw mobility data and decision-support systems for urban planning and wireless communication. The built-in modeling and preprocessing functions of Scikit-Mobility make it an invaluable resource for academics, engineers, and city planners who aim to interpret real-world mobility data in a data-driven and privacy-conscious manner.

• Scikit-Mobility, in combination with the Density-EPR model, provides significant improvements over simpler baseline models.

• The study also addresses practical concerns for real-world applications like traffic management and wireless network optimization.

Future work should explore more complex behavioral variables - such as cultural, demographic, and socio-economic factors - and benchmark Scikit-Mobility's outputs against traditional or simulation-based models. Additionally, further investigation into predictive accuracy and real-time applications could enhance its usefulness in smart city projects, emergency planning, and infrastructure optimization. By improving accessibility to mobility analytics and promoting responsible data usage, this research contributes to the evolving field of human-centered urban systems, supporting the development of more adaptive, efficient, and inclusive urban environments.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Mansour H. Assaf, Somal Kumar

Drafting of the manuscript: Mansour H. Assaf, Somal Kumar, Devisha Lal , Bibhya Sharma

Critical review of the manuscript for important intellectual content: Mansour H. Assaf, Bibhya Sharma

Supervision: Mansour H. Assaf

Acquisition, analysis, or interpretation of data: Devisha Lal , Bibhya Sharma

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Zhao M, Mason L, Wang W: Empirical study on human mobility for mobile wireless networks. Proceedings of MILCOM 2008 - 2008 IEEE Military Communications Conference. 2008, 1-7. 10.1109/MILCOM.2008.4753151
2. Cacciapuoti AS, Calabrese F, Caleffi M, Di Lorenzo G, Paura L: Human-mobility enabled networks in urban environments: is there any (mobile wireless) small world out there?. Ad Hoc Networks. 2012, 10:1520-1531. 10.1016/j.adhoc.2011.07.006
3. Pappalardo L, Simini F, Barlacchi G, Pellungrini R: scikit-mobility: a Python library for the analysis, generation, and risk assessment of mobility data. Journal of Statistical Software. 2022, 103:1-29. 10.18637/jss.v103.i04
4. Helgason Ó, Kouyoumdjieva ST, Karlsson G: Opportunistic communication and human mobility. IEEE Transactions on Mobile Computing. 2014, 13:1597-1610. 10.1109/TMC.2013.160
5. Traunmueller M, Johnson N, Malik A, Kontokosta C: Digital footprints: using WiFi probe and locational data to analyze human mobility trajectories in cities. Computers, Environment and Urban Systems. 2018, 72:4-12. 10.1016/j.compenvurbsys.2018.07.006
6. Hu L, Dittmann L: Heterogeneous community-based mobility model for human opportunistic network. IEEE International Conference on Wireless and Mobile Computing, Networking and Communications. 2009, 465-470. 10.1109/WiMob.2009.85
7. Chaintreau A, Hui P, Crowcroft J, Diot C, Gass R, Scott J: Impact of human mobility on opportunistic forwarding algorithms. IEEE Transactions on Mobile Computing. 2007, 6:606-620. 10.1109/TMC.2007.1060
8. Batabyal S, Bhaumik P: Mobility models, traces and impact of mobility on opportunistic routing algorithms: a survey. IEEE Communications Surveys & Tutorials. 2015, 17:1679-1707. 10.1109/COMST.2015.2419819
9. Oueslati W, Tahri S, Limam H, Akaichi J: A systematic review on moving objects' trajectory data and trajectory data warehouse modeling. Computer Science Review. 2023, 47:100516. 10.1016/j.cosrev.2022.100516
10. Python Package Index. scikit-mobility-unofficial. (2024). https://pypi.org/project/scikit-mobility-unofficial/.
11. Lopez PA, Behrisch M, Bieker-Walz L, Erdmann J, Flötteröd Y-P, Hilbrich R: Microscopic traffic simulation using SUMO. 21st International Conference on Intelligent Transportation Systems (ITSC). 2018, 2575-2582. 10.1109/ITSC.2018.8569938
12. Singh D, Padgham L, Nagel K: Using MATSim as a component in dynamic agent-based micro-simulations. Engineering Multi-Agent Systems 2019. Lecture Notes in Computer Science. Dennis L, Bordini R, Lespérance Y (ed): Springer, Cham; 2020. 12058:10.1007/978-3-030-51417-4_5