The impact of urban road network morphology on pedestrian wayfinding behavior

Debjit Bhowmick[1], Stephan Winter[1], Mark Stevenson[1,2], and Peter Vortisch[3]

[1]Department of Infrastructure Engineering, The University of Melbourne, Australia
[2]Transport, Health and Urban Design Research Lab, The University of Melbourne, Australia
[3]Institute for Transport Studies, Karlsruhe Institute of Technology (KIT), Germany

Abstract: Pedestrians do not always choose the shortest available route during the process of wayfinding. Instead, their route choices are influenced by strategies, also known as wayfinding heuristics. These heuristics aim to minimize cognitive effort of the pedestrian and their application usually leads to satisfactory route choices. Our previous study evaluated and analyzed resultant routes from the application of four well-known pedestrian wayfinding heuristics across nine distinct network morphologies via simulation. It was observed that the variation in the cost (difference in route length between a heuristic route and the shortest route, expressed as a percentage of the shortest route length) across the four heuristics increased with an increase in the irregularity of the network. Based on these results, we claimed that, people may opt for more diverse heuristics while walking through relatively regular networks, as route cost across heuristics are more similar in magnitude and thus applying any one of them would not result in a substantial difference in the travelled distance. Likewise, they may prefer specific heuristics in the relatively irregular networks, as some heuristics are significantly costlier than others, thus creating greater variation in cost across heuristics and hence would result in significantly greater travelled distances. In this study, we investigated this claim by comparing simulated routes with observed pedestrian trajectories in Beijing and Melbourne, two cities at opposite ends of the regularity spectrum, as established in the literature. Using statistical tests, we claim with confidence that on an average, heuristic choice distribution is uniform in Melbourne, a city having a regular network morphology. On the contrary, heuristic choice distribution was skewed in Beijing, a city with more irregularities in its street network morphology.

This novel finding could help urban planners and future researchers in producing more accurate patterns of aggregate pedestrian movement in outdoor urban spaces.

---

1   Introduction

Human wayfinding in outdoor spaces involves the process of selecting segments of an existing real-world network to find a viable route between an origin and a destination [13]. During wayfinding, pedestrians do not always choose the shortest possible route [8] as they may not be able to discern it, especially when the shortest routes are complex. Hence, they apply certain strategies or wayfinding heuristics that attempt to minimize their cognitive effort [3]. For example, pedestrians may seek routes with fewer turns—routes that are simpler in nature, hence require less cognitive effort or are shorter to communicate and memorize—even if this route is not geometrically the shortest one. This strategy of wayfinding in a street network and reaching the destination with the fewest number of turns is one wayfinding heuristic. Like this 'Fewest turns strategy', there exists multiple well-established wayfinding heuristics that are known to be applied by pedestrians. These wayfinding heuristics are applied by pedestrians irrespective of their level of spatial aptitude or familiarity with a given road network.

Empirical studies have revealed that pedestrians switch between wayfinding strategies with a change in the ambient environment. Through his experiments, Golledge [13] inferred that "perceptions of the configuration of the environment itself (particularly different perspectives as one changes direction) may influence route choice." This gives us the impression that may be human beings are able to understand that, given a type of network morphology, certain heuristics are better at optimizing not just cognitive effort, but physical effort (in terms of distance travelled) as well. We say that certain heuristics are (on average) *less costly* than others in certain types of road networks, taking into account the difference in route length between the heuristic route and the shortest possible route.

In this regard, our previous work [5] showed through simulation that although some heuristics are consistently cheaper and some are consistently costlier across nine different types of network morphologies, the variation in cost across these wayfinding heuristics is dependent on the regularity of the network structure, as inferred from visual assessment. It was observed that more regular networks had lesser variation in cost across heuristics while more irregular networks experienced more variation. For example, in Melbourne, the observed standard deviation in route cost was 6.96% while the corresponding statistic in Beijing was 9.39% (these numbers although not present in [5] are derived from the same analysis). Regularity of network morphologies was based on the analysis and findings by [33]. The results supported the argument that pedestrians possibly opt for a variety of heuristics in regular networks while opting for specific heuristics (or avoiding them) in irregular ones. While we arrived at this conclusion by thoroughly simulating four wayfinding heuristics in nine network morphologies following a systematic methodology, the simulation approach had to use some assumptions. While analysis of the simulated routes across different network structures helped us formulate this hypothesis, yet we could not claim with confidence that this is representative of actual pedestrian behavior. For ground-

ing of results from our previous study, we had to analyze actual pedestrian routes and then draw comparisons with our previous results.

Shortest paths have been the foundation of a host of well-established studies involving computational movement analysis of pedestrians. But pedestrians do not always follow (or even know) the shortest available paths. In the literature, we have found studies which reveal that pedestrians move in outdoor spaces by relying on heuristics and not shortest paths (these are discussed in detail in Section 2.1). Studies have also proven that most people use such strategies by observing commonalities in their movement patterns [8]. Thus, to accurately predict actual pedestrian movement, shortest path algorithms are not appropriate as the computed routes are not representative of actual pedestrian routes for most cases. Studies involving wayfinding heuristics are aimed at understanding actual movement patterns of people (for example, how and where people actually move in a city or in a neighborhood). Disregarding shortest path strategies, these studies also form the basis of producing aggregate patterns of pedestrian movement in urban areas [9] and can serve as a useful tool for urban planners and facility planners, especially when designing more walkable neighborhoods. But there exists a host of heuristics (and not just one) which pedestrians make use of during outdoor wayfinding. We also know that the choice of heuristic is dependent on the environment [13]. Thus, to reproduce actual movement patterns, there needs to be an understanding on whether heuristic choice distribution changes with a change in the environment configuration. We have chosen four well-known heuristics, all geometric in nature, and thus dependent on the pedestrian network configuration (morphology). Hence, this study aims at investigating actual pedestrian movement to ground the simulation finding, i.e. to check whether pedestrian heuristic choice is dependent on network morphology. Thus, we chose two cities with contrasting pedestrian network morphologies, and analyzed pedestrian trajectory data for these cities, to understand whether network morphology has an influence on pedestrian wayfinding behavior, or in other words, the choice of heuristics.

Cost, in this context, means route cost which is the difference in route length between the shortest route and the heuristic route, expressed as a percentage of the shortest route. The use of this term is consistent with our terminology from our previous work. Route cost is not the resultant of any detour. It is a measure representing the heuristic routes relative to the shortest available routes. Heuristic routes are an attempt to simulate usual routes of pedestrians, albeit approximately. Hence, while walking along their usual route, the pedestrian does not intend to take a detour from the shortest route because they may not be aware of the shortest route in the first place. The term 'detour' has been reserved for cases where pedestrians deviate from their usual route for other reasons, such as personal safety concerns or avoid a busy route during rush hour. We do not possess knowledge of the presence of detours in our datasets. Additionally, enriching our dataset with semantic information on detours is not within the scope of this paper. Hence, we believe that 'detour cost' is not an appropriate term.

We hypothesized that in Melbourne, a city with a relatively regular pedestrian network morphology, the choice of heuristics will be uniformly distributed, while in Beijing, a city with a relatively irregular network, this distribution will be skewed. This paper attempts to answer the following research question: *Does distribution of heuristic choice vary across network morphology?*

Our study makes use of GPS datasets from Melbourne and Beijing, two cities at opposite ends of the regularity spectrum, as described in [33]. As the records in these datasets

were labelled with their corresponding transportation mode, we filtered walking points, employed trip segmentation thresholds to differentiate between individual trips, and performed map matching (matching raw GPS trajectories to appropriate segments of the underlying pedestrian network) to obtain the actual traversed routes. For the same origin-destination pairs, we obtained the theoretical heuristic routes using simulation of four wayfinding heuristics. Consequently, we used Network Hausdorff Distance (NHD) to derive (dis)similarity between actual and simulated routes to infer heuristics chosen by pedestrians, either partially or fully.

The paper is organized as follows. Section 2 contains a review of the existing literature along with the heuristic algorithms proposed in our previous study while Section 3 talks about the datasets used in this study. Section 4 contains the detailed methodology followed in this study. Section 5 presents some preliminary findings and Section 6 discusses the findings and presents relevant arguments in relation to the same.

2   Related work

2.1   Wayfinding heuristics

Several studies have explored human wayfinding strategies in outdoor spaces. A review of existing wayfinding literature reveals the existence of multiple heuristics that are applied by pedestrians. These heuristics have been theorised based on observations of actual and probable pedestrian behavior in relatively small environments [3, 8–10, 13, 19]. Comparison between wayfinding heuristics has been done on a small scale by [20]. In contrast, in our previous work heuristic routes were simulated in a relatively larger, city-wide scale to investigate the impact of network morphology on pedestrian wayfinding decisions [5]. These simulated routes represented theoretical routes chosen by pedestrians applying a single heuristic consistently during their wayfinding exercise. The heuristics chosen were *modified least angle strategy*, *longest-leg first strategy*, *shortest-leg first strategy* and *fewest turns strategy*.

Although there exists a host of other wayfinding heuristics, only the aforementioned ones are geometric in nature and thus dependent on network morphology. In these heuristics, the location of taking a turn or the number of turns taken during wayfinding determine the route choice. Human perceptions and conceptualizations vary, so what accounts to form a turn is vague from a cognitive perspective. But also representations of walkable features in databases vary in their level of abstraction and detail, challenging additionally to define what constitutes a turn. Accordingly, our previous study [5] defined a 'turn' as follows: "If two consecutive road segments in a route have a deflection angle (difference in bearing) of 45° or more, the move from one to the other is considered a turn." This definition was applied to appropriate levels of geometric abstraction. It led to satisfactory outputs according to visualizations of randomly sampled routes. But any research is sensitive to the chosen threshold value. The four chosen heuristics and the implemented algorithms are discussed briefly as follows.

Modified Least angle strategy: [19] proposed a real-world wayfinding heuristic called 'least angle strategy' which can be applied in an unknown environment if the destination can be perceived directly by the navigator, at least at the beginning of the navigation process. At each decision point, the pedestrian prefers the road segment which has the least deviation from the direction of the intended destination. However, the original least angle

strategy [19] has a significant shortcoming. The algorithm resulted in significantly longer routes more often by taking impractical detours in real street networks, meaning that these routes would not be chosen by a pedestrian during wayfinding. For example, in cases where the algorithm chooses a road segment over others based on least angle, and then the consequent roads led to detours, the results were not representative. In this paper we modified the least angle strategy as shown in Algorithm 1 to avoid similar shortcomings. It preserves the principle philosophy without running into large outliers, making it more competitive. In other words, this modified version resulted in more realistic routes, more often. It makes use of the A-star algorithm where the difference between two bearings, one between the origin and the destination, and the other between any node and the destination, has been selected as the heuristic. This is termed as deflection angle. A large positive number has been multiplied with deflection angle so that route selection by A-star algorithm depends, almost entirely, on selecting nodes that minimize the deflection angle and not the length of the edges of the road network. While this algorithm is not fully robust, it results in appropriate routes similar to what the original least angle heuristic should have resulted in under practical circumstances. Hence, we decided to implement this algorithm for our study and refer to this as the Modified Least Angle strategy in this paper henceforth.

Algorithm 1 Modified Least Angle strategy algorithm after [5]

Require: An undirected graph $G = (N, E)$, where N is the set of nodes and E is the set of edges in the network
　　with $edge\_weight \leftarrow edge\_length$
　　$origin, destination \in N$
1: **Define** $heuristic$**:**
　　　$target\_angle \leftarrow$ bearing($origin$,$destination$)
　　　$node\_angle \leftarrow$ bearing($node$,$destination$)
　　　$deflection\_angle \leftarrow$ absolute_value($target\_angle$ - $node\_angle$)
　　　return 100000 * $deflection\_angle$ (so that edge_length has minimum influence on chosen route)
2: Compute heuristic for all $node \in N$
3: $route \leftarrow$ A-Star_shortest_route ($origin$,$destination$, $heuristic$)
4: **Return** $route$

Longest Leg First strategy: The longest leg first strategy involves basing decisions disproportionately on the straightness of the initial segments of the routes [3]. The pedestrian chooses to prefer longer and straighter initial segments to reach as close as possible to their destination, without taking a 'turn' and thereby reducing the cognitive effort spent during wayfinding. This heuristic is also popularly known as the 'initial segment strategy'. The algorithm has been provided in Algorithm 2.

Algorithm 2 Longest Leg First strategy algorithm after [5]

Require: An undirected graph $G = (N, E)$, where N is the set of nodes and E is the set of edges in the network
　　with $edge\_weight \leftarrow edge\_length$
　　$origin, destination \in N$
　　Nomenclature: NT_nodes = nodes which can be traversed from origin without taking a turn
1: Search for all NT_nodes in the graph using Breadth-First Search
2: Derive shortest path from destination to all $node \in NT\_nodes$ using dijkstra_path($destination$,$node$)
3: $route\_node \leftarrow node$ which satisfies min(dijkstra_path_length($destination$,$node$))
4: $final\_segment \leftarrow$ dijkstra_path($route\_node$,$destination$)
5: $initial\_segment \leftarrow$ traversed_path($origin$,$route\_node$)
6: $route \leftarrow$ append($initial\_segment$,$final\_segment$)
7: **Return** $route$

Shortest Leg First strategy: Although [13] and [9] have mentioned the shortest leg first strategy as one of the least preferred wayfinding heuristics by pedestrians, there was no formal definition found in the literature. Hence, for this study, we have assumed that this strategy involves taking turns in the initial portion of the route to keep the latter portions as straight as possible. [20] stated that shorter initial legs provide pedestrians with the choice to explore further alternatives quickly at the next decision point, to reduce the cost of potentially required backtracking when compared to long initial segments. Based on our understanding, we have obtained the shortest leg first route for an OD pair by swapping the positions of origin and destination in Algorithm 2.

Fewest Turns strategy: [13] observed that the fewest turns strategy is the most popular wayfinding strategy and ranked it just after shortest distance and least time criteria. [46] developed modified wayfinding algorithms based on this heuristic. Pedestrians tend to choose routes involving the fewest number of turns that result in so called simpler routes, since turns involve decision making and increased cognitive effort. Our algorithm involves reaching a set of nodes from the origin that do not require taking a turn, and then selecting from that set, the node closest to the destination, and repeating the entire process at every turn until the destination is reached.

A visual illustration of typical heuristic routes for a fixed origin-destination pair on an urban pedestrian network has been shown Figure 1. The example routes were simulated on the pedestrian network of New Orleans, a city that was included in our previous study. As the city has a grid-like network, the contrast between the heuristic routes are apparent as the heuristics tend to show their typical route choice outcomes.

Algorithm 3 Fewest Turns strategy algorithm after [5]

Require: An undirected graph $G = (N, E)$, where N is the set of nodes and E is the set of edges in the network
    with $edge\_weight \leftarrow edge\_length$
    $origin, destination \in N$
    Nomenclature: NT_nodes = nodes which can be traversed from origin without taking a turn
1: $temp\_route\_node \leftarrow origin$
2: **while** $temp\_route\_node \neq destination$ **do**
3:     Search for all NT_nodes in the graph using Breadth-First Search
4:     Calculate shortest path from all NT_nodes to the destination using Dijkstra's shortest path algorithm
5:     $route\_node \leftarrow node \in NT\_nodes$ which satisfies min(dijsktra_path_length($destination,node$))
6:     $temp\_route\_segment \leftarrow$ traversed_path($temp\_route\_node,route\_node$)
7:     $route \leftarrow$ append($route,temp\_route\_segment$)
8:     $temp\_route\_node \leftarrow route\_node$
9: **end while**
10: **Return** $route$

2.2   Regularity of urban road network morphology

Thompson et al. [33] used convolutional neural network (CNN) to study precinct-level images of maps of 1667 cities around the world. The images (1,000 images for each city, making a total 1.667 million images) provided a high-level abstraction of the urban characteristics of interest, primarily road networks and rail transit networks. Through this visual classification technique, this study was able to capture the diversity of urban design and morphology in relation to land transport on a global scale. Nine distinct city types were identified based on the shape and extent of road and rail infrastructure networks. Melbourne, a city that evolved post-motorization, was classified as a 'Motor' city characterized
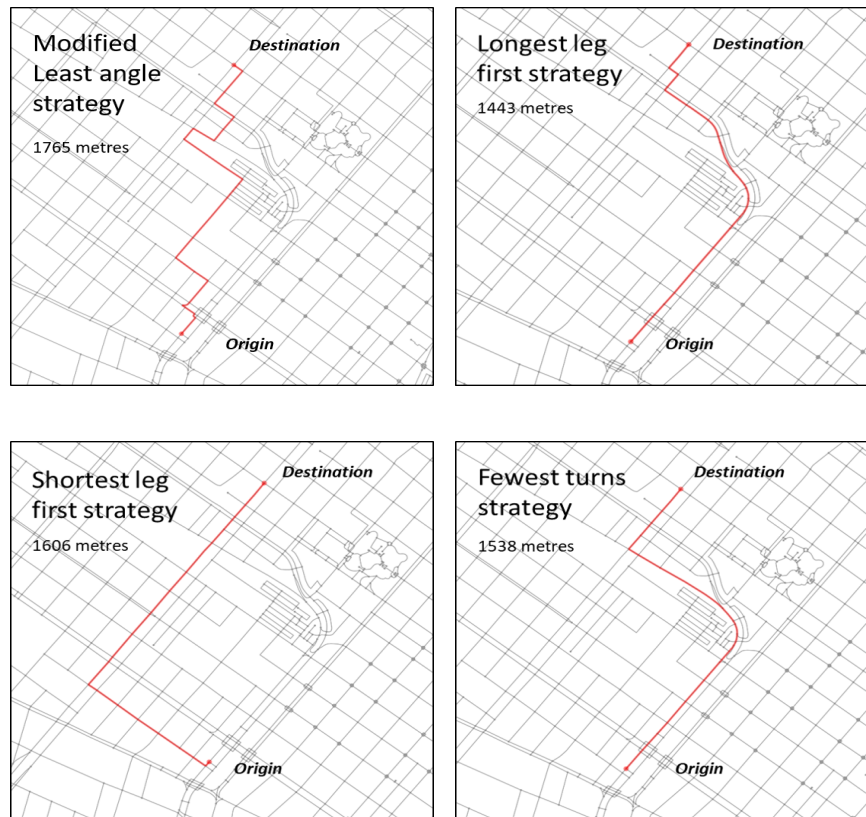
Figure 1: Comparison of typical heuristic routes.

by highly organized, medium to low density, grid-based road networks. On the other hand, Beijing was classified as 'Irregular' based on the more irregular morphology of their road and rail network that has been influenced by historic planning regimes. Hence we selected the two cities, Melbourne and Beijing, for this study as their road network morphology has been established to be contrasting [33].

2.3   Map matching

Map matching is referred to the process of matching observed GPS points (latitude, longitude, timestamp) to a sequence of existing road segments. Raw GPS traces are often inaccurate with the accuracy varying from a few metres to sometimes 1–2 kilometers. These inaccuracies are due to a range of reasons, including atmospheric influences on GPS signals and the presence of urban canyons and other terrestrial features that are likely to affect GPS signals [34]. Due to the level of noise in the GPS signals simple map matching of the observed points to their nearest street segment may result in inaccurate results. Hence, geometrical and topological constraints of the road network are necessary to build a path with an acceptable level of probability that it was traversed.

Multiple solutions of the map matching problem under various ground conditions have been suggested [7,16,24,38]. Newson and Krumm [26] proposed a map matching algorithm based on the principles of hidden-Markov models (HMM). They stated that the HMM was found to be successful in accounting for measurement noise and road network layout. To overcome some limitations of the aforementioned approach, Meert and Verbeke [25] proposed a new map matching approach by implementing HMMs with non-emitting states. In this study, we have made use of their algorithm in the form of Python codes publicly shared in GitHub (https://github.com/wannesm/LeuvenMapMatching).

2.4 Route similarity

One important aspect of trajectory data analysis is the similarity measurement of trajectories. Trajectories are composed of "a sequence of time-stamped locations" [17]. Past studies have made use of Euclidian space and calculated trajectory similarity based on Euclidian distance [36,39,45]. But Euclidian distance is not an appropriate measurement tool in road network space where topological constraints exist. Hence, more recent studies have used network distance instead of Euclidian distance for measuring the similarity between a pair of trajectories [11, 18, 22]. Furthermore, there exists noise in GPS data which results in the points not coinciding with the underlying road network for which map matching was done, as mentioned in Section 2.3. Hence, to compare network-based trajectories which have been mapped to the underlying road network (to form a sequence of nodes traversed), it is essential to use appropriate similarity metrics based on network constraints and not the ones based on Euclidian space. Thus, we employ Hausdorff distance, a commonly used similarity measure used in computational geometry [21] with recent advances using it for inferring trajectory similarity [11]. In our study, we use the definition of network Hausdorff distance (NHD) between two trajectories, a version of the original Hausdorff distance modified for applications on networks, as described in [11]. Calculation of NHD has been based on Equation 1:

$$NHD\left(t_i, t_j\right) = \max_{n \in t_i} \min_{m \in t_j} dist\left(n_{t_i}, m_{t_j}\right) \tag{1}$$

where t_i and t_j are two trajectories, n and m are nodes belonging to t_i and t_j respectively, and $dist$ indicates Dijkstra's shortest-path distance between points n and m. Thus, to compute NHD between t_i and t_j, one needs to

- compute Dijkstra's shortest path with edge length as weights between a node in t_i and all the nodes in t_j,
- choose the minimum value among all the computed shortest route lengths,
- repeat the process for all other nodes of t_i, and
- retrieve minimum values for all other nodes of t_i.
- The maximum value from the set of obtained minimum values gives the NHD.

As has been shown in [11], NHD between t_i and t_j and t_j and t_i may not be the same, meaning NHD could result in assymetric distances depending on network configuration. Hence, during computation, NHD has been calculated between the actual (map-matched) route and the simulated heuristic route and not the other way around, for the sake of consistency. Also, the relationship between NHD and lengths of two routes is not trivial, in the sense that they may not be directly proportional.

NHD (in meters) is a measure of how similar (or dissimilar) two routes in a road network are. The greater the magnitude of NHD, the more is the dissimilarity. For example, if NHD between the actual route and theoretical route followed by heuristic A is 50 meters and that with heuristic B is 90 meters, it indicates that the similarity between the actual and heuristic A route is more than that with heuristic B route. A positive NHD value shows that there exists some difference between two routes and the similarity is approximate. A zero NHD value indicates that the two routes are one and the same, only in cases where the start and end point of two routes are the same (as is in this study). Thus, from the above example, we infer that the actual route follows heuristic A approximately more closely than heuristic B.

2.5   OpenStreetMap data quality

The assessment of the data quality of OpenStreetMap (OSM) has caught the attention of researchers over the recent years, given its massive increase in patronage. OpenStreetMap is volunteered geographic information (VGI) wherein volunteers acquire spatial information and upload it for public use. Past OSM data quality analyses against conventional geographic information sources have revealed that the completeness of data varies with land use (urban vs rural), country (developed vs developing) and road type (motorways vs pedestrian ways) [42] as OSM is dependent on the contribution of data from volunteers in a given area. Hence, concerns about the credibility of research using OSM data must be carefully addressed. A study conducted in all the states in the US revealed that the coverage of pedestrian network data in OSM was higher than the US Census TIGER/Line data contrary to motorways [48]. Furthermore, Zielstra and Hochmair [47] in 2012 compared OSM with different proprietary geo-datasets in the US and Germany and concluded that the OSM database was relatively complete and can be used effectively for pedestrian routing. To further strengthen the argument in favor of OSM's pedestrian data completeness, Novack et al. [27] relied entirely on OSM data for proposing a system that generates pleasant pedestrian routes, and Gil [12] proposed a multimodal urban network model using OSM network data including pedestrian ways. Australia is among the top countries in terms of the ever-increasing OSM data completeness [23] where studies have focused on routing based on OSM street network data [30]. In China, OSM data related to Beijing has been reported to be fairly complete [41, 42]. Based on these evidences, and the fact that the coverage and quality of OpenStreetMap data is growing day by day, we argue that the use of OpenStreetMap data for this study is justified, although we concede that occasionally OpenStreetMap may suffer from incompleteness and hence cannot be considered to be robust. For our study, we import pedestrian networks of Beijing and Melbourne from OpenStreetMap [6] which have been illustrated in Figure 2. We have used the Python package OSMnx [6] for extracting network information from OpenStreetMap.

The overall road network structure between Beijing and Melbourne may not appear to be too dissimilar only when looking at it at a large scale, like Figure 2. On a closer look, Melbourne is a designed modern city, and Beijing an old city, with an elaborate pedestrian network crowded with dead ends. So, while looking at the two cities at a micro-scale, namely at a scale akin to a pedestrian's average walking route, it can be observed that Melbourne retains its regular grid-like pattern (even as we move into the suburbs) while Beijing does not. There are a host of studies which analyze aggregate city networks using complexity measures such as average circuity, entropy, and centrality. These complexity

Figure 2: Pedestrian networks obtained via OpenStreetMap (a) Beijing and (b) Melbourne. Each dimension of the figures represents 20 kilometers on the ground.

measures (conducted on a large city-wide scale) do not always reveal the true nature of street network orientation. In this study, we are interested in studying pedestrian movements. Pedestrian movements are very different from movement via other transportation modes as (a) pedestrian movements are mostly limited to shorter trip distances and (b) pedestrian movements do not always conform to the major roads, but mostly are concentrated within the arterial and sub-arterial streets. Hence, we felt that the complexity measures at the city-scale are not entirely appropriate for our study. The original study which we rely on for our choice of study areas [33], analyzed 1,000 map images for each of the cities at smaller scales (400m x 400m, which is a relevant scale for pedestrian movement) and concluded that Melbourne and Beijing street network morphologies are of contrasting nature. We present sample figures (Figure 3 and Figure 4) of typical pedestrian network structure in both the cities at a much smaller scale. Here the contrast between the two cities becomes more apparent.
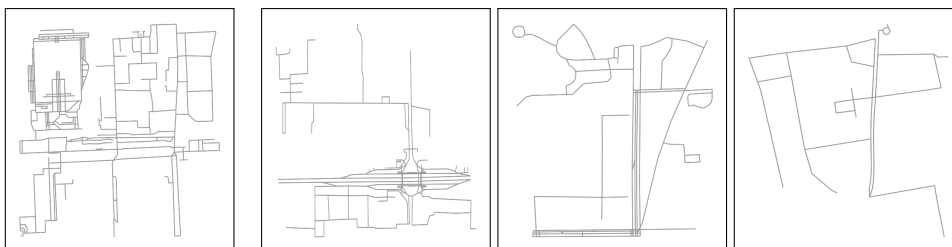

Figure 3: Typical pedestrian network structure within urban blocks in Beijing. Each dimension of the figures represent 1 kilometer on the ground.

Figure 4: Typical pedestrian network structure within urban blocks in Melbourne. Each dimension of the figures represent 1 kilometer on the ground.

3   Data

3.1   Beijing dataset

This GPS trajectory dataset was collected in Microsoft Research Asia's Geolife project by 182 users in a period of over five years (from April 2007 to August 2012) [43, 44]. The raw dataset contains 17,621 trajectories with a total distance of 1.2 million kilometers and a total duration of more than 50,000 hours. These trajectories were recorded by different GPS loggers and GPS-phones, and have a variety of sampling rates. 91.5 percent of the trajectories are logged in a dense representation, e.g. every 1 to 5 seconds or every 5 to 10 meters per point. Although this dataset is distributed over 30 cities of China and in some cities located in the USA and Europe, the majority pertains to Beijing, China. A substantial portion of the data was labelled by the users generating the data with the corresponding travel mode. In our study, we have limited our algorithms to the labelled portion of this large dataset (10.4 million GPS points, 9,070 trajectories from 70 users).

3.2   Melbourne dataset

Data for Melbourne was generated from the Victorian Future Mobility Sensing Project which was part of a new Urban Mobility and Intelligent Transportation initiative by the University of Melbourne, in partnership with Department of Economic Development, Jobs, Transport and Resources (DEDJTR), Massachusetts Institute of Technology (MIT), and Singapore-MIT Alliance for Research and Technology (SMART). The project collected personal travel data using a download-able smartphone application developed by SMART. Mode detection techniques were applied on the raw data to infer the transportation mode. The inferred modes were validated from the survey participants by asking them at the end of each day. Survey respondents were typically asked to complete the survey for 14 days, including five continuous days [31]. The raw dataset contains 1.2 million GPS points contributed by 84 users.

4 Methodology

4.1 Trip segmentation

In the first step, for each user, raw GPS points having transportation mode label as 'walk' or equivalent were filtered. Consequently, we obtained a series of GPS data points for each user in chronological order. These GPS points needed to be clustered into separate walking trips which would then be further analyzed. Thus, in the second step, trip segmentation criteria were applied to the filtered set of GPS points. A review of existing trip identification literature indicated that trip segmentation thresholds (also known as 'dwell time') are applied under two conditions: GPS signal-available situation and GPS signal-lost situation [14, 15]. It can be observed in [14] that the signal-available dwell time thresholds are consistently smaller than the signal-lost dwell time thresholds. This dwell time thresholds tend to vary with characteristics of local activity and ranges between 45 and 900 seconds [32]. Trip Identification and Analysis System (TIAS) concludes 'confident' trip ends for dwell time greater than 300 seconds [2]. For our study, we have selected a threshold of 300 seconds for differentiating between consecutive walking trips.

Although the participants in the datasets had labelled their data by stating the duration of travel in certain transportation modes, plotting GPS points clearly indicated unreasonable spatial gaps between two clusters of points inside the same walking trip. This indicated that using only a time-based threshold was not appropriate for trip segmentation due to occasional erroneous labelling of transport mode by the survey participants. For example, there could be a chance that the participant took a motorized mode of transport for a very short duration (less than 300 seconds) and instead of differentiating that non-walking trip, incorporated it under the encompassing walking trip by mistake. This resulted in erroneous map matching, as observed from trials. One such instance from the Beijing dataset is illustrated in Figure 5. But such observations could stem from noisy GPS points as well. To remove such potentially erroneous labelling and avoid trip segmentation due to noisy GPS data points (outliers) at the same time, we have supplemented the first trip segmentation threshold with an additional threshold. Here, we check whether the time difference between two consecutive data points is greater than 20 seconds. If not, then we do not consider trip segmentation and thus try and avoid trip segmentation due to outliers. Otherwise, we calculate the velocity between the two points by dividing the great circle distance by the time gap. If the velocity is unreasonable (greater than 2 meters/second) in terms of human walking speeds, the trip is segmented. The flowchart for this method has been illustrated in Figure 6. While the aforementioned trip segmentation does not guarantee robust results, from our observations on the datasets (with sampling rate less than 20 seconds), these thresholds provide satisfactory outcomes.

4.2 Activity locations

Apart from trip segmentation, there is the aspect of identification of activity locations that reside at the end points of trips [15]. Observation of plots of some trajectories revealed that their points were clustered in a small geographic area, indicating the occurrence of an activity, rather than a trip. It was necessary to remove such instances to obtain more representative results, since our study is interested in routes and their characteristics and not the origins and destinations where activities take place. One study showed the use of a

Figure 5: Example of two spatially far apart clusters, inappropriately labelled under the same walking trip in the raw Beijing dataset.

sophisticated algorithm for inferring activity locations by employing density-based spatial clustering of applications with noise (DBSCAN) and support vector machines (SVM) [15], while another study applied distance and time thresholds to do the same [37]. We have applied the distance and time thresholds of 200 metres and 20 minutes to remove such instances, following [37] as mentioned in Equation 2:

$$Dist\,(p_1, p_n) < 200\ m \quad \&\& \quad T_d\,(p_1, p_n) > 20\ min \tag{2}$$

where $Dist\,(p_1, p_n)$ refers to the Haversine distance between the first point p_1 and the last point p_n of the inferred trip and T_d is the time duration. In addition to the above criteria, after map matching, we have checked whether the map matched route distance exceeds twice the length of the corresponding shortest route or whether the length of the shortest route is equal to zero, indicating a possible round-trip. We have consequently removed such activity-based trips and round-trips, which are not relevant for this study as including them in our analysis will lead to our results being less representative of ground truth.

4.3 Filtering walking trips based on trip duration

In our previous work [5], we simulated heuristic routes between a pair of origin and destination only if the length of the shortest route between them fell inside the range of 400 metres (equivalent to a 5-minute walk) to 2,000 metres (equivalent to a 25-minute walk), based on the reviewed literature [1, 28, 29, 35, 40]. In this study, we have only considered walking trips where the duration of the trip lasts at least five minutes and not more than 25
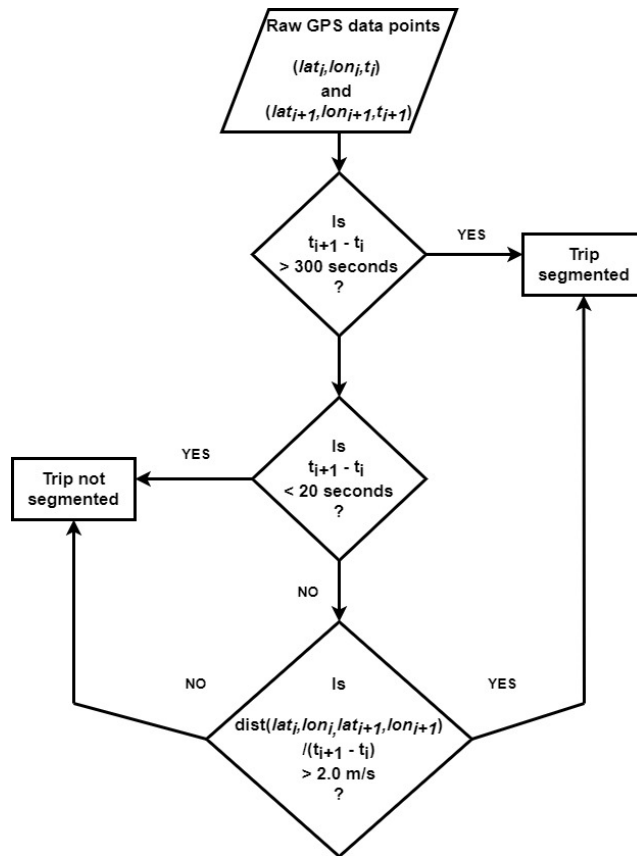
Figure 6: Flowchart of trip segmentation procedure.

minutes. Trips shorter than five minutes rarely deviate from the shortest route with actual routes and wayfinding heuristics coinciding with the same. Trips longer than 25 minutes are rarely non-activity-based trips and have a high chance of having multiple destinations instead of just one.

4.4 Removing trips made outside the cities

As mentioned in Section 3.1, the Geolife dataset contains trips made outside Beijing as well. Since the scope of our study is limited to analyzing walking trips made within Beijing and Melbourne, it was necessary to remove trips that were made outside the city. We defined a bounding box encompassing Beijing as *[39.75872, 40.159191, 116.04142, 116.638641]* and Melbourne as *[-37.540112, -38.411251, 144.553207, 145.507736]*. These limits were obtained from Yahoo! GeoPlanet database as was also reciprocated in a public GitHub repository (https://github.com/amyxzhang/boundingbox-cities). We considered a trip to be made outside the city if the trip origin did not fall within the aforementioned bounding box.

4.5   Map matching

As mentioned in Section 2.3, we have made use of a public GitHub repository based on [25] for map matching. For searching for probable consecutive road segments, we have set the search radius parameter at 300 metres. A greater value is computationally more expensive and sometimes results in more inaccurate outcomes, at least in case of walking trajectories where the points are closely spaced as compared to its motorized counterparts. A lesser value of the search radius often results in impossible map matching, as was experienced from values of 200 and 250 metres. Map matching resulted in the algorithm returning a sequence of OSM nodes that were traversed. We have considered the first and last point of the obtained sequences as the origin and the destination for each trip, respectively. This was necessary to simulate shortest route using Dijkstra's shortest-path algorithm and heuristic routes using algorithms mentioned in Section 2.1.

The preprocessing methodology has been illustrated in Figure 7 and the data has been described in Table 1.

	Raw labelled dataset			Preprocessed dataset	
City	Labelled points	GPS	Users in labelled dataset	Number of walking trips	Number of users
Beijing	10.4 million		70	1458	45
Melbourne	1.2 million		84	178	48

Table 1: Data description.

5   Preliminary findings

Figure 8, illustrating the temporal distribution of the number of trips, shows two distinct peaks, one in the morning and one in the evening, in both the datasets. Trips made during the night and early morning are significantly lower than the other times of the day. Also, the evening peak in Melbourne (5 p.m.) occurs earlier than Beijing (6–7 p.m.), while the morning peak is similar (8–9 a.m.) dropping possible hints at the difference in usual working-hours in both the cities. Furthermore, given the temporal distribution revealed by the visualizations is typical for the population, we assume some representativeness of our datasets.

The mean route lengths of the actual (map-matched) route, the shortest possible route and the routes simulated based on the four wayfinding heuristics have been illustrated in Figure 9. It can be observed that the mean route lengths (both actual and simulated) in Melbourne are consistently lower than those in Beijing, even though we had filtered trips that had a duration between 5 and 25 minutes, as mentioned in Section 4.3. This was also observed in our previous study, where actual routes had not been analyzed but rather simulations were undertaken. If anything, the contrast appears even more between Beijing and Melbourne, in comparison to our previous study. The mean route costs *(the difference in route length between a given route and the corresponding shortest route expressed in terms of a percentage of the shortest route length)* of the actual route and the simulated heuristic routes have been illustrated in Figure 10. The variation of cost across heuristics is less in Melbourne (Standard Deviation = 3.33% and Coefficient of Variation = 61.62%) as compared to Beijing (Standard Deviation = 6.20% and Coefficient of Variation = 90.04%), a pattern that
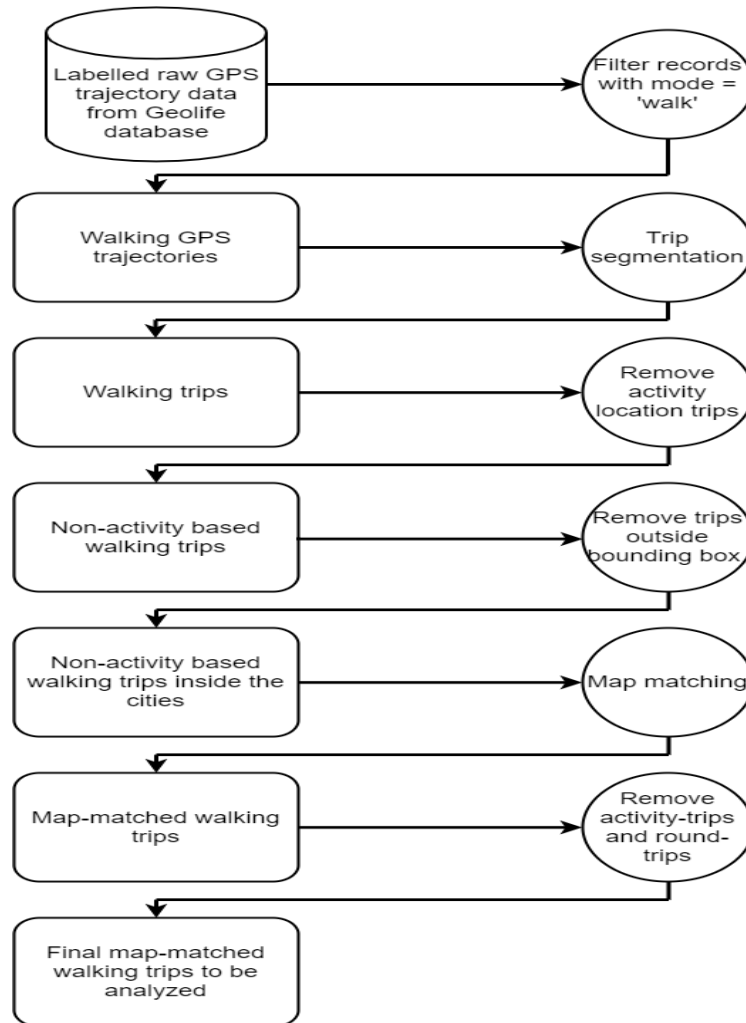
Figure 7: Methodology flowchart.

is in line with the conclusion from our previous study (Melbourne : Standard Deviation = 6.96% and Coefficient of Variation = 87.05% and Beijing : Standard Deviation = 9.42% and Coefficient of Variation = 101.80%).

These route length and route cost results show that our previous study (which only used simulations) and our current study (which analyzes actual observations), both follow a similar pattern and are not contradictory. They reveal two important things. One, these preliminary findings on route length and route cost validate the results of our previous study. We do not say that the results are the same, but the pattern is apparently similar (Melbourne's more consistent than Beijing), and they support the argument of contrasting morphologies to a greater extent. It must be noted that the mean cost of Melbourne's actual route in Figure 10 is more than Beijing, because of comparison with shorter 'shortest available routes' than Beijing. Two, even though the spatial extent of our study areas are
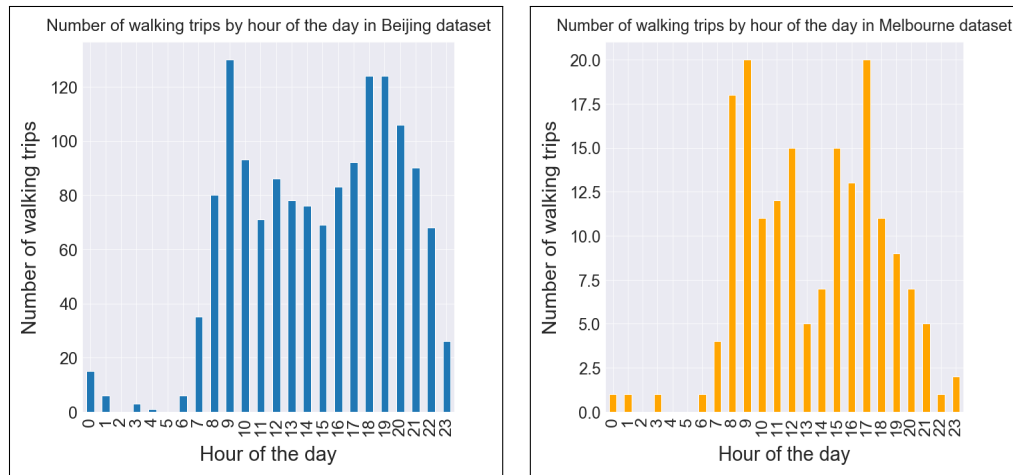
Figure 8: Walking trips by hour of the day in: (a) Beijing dataset and (b) Melbourne dataset.
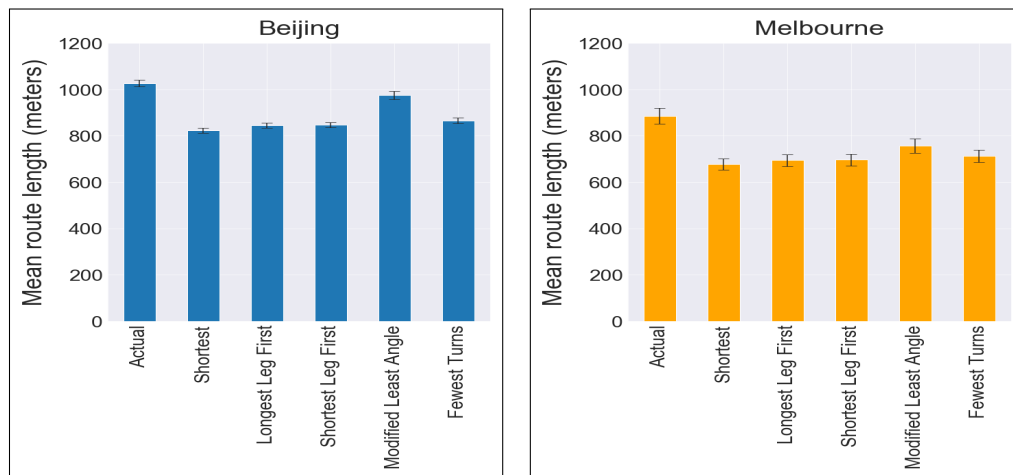

Figure 9: Mean route length in: (a) Beijing dataset and (b) Melbourne dataset.

not confined to a 5-kilometer bounding box (like in our previous study), the contrast between the morphologies of Beijing and Melbourne remain consistent (if not increased) even at a larger scale. In our previous study, we had selected the smaller study area so that it preserved the unique morphological characteristics of the pedestrian network of each city without diminishing the morphological differences between cities. Usually, as we move further into the suburbs of a city, the morphology tends to lose its uniqueness (usually, by becoming more irregular) and the density of their pedestrian network also reduces drastically. As we had to consider larger study areas (for the sake of not depleting our sample sizes), we felt that we might lose the significant contrast in network structure between Beijing and Melbourne. But a closer assessment of Figure 2 shows that Melbourne's suburban
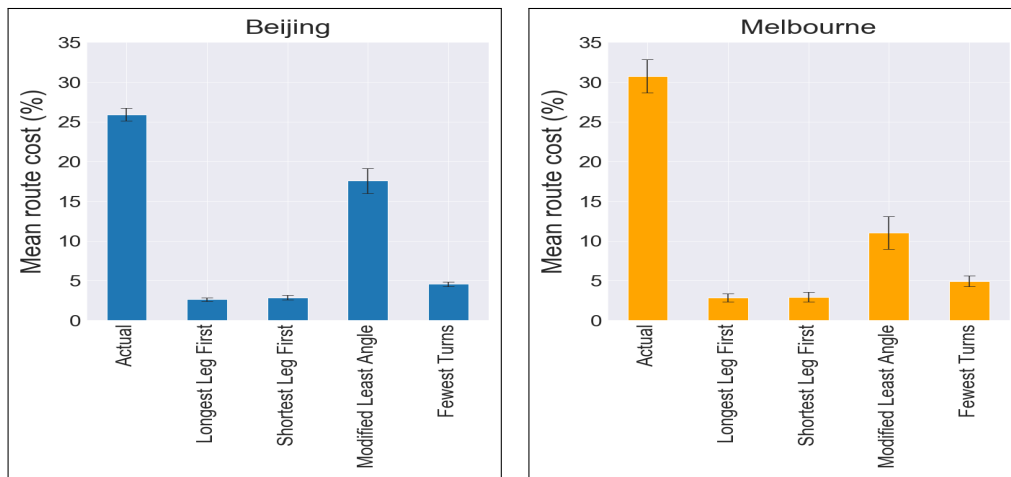
Figure 10: Mean route cost: The difference in length between a route and the corresponding shortest route, for a given OD pair, expressed in terms of percentage of the shortest route in: (a) Beijing dataset and (b) Melbourne dataset.

pedestrian network maintains its grid-like structure, much more consistently than Beijing. That is, even at a larger-scale (larger than the 5-kilometer bounding box), Melbourne is much more regular than Beijing. And this is supported by our our preliminary results.

6   Results and discussion

To investigate the relationship between heuristic choice distribution and network morphology we have made use of one-way analysis of variance (ANOVA) test, which tests the null hypothesis that two or more independent groups have the same population mean. As mentioned in Section 2.4, we have opted to use NHD as our route similarity metric. NHD (in meters) is a measure to quantify the dissimilarity between two routes in a road network. The greater the magnitude of NHD, the more is the existing dissimilarity. We compare the all the actual (map-matched) routes with their corresponding theoretical (heuristic) routes from both the datasets based on NHD.

In the context of this study and our stated hypothesis, we found that the variation of NHD across the heuristics is far more apparent in Beijing (standard deviation = 16.58 m and coefficient of variation = 11.77%) as compared to Melbourne (standard deviation = 7.11 m and coefficient of variation = 6.85%). The one-way ANOVA test was applied to both the datasets to check whether the mean NHD obtained from the four heuristics in each city was statistically significantly different from each other. In Melbourne, this difference was statistically not significant at 95% confidence interval. That is, the difference in mean NHD across the four heuristics is probably random in nature. In contrast, this difference was found to be statistically significant at 99% confidence interval in Beijing. This indicates a strong evidence against the null hypothesis (that all four mean NHDs were equal in Beijing), which leads to its rejection. The detailed results are as follows.
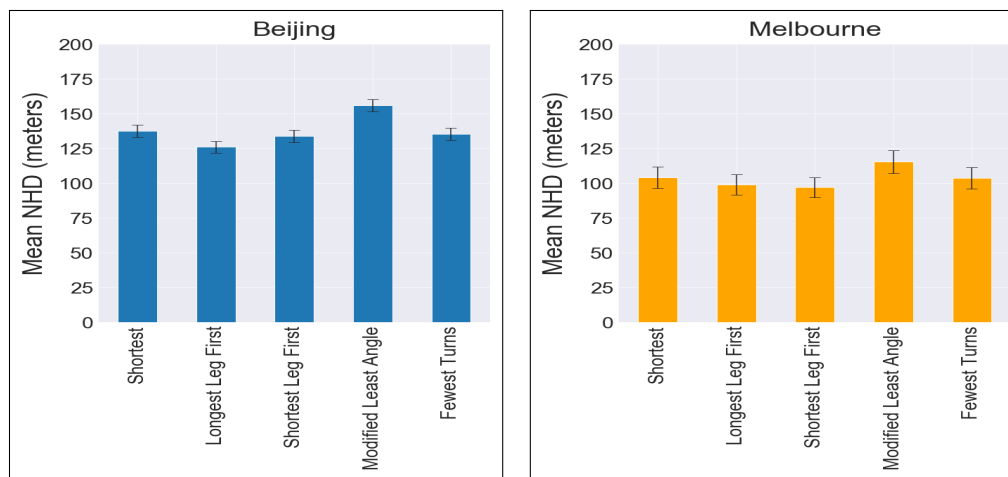
Figure 11: Mean Network Hausdorff Distance (NHD) in: (a) Beijing dataset and (b) Melbourne dataset.

- Melbourne: F = 1.15, p = 0.3293
- Beijing: F = 8.53, p = 0.0000

Results from the one-way ANOVA test make for interesting interpretations with respect to our hypothesis. Based on the findings from our previous study, we had argued that pedestrians choose heuristics by morphology as it was rational to disregard costly heuristics in irregular networks (thus creating a skewed heuristic choice distribution) and choose any heuristic in regular networks as all were equally costly (uniform heuristic choice distribution). Thus, we hypothesized that in Melbourne, the choice of heuristics will be uniformly distributed, while in Beijing, this distribution will be skewed. In this study, the choice of heuristic, rather the extent of compliance of the actual route with a heuristic route, was measured using NHD. So the distribution of heuristic choice was inferred by statistically measuring the uniformity of mean NHD values (average over all routes in the dataset) across all four heuristics. The ANOVA results suggest that this extent of compliance across heuristics is uniform in Melbourne. On the contrary, in Beijing the extent of compliance varies significantly across heuristics. In other words, actual routes, on an average, had uniformly complied with all four heuristics in Melbourne i.e. not one heuristic is significantly more (or less) dissimilar from the actual routes. But such was not the case in Beijing. This strengthens our argument that some heuristics are more (or less) popular in Beijing while all four heuristics are equally popular in Melbourne, owing to its more grid-like regular pattern of pedestrian network. From Figure 11, it is evident that in Beijing, Modified Least Angle heuristic is significantly less popular as it has the least average compliance (highest mean NHD value among heuristics) with the actual routes. On the basis of these statistical validations, it can be argued with some confidence, that if mean NHD is considered a proxy for choice of heuristics, and Melbourne and Beijing are representative of their respective network morphologies, pedestrians are unbiased towards wayfinding heuristics in regular networks while being biased in irregular networks.

7   Conclusion

We investigated whether network morphology of an urban pedestrian network has an impact on wayfinding heuristic choice distribution. In our previous work, we had shown via simulation that the variation in the cost of heuristic routes was greater in irregular networks as compared to regular ones. In regular grid-like networks, all heuristics were uniformly costly and not significantly longer than the shortest available route. On the contrary, in irregular networks, some heuristics were consistently resulting in significantly costlier alternatives in comparison to the shortest available routes. Based on this rationale, we hypothesized that pedestrian actions on the ground would be in line with these findings. In other words, we had argued that pedestrians choose heuristics by morphology as it was rational to disregard costly heuristics in irregular networks (thus creating a skewed heuristic choice distribution) and choose any heuristic in regular networks as all were equally costly (uniform heuristic choice distribution). We chose Beijing and Melbourne as the two cities for our study as they were deemed to have contrasting pedestrian network morphologies (as suggested by literature). We also concluded the same via close-up visual observations of the networks, especially inside the urban and suburban blocks, where Melbourne clearly had more regular patterns than Beijing. Our preliminary findings (in terms of route length and route cost) suggested the same.

In this paper, we demonstrated the use of raw GPS trajectories from both the cities in conjunction with heuristic route simulation to investigate whether these claims can be augmented with actual observations of pedestrian wayfinding behavior. Network Hausdorff Distance (NHD) was used as a measure of comparing actual routes with heuristic routes and compute the extent of compliance with our four studied heuristics. Using one-way ANOVA test on NHD values across heuristics, we established statistically that the mean NHD values for all four heuristics were not significantly different in Melbourne, but were significantly different in Beijing. This meant that actual routes had uniformly complied with all four heuristics in Melbourne but not in Beijing. In other words, heuristic choice distribution is different between the chosen cities, uniform in Melbourne and skewed in Beijing. This provided sufficient statistical evidence towards proving our hypothesis. Considering Melbourne and Beijing to be representative of regular and irregular network morphologies respectively, we generalized our conclusions and argued in favor of our hypothesis with requisite statistical evidence. As wayfinding heuristics help generate realistic aggregate movement patterns of people in urban spaces, relevant future studies should be able to make informed decision on the choice distribution of these heuristics (with multiple strategies under consideration) across the pedestrian population considered for the study, based on network morphology of the urban space studied.

There were certain considerations and assumptions made in this study that need to be highlighted as well. First, we used map matching to infer actual routes from sets of raw timestamped GPS records. Map matching results in the most probable route, given the fact that GPS data often suffers from positioning errors. We have made use of a sophisticated algorithm to overcome these challenges, yet care must be taken while interpreting actual routes. Second, we have employed multiple space-time-based criteria to filter out activity-based trips, round trips, trips that fall outside the usual walking trip lengths, and trips where the effect of heuristics will not be pronounced. While one can argue about the appropriateness of the thresholds and their values, our judgments were based on consultation of existing literature and observations of randomly sampled results from our datasets.

Third, the Melbourne dataset used for the analysis was smaller in comparison to Beijing's. Although, two datasets with closer sample sizes would have been more desirable, the usual temporal pattern of pedestrian volume in urban spaces was mirrored precisely by both the datasets. Also, from our preliminary findings, results from the Melbourne dataset had intuitive comparisons with the Beijing data, even though its sample size was considerably less. Hence, we believe that the findings of our study are not questionable in this regard.

Another important consideration in reference to the datasets is the existence of super-users (users contributing heavily in the datasets). This is evident from looking at Figure 12 where clearly some users have contributed more than the rest (users #153 and #86 in Beijing and users #73 and #153 in Melbourne). As they are present in both the datasets, super-users influencing the results and acting as the differentiating factor between the two cities seems highly unlikely. While user bias can produce misleading results [4], it is important to note the context of the study, which in this case, is heuristic choice popularity distribution, and not popularity of any specific route or street segment. In relevance to this study, there could be cases where super-users, by recording their weekday walking trips using the same route (and thus the same heuristic), influence one heuristic greatly than others. But these super-users have not only shared their weekday walking trips, but also other recreational trips with varying heuristics much more than other participants. People do not apply the same heuristics in every situation and they tend to switch depending upon the environment. From visual assessment of individual heuristic choice distribution, we observed that these super-users were not disproportionately adhering to any one heuristic. Furthermore, it must be kept in mind that we used NHD, a continuous variable, to measure route (dis)similarity. In most cases, there is no absolute compliance with ideal heuristic routes. We cannot claim that one route follows one heuristic absolutely, and not the others (no binary outcome) and that was not the goal of our study. Thus, there are positive NHD values, based on which we supported our hypothesis on choice distribution of heuristics. By using a continuous variable such as NHD and not any binary outcome, the problem of bias of super-users reduces significantly. While there may be arguments in favor of random undersampling of data to remove user bias, reducing a small dataset further would not have necessarily yielded more representative results and reduced the credibility of statistical claims.

Finally, there are a host of other factors that can influence the wayfinding decisions of pedestrians. Our study was confined to geometric heuristics, ones that are dependent on pedestrian network structure. But people are not limited solely by these four heuristics, or just geometric heuristics, and urban areas offer much more than just their street orientation (in terms of land-use and infrastructure). Pedestrians may select routes with most landmarks, maximum weather protection, maximum perceived safety, least crowded and least pollution. Also, pedestrians may apply multiple heuristics at multiple stages of a single walking trip, and they are not always strictly adhering to their chosen heuristic. This is also reflected in the positive NHD values for the heuristics, meaning that compliance with ideal heuristic routes is partial in most cases. But these non-geometric heuristics are not relevant for this study, as our intention was to test heuristic choice distribution across network morphologies. For example, when analyzing two urban spaces vastly different in terms of green space proportion, it will reveal contrasting heuristic choice distributions. Then, of course, the heuristics in consideration have to be relevant to land-use and not network morphology. Yet, one could argue about the relevance of the four heuristics used in this study. It must be noted that the context of this study was comparing heuristic choice
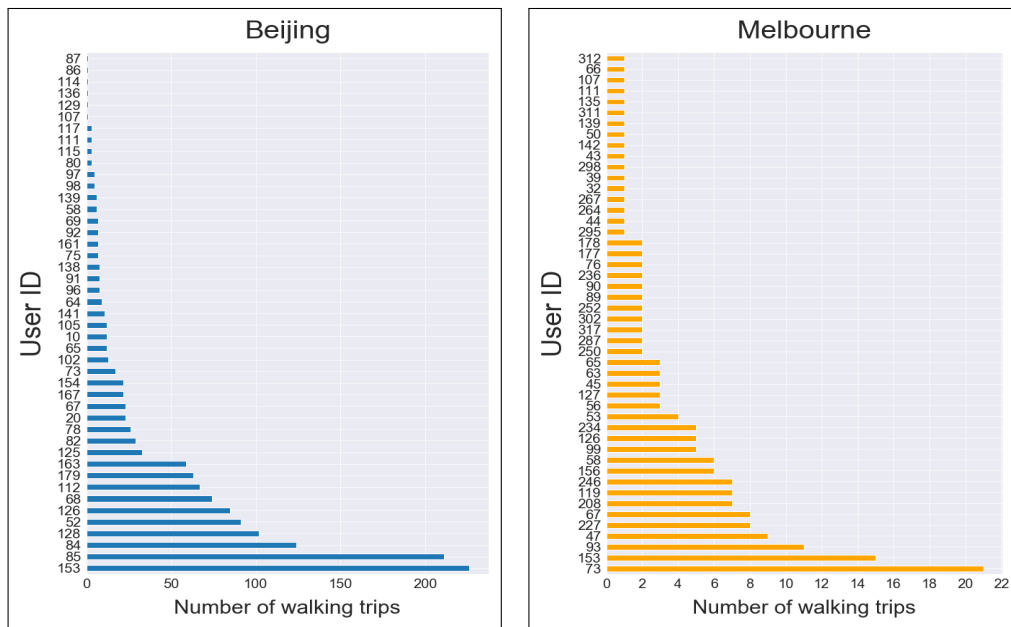
Figure 12: Walking trips by user in: (a) Beijing dataset and (b) Melbourne dataset.

distribution between two contrasting network morphologies. The intention was not to check extent of compliance for any individual heuristic. Hence, we investigated heuristic choice over all heuristics and across two contrasting morphologies. So, even though other heuristics have been applied by pedestrians, quantifying dissimilarity with actual routes using NHD meant that we had a continuous variable to compare all the four heuristics (instead of *fully complied* or *not complied at all*), and judge the extent of compliance. This helped us disregard the effect of other heuristics not included in this study that may have been partially applied. Hence, the findings of this study hold true. Overall, the findings from our previous study made us argue that in regular grid-like networks, where heuristic choice does not matter and almost all strategies lead to a route not substantially different from the shortest available route, heuristic choice distribution would be uniform. In this study, we gather enough statistical evidence to suggest the same.

References

[1] ARASAN, V. T., RENGARAJU, V., AND RAO, K. K. Characteristics of Trips by Foot and Bicycle Modes in Indian city. *Journal of Transportation Engineering 120*, 2 (1994), 283–294. doi:10.1061/(ASCE)0733-947X(1994)120:2(283).

[2] AXHAUSEN, K., SCHONFELDER, S., WOLF, J., OLIVERIA, M., AND SAMAGA, U. Eighty Weeks Of GPS Traces: Approaches To Enriching Trip Information. In *Transportation Research Board Annual Meeting* (2004), pp. 1870–06. doi:10.3141/1870-06.

[3] BAILENSON, J. N., SHUM, M. S., AND UTTAL, D. H. The initial segment strategy: A heuristic for route selection. *Memory & Cognition 28*, 2 (2000), 306–318. doi:10.3758/BF03213808.

[4] BERGMAN, C., AND OKSANEN, J. Estimating the Biasing Effect of Behavioural Patterns on Mobile Fitness App Data by Density-Based Clustering. In *Geospatial Data in a Changing World*. Springer, 2016, pp. 199–218. doi:10.1007/978-3-319-33783-8_12.

[5] BHOWMICK, D., WINTER, S., AND STEVENSON, M. Comparing the costs of pedestrian wayfinding heuristics across different urban network morphologies. In *GeoComputation 2019* (2019). doi:10.17608/k6.auckland.9846137.v1.

[6] BOEING, G. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems 65* (2017), 126–139. doi:10.1016/j.compenvurbsys.2017.05.004.

[7] BRAKATSOULAS, S., PFOSER, D., SALAS, R., AND WENK, C. On Map-Matching Vehicle Tracking Data. In *Proceedings of the 31st international Conference on Very Large Databases* (2005), VLDB Endowment, Trondheim, Norway, pp. 853–864. doi:10.5555/1083592.1083691.

[8] CHRISTENFELD, N. Choices from identical options. *Psychological Science 6*, 1 (1995), 50–55. doi:10.1111/j.1467-9280.1995.tb00304.x.

[9] DALTON, R. C. The secret is to follow your nose: Route path selection and angularity. *Environment and Behavior 35*, 1 (2003), 107–131. doi:10.1177/0013916502238867.

[10] DUCKHAM, M., AND KULIK, L. "Simplest" Paths: Automated Route Selection for Navigation. In *Spatial Information Theory. Foundations of Geographic Information Science* (2003), Springer Berlin Heidelberg, pp. 169–185. doi:10.1007/978-3-540-39923-0_12.

[11] EVANS, M. R., OLIVER, D., SHEKHAR, S., AND HARVEY, F. Fast and exact network trajectory similarity computation: A case-study on bicycle corridor planning. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing - UrbComp '13* (2013), ACM, p. 9. doi:10.1145/2505821.2505835.

[12] GIL, J. Building a multimodal urban network model using openstreetmap data for the analysis of sustainable accessibility. *OpenStreetMap in GIScience* (2015), 229–251. doi:10.1007/978-3-319-14280-7_12.

[13] GOLLEDGE, R. G. Path selection and route preference in human navigation: A progress report. In *Spatial Information Theory A Theoretical Basis for GIS* (1995), Springer, pp. 207–222. doi:10.1007/3-540-60392-1_14.

[14] GONG, L., MORIKAWA, T., YAMAMOTO, T., AND SATO, H. Deriving Personal Trip Data from GPS data: A Literature Review on the Existing Methodologies. *Procedia-Social and Behavioral Sciences 138* (2014), 557–565. doi:10.1016/j.sbspro.2014.07.239.

[15] GONG, L., SATO, H., YAMAMOTO, T., MIWA, T., AND MORIKAWA, T. Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines. *Journal of Modern Transportation 23*, 3 (2015), 202–213. doi:10.1007/s40534-015-0079-x.

[16] GREENFELD, J. S. Matching GPS Observations to Locations on a Digital Map. In *81st Annual meeting of the Transportation Research Board* (2002), vol. 1, Transportation Research Board, Washington DC., pp. 164–173.

[17] GUDMUNDSSON, J., LAUBE, P., AND WOLLE, T. Computational Movement Analysis. In *Springer Handbook of Geographic Information*. Springer, 2011, pp. 423–438. doi:10.1007/978-3-540-72680-7_22.

[18] GÜTING, R. H., DE ALMEIDA, T., AND DING, Z. Modeling and querying moving objects in networks. *The VLDB Journal – The International Journal on Very Large Databases 15*, 2 (2006), 165–190. doi:10.1007/s00778-005-0152-x.

[19] HOCHMAIR, H. Least angle heuristic: Consequences of errors during navigation. In *GIScience* (2000), Citeseer, pp. 282–285.

[20] HOCHMAIR, H. H., AND KARLSSON, V. Investigation of Preference Between the Least-Angle Strategy and the Initial Segment Strategy for Route Selection in Unknown Environments. In *Spatial Cognition IV. Reasoning, Action, Interaction* (2005), Springer Berlin Heidelberg, pp. 79–97. doi:10.1007/978-3-540-32255-9_5.

[21] HUTTENLOCHER, D. P., AND KEDEM, K. Computing the minimum Hausdorff distance for point sets under translation. In *Proceedings of the sixth annual symposium on Computational geometry - SCG '90* (1990), ACM, pp. 340–349. doi:10.1145/98524.98599.

[22] HWANG, J.-R., KANG, H.-Y., AND LI, K.-J. Spatio-temporal Similarity Analysis Between Trajectories on Road Networks. In *International Conference on Conceptual Modeling* (2005), Springer, pp. 280–289. doi:10.1007/11568346_30.

[23] JOKAR ARSANJANI, J., ZIPF, A., MOONEY, P., AND HELBICH, M. An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications. *OpenStreetMap in GIScience* (2015), 1–15. doi:10.1007/978-3-319-14280-7_1.

[24] LOU, Y., ZHANG, C., ZHENG, Y., XIE, X., WANG, W., AND HUANG, Y. Map-matching for low-sampling-rate GPS trajectories. In *17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2009), ACM, pp. 352–361. doi:10.1145/1653771.1653820.

[25] MEERT, W., AND VERBEKE, M. HMM with non-emitting states for Map Matching. In *European Conference on Data Analysis (ECDA), Paderborn, Germany* (2018).

[26] NEWSON, P., AND KRUMM, J. Hidden Markov Map Matching Through Noise and Sparseness. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2009), ACM, pp. 336–343. doi:10.1145/1653771.1653818.

[27] NOVACK, T., WANG, Z., AND ZIPF, A. A System for Generating Customized Pleasant Pedestrian Routes Based on OpenStreetMap Data. *Sensors 18*, 11 (2018), 3794. doi:10.3390/s18113794.

[28] RAHUL, T., AND VERMA, A. A study of acceptable trip distances using walking and cycling in Bangalore. *Journal of Transport Geography 38* (2014), 106–113. doi:10.1016/j.jtrangeo.2014.05.011.

[29] ROBERTSON, S., AND THOREAU, R. Usability of pedestrian crossings: Further results from fieldwork. In *Proceedings of the International Conference on Contemporary Ergonomics (CE2005), 5-7 April 2005, Hatfield, UK* (2005), pp. 608–612.

[30] ROCHA, C. M., KRUGER, E., MCGUIRE, S., AND TENNANT, M. Role of public transport in accessibility to emergency dental care in Melbourne, Australia. *Australian Journal of Primary Health 21*, 2 (2015), 227–232. doi:10.1071/PY13102.

[31] RODDIS, S., WINTER, S., ZHAO, F., AND KUTADINATA, R. Respondent preferences in travel survey design: An initial comparison of narrative, structured and technology-based travel survey instruments. *Travel Behaviour and Society 16*, 7 (2019), 1–12. doi:10.1016/j.tbs.2019.03.003.

[32] SCHUESSLER, N., AND AXHAUSEN, K. W. Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board 2105*, 1 (2009), 28–36. doi:10.3141/2105-04.

[33] THOMPSON, J., STEVENSON, M., WIJNANDS, J. S., NICE, K. A., ASCHWANDEN, G. D., SILVER, J., NIEUWENHUIJSEN, M., RAYNER, P., SCHOFIELD, R., HARIHARAN, R., ET AL. A global analysis of urban design types and road transport injury: an image processing study. *The Lancet Planetary Health 4*, 1 (2020), e32–e42. doi:10.1016/S2542-5196(19)30263-3.

[34] VICEK, C., MCLAIN, P., AND MURPHY, M. GPS/dead reckoning for vehicle tracking in the "urban canyon" environment. In *Proceedings of VNIS'93-Vehicle Navigation and Information Systems Conference* (1993), IEEE, pp. 461–34. doi:10.1109/VNIS.1993.585671.

[35] VICTORIA WALKS. Pedestrian access strategy : A strategy to increase walking for transport in Victoria. Tech. rep., State of Victoria, 2010.

[36] VLACHOS, M., KOLLIOS, G., AND GUNOPULOS, D. Discovering Similar Multidimensional Trajectories. In *Proceedings 18th International Conference on Data Engineering* (2002), IEEE, pp. 673–684. doi:10.1109/ICDE.2002.994784.

[37] WEI, Q., SHE, J., ZHANG, S., AND MA, J. Using Individual GPS Trajectories to Explore Foodscape Exposure: A Case Study in Beijing Metropolitan Area. *International Journal of Environmental Research and Public Health 15*, 3 (2018), 405. doi:10.3390/ijerph15030405.

[38] WHITE, C. E., BERNSTEIN, D., AND KORNHAUSER, A. L. Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies 8*, 1-6 (2000), 91–108. doi:10.1016/S0968-090X(00)00026-7.

[39] YANAGISAWA, Y., AKAHANI, J.-I., AND SATOH, T. Shape-Based Similarity Query for Trajectory of Mobile Objects. In *International Conference on Mobile Data Management* (2003), Springer, pp. 63–77. doi:10.1007/3-540-36389-0_5.

[40] YANG, Y., AND DIEZ-ROUX, A. V. Walking Distance by Trip Purpose and Population Subgroups. *American Journal of Preventive Medicine 43*, 1 (2012), 11–19. doi:10.1016/j.amepre.2012.03.015.

[41] ZHAO, P., JIA, T., QIN, K., SHAN, J., AND JIAO, C. Statistical analysis on the evolution of OpenStreetMap road networks in Beijing. *Physica A: Statistical Mechanics and its Applications 420* (2015), 59–72. doi:10.1016/j.physa.2014.10.076.

[42] ZHENG, S., AND ZHENG, J. Assessing the Completeness and Positional Accuracy of OpenStreetMap in China. In *Thematic Cartography for the Society*. Springer, 2014, pp. 171–189. doi:10.1007/978-3-319-08180-9_14.

[43] ZHENG, Y., XIE, X., AND MA, W.-Y. Understanding Mobility Based on GPS Data. In *Proceedings of the 10th International Conference on Ubiquitous Computing - UbiComp '08* (2008). doi:10.1145/1409635.1409677.

[44] ZHENG, Y., XIE, X., AND MA, W.-Y. Mining Interesting Locations and Travel Sequences From GPS Trajectories. In *Proceedings of the 18th international conference on World wide web - WWW '09* (2009). doi:10.1145/1526709.1526816.

[45] ZHENG, Y., AND ZHOU, X. *Computing with Spatial Trajectories*. Springer Science & Business Media, 2011. doi:10.1007/978-1-4614-1629-6.

[46] ZHOU, Y., WANG, W., HE, D., AND WANG, Z. A fewest-turn-and-shortest path algorithm based on breadth-first search. *Geo-spatial Information Science 17*, 4 (2014), 201–207. doi:10.1080/10095020.2014.988198.

[47] ZIELSTRA, D., AND HOCHMAIR, H. H. Using Free and Proprietary Data to Compare Shortest-Path Lengths for Effective Pedestrian Routing in Street Networks. *Transportation Research Record: Journal of the Transportation Research Board 2299*, 1 (2012), 41–47. doi:10.3141/2299-05.

[48] ZIELSTRA, D., HOCHMAIR, H. H., AND NEIS, P. Assessing the Effect of Data Imports on the Completeness of OpenStreetMap–A United States Case Study. *Transactions in GIS 17*, 3 (2013), 315–334. doi:10.1111/tgis.12037.