Route Selection - From Replication to Recreation

BARTOSZ MAZURKIEWICZ, TU Wien, Austria
IOANNIS GIANNOPOULOS, TU Wien, Austria

The choice of a route from an origin to a destination depends on several criteria. These criteria can range from route length to environment type. In several situations, we are not only interested in finding a route between two points, but to find a route between all possible origin-destination points in a specific geographic area. This is very common during experimental design, when one is seeking for a generalizable route to evaluate a navigation system. For this case, the selected route should be representative for the area, and not an exception with peculiarities. In this work we demonstrate (1) how to choose an *average* route for a bike navigation study in Vienna, Austria and (2) how to find similar routes in Florence, Italy and Bremen, Germany in order to replicate the study. The selection is based on route features and associated weights. They can be highly customized according to the needs. We demonstrate our approach and introduce four application scenarios to exemplify the benefits of a systematic route selection.

CCS Concepts: • **Information systems** → *Geographic information systems*; *Location based services*; *Decision support systems*; • **General and reference** → *Empirical studies.*

Additional Key Words and Phrases: Route Selection, Similar Routes, Human Wayfinding, Navigation, Experiments, Replicability

ACM Reference Format:
Bartosz Mazurkiewicz and Ioannis Giannopoulos. 2021. Route Selection - From Replication to Recreation. In *Cycling@CHI: Towards a Research Agenda for HCI in the Bike Lane at CHI '21, May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 5 pages.

1 INTRODUCTION

Navigation systems have found their way in our daily lives and are used in order to navigate us while walking, driving by car or even by bicycle. There are many possible routes between an origin and a destination and these can have different properties, such as total length and number of intersections, amongst other. Navigation systems, utilize these properties in order to compute a desired result, e.g., shortest or cognitively easiest route. Navigation systems are evaluated in order to prevent undesired effects, such as user frustration, but also for the investigation of novel systems. Researchers strive to design their experiments as valid and replicable as possible. The route selection is one of the most important steps towards these experimental goals. Unfortunately, this selection is mostly performed by the rule of thumb, trying to mix different types of street segments and intersections.

Mazurkiewicz et al. [4] introduced a framework that is able to rank routes based on given criteria, such as length, number of intersections, slope and number of turns, amongst other. This ranking is based on the average route of the considered area. For instance, the highest ranked route will be the most representative/average route in the given environment fulfilling the given weighted criteria. This is the first systematic approach towards route selection for experimental design. The route selection is rather a universal problem and relevant for pedestrians, cyclists [5] and cars. There have been some approaches trying to tackle this problem. Spretke et al. [6] derive representative driving

---

Manuscript submitted to ACM

routes for cars based on car fleet trajectory data. Unfortunately, the data is not always available, either in the necessary amount or at all.

Through such a systematic approach [4], a novel bike navigation system can be evaluated on routes resembling other routes in the given area without introducing peculiarities (e.g. 8-way intersection in an area where 3- and 4-way intersections are prevalent). Finding and selecting routes in one environment that resemble routes in another distant environment (e.g., finding the most similar route) can yield multiple and interesting benefits. In this work we adopt the framework for cycling routes and introduce and address four relevant application scenarios.

2 APPLICATION SCENARIOS FOR CYCLISTS

The route selection framework [4] can serve multiple purposes. Four application scenarios become eminent:

(1) **Ecological validity.** When designing experiments for the evaluation of a novel bike navigation system, one of the most important experimental decisions concerns the proper selection of test routes. The selected routes should represent the relevant environment in order to allow to generalize the findings as much as possible.
(2) **Replication.** Replication of research in different geographic regions is crucial for multiple reasons, e.g., for validation purposes, but also for measuring effects other than the ones resulting from the environmental conditions, e.g., cultural effects.
(3) **Comparison.** Being able to run experiments on similar routes might allow to compare different navigation systems. This can be achieved even though the experiments have been performed at different geographic areas, without having to replicate the experiment of the other system entirely in order to get the results for comparison.
(4) **Recreation.** Apart from the introduced scientific and experimental purposes, this adapted framework can also be utilized for recreational, entertainment and sports purposes. For instance, a cyclist can select a route for her training fulfilling personal criteria, find comparable routes to challenge a peer cyclist from an other city or country, or even prepare for a race competition by training in her local surroundings.

3 ROUTE SELECTION FOR EXPERIMENTAL AND RECREATIONAL PURPOSES

To address the application scenarios introduced in section 2 we consider Vienna, Austria as our baseline. Data from Florence, Italy and Bremen, Germany will be used in order to exemplify the scenarios in different geographic areas. The route features have to be defined, then the data have to be acquired and processed, and finally, the routes have to be selected by the framework [4].

3.1 Route Features

While designing a navigation experiment or preparing for a bike tour, the route selection is crucial. Each route is characterized by a large number of features, and a subset of them can be selected to extract a relevant route. In order to address the application scenarios and exemplify our approach, the following feature categories were selected:

(1) ***Number of decision points***, i.e., intersections. Start and end points are not considered as decision points. This is a hard criterion which is set before the actual analysis starts. Hence, it has no associated weight.
(2) ***Cardinality of decision points***, i.e., the average number of options a decision point has and the number of n-way intersections on the route, see Table 1.
(3) ***Frequency of turn types***, i.e., the number of left and right turns, as well as the number of non-turns.

(4) ***Regularity of decision points***, i.e., the sum of angles the options at a decision point need to be rotated in order to create a regular intersection [2]. This way the type of street network (e.g., gridded) can be approximated.
(5) ***Bearing of the route***, i.e., the orientation of street segments with respect to true north.
(6) ***Length-related features***, i.e., the total route length as well as the mean, standard deviation and median for the segment lengths of the route.

The above features were considered in order to find an average route in Vienna. The number of decision points was set to 12 (excluding start and end point) in order to avoid trivial route length, in terms of decision points. The route length was limited to be between 2 and 3 km, which provides rides of around 10 minutes [3]. The features were considered equally important for our experiment and were therefore equally weighted (for more details see [4]). It is important to stress that these feature categories and weights can be extended (e.g., by slope or route safety) or changed according to needs and data availability.

3.2 Data Acquisition and Preparation

The directed bike networks were downloaded via the OSMnx python package [1]. This data was the basis for four of the feature categories, except for *Number of decision points* which was set a-priori to 12 and the *Regularity of decision points* which was calculated according to [2]. In order to compute all possible routes with a given number of decision points Sagemath 9.1 with its SubgraphSearch function[1] was used. All routes containing at least one 2-way node were excluded.

3.3 Baseline - Average Route in Vienna

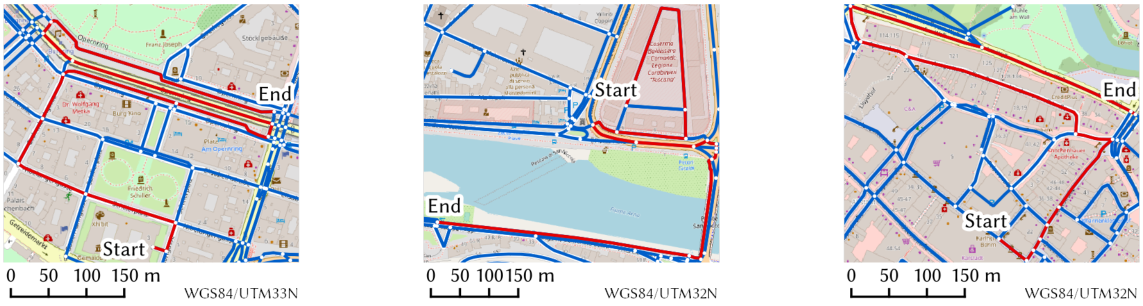
The bike network of Vienna (*"Wien, innere Stadt"*) consists of 938 nodes and 2 091 edges. There are 2 617 610 possible routes of 12 decision points in this area, from which 18 146 have a length between 2 and 3 km. According to Mazurkiewicz et al. [4] the average route (ranked highest according to weighted euclidean distance based on z-scores), will be considered as representative route. First, the *best possible route* is computed, which is likely a non-existent and hypothetical route. The following steps are necessary: (1) all absolute values for each route are transformed into positive z-scores (whether these values are over- or undershooting the mean is irrelevant); (2) the *best possible route* is created, which is a hypothetical route containing the lowest z-score for each subcategory. In a perfect world this hypothetical route would have a z-score of 0 for each subcategory.

Having the *best possible route*, the *weighted euclidean distance* between this route and any possible route is calculated. All feature categories are equally weighted with 0.2 (the category *Number of decision points* is not weighted since it is a hard criterion). This value gets split equally over all subcategories of a category, example: The category *Frequency of turn types* has a weight of 0.2. Therefore, all three subcategories, number of right, left and non-turns get a weight equal to $\frac{0.2}{3} = 0.06$. The route with the smallest weighted euclidean distance is the most average one (see Figure 1a). It has to be noted, that the results of this framework should be considered as suggestions, serving as a recommendation that requires human inspection, since we might not have all the relevant data in order to avoid unusual cases (see Section 4).

Having computed the average route in the city of Vienna, the four application scenarios can be addressed. For all four cases, the most similar route in Bremen and Florence have to be computed. The only thing that has to be adapted in our framework is *the best possible route*. We will do the same computations for the other two cities as we did for Vienna, but this time, the average route computed for Vienna will serve as *the best possible route* instead of the hypothetical route with the minimum z-score for each subcategory.

---

[1] http://sage-doc.sis.uta.fi/reference/graphs/sage/graphs/generic_graph_pyx.html#sage.graphs.generic_graph_pyx.SubgraphSearch, last access 04.02, 2021



(a) (b) (c)

Fig. 1. Routes generated by the framework (in red). White dots indicate intersections in the bike network (in blue). Base layer : OpenStreetMap. (a) The average route in Vienna, scale 1:10 000 (b) The most similar route in Florence, scale 1:13 000 (c) The most similar route in Bremen, scale 1:10 000.

Table 1. Characteristics of the average route in Vienna and the most similar routes in Florence and Bremen. Abbreviations used: Seg. - segment, M. - mean, St. dev. - standard deviation, Med. - median, W. - weighted

City	Avg. Options	# Intersec.			Regularity			# Turns			Length-related Features				W. M. Bearing
		3	4	5	3	4	5	r	l	s	Total Length	M. Seg. Length	St. dev. Seg. Length	Med. Seg. Length	
Vienna	3.67	5	6	1	55.77	15.69	83.17	4	4	4	2259.58	173.81	169.98	88.51	174.86
Florence	3.58	7	3	2	60.11	109.39	78.10	2	4	6	2203.30	169.48	185.06	92.17	174.13
Bremen	3.42	7	5	0	68.18	37.30	NA	2	4	6	2178.39	167.59	211.47	74.93	195.33

3.4 Most Similar Routes for the Application Scenarios

The corresponding bike networks for Florence (*"Firenze, centro storico"*) and Bremen (*"Bremen, Altstadt"*) consist of 1 960 and 442 nodes and 3 901 and 1 018 edges, respectively. In Florence there are 70 726 routes of 12 decision points and a length between 2 and 3 km, whereas Bremen has 55 548 of those. The street networks differ concerning the intersection types, e.g., in Florence there are 7-way intersections which are not present in Vienna (see Section 4). If subcategories in either Florence or Bremen were not identical with those ones in Vienna we excluded routes where one of those subcategories was not 0, e.g., all routes with at least one 7-way intersection from Florence were excluded (the subcategories for 6-way intersections are missing in Table 1 because the most average route in Vienna had zero 6-way intersections and the most similar routes in Florence and Bremen too). Again, the absolute values were transformed into z-scores for each city respectively. Now, the *best possible route* is the average route in Vienna, which was calculated previously. The absolute values of this route were converted into z-scores based on all other routes in Florence and Bremen, respectively. Next, the *weighted euclidean distance* was calculated for each route. The route with the smallest distance is the most similar one to the average route in Vienna, given these subcategories and weights (see Figures 1b and 1c). Several similarities can be recognized. All routes have a part which goes back and forth. Furthermore, all routes have one segment oriented towards south/south-west which comes after the back and forth part of the route.

4 DISCUSSION AND CONCLUSION

The presented application scenarios in Section 2 were addressed in the following way. The selected route in Vienna, can serve as a representative route for performing a cycling experiment that can generalize to a certain degree, thus addressing the fist application scenario. The computed routes in Bremen and Florence are the most similar to the one selected in Vienna, thus allowing to replicate the experiment, addressing the second application scenario. Similarly, the same routes could be used for the evaluation of a novel navigation system, allowing to compare the results against the results obtained by evaluating the navigation system in Vienna, addressing the third application scenario. Finally, in the same manner, setting the *best route* to fulfill certain criteria might allow to optimize training sessions as well as be used to satisfy recreational purposes, e.g., by weighting land cover parameters higher in order to compute a scenic route.

In this work we demonstrate (1) a systematic approach for choosing a route for a cycling experiment and (2) how to find similar routes in other geographical areas. The framework should be utilized as a recommendation system requiring human rating before choosing a route in order to avoid selecting unrealistic routes due to data limitations. Therefore, it is important to validate several suggestions given by our framework in order to choose an appropriate one. On the other hand, if the resulting routes are generally unusual, the categories and/or weights should be adapted accordingly.

Another aspect which should be discussed is how to handle differences in city properties. We excluded those routes which had other properties that the average route in Vienna. There are several ways to handle this case. One possible way of handling this issue, assuming that the missing property (in the average route) is important for the user, is adding the mean value from the population of all routes to the target route. Another possibility would be to adjust the weights, under the assumption that 7-way intersections are more similar to 6-way intersections than 3-way intersections considering special aspects and therefore preferred. In our work, we presented only routes with the smallest weighted euclidean distance. It is still uncertain how big a difference must be in order to yield different study results. Therefore, it would be interesting to use the route as an independent variable in future experiments in order to see if indeed different results originate from choosing different routes in terms of weighted euclidean distance. The presented framework is flexible and new categories can be added. Furthermore, the weights allow to favor routes with specific characteristics. Our approach depends greatly on data availability and processing. In order to lower the entry barrier an API[2] will be provided which will help to select and compare routes from different geographic areas for pre-computed attributes.

REFERENCES

[1] Geoff Boeing. 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65 (2017), 126 – 139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004

[2] Paolo Fogliaroni, Dominik Bucher, Nikola Jankovic, and Ioannis Giannopoulos. 2018-08. Intersections of Our World. In *10th International Conference on Geographic Information Science*, Winter et. al (Ed.), Vol. 114. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 3. https://doi.org/10.3929/ethz-b-000283277 10th International Conference on Geographic Information Science (GIScience 2018).

[3] Pablo Jensen, Jean-Baptiste Rouquier, Nicolas Ovtracht, and Céline Robardet. 2010. Characterizing the speed and paths of shared bicycle use in Lyon. *Transportation Research Part D: Transport and Environment* 15, 8 (2010), 522 – 524. https://doi.org/10.1016/j.trd.2010.07.002

[4] Bartosz Mazurkiewicz, Markus Kattenbeck, Peter Kiefer, and Ioannis Giannopoulos. 2020. Not Arbitrary, Systematic! Average-Based Route Selection for Navigation Experiments. In *11th International Conference on Geographic Information Science - Part I (GIScience 2021, Vol. 177)*, Krzysztof Janowicz et. al (Ed.). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 8:1–8:16. https://doi.org/10.4230/LIPIcs.GIScience.2021.I.8

[5] Gian-Luca Savino, Laura Meyer, Eve Emily Sophie Schade, Thora Tenbrink, and Johannes Schöning. 2020. Point Me In the Right Direction: Understanding User Behaviour with As-The-Crow-Flies Navigation. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–11.

[6] D. Spretke, M. Stein, L. Sharalieva, A. Warta, V. Licht, T. Schreck, and D. A. Keim. 2015. Visual Analysis of Car Fleet Trajectories to Find Representative Routes for Automotive Research. In *2015 19th International Conference on Information Visualisation*. 322–329. https://doi.org/10.1109/iV.2015.63

---

[2] See https://geoinfo.geo.tuwien.ac.at/index.php/resources/ for updates