Check for updates

Special Issue: Advances in Spatial and Transport Network Analysis

B Urban Analytics and City Science

EPB: Urban Analytics and City Science
2022, Vol. 0(0) 1–16
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23998083221135611
journals.sagepub.com/home/epb
SAGE

Data-driven micromobility network planning for demand and safety

Pietro Folco and Laetitia Gauvin
ISI Foundation, Italy

Michele Tizzoni
ISI Foundation, Italy
University of Trento, Italy

Michael Szell
IT University of Copenhagen, Denmark

Abstract

Developing safe infrastructure for micromobility like bicycles or e-scooters is an efficient pathway towards climate-friendly, sustainable, and livable cities. However, urban micromobility infrastructure is typically planned ad-hoc and at best informed by survey data. Here, we study how data of micromobility trips and crashes can shape and automatize such network planning processes. We introduce a parameter that tunes the focus between demand-based and safety-based development, and investigate systematically this tradeoff for the city of Turin. We find that a full focus on demand or safety generates different network extensions in the short term, with an optimal tradeoff in-between. In the long term, our framework improves overall network quality independent of short-term focus. Thus, we show how a data-driven process can provide urban planners with automated assistance for variable short-term scenario planning while maintaining the long-term goal of a sustainable, city-spanning micromobility network.

Keywords

urban data science, micromobility infrastructure, sustainable mobility, road safety

Introduction

In the transport sector, micromobility like cycling is one of the most promising and economic solutions to the climate crisis (Creutzig et al., 2015; Gössling et al., 2019; Lamb et al., 2021; Nieuwenhuijsen 2020). At the same time, the most ideal environments to push for more active travel

Corresponding author:
Michael Szell, NEtwoRks, Data, and Society (NERDS), ITU Copenhagen, Rued Langgaards Vej 7, Copenhagen 2300, Denmark.
Email: misz@itu.dk

are cities because most trips are intraurban and of short length (Alessandretti et al., 2020). However, to increase the generally low modal share of cycling and other forms of micromobility, well-designed and cohesive urban micromobility infrastructure networks are needed (CROW 2016; Nieuwenhuijsen 2020).

Traditionally, the development of such networks follows a local, ad-hoc approach without a systematic understanding of network effects (Szell et al., 2022; Natera Orozco et al., 2020; Zhao et al., 2018). Those approaches that are data-driven are typically based on survey data which provide only a rough proxy for trip demands (Larsen et al., 2013; Lovelace et al., 2017). However, with the increased availability of high-granularity and high-frequency data sets, new data-driven approaches are emerging. On the one hand, detailed infrastructure data from crowdsourced platforms like OpenStreetMap (OSM) and their eased access via novel computational tools (Boeing 2017) enable scientific inquiry of structure-based strategies and limitations to bicycle network growth (Szell et al., 2022; Natera Orozco et al., 2020; Vybornova et al., 2022). These approaches place value on structural network properties while sacrificing specifity such as concrete demand models. The idea is that independently of the current demand patterns, eventually a cohesive, city-spanning network should be established in which long-term effects such as induced demand will push the transport system into a sustainable equilibrium (ITF, 2021, OECD Publishing, 2021; Lyons and Davidson, 2016; Nelson and Allen, 1997). On the other hand, modern approaches that incorporate a variety of empirical data sets can improve the estimation of flows or potential demand to better inform priorities for short-term investments (Olmos et al., 2020).

Here, we follow a modern data-driven approach by extending the structural-only network growth model of Szell et al. (2022) to account for: 1) the existing bicycle network and 2) empirical micromobility data sets of e-scooter trips and bicycle crashes, collected for the city of Turin, Italy. Our approach combines the strengths of a long-term structural growth process that aims to develop a city-spanning network (Szell et al., 2022), with flow (Olmos et al., 2020) and crash data (Larsen et al., 2013) which are important on the short-term for considering demand and avoiding deaths and injuries from crashes. Thus, our framework provides a balance on two scales: first, on the short-term task of "putting out fires" with surgical investments while maintaining the long-term focus on a cohesive network, and second on deciding how much to focus on one versus another data set. As we show in this paper, our framework consolidates short-term and long-term development goals, and establishes a data-informed trade-off between demand and safety. Although our main goal is to study how a network planning process could be informed by data sets in general, we also conclude with concrete development scenarios of new protected bicycle tracks for the city of Turin which plans to implement 80 km (Reyneri 2020) of traffic calming measures. We apply our method to the case of Turin due to data availability, but it is applicable to any city with multiple data sets to inform tradeoffs between different development goals.

Data acquisition and processing

Infrastructure networks

We downloaded existing transport infrastructure networks of Turin from OpenStreetMap (OSM) using the Python library *OSMnx* (Boeing 2017). The street network and the cycling network have been downloaded and topologically simplified, where the bicycle network is the union of OSM data structures that encode on-street and off-street protected bicycle infrastructure. In these networks, nodes correspond to intersections, and links correspond to streets between them. Figure 1(a) displays the street network of Turin and the existing cycling infrastructure (in green), as of July 2021. For additional details on the cycling network see Section S1 in the Supplementary Information.

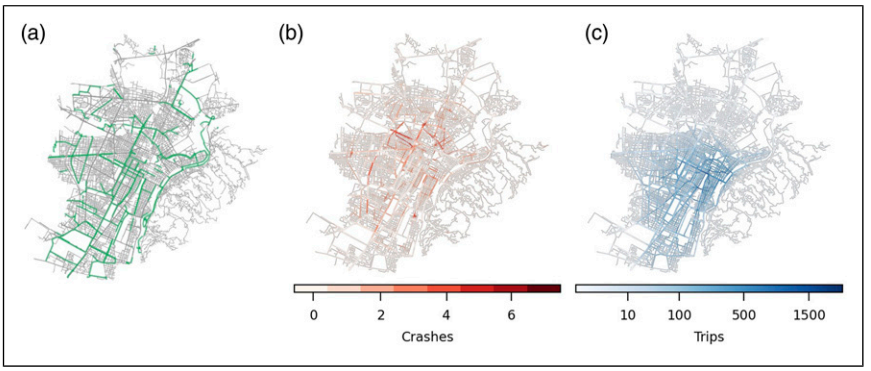

Figure 1. Street and cycling infrastructure, crash and trip data for the city of Turin. (a) Existing cycling infrastructure is shown in green, the street network is shown in gray. (b) Data of road crashes that involve at least one bicycle, in the year 2019. (c) E-scooter trip data from 26 May 2021 to 28 October 2021. Trips are estimated using the locations (origins and destinations) of Bird e-scooters.

Bicycle crashes in Turin in 2019

We used a dataset of bicycle crashes in Turin in 2019[1] providing a variety of information, including date and time, number, and type of vehicles involved, or geographical coordinates of the crash. We have only considered crashes involving at least one bicycle, yielding 314 crashes in total. Figure 1(b) shows the distribution of crashes across the segments of the street network of Turin, where a crash is assigned to a street if it happened within 50 m of it.

E-scooter real-time positions

We collected real-time positions of e-scooters of the e-scooter sharing company *Bird* in Turin by accessing its public API.[2] Each query returns the position of all e-scooters that are parked within a given radius from a pair of coordinates. To cover the entire city, we divided it into a grid consisting of 71 grid cells of size 1506 m and ran a query for each centroid of the grid cells. We set the side of the cells to 1506 m so that all adjacent queries overlap and we do not miss data in between. Duplicates resulting from overlaps are removed. Time resolution of each query is 50 min. We discarded cells of all the industrial, suburban, or hilly areas of the city where no e-scooters are present because parking is not allowed by the service (see Section S2 in the Supplementary Information for more details).

We collected data of the location of *Bird* e-scooters from 26 May 2021 until 28 October 2021. The relevant features returned by each query for each e-scooter are as follows: geographical coordinates of vehicle position, an identifying label allowing to track the movements of the vehicle, and battery charge level. We used information on battery charge levels to identify e-scooter movements made by the company to re-locate the vehicles and charge them, and discarded them from our dataset. Based on the location data of e-scooters, we computed the density of e-scooters in the city and built an origin-destination (OD) matrix. Each origin and destination corresponds to the start and end point of an e-scooter's movement. We define a movement if:

(i) An e-scooter changes its position by at least 100 m.
(ii) The time between geolocation at the origin and at the destination is less than 90 min. We add this condition because sometimes the API service has crashed and therefore some queries ($\approx 8\%$) are separated by more than 50 min. We do not add movements when the queries have

a separation of more than 90 min because the more time passes between one query and the next, the more there is the risk that a change in the position of an e-scooter is the sum of more than one movement.

Before creating the OD matrix, we cleaned the OD data from movements made by the company: Vehicles are sometimes relocated to high-demand areas (railway stations, office areas, etc.), additionally the vehicle's battery must be charged. To clean up the dataset from this kind of movement we used the following rule. A movement of an e-scooter is removed if one of the following two conditions is satisfied: (i) The vehicle "disappears" from the city and when it reappears the battery level has increased (which should happen when the company takes the vehicle to a warehouse to charge its battery), or (ii) the vehicle "appears" somewhere, but the battery level did not decrease with respect to the previous queries (which should happen when the company relocates the vehicle).

Finally, we created a trip dataset from the OD matrix. We used each OD sample to define a *trip* by computing the shortest path on the combined street and bicycle network between the origin and destination coordinates. Each calculated shortest path is a trip, yielding a dataset containing 40, 694 trips. Figure 1(c) shows the spatial distribution of trips over the street network of Turin.

As of 2022, there are seven e-scooter companies operating in Turin. We have no information on how much of the *Bird* data represents the overall demand and use of e-scooters for travel. However, the operational areas of all e-scooter providers are essentially overlapping.

Model description

Our network development model extends the growth framework developed by Szell et al. (2022) by integrating existing bicycle infrastructure and data on micromobility trips and bicycle crashes. The goal of this model is a general development towards a cohesive network, but also to account for safety and demand using the crash data (proxy for safety) and the e-scooter trip data (proxy for demand). We can summarize our model in six steps:

Step 0—Definition of area of study

The area of study we consider is the entire municipality of Turin with the exception of some underpopulated parts. We have chosen to remove from our analysis the underpopulated areas of the city by setting a threshold on the number of inhabitants:[3] All the statistical areas with population density below 1, 185/km^2 were removed from the analysis. This effectively removes the Eastern hilly part of the city which is a low demand residential area. We have added an exception to the above rule by including parks and green areas, even though the number of residents is obviously below the threshold as parks might include potentially safe cycling areas. Figure 2(a) shows the portion of the road network of Figure 1(a) that was included in the study.

Step 1—Definition of seed points

We define a set of seed points snapped to the intersections of the entire infrastructure network (composed by the road network and the existing bicycle network). We also define a parameter δ, as the length in meters that sets the minimum distance between two seeds. The process of defining the seeds is the following: i) we add the seeds to the intersections of the existing infrastructure such that all seeds are at least at distance δ m from each other, ii) we add the seeds on the intersections of the entire network from a grid of side δ m. If the distance between a seed on the existing infrastructure and a potential new seed on the grid is less than δ m, the seed is discarded (see Figure 2(b) for

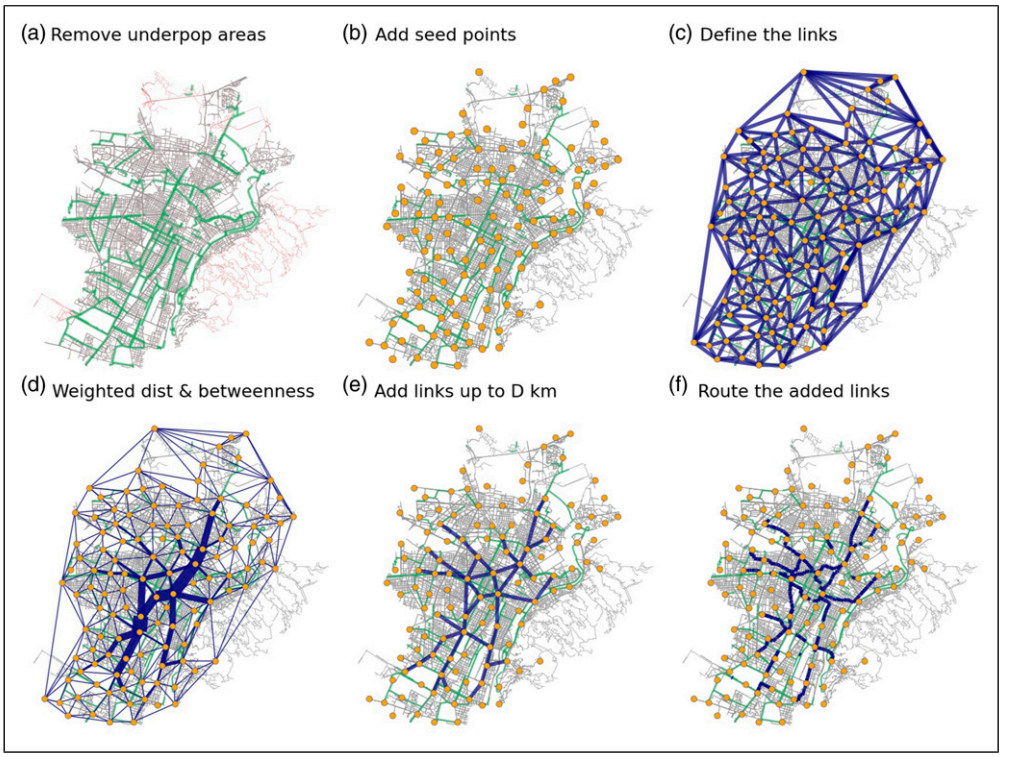

Figure 2. Pipeline of the model. (a) Step 0: the underpopulated and hilly areas of the city (red links) are removed: no seed points will be added in these areas. (b) Step 1: Seed points are added following a grid structure and the existing cycle network. (c) Step 2: The link of the abstract network is defined. (d) Step 3: For each link, the weighted distance d_W is defined, then the edge betweenness (represented with the width of the links) is calculated. (e) Step 4: The links are ranked by betweenness, then the top ranked links are selected up to a total of D km. (f) Step 5: The selected links are routed on the street network. The blue links represent the solution proposed by the model, they integrate the existing cycle network (green links).

details). In the remainder of the paper, we set $\delta = 300$ m which is reasonable with respect to the scale of the road network in Turin (length of the links, frequency of intersections). We also considered other values of δ in the range 200–700 m, and our results proved to be robust against variations in δ (see Section S3 of the Supplementary Information).

Step 2—Definition of potential links

Next, we choose the potential links we will create between the seed points. We sort all pairs of seeds in ascending order of route distance, then we connect them stepwise with a straight link. A potential link is added if it does not cross another previously added potential link. Figure 2(c) shows the result of this process. This condition, combined with the ascending orders of the pairs of seeds, follows the considerations of Szell et al. (2022) to create a cohesive network via a greedy triangulation (Cardillo et al., 2006), in which the closest triangle-creating pairs of seeds are connected locally. Note that this process ignores links of the existing bicycle network: if a new link would cross only edges from the existing infrastructure, it will be added nevertheless.

Szell et al. (2022) motivate use of the greedy triangulation as optimizing investment costs (Cardillo et al., 2006), that is, minimizing total length, however, geometric optimization such as

Delaunay triangulation could also be a possibility (Barthelemy 2022). We compare these two methods in Section S7 in the Supplementary Information; we find high overlap between the different networks generated (Figure S7), qualitatively identical ranking and growth behavior of the metrics (Figure S8), and slightly better results for the greedy triangulation. Boundary effects are not an issue for the range of growth studied.

Step 3—Weighted distance

Assuming we have a budget to build D km of potential links, we have to choose which potential links to build among those defined in step 2. For this purpose, we define a weighted distance d_W as the route distance adjusted with the number of crashes covered by the routed link and the number of trips passing through the routed link (to see details on how we count crashes and trips, see section S4 in the SI). We define the weighted distance d_W for each potential link from step 2 and for each link of the existing infrastructure. For a link e (A, B) we define d_W (A, B) as follows

$$d_W(A,B) = \alpha\, d_{\text{trip}} + (1-\alpha)\, d_{\text{crash}} \tag{1}$$

where:
(i) $\alpha \in [0, 1]$
(ii) $d_{\text{trip}} = \frac{(d+1)}{(1+9N_{\text{trip}})}$ with $N_{\text{trip}} = \frac{n_{\text{trip}}}{\max(n_{\text{trip}})}$
(iii) $d_{\text{crash}} = \frac{(d+1)}{(1+9N_{\text{crash}})}$ with $N_{\text{crash}} = \frac{n_{\text{crash}}}{\max(n_{\text{crash}})}$

The variable d denotes the length of e (A, B) in meters, n_{crash} is the number of crashes per km covered by link e (A, B) and n_{trip} is the number of trips per km passing through link e (A, B). The range of the denominator of d_{crash} and d_{trip} is the interval [1,10]. In this way, the range of variability of d_{crash} and d_{trip} is the same. We chose this definition for the denominator because it is the easiest way to set the same range of variability for both crashes and trips without adding nonlinear effects. The weighted distance affects the edge betweenness centrality calculated for the choice of links in the next step 4: A smaller d_W (A, B) contributes to a higher betweenness for link e (A, B). In Figure 2(d), the width of potential links is proportional to their betweenness.

Using this weighting, the parameter α controls the influence of safety versus demand. A smaller α makes the model more "safety oriented" while a higher α makes the model more "demand oriented." The extreme values 0 and 1 define two models that use, respectively, only crash and trip data.

Step 4—Choose the links

A criterion is needed to choose a number of km of links equal to D km from among the potential links. Following the considerations of Szell et al. (2022), we choose edge betweenness as a simple proxy for flow, and as an effective way to build a functional bicycle network quickly, to rank the potential links. A link with a small weighted distance is more likely to be ranked in the top positions. We add the links iteratively to the proposed solution, starting from the one with the highest betweenness, until the length of the proposed solution has reached D km. The result is the network shown in Figure 2(e).

Step 5—Routing

The links selected in step 4 are routed on the street network. The output of the model is a bicycle network made up of the existing infrastructure and the new D km of the proposed solution (Figure 2(f)).

Evaluation metrics

To monitor the network growth process, we generate the network up to $D = 295$ km and store a snapshot of the network every new 5 km of investment. Recently, the city of Turin announced 80 km of new investments in traffic calming measures (Reyneri 2020). In this work, we assume that the 80 km of announced infrastructure are equivalent to 80 km of new cycling paths, and we set as main target the value of $D = 80$ km of new investments.

To assess how our network growth model uses micromobility data to drive investments, we evaluate the potential improvement of two different metrics, one representing safety and the other demand: the *crash coverage* and the *trip coverage*. For both metrics, we compute their potential improvement as

$$\pi(D) = \mu(D) - \mu(0), \tag{2}$$

where $\mu(D)$ is the metric μ calculated with D km of new added links, and $\mu(0)$ is the metric μ calculated on the existing bicycle infrastructure (i.e., when $D = 0$ km).

Crash coverage

To assess how many crashes (involving bicycles) would be covered by a proposed bicycle network, we define a crash coverage metric. We first assume that a road segment is safer if there is a protected cycle track, so we expect a reduction in the number of crashes near it. Since we cannot compute how many crashes would have been avoided with a cycle track, we count how many crashes are covered by the proposed cycling infrastructure, assuming that higher coverage increases road safety (Teschke et al., 2012). We label a crash as "covered" if located within 50 m of a bicycle infrastructure element (see Figure S3 in the SI for technical details). In this way, we can see how the number of crashes covered in the city increases as the network grows and compare these results between different settings of our model. Mathematically, we define the crash coverage C as

$$C = \frac{N_{\text{cov}}}{N_{\text{tot}}}, \tag{3}$$

where N_{cov} is the number of crashes geolocated within 50 m of a bicycle infrastructure link, and $N_{\text{tot}} = 314$, the total number of crashes involving bicycles in 2019.

Trip coverage

Parallel to the crash coverage metric, we define a trip coverage metric to assess how the proposed solution fits the demand. Given M trips chosen randomly, we calculate the percentage of km traveled on the cycle infrastructure with respect to the total km traveled. Mathematically, we define the trip coverage T as

$$T = \frac{\sum_m l_{\text{bike}}(m)}{\sum_m l_{\text{tot}}(m)}, \tag{4}$$

where $l_{\text{bike}}(m)$ is the number of km on the bicycle path of trip m, and $l_{\text{tot}}(m)$ is the total length (in km) of trip m.

In defining our metric, we calculate the shortest path between origin and destination for each trip. As a baseline, we do not take into account that a user may accept a certain percentage of detour if it allows to cycle on safer infrastructure, for example, a protected cycle track. Our way of calculating the trip (shortest path) therefore gives us the trip coverage if users accept a 0% detour, which can be

considered as a lower bound. For a more realistic analysis, we also evaluated trip coverage for an accepted detour level of 25%, which yields similar results in terms of potential improvement (see Section S5 in the Supplementary Information for details).

Results

Using our development model, we generated networks of new cycling paths in Turin by varying the parameter α in the range [0, 1], and for increasing values of realistic investments D, representing the total length in km to be added to the existing infrastructure. Figure 3 shows the performance of the model in terms of potential improvement in trip coverage (Figure 3(a)) and crash coverage (Figure 3(b)), as $D \geq 5$ km of new cycling paths are added to the network, for three characteristic values of α. As expected, we obtain the largest improvement in crash coverage for $\alpha = 0$ and the largest improvement in trip coverage for $\alpha = 1$. For any value of D, trip coverage for $\alpha = 1$ is consistently higher than for $\alpha = 0.5$ which in turn is higher than for $\alpha = 0$. On the other hand, we see the best performance in crash coverage for $\alpha = 0$ for each D (with the exception of $D \leq 10$ km where statistical noise dominates over potential negligible improvement around 1%, insensitive to the choice of α). The differences in performance as α varies are evident at $D = 80$ km but also for lower values of D, showing that the model is reliable also at intermediate steps of the development plan. Therefore, our results confirm that the response of the model with respect to the parameter α provides urban planners with a choice between directing an investment plan towards the optimization of travel demand or safety simply by varying α, in any reasonable range of investment D. In general, for the data at hand the potential improvement achieved by the model in trip coverage (Figure 3(a)) is higher than in crash coverage (Figure 3(b)). For instance, considering the results at $D = 80$ km, trip coverage displays a potential improvement of 22% (corresponding to a trip coverage of T = 29%) for $\alpha = 1$, while crash coverage improves by only 16% (corresponding to a crash coverage of C = 54%) for $\alpha = 0$. Such a performance gap decreases with larger values of new investments D, as shown in Figure S6 of the Supplementary Information. In the extreme case of $D = 295$ km, the potential improvement of trip and crash coverage achieved by the model is about 30%, in both cases. Tables S1, S2, and S3 of the Supplementary Information report numerical values of improvements for all metrics, and for $\alpha = 0, 0.5, 1$.

Different choices of α values lead to different structures of the generated cycling network. As an example, Figure 4 shows the first 25 km of new cycle infrastructure generated by the model in Turin

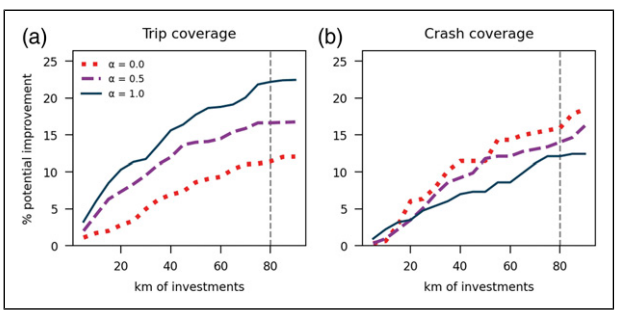

Figure 3. Network metrics improvement as a function of added cycling infrastructure in kilometers. (a) Potential improvement of trip coverage for α = 0.0, 0.5, 1.0. (b) Potential improvements of crash coverage for α = 0.0, 0.5, 1.0. As expected, lower α values lead to higher improvements for a safety-focused growth; while higher α values correspond to higher improvements for a demand-focused growth. The dashed vertical line indicates the value D = 80 km.

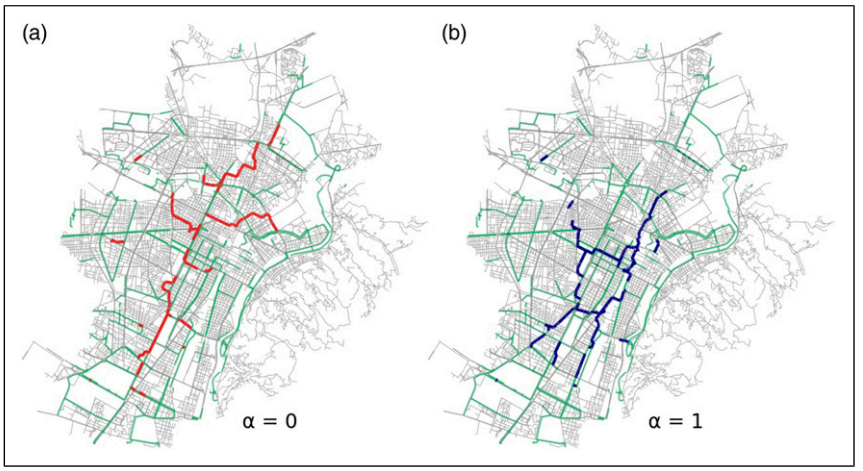

Figure 4. The proposed solutions at 25 km of investments using the model described in section Model description. (a) The proposed solution for $\alpha = 0$, that is, using only crash data. (b) The proposed solution for $\alpha = 1$, that is, using only trip data.

for $\alpha = 0$ (panel A) and $\alpha = 1$ (panel B). For $\alpha = 0$, when only crash data are taken into account, new cycling paths along the North–South direction of the city are initially prioritized, while for $\alpha = 1$ most of the new links are added in the central area of the city. In particular, we can observe that for $\alpha = 1$ the model generates a cycling ring around the main train station of the city. Such different prioritizations of new links are mainly related to the spatial distribution of the crash and trip input datasets shown in Figure 1, confirming that the model is well adapted to take into account different inputs. Crashes are more uniformly distributed throughout the city with some hotspots, while trip demand is highly concentrated in the city center, the area where most of the e-scooter trips are concentrated and the service is available (see Figure S1 in the Supplementary Information for a map of the area where e-scooters can be parked). Figure 4 also shows how the model integrates with the existing infrastructure by connecting its disconnected components. As expected, the number of disconnected components of the network decreases with increasing D, however, the trend of the number of components does not depend on α: as α varies, for a given D, the number of components remains approximately the same (see Section S7 in the Supplementary Information for details).

So far, we have presented the extreme cases of our network growth model: namely, $\alpha = 0$ and $\alpha = 1$. However, an urban planner may want to look for solutions that improve both safety and demand at the same time, striking a balance between the two. By setting intermediate values of α, our model takes into account both crash and trip data providing a solution that balances both inputs, following the weighting approach of equation (1). Thus, for a given investment in new cycling paths, D, it is possible to define a trade-off value of α, $\alpha^*(D)$, for which both metrics, trip coverage and crash coverage, achieve the same potential improvement. By setting $\alpha = \alpha^*$, the model provides the most balanced solution between demand and safety. It is important to notice that an equal balance between demand and safety would correspond to $\alpha = 0.5$, in terms of weighted distance d_W, however, this does not immediately translate into an equal improvement on both metrics, due to the different spatial distributions of trips and crashes.

As an example, Figure 5(a) shows how trip coverage and crash coverage respond to the variation of α, at $D = 80$ km. In this case, we find $\alpha^* = 0.38$, indicated with a dashed line on the plot. However, it must be taken into account that α^* is a quantity that depends (weakly) on D. Figure 5(b) shows the potential improvement in trip coverage and crash coverage at the trade-off value, for increasing D.

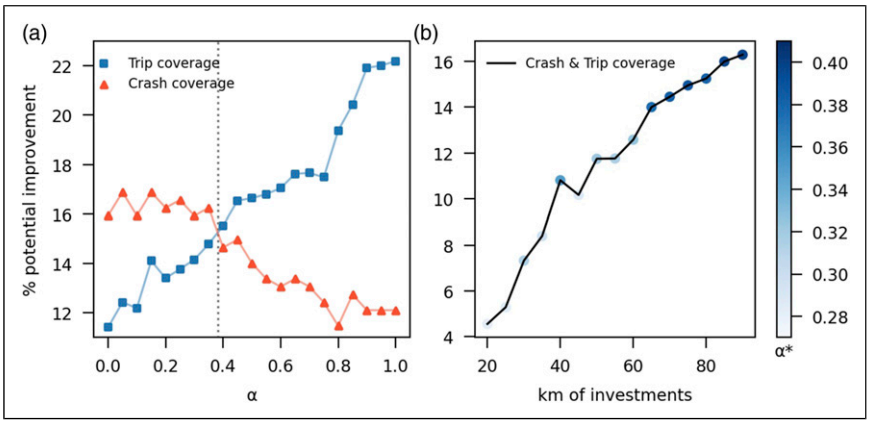

Figure 5. Trade-off between crash coverage and trip coverage. (a) The potential improvement at 80 km of new investments as a function of the parameter α. Crash coverage is decreasing with α, while trip coverage is increasing with α. At the value α^*, marked by a dashed line, both metrics achieve the same relative improvement. (b) The potential improvement as a function of km of investments, at the intersection α^*. The colorbar represents the values of α^*.

By measuring α^* for varying D in the range 20–90 km, we observed that α^* varies in a relatively small interval: $\alpha^* \in [0.28, 0.40]$. For these values of D and α^*, the potential improvement in trip coverage and crash coverage increases from about 4% up to 16%. This is a theoretical trend, since, as D varies, the potential improvement is calculated for $\alpha^*(D)$. This dependence on investment length helps us to understand which trend is obtained at the trade-off value, but in a practical context where the final D might be unknown, α must be set at the beginning of the process. In such a case, in Turin, a value of α between 0.28 and 0.40 would provide a balanced solution to generate new cycling paths that take into account both demand and safety.

Discussion

In this study, we have presented a new data-driven cycling network growth model, which aims to build up a cohesive network while incorporating multiple empirical data sets, optimizing different development targets at the same time. In particular, we developed the model to generate new cycling tracks, to be integrated into existing infrastructure, and to prioritize travel demand and cycling safety. Through one tunable parameter α, the model generates networks by either focusing on maximizing trip coverage to satisfy existing travel demand, or by focusing on crash coverage to increase safety, or by balancing a mix of the two targets. Our work extends previous network development approaches (Natera Orozco et al., 2020; Olmos et al., 2020; Szell et al., 2022) by including multiple dimensions into the model, in the effort of providing urban planners with a tool that can address different mobility needs of a city. To assist this effort, we released the code of our model publicly.

In this paper, we used the micromobility modes of e-scooters and bicycles interchangeably, being fully aware that these are not the exact same modes of transport. They could have potentially different user profiles, usage patterns, or risks due to the novelty of e-scooters (Sanders and Nelson 2022). We used on one hand bicycle crashes as a proxy for micromobility safety, on the other hand e-scooter trips as a proxy for micromobility demand. It would have been preferable to have also data of bicycle trips or e-scooter crashes available—but to the best of our knowledge those data are not collected or being made available publicly in Turin. We justify this choice as follows: First, our

study does not aim to generate a final, concrete network that can be implemented 1-to-1 by the city, but we aim to study the general trade-offs between different data sets when used concurrently for data-informed network growth. From this broad perspective, it is most important that the two data sets, despite being correlated (Section S9 and Figure S9 in the Supplementary Information), are different enough to produce different results, which is the case (see e.g., Figures 3–5). Second, e-scooter and cycling crashes (Shah et al., 2021) and their usage patterns in Turin (Chicco and Diana 2022) and elsewhere (Reck et al., 2021) have been shown to display many similarities. Third, Italian law (Città di Torino, 2021) considers e-scooters and bicycles as equivalent concerning traffic rules—they share the same network infrastructure. Fourth, in other Italian cities where both e-scooter and bicycle trip data are available, there is a strong correlation (see Section S8 in the Supplementary Information). Lastly, the approaches to non-cycling micromobility network planning are structurally identical or very similar to bicycle network planning which we used here (Ignaccolo et al., 2022; Fazio et al., 2021; Comi et al., 2022). Many modes of micromobility are relatively understudied due to their novelty; therefore, research on the topic is highly fluid and expected to provide new insights in the near future (Cook et al., 2022).

We applied our modeling framework to the specific case of the city of Turin, but the model can be easily extended to other cities for which micromobility travel demand such as bicycle or e-scooter trips, and crash data are available. From a theoretical perspective, our results do not depend on the city under study: the model is defined to balance the trade-off between two different mobility targets, here travel demand and safety, and it allows to identify a solution that prioritizes either one through a single parameter. Urban planners could adapt the model to include different metrics to be prioritized by the proposed solution, such as the number of speed-limited streets across the cycling path, street lightning, perception of urban environment, and other measures of comfort (Quercia et al., 2014, 2015). Of course, such model extensions will require the availability of geo-located data representing the metrics of interest that could be mapped onto the existing transport infrastructure.

Our work comes with limitations and possible extensions. In particular, specific instances of the cycling networks generated by our model may be sensitive to the input data representing the target metrics. In our study, we examined the case of Turin for which we had crash data from 2019 and trip data from 2021. The crash dataset contains 314 geo-located events in total, which is large enough to allow our data-driven planning approach, but also leaves open the desire for extension to reduce possible statistical noise. Crash data are notoriously underreported, especially for vulnerable road users (Adminaité-Fodor and Jost 2020; Olszewski et al., 2019), so a change in reporting procedures could improve the extent and statistical expressiveness of crash data ground truth. Despite general issues of reporting bias, we are not aware of concrete significant biases in the datasets we used here. Data from 2020 were also available but we did not include them into the model because urban mobility was strongly affected by COVID-19 restrictions in 2020 (Gauvin et al., 2021), and 2020 crash data cannot be considered a representative baseline. Unfortunately, crash data before 2019 were not geo-located. A more systematic mapping of bicycle safety in Turin, through a multi-year data collection, could improve our results by providing a more precise picture of crash hotspots across the city. Also, we considered crash locations as the most immediate proxy for street safety, however, other street-level features could be used to model risk, such as street width or number of intersections (Daraei et al., 2021).

As a proxy for travel demand, we considered e-scooter trip data. The dataset under study is quite large, including more than 40 thousand trips; however, we assumed that e-scooter riders have similar mobility needs of cyclists, but this may not be always the case. Recent studies have shown that micromobility modes display different characteristics: bike sharing services are predominantly used for commuting, while e-scooter sharing is more related to recreational activities (Reck et al., 2021; McKenzie 2019). We do not expect such differences to significantly affect our model, since they are

mainly reflected in temporal patterns of ridership rather than their spatial distribution, but a more specific disaggregation of trip demand by micromobility mode could improve our results and would be a natural extension of the model. Also, the e-scooter sharing service is only available within a part of the city area (see Figure S1), therefore, trip data are limited to a portion of the city. Combining data from different e-scooter/bike sharing providers would be a solution to overcome this limitation.

It is important to note that the relative improvement of target metrics is computed with respect to a baseline value defined at $D = 0$, that is, with the existing cycling infrastructure. Therefore, our model assumed that the existing infrastructure already satisfies trip demand and cycling safety at the baseline values of coverage. Such an assumption could be relaxed to improve model evaluation, for instance by defining different levels of trip and crash coverage, also at the baseline, depending on the type of infrastructure that already exists and that is planned. Further, our proposed approach will always result in initially centralized networks, see also Szell et al. (2022). However, a centralized initial solution generally reflects the typically centralized mobility flows in cities (Schläpfer et al., 2021), and the algorithm already places the seed points in a way to cover the whole city so that eventually all points will be reached. We consider implementing a non-centralized initial solution conceptually too different and outside of the scope of this paper, but we refer to Ospina et al. (2022) who develop a maximum coverage design from the beginning.

Moreover, we assumed that the addition of new bicycle tracks will automatically increase cycling safety, but this is a debated issue in the literature. Although there exist empirical guidelines (CROW 2016), there is a lack of scientific consensus on which specific type of infrastructure provides the best safety, and under which conditions it should be installed. Generally, exposure data can provide evidence for greater protection through physically separated infrastructure (Teschke et al., 2012), but safety is such a complex, yet unresolved topic requiring a deep discussion of a multitude of variables (Klanjčić et al., 2022) or elaborate tools and analysis (Kondo et al., 2018), that we consider it outside the scope of this work. It is, however, important to note that the total number of crashes is just a reflection of infrastructure safety versus route choice since we were not able to consider exposure data. For example, it might happen that unsafe roads are not classified as such, because no cyclists are using them and thus, no crashes occur there. Adequately corresponding data of bicycle trips, which we do not have unfortunately, would allow to calculate crash risk per travelled length. For this, it would be important to have data over a longer observation period with a larger number of crashes, possibly assisted by state-of-the-art methods like Bayesian analysis (Kondo et al., 2018), towards adequate statistical robustness. Nevertheless, the absolute number of crashes is important to consider for short-term investments, that is, in the realistic scenario where the city wants to "put out fires."

To summarize, our approach is as an extension of the network growth model of Szell et al. (2022), which overcomes two previous limitations: Our new model 1) does not ignore existing cycling infrastructure but snaps seed points explicitly to existing bicycle tracks, and 2) it incorporates empirical traffic data to allow meeting short-term goals. These are two important steps away from a purely theoretical growth model towards a more realistic approach that could be useful for both underdeveloped and developed cities. By combining the goals of satisfying demand/safety and structural cohesion, we unite a short-term engineering approach of accounting for acute mobility circumstances with a long-term systemic approach of designing an accessible, sustainable mobility system (oec 2021). Such a flexible, data-driven process can thus provide urban planners with a well-rounded, automated assistance for variable short-term scenario planning while maintaining the long-term goal of a sustainable, city-spanning micromobility network.

Acknowledgements

PF, LG, MT gratefully acknowledge the Lagrange Program of the ISI Foundation funded by CRT Foundation. MS gratefully acknowledges support by the Danish Ministry of Transport. We gratefully acknowledge the open data that this article is based on, from https://www.openstreetmap.org, copyright OpenStreetMap contributors.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Danish Ministry of Transport and by the CRT Foundation.

Data and code availability

Data and code to fully reproduce the results of the study are available at the repository: https://github.com/pietrofolco/Data-driven_bicycle_network_planning_for_demand_and_safety.

ORCID iDs

Pietro Folco https://orcid.org/0000-0002-9346-8045
Michele Tizzoni https://orcid.org/0000-0001-7246-2341
Michael Szell https://orcid.org/0000-0003-3022-2483

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Crashes in Turin (2019) available at: http://aperto.comune.torino.it/dataset/elenco-incidenti-nell-anno-2019-nella-citta-di-torino
2. Access instructions to the Bird API are available at: https://github.com/ubahnverleih/WoBike/blob/master/Bird.md
3. Number of residents in Turin (2020) by statistical areas available at: http://aperto.comune.torino.it/en/dataset/popolazione-per-sesso-e-zona-statistica-2020. Shapefiles of statistical areas of Turin available at: http://geoportale.comune.torino.it/geodati/zip/zonestat_popolazione_residente_geo.zip

References

Adminaité-Fodor D and Jost G (2020) How safe is walking and cycling in Europe? PIN Flash Report 38, European Transport Safety Council.

Alessandretti L, Aslak U and Lehmann S (2020) The scales of human mobility. *Nature* 587(7834): 402–407.

Barthelemy M (2022) *Spatial Networks: A Complete Introduction: From Graph Theory and Statistical Physics to Real-World Applications*. Springer Nature.

Boeing G (2017) OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65: 126–139.

Cardillo A, Scellato S, Latora V, et al. (2006) Structural properties of planar graphs of urban street patterns. *Physical Review E* 73(6): 1–7. DOI: 10.1103/PhysRevE.73.066107.

Chicco A and Diana M (2022) Understanding micro-mobility usage patterns: a preliminary comparison between dockless bike sharing and e-scooters in the city of turin (italy). *Transportation Research Procedia* 62: 459–466.

Città di Torino (2021) http://www.comune.torino.it/torinogiovani/vivere-a-torino/sharing-di-monopattini-elettrici-a-torino#normativa

Comi A, Polimeni A and Nuzzolo A (2022) An innovative methodology for micro-mobility network planning. *Transportation Research Procedia* 60: 20–27.

Cook S, Stevenson L, Aldred R, et al. (2022) More than walking and cycling: What is 'active travel'? *Transport Policy* 126: 151–161.

Creutzig F, Jochem P, Edelenbosch OY, et al. (2015) Transport: a roadblock to climate change mitigation? *Science* 350(6263): 911–912.

CROW (2016) *Design Manual for Bicycle Traffic*.

Daraei S, Pelechrinis K and Quercia D (2021) A data-driven approach for assessing biking safety in cities. *EPJ Data Science* 10(1): 11.

Fazio M, Giuffrida N, Le Pira M, et al. (2021) Planning suitable transport networks for e-scooters to foster micromobility spreading. *Sustainability* 13(20): 11422.

Gauvin L, Bajardi P, Pepe E, et al. (2021) Socio-economic determinants of mobility responses during the first wave of covid-19 in italy: from provinces to neighbourhoods. *Journal of The Royal Society Interface* 18(181): 20210092.

Gössling S, Choi A, Dekker K, et al. (2019) The social cost of automobility, cycling and walking in the European Union. *Ecological Economics* 158: 65–74.

Ignaccolo M, Inturri G, Cocuzza E, et al. (2022) Developing micromobility in urban areas: network planning criteria for e-scooters and electric micromobility devices. *Transportation Research Procedia* 60: 448–455.

ITF (2021) *Travel transitions: How transport planners and policy makers can respond to shiftingmobility trends*. OECD Publishing.

Klanjčić M, Gauvin L, Tizzoni M, et al. (2022) Identifying urban features for vulnerable road user safety in Europe. *EPJ Data Science* 11: 27, (in print).

Kondo MC, Morrison C, Guerra E, et al. (2018) Where do bike lanes work best? a bayesian spatial model of bicycle lanes and bicycle crashes. *Safety Science* 103: 225–233.

Lamb WF, Wiedmann T, Pongratz J, et al. (2021) A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018. *Environmental Research Letters* 16: 073005.

Larsen J, Patterson Z and El-Geneidy A (2013) Build it. but where? the use of geographic information systems in identifying locations for new cycling infrastructure. *International Journal of Sustainable Transportation* 7(4): 299–317.

Lovelace R, Goodman A, Aldred R, et al. (2017) The propensity to cycle tool: an open source online system for sustainable transport planning. *Journal of Transport and Land Use* 10(1): 505–528.

Lyons G and Davidson C (2016) Guidance for transport planning and policymaking in the face of an uncertain future. *Transportation Research Part A: Policy and Practice* 88: 104–116.

McKenzie G (2019) Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in washington, dc. *Journal of Transport Geography* 78: 19–28.

Natera Orozco LG, Battiston F, Iñiguez G, et al. (2020) Data-driven strategies for optimal bicycle network growth. *Royal Society Open Science* 7(12): 201130. DOI:10.1098/rsos.201130.

Nelson AC and Allen D (1997) If you build them, commuters will use them: association between bicycle facilities and bicycle commuting. *Transportation Research Record* 1578(1): 79–83.

Nieuwenhuijsen MJ (2020) Urban and transport planning pathways to carbon neutral, liveable and healthy cities; a review of the current evidence. *Environment International* 140: 105661.

OECD Publishing (2021) Transport strategies for net-zero systems by design.

Olmos LE, Tadeo MS, Vlachogiannis D, et al. (2020) A data science framework for planning the growth of bicycle infrastructures. *Transportation Research Part C: Emerging Technologies* 115: 102640.

Olszewski P, Szagala P, Rabczenko D, et al. (2019) Investigating safety of vulnerable road users in selected EU countries. *Journal of Safety Research* 68: 49–57.

Ospina JP, Duque JC, Botero-Fernández V, et al. (2022) The maximal covering bicycle network design problem. *Transportation Research Part A: Policy and Practice* 159: 222–236.

Quercia D, Schifanella R and Aiello LM (2014) The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In: Proceedings of the 25th ACM conference on Hypertext and social media., pp. 116–125.

Quercia D, Schifanella R, Aiello LM, et al. (2015) Smelly maps: the digital life of urban smellscapes. In: Ninth International AAAI Conference on Web and Social Media.

Reck DJ, Haitao H, Guidon S, et al. (2021) Explaining shared micromobility usage, competition and mode choice by modelling empirical data from zurich, switzerland. *Transportation Research Part C: Emerging Technologies* 124: 102947.

Reyneri A (2020) Breaking: Over 2000 km pro-cycling corona measures announced in europe. https://ecf.com/news-and-events/news/breaking-over-2000-km-pro-cycling-corona-measures-announced-europe

Sanders RL and Nelson TA (2022) Results from a campus population survey of near misses, crashes, and falls while e-scooting, walking, and bicycling. *Transportation Research Record*: 03611981221107010.

Schläpfer M, Dong L, O'Keeffe K, et al. (2021) The universal visitation law of human mobility. *Nature* 593(7860): 522–527.

Shah NR, Aryal S, Wen Y, et al. (2021) Comparison of motor vehicle-involved e-scooter and bicycle crashes using standardized crash typology. *Journal of Safety Research* 77: 217–228.

Szell M, Mimar S, Perlman T, et al. (2022) Growing urban bicycle networks. *Scientific Reports* 12: 6765. DOI:10.1038/s41598-022-10783-y.

Teschke K, Harris MA, Reynolds CC, et al. (2012) Route infrastructure and the risk of injuries to bicyclists: a case-crossover study. *American Journal of Public Health* 102(12): 2336–2343.

Vybornova A, Cunha T, Gühnemann A, et al. (2022) Automated detection of missing links in bicycle networks. *Geographical Analysis* 0: 1–29.

Zhao C, Carstensen TA, Nielsen TAS, et al. (2018) Bicycle-friendly infrastructure planning in Beijing and copenhagen-between adapting design solutions and learning local planning cultures. *Journal of Transport Geography* 68: 149–159.

Author Biographies

Pietro Folco is a junior data science researcher who have just completed a 1-year fellowship funded by the Lagrange Applied Research Program at the ISI Foundation. At the University of Turin, he obtained a Bachelor's Degree in Physics and a Master's Degree in Physics of Complex Systems with a master's thesis entitled "Deep Learning methods for predictions and integrated Feature Selection". In May 2021 he started working at the ISI Foundation where he carried out applied research in Data Science for Social Impact. His scientific interests are mainly in urban data science, machine learning and data visualization.Laetitia Gauvin is Senior Research Scientist in the Data Science for Social Impact and Sustainability team at ISI Foundation. She completed her Master's at Imperial College, London and at the Centre de Physique Théorique, Marseille. She then pursued a Ph.D. on the study of socio-economic systems using tools from statistical physics at the Laboratoire de Physique Statistique of the Ecole Normale Supérieure in Paris. As a postdoctoral researcher, she has been involved in a project on urban Collective Dynamics. Her research interests regards on the study of socio-economic systems, with a particular focus on Sustainable Development Goals.

Michele Tizzoni is a Senior Assistant Professor at the Department of Sociology and Social Research of the University of Trento. Prior to that, he held the position of Research Scientist at the ISI Foundation in Italy and a post-doctoral fellow at the BC Center for Disease Control in Canada. In 2012, he obtained his Ph.D. in Physics from the Politecnico of Turin. He is conducting research on

the use of digital traces and network science approaches to infectious disease epidemiology and public health. More recently, he has been working in the Data for Good domain, investigating how data mining and machine learning can support non-profit organizations' work in social innovation, international development, and humanitarian action.

Michael Szell is associate professor in Data Science at IT University of Copenhagen, and external researcher at ISI Foundation and at the Complexity Science Hub Vienna. His research quantifies the patterns behind interlinked human behavior and human-built structures through mining large-scale data sets. He follows an anti-disciplinary approach using methods from data science and network science. Michael's current focus is on sustainable urban mobility and urban data science. He has also developed several interactive data visualization platforms such as What the Street!?, and the award-winning massive multiplayer online game Pardus.