9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Understanding how Road Network Safety Impacts Cycling Behaviour

Keywords: Mobility, Urban Planning, GPS Traces, Human Behaviour

Extended Abstract

Promoting sustainable mobility brings beneficial effects to cities. Bike2Work, uses gamification and economic incentives to encourage citizens to use bicycles for home-work commuting [1]. An important aspect that may hinder the regular use of bicycles may be represented by the lack of safe routes to ride. Through the data collected by a specifically designed app (Ferrara Play&Go) [1], we analyze the bicycle trajectories to understand how much the paths traveled by users diverge from the optimal ones (i.e., shortest paths in terms of length). With users' mobility data and street-level information, we aim to understand whether the underline street network's safety level plays a role in the paths selected by cyclists.

To access street-level information, we download the street network from OpenStreetMap (OSM) keeping only the roads on which bikes are allowed (e.g., we remove highways), and then analyze and process the anonymized GPS traces collected during the Bike2Work campaign. The dataset includes 26,221 trajectories generated by 605 users over six months from May 2021 to September 2021.

Formally, we define a trajectory as follows. A spatio-temporal point $p = (t,l)$ is a tuple where t indicates a timestamp and l a geographic location.A trajectory $P_i^u = p_1, p_2, \ldots, p_n$ is a time-ordered sequence of n spatio-temporal points visited by a user u, who may have several trajectories, $P_1^u, \ldots, P_k^u$, where all the locations in P_i^u are visited before locations in P_{i+1}^u.In our dataset, l is a tuple of latitude and longitude with the GPS points sampled every five seconds.
The street network downloaded from OSM can be formalized as follows: A street network $SN = (V,E)$ is a directed graph where the vertices $v \in V$ are intersections or initial/final points of a road and the edges $e \in E$ are the streets. Each vertex v_i has an associated latitude and longitude, while each edge $e_{(v_i,v_j)}$ has a set of properties $a_{e_{(v_i,v_j)}}$ (e.g., speed limit, size, etc.) and is connected with two vertices. To unveil potential issues on the road network, assuming that a study participant tries to reach their workplace/home as fast as possible, we evaluate how much a participant's observed trajectory deviates with respect to its corresponding shortest paths (both in terms of time and length). For all users' trajectories, we compute the shortest path between trajectories' origins ($p_1 \in P$) and destinations ($p_n \in P$). First, we map the origins and the destinations with the nearest $v_i \in SN$ using the ball tree algorithm for Haversine nearest neighbor search implemented in *osmnx*.

Then, we apply the Dijkstra algorithm to compute the shortest path on SN. We generate two different shortest paths that we use to create two weighting schemes for the edges: (i) the *time* needed to commute on an edge, and (ii) the *length* (in meters) of the edge. The lower the time or the length, the more similar the observed trajectory is to the shortest one (see Figure 1).

A reason for cyclists to deviate from the optimal path may not just be related to the distance to commute. The street network may play an important role. Suppose that a cyclist has to commute between two destinations nearby, but the shortest path is a dangerous road. It is likely that a cyclist will use a longer but safer path. To validate this hypothesis, we computed the so-called Level of Traffic Stress (LTS) [2], an index that, given some meta-information about the street (e.g., speed limit, street size, type of cycle lane), classifies the streets into four different levels of danger for cyclists. LTS 1 represents streets with no or little stress, suitable for children (e.g., a large cycle lane completely separated by other streets), while LTS 4 represents streets for expert cyclists which are more dangerous to travel (e.g., a road where speed limits for cars are significantly high and there are no dedicated cycle lanes). We compute the LTS for the city of Ferrara using the algorithm provided by BikeOttawa. This algorithm takes into

[1]https://airbreakferrara.net/ferrara-playngo/

consideration the characteristics of a street provided by OSM and leverages them to compute the LTS score. After discarding the streets on which bikes are not permitted (e.g., highways), the model retrieves the information on the remaining streets such as the number of lanes, the maximum permitted speed, the presence of parking lots on the side of the street, and whether or not the bike line is separated from the rest of the street. A combination of the aforementioned data corresponds to the LTS score of a street. For example, the LTS is equal to 1 when there is a separated bike line. When the bike lane is adjacent to a road with a maximum speed greater than 65 km/h, the LTS score is 4. Before computing an LTS score for each edge $e \in E \in SN$, we performed a map-matching process to infer the path traveled on the road network from a participant's GPS trajectory [3]. To compute the stress level of a bike trip we average the LTS scores of each of the road segments included in a trajectory. In particular, a trajectory T is now a set of edges where each edge e_i, among the other attributes, have an associated LTS. We indicate the LTS of a specific edge as e_i^{LTS}. The score for a trajectory is the average LTS score of all the street segments used to travel from the origin to the destination. Comparing the observed original trajectories with the corresponding optimal trajectories (i.e., shortest paths), we observed that cyclists in Ferrara not only tend to travel longer paths (original trajectories (median) $= 3451.82$ meters; optimal trajectories (median) $= 2948.81$ meters) but also travel on streets that display lower stress levels. Regarding the latter, in Figure 2, for each trajectory, we plot a point and we report the observed LTS score (on the y-axis) and the score associated with the relative shortest path (on the x-axis). Points on the diagonal (the grey dotted line) are the trajectories with an observed LTS score equal to one of the observed trajectories. Points below the diagonal are travels that have an observed score lower than the one of the optimal path. This means that the individual who registered the trajectory decided to take a safer path, even if longer as suggested by the trend line (orange line) in Figure 2 (slope $= 0.3437$, intercept $= 1.3851$). The trend line was computed with the ordinary least squares method. Similar results were obtained using shortest paths based on time. Looking at Figure 2, it is not clear if the safeness of the selected path is also related to the length of the travel. To answer this question, we consider all the travels shorter than 10 Km and we divided them into deciles based on the distance between the origins and destinations. In other terms, trajectories in the first bin are the shortest while trajectories in the last bin are the longer. For all the trajectories we computed the observed LTS and the LTS of the optimal path. In Figure 3, we can see two boxes for each bin representing the median and standard deviation of the LTS scores of the trajectories in each bin. The red boxes are the scores of the optimal trajectories while the blue boxes are the scores of the observed trajectories. It clearly emerges that regardless of the distance traveled, on average bikers tend to travel on safer roads (observed LTS is lower than the optimal LTS). Moreover, we observe that the longer the distance traveled, the larger the difference between the optimal and observed LTS. These analyses could potentially be used to find the trajectories that show a large divergence in the average LTS score between optimal and observed trajectories. Therefore, based not only on the stress level but also on how much a street is used by cyclists, we could locate the more problematic streets. This, in turn, could be used to inform municipalities about improving the cycling road network.

References

[1] Antonio Bucchiarone, Annapaola Marconi, Piergiorgio Cipriano, and Luca Giovannini. BIKE2WORK: A shift towards sustainable mobility. In Cornel Klein and Matthias Jarke, editors, *Proceedings of the 11th International Conference on Smart Cities and Green ICT Systems, SMARTGREENS 2022, Online Streaming, April 27-29, 2022*, pages 147–156. SCITEPRESS, 2022.

[2] Maaza C Mekuria, Peter G Furth, and Hilary Nixon. Low-stress bicycling and network connectivity. *Mineta*, 2012.

[3] Can Yang and Gyozo Gidofalvi. Fast map matching, an algorithm integrating hidden markov model with precomputation. *International Journal of Geographical Information Science*, 32(3):547 – 570, 2018.

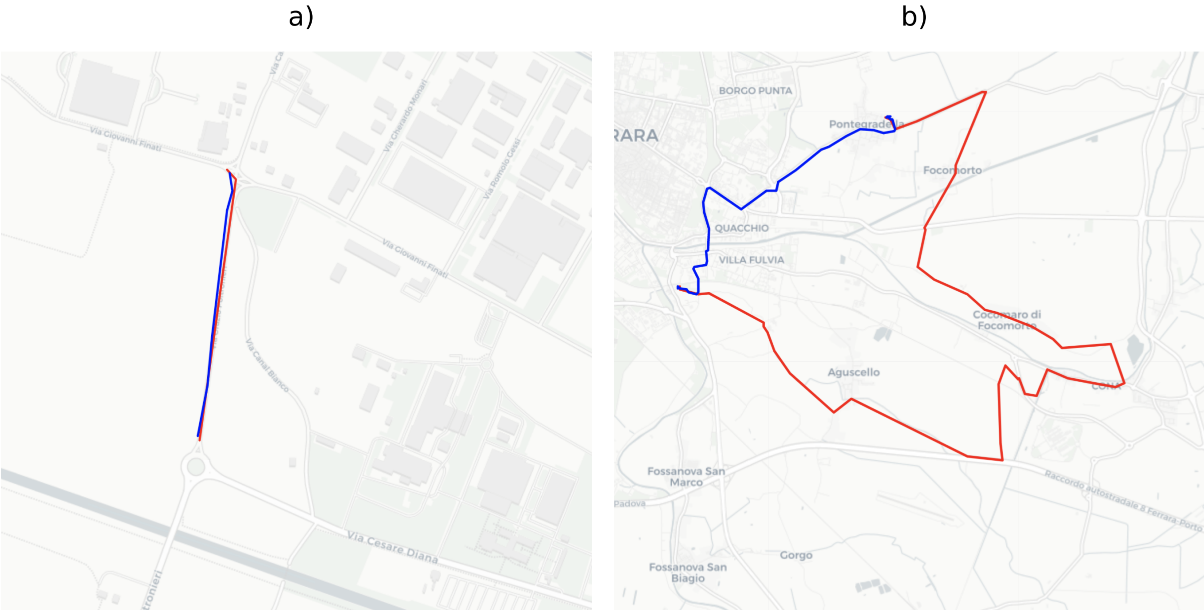

Figure 1: Users' trajectories example. (A) A user's short trajectory difference between the original trajectory and the shortest one; (B) a long trajectory difference between the original trajectory and the shortest one.

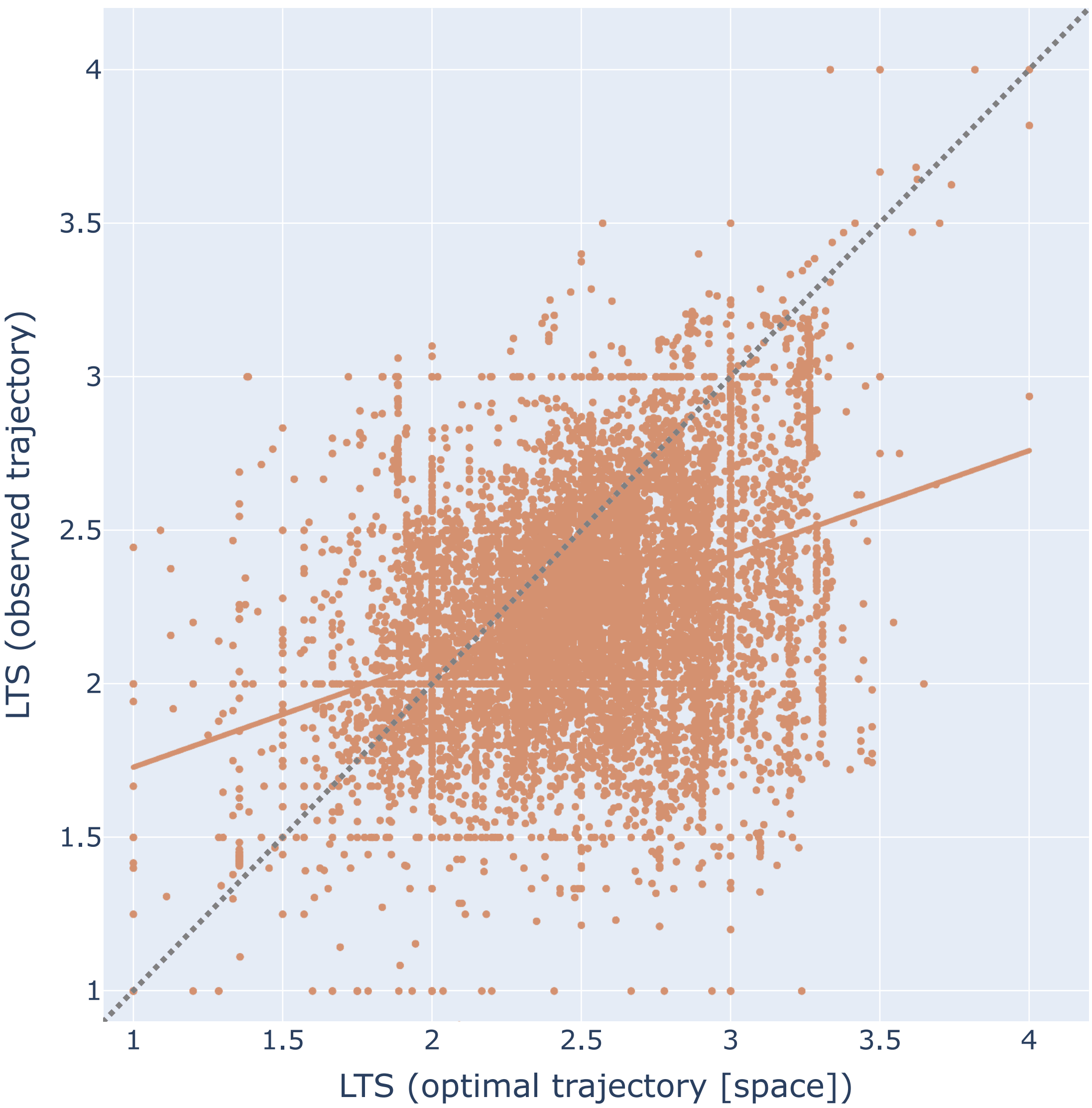

Figure 2: LTS score of observed trajectories (y-axis) and LTS score of optimal trajectories in terms of space (x-axis). Each point represents a trajectory, while the orange line represents the trend line computed using Ordinary Linear Square regression (orange line). Finally, the diagonal gray dotted line represents the case in which both the shortest path and the observed paths have an equal LTS. The majority of points lie under the diagonal which means that cyclists tend to travel on roads with lower LTS (safer roads) despite not being the shortest ones.

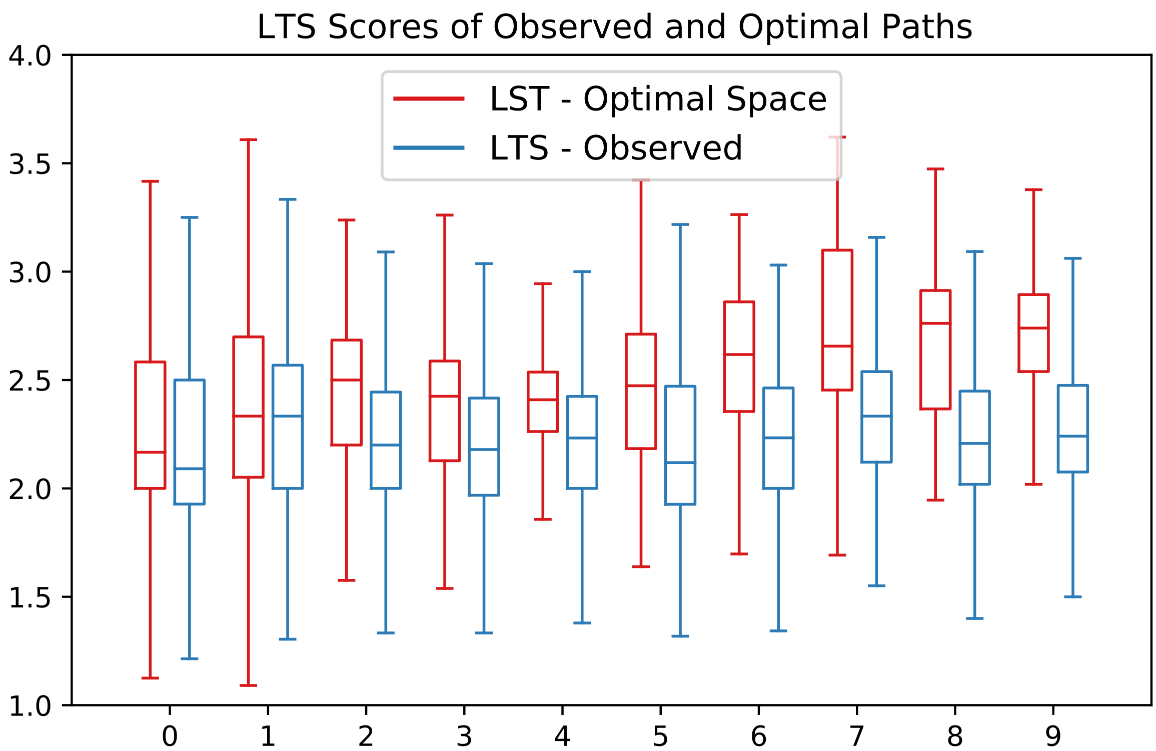

Figure 3: We divided the trajectories sorted by growing distance between origin and destination into ten equal-sized bins. For each group, we estimate the median (line in each box) and the variance (box-tails) of the LTS scores observed (in blue) and one of the optimal paths (in red). It is possible to see how, in general, the LTS of the optimal path augment (i.e., more dangerous path) as the distance between origin and destination increases. On the other hand, the LST scores of the observed paths tend to be stable regardless the distances commuted.