Decision-making in Urban Trajectories of Bike Users: a Preliminary Study

Merlin Raud[1,†], Amnir Hadachi[1,†], Rajesh Sharma[1,†] and Flavio Bertini[2,*,†]

[1]Institute of Computer Science, University of Tartu, Narva mnt 18, 51009 Tartu, Estonia
[2]Department of Mathematical, Physical and Computer Sciences, University of Parma, Parco Area delle Scienze 53/A 43124 Parma, Italy

Abstract
Developed countries have invested considerable efforts in promoting sustainable transportation modes, identifying bicycle use as a more environmentally friendly alternative. Nevertheless, this transition necessitates a deeper understanding of bicycle user behaviour to facilitate the process. This study investigates bike users' route selection behaviour based on real-world data. Data collected over six months from Bologna, Italy, were analysed to understand the flow and choices of bike users in a medium-small historical city. We employed a tessellation algorithm combined with Open Street Map street tags to examine bike users' preferences regarding urban trajectories. The dataset comprises of 290,117 unique trips consisting of 60,414,481 Global Positioning System (GPS) points. The results indicate that bicycles are predominantly used for short- to medium-distance trips. Specifically, in suburban regions, bike users tend to use larger roads leading to the city centre, whereas, in the city centre, they prefer routes that bypass the narrow streets of the historic area. These findings offer insights into route choice factors and can guide improvements to cycling infrastructure for a safer, more efficient urban environment.

Keywords
Mobility data analysis, cycling mobility, urban mobility, real-world data mining

1. Introduction

In response to climate change, energy crises, fossil fuel depletion, overpopulation, and traffic congestion, sustainable living has garnered increased attention. In densely populated urban areas, promoting sustainable transportation is vital. Encouraging cycling can reduce the use of motorised vehicles and decrease traffic volumes. Cycling is a sustainable, healthy, and economical transport mode, particularly suitable for short-distance trips in historic and narrow urban spaces. Facilitating bicycle use for commuting and leisure as a sustainable transport option requires a thorough understanding of urban mobility and travel behaviour. The role of data mining in mobility research and urban cycling studies is crucial. To assist in this cause, mobility studies can uncover patterns in daily and routine human movements [1]. This data is vital for urban planning, the development of transportation models, and the promotion of healthy lifestyles [2], and it supports traffic managers and policymakers in making targeted decisions at various levels [3]. A key research area is bike users' route choice behaviour. Analysing the factors that influence bike users' route selection can lead to improvements in the transport network, thereby encouraging urban cycling. The recent increase in mobile phone records, GPS data, and other datasets have facilitated more accurate human mobility analyses [4]. Particularly, the use of data mining methods on GPS data provides essential information on the movement patterns and behaviours of bike users.

Although various models for analysing bike users' behaviour have been proposed, a detailed investigation of route selection between two locations and multiple alternative trajectories is still lacking. Moreover, despite the precision offered by GPS data, it is crucial to preserve users' privacy. This paper addresses this gap by utilising mobility data from the *Bella Mossa* initiative, provided by the Public Transport Authority *SRM Reti e Mobilità Srl*, and the Open Street Map (OSM) data to examine cycling patterns in Bologna, a historic university city with narrow streets in northern Italy. We conducted an analysis of bicycle usage based on 290,117 self-reported unique trips, consisting of 60,414,481 GPS points, spanning six months, from April 2017 to September 2017. Specifically, we applied tessellation and clustering techniques to analyse mobility flows in urban areas based on origin-destination tiles and to study the trajectory selections of bike users.

This study aims to investigate how cycling infrastructure influences bike users' route choices by utilising OSM street-type tags while preserving users' privacy. To achieve this, the exact routes from GPS tracks are transformed into coarser shifts between origin and destination, retaining useful information for analysis. The following are the highlights of this study. Firstly, we investigate the utilisation of various types of infrastructure, including urban highways, different bike lane configurations, and streets, to understand bike users' preferences regarding route choice. Secondly, we identify the most frequently used flows and routes for transportation purposes in Bologna, shedding light on preferred routes. Lastly, we examine how road types and their characteristics influence bike users' route choices, providing insights into the factors influencing route selection.

The obtained results reveal that cyclists tend to favour streets with lower traffic volumes and dedicated cycling infrastructure. Furthermore, cyclists prefer to avoid narrow streets in the old town when possible. Understanding these patterns helps identify opportunities to improve infrastructure and create a safer urban environment. For instance, identifying the most frequently used origin-destination tiles and the preferred paths between them can help guide infrastructure development and promotion. Also, the findings will guide recommendations to promote cycling as a viable urban transportation option, and the proposed approach, despite the conclusions being specific for the dataset, the pipeline can be applied in similar urban contexts.

Published in the Proceedings of the Workshops of the EDBT/ICDT 2025 Joint Conference (March 25-28, 2025), Barcelona, Spain
*Corresponding author.
†These authors contributed equally.
merlin.raud@gmail.com (M. Raud); amnir.hadachi@ut.ee (A. Hadachi); rajesh.sharma@ut.ee (R. Sharma); flavio.bertini@unipr.it (F. Bertini)
0000-0001-9257-3858 (A. Hadachi); 0000-0003-3581-1332 (R. Sharma); 0000-0001-6925-5712 (F. Bertini)

2. Related Work

Cycling offers a sustainable, eco-friendly, cost-effective, and inclusive solution for urban mobility [5]. Compact cities support active transport but face challenges in designing efficient cycling infrastructure due to space constraints [6]. Urban areas have diverse infrastructure for different users, with dedicated bike lanes significantly promoting bicycle use [7].

The pervasive collection of tracking data has enabled extensive trajectory data collection for applications like intelligent transport systems, urban management, and environmental protection [8]. However, challenges persist due to privacy issues, commercial concerns, and data inaccuracies [9]. GPS trajectory analysis then categorises trajectories or uses data mining to uncover movement patterns and predict future behaviours [10].

Trajectory analysis is essential for understanding route selection in urban environments. This process involves choosing a specific path from an origin to a destination based on various characteristics [11]. A deeper insight into route choice can enhance the comprehension of infrastructure usage and facilitate the cycling mobility promotion [12].

In this paper, we propose a method based on tessellation and flow analysis algorithms to study the decision-making process of urban bike users. Specifically, we leveraged information from OSM and addressed privacy concerns related to the use of GPS data through a clustering algorithm.

3. The *Bella Mossa* Dataset

The *Bella Mossa*[1] data was collected from the municipal area of Bologna, a historical city in northern Italy. The *Bella Mossa* initiative was a 6-month program to promote a healthy lifestyle and sustainable mobility from 01/04/2017 to 30/09/2017. Around 15,000 unique participants recorded 895,000 trips for a total of 3.7 million travelled kilometres. Participants earned redeemable points for cycling, walking, and using public transport by using a GPS-enabled smartphone app that recorded trips and anonymously sent the data to a database. Each dataset entry includes a unique activity ID, a user-selected activity type descriptor, and details about the timestamp, GPS position, and accuracy. The activity ID enables trip reconstruction without being traceable to the participant. In this study, we specifically focus on data points labelled as *cycling* activities. The dataset is anonymised; individual user data has not been analysed, and only aggregated analysis results are presented.

We conducted pre-processing to clean and correct data, addressing incomplete or inaccurate values. Preliminary analysis identified extremely short or long activities. Short activities with fewer than two *GPS* points were removed. Long activities with pauses longer than six minutes or where the bike user stayed in one place for over six minutes were split into multiple segments. Activities longer than two hours or over 30km after segmenting were excluded, as they likely included multiple segments or non-urban activities (*e.g.*, sports). Activities under 100m or less than three minutes were also removed. Finally, activities with *GPS* points outside Bologna or timestamps outside the study period were excluded. The initial dataset of 320,109 unique activities with 72,396,179 *GPS* points was reduced to 290,117 unique activities consisting of 60,414,481 *GPS* points, recorded every 3-10 seconds.

[1]Due to Italian legislation restrictions, we are unable to publicly release the datasets. Any requests can be submitted to *SRM Reti e Mobilità Srl*.

4. The Proposed Method

In this section, we outline the proposed method using the *Bella Mossa* dataset and the OSM street network data. First, we employed a map-matching algorithm to refine the bike users' trajectories, converting sequences of GPS points into sequences of road segments. Next, we applied tessellation and clustering algorithms for trajectory and road type analysis, while preserving users' privacy.

The map-matching algorithm assigns original GPS coordinates to the most plausible road network segments, improving the accuracy of GPS data and linking these coordinates to the corresponding road network information. In practice, we used the Open-Source Routing Machine (OSRM) project and the Hidden Markov Model (HMM) approach with the Viterbi path-finding algorithm [13]. This makes it possible to check all combinations of nodes and transitions (*i.e.*, intersections and road segments) to determine the optimal route between two origin-destination GPS points, representing it through road segments (Figure 1). Furthermore, we took into account the time intervals between consecutive GPS points and the typical speed of the transportation mode to ascertain the most efficient route. The algorithm also eliminates outliers that cannot be matched and divides traces when there are substantial gaps in timestamps indicating unlikely transitions [14]. The OSM street network data -a directed graph where edges refer to road segments and nodes refer to intersections- was acquired from Geofabrik, with a timestamp set to January 1st, 2018. To handle anomalous behaviours, such as bike users travelling in the opposite direction on one-way streets, we added reversed edges to facilitate the map-matching algorithm. The resulting road network consists of 106,552 edges with a total length of 3,545.12km.

We used a spatial tessellation algorithm to subdivide the city area into tiles according to the GPS points densities [15]. In particular, we first applied the K-means algorithm using the origin GPS points of the bike user's trajectories. Then the identified clusters' centre was used to create the Voronoi diagrams (*i.e.*, the tessellation) using the Tesspy library. Experimentally, the ultimate k-value chosen was 300. A lower k-value produces larger tiles, particularly in suburban areas, where distinct trajectories might erroneously be deemed similar. Conversely, a higher k-value yields clusters with too few trajectories for meaningful analysis. A tessellation incorporating both origin and destination points was evaluated but did not produce a significant alteration in the resulting tile topology. This irregular tessellation method using the clusters' centre as points of interest (POI) to drive the subdivision of areas results in more realistic tiles com-

Figure 1: The map-matching outcome: the GPS points in *blue* and the corresponding road segments in *red*.

pared to regular tessellation [15].
Combining the representation of users' trips with road segments –coarser than GPS points– and map tessellation –which identifies areas rather than precise points– makes it possible to protect users' privacy. Moreover, as shown in Figure 2, the irregular tessellation enables a more focused analysis of the city's most frequently used areas by increasing the density of the tiles.

Flow analysis was utilised to explore the decision-making process of bike users regarding urban trajectories. We examined cyclists' trajectories from an origin Voronoi tile to a designated destination one, with the origin and destination tiles determined using the initial and final GPS points. This approach resulted in a comprehensive table summarising all movements between all Voronoi tile pairs. Then detailed trajectory analysis was conducted on the top twenty most frequent flows, merging flows from opposite directions to simplify distance calculation and grouping.
Given the complexity of the road network, multiple routes exist between origin-destination pairs. To identify the most common paths, we performed a trajectory clustering analysis using the Fréchet distance [16] as the clustering metric and Density-based Spatial Clustering of Applications with Noise (DBSCAN) as the clustering method. The Fréchet distance, which measures the similarity between two curves, can be intuitively explained using the analogy of a person walking a dog on a leash: both move independently along their respective paths from start to finish, adjusting their speeds but without retracing their steps. Formally, the Fréchet distance $\delta_F(A, B)$ represents the minimum leash length required for the two to complete their paths simultaneously and can be formalised as follows

$$\delta_F(A, B) = inf_{\mu}\ max_{a \in A}\ dist(a, \mu(a)) \tag{1}$$

where A and B are the two curves, $dist()$ denotes the Euclidean distance, and $\mu()$ is a continuous bijective mapping from points on curve A to points on curve B. DBSCAN, a widely used density-based method, was chosen for its ability to determine the number of clusters autonomously, handle clusters with varying densities, and identify outliers as noise, making it particularly suitable for analyzing urban travel patterns.
The clustering process divided all trajectories within a flow into distinct clusters based on the routes taken, with cluster sizes ranging from 1 to over 200 trajectories. Clusters containing fewer than 5 trajectories were deemed less significant for the final analysis. Finally, leveraging the OSM street tags, the road type analysis allows us to explore the road infrastructure primarily used by cyclists. For computational efficiency, the road type analysis focused on data for the month of April, while the entire dataset was utilised for flow analysis. We specifically computed the road type usage proportion by dividing the lengths of all road types in the activity by the total activity length. Similarly, we extended the road type analysis to the trajectory clusters to ascertain whether street type influences bike users' route selection.

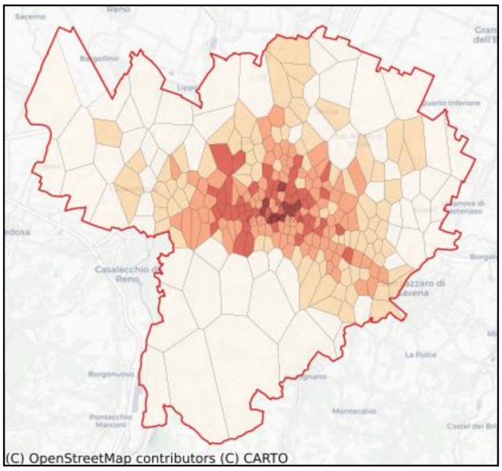


Figure 2: Tessellation based on trajectory origin density.

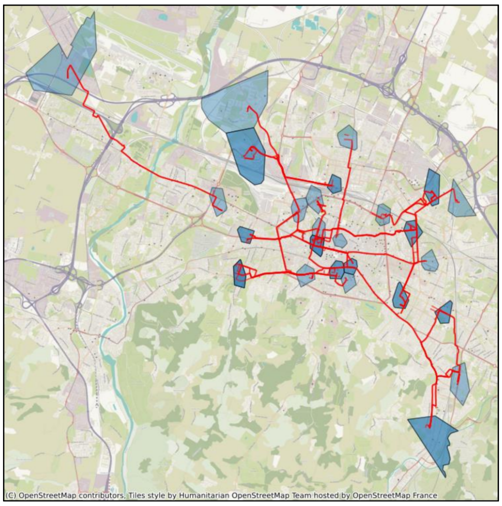


Figure 3: The 20 most popular origin-destination flows and their most used trajectories.

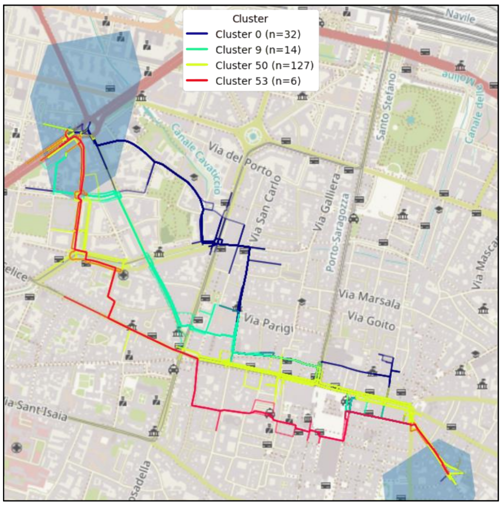
5. Experimental Result

In this section, we briefly present the results discussing the main outcome of the route choices and road type analysis.

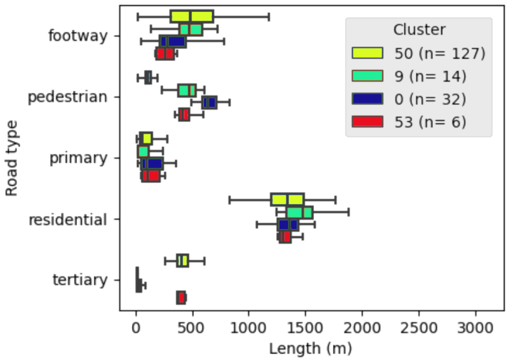
Highlight 1: Road Network Usage - Figure 1 depicts the result of the map-matching process. Notably, the cleaning procedure enables the extraction of noise-free trajectories represented as sequences of road segments. This preprocessing activity is of the utmost importance to refine subsequent analyses and obtain meaningful results. In particular, the origin GPS points of the cleaned trajectories were used to tessellate the city area into tiles according to the GPS points densities. The resulting irregular tessellation with Voronoi diagrams is shown in Figure 2. The tiles with the highest number of activity GPS points were located in the city centre.

Highlight 2: Frequent Flows - Flow analysis utilised tessellation tiles, assigning each activity an origin and destination tile. Activities with the same origin-destination pair were grouped and counted as one flow. We performed trajectory analysis on the 20 most frequent flows, employing trajectory clustering to identify different routes between locations. Figure 3 shows the most popular routes among the 20 top flows.

Highlight 3: Route Choices - Each flow was analysed to identify various routes between tile pairs. Trajectories were clustered into similar route groups using the DBSCAN algorithm, and a street-type analysis was conducted –based on the OSM road type labels– to assess its influence on route selection. For instance, Figure 4a illustrates the different



(a)



(b)

Figure 4: A pair of tiles showing identified routes (*a*) and road type analysis results (*b*).

routes chosen by bike users, correlated with their respective road types from the OSM informational tags, as depicted in Figure 4b.

6. Conclusion

Due to environmental concerns and urbanisation, there has been a growing emphasis on alternative modes of transportation, such as cycling. This paper presents a framework for analysing mobility flows within urban settings, examining the route choices between pairs of origin-destination tiles. An in-depth analysis of the top 20 flows showed that route choices were notably affected by one-way streets. According to the OSM road type labels and the results, cyclists typically avoid travelling against the designated direction, even when permitted. Furthermore, streets with cycling infrastructure are preferred, while narrow ones are generally avoided.

For future work, we plan to compare the real trajectories with potential shortest and quasi-shortest paths to better understand why cyclists choose specific routes. Additionally, we intend to investigate whether the effectiveness of the clustering algorithm may impact the final inference. Also, we plan to investigate the generalisability of the proposed framework to other transportation modes, exploring different tessellation strategies or clustering techniques to extract the sub-network of each mode.

References

[1] V. A. Paz-Soldan, R. C. Reiner Jr, A. C. Morrison, S. T. Stoddard, U. Kitron, T. W. Scott, J. P. Elder, E. S. Halsey, T. J. Kochel, H. Astete, et al., Strengths and weaknesses of global positioning system (gps) data-loggers and semi-structured interviews for capturing fine-scale human mobility: findings from iquitos, peru, PLoS neglected tropical diseases 8 (2014) e2888.

[2] R. Wang, Q. Wang, N. Li, Percolation transitions in urban mobility networks in america's 50 largest cities, Sustainable Cities and Society 91 (2023) 104435.

[3] R. Goel, L. Bonnetain, R. Sharma, A. Furno, Mobility-based sir model for complex networks: with case study of covid-19, Social Network Analysis and Mining 11 (2021) 1–18.

[4] W. Oueslati, S. Tahri, H. Limam, J. Akaichi, A systematic review on moving objects' trajectory data and trajectory data warehouse modeling, Computer Science Review 47 (2023) 100516.

[5] L. Antón-González, M. Pans, J. Devís-Devís, L.-M. Gonzalez, Cycling in urban environments: quantitative text analysis, Journal of Transport & Health 32 (2023) 101651.

[6] C. Ding, D. Wang, C. Liu, Y. Zhang, J. Yang, Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance, Transportation Research Part A: Policy and Practice 100 (2017) 65–80.

[7] L. Wysling, R. S. Purves, Where to improve cycling infrastructure? assessing bicycle suitability and bikeability with open data in the city of paris, Transportation research interdisciplinary perspectives 15 (2022) 100648.

[8] X. Zhou, K. Zheng, H. Jueng, J. Xu, S. Sadiq, Making sense of spatial trajectories, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015, pp. 671–672.

[9] X. Kong, Q. Chen, M. Hou, H. Wang, F. Xia, Mobility trajectory generation: a survey, Artificial Intelligence Review 56 (2023) 3057–3098.

[10] Y. Zheng, Trajectory data mining: an overview, ACM Transactions on Intelligent Systems and Technology (TIST) 6 (2015) 1–41.

[11] D. M. Scott, W. Lu, M. J. Brown, Route choice of bike share users: Leveraging gps data to derive choice sets, Journal of Transport Geography 90 (2021) 102903.

[12] A. Meister, M. Felder, B. Schmid, K. W. Axhausen, Route choice modeling for cyclists on urban networks, Transportation research part A: policy and practice 173 (2023) 103723.

[13] S. Saki, T. Hagen, A practical guide to an open-source map-matching approach for big gps data, SN Computer Science 3 (2022) 415.

[14] R. Pritchard, D. Bucher, Y. Frøyen, Does new bicycle infrastructure result in new or rerouted bicyclists? a longitudinal gps study in oslo, Journal of transport geography 77 (2019) 113–125.

[15] T. Hagen, J. Hamann, S. Saki, Discretization of urban areas using POI-based tesselation, Technical Report, Working Papers, 2022.

[16] M. De Berg, A. F. Cook IV, J. Gudmundsson, Fast fréchet queries, Computational Geometry 46 (2013) 747–755.