Extracting Realistic Pedestrian, Cycling, and Driving Street Networks from OpenStreetMap

Juan Pablo Duque[1], Maria Brovelli[1]

[1] Department of Civil and Environmental Engineering (DICA), Politecnico di Milano – (juanpablo.duque, maria.brovelli)@polimi.it

Keywords: OpenStreetMap, street networks, data quality, open source.

Abstract

This paper presents a methodology for extracting realistic and usable pedestrian, cycling, and driving street networks derived from OpenStreetMap (OSM) data. While OSM is a valuable source of information, it is not exempt from inconsistencies and errors, particularly affecting pedestrian and cycling networks, which can lead to inaccurate navigation or unsafe routes. To address this, we propose a methodology which employs a set of filters and post-processing algorithms in a selective removal process to mitigate such inconsistencies and generate networks that are realistic and usable for various applications like routing and analysis. The methodology is designed to be general and globally applicable at the city level. Particular attention is paid to pedestrian networks, addressing the inconsistency where separately mapped sidewalks are not explicitly specified in their corresponding driving streets. An algorithm was designed to detect these cases based on the geometry of the street segments. The methodology also offers an assessment functionality to identify potential inconsistencies in OSM extracted networks, providing support for mapping campaigns seeking to improve its quality. The methodology was applied to five capital cities showing its effectiveness in refining the networks, with larger changes observed in cities with higher mapping quality. Code and examples are available in a GitHub repository.

1. Introduction

Geospatial information is key to provide context to the world we live in. Accurate, realistic, and complete maps are a first-hand necessity, specially for applications involving routing, autonomous navigation, or disaster response. In this context, the role of crowdsourced maps is to provide a fast, reliable, and affordable alternative for obtaining geospatial information, specially at the global level. On the internet, crowdsourced information can be found related to Points of Interest (POI), street networks, landscape data, and even 3D models.

In the crowdsourced map scene, in particular for street networks, a project that has dominated for its reasonable accuracy, extensive community, and openness, is OpenStreetMap (OSM). OSM is a collaborative project to build a free and editable map of the world, maintained by the community. It is widely used in research, humanitarian efforts, and even for commercial applications.

However the utility of OSM for obtaining street networks, as a crowdsourced dataset it is not exempt from inconsistencies and errors. In particular when building pedestrian and cycling street networks, many inconsistencies can be found due to the different levels of mapping detail of each city, different tagging conventions among countries with respect to the mapping of pedestrian and cycling infrastructure, outdated, and contradictory information. This can lead to incorrect maps that may cause wrong navigation information and travel times for users, or even suggest insecure routes that are not apt for pedestrians and/or bicycles.

To overcome these limitations, we propose a methodology that attempts to mitigate some of the inconsistencies when deriving pedestrian, cycling, and driving networks from OSM. This methodology is focused on producing realistic and usable street networks. By realistic, we mean that they should be as close to reality as possible, while usable is related to the fact that the derived networks should allow functional and analytical aspects, such as routing, simulations, or integration with other data. To do so, the methodology features a set of filters and post-processing algorithms applied to street networks extracted from OSM in a selective removal process, where certain segments of the network are removed based on their properties and/or their surroundings, mitigating known inconsistencies and redundancies. Our proposed methodology is intended to be as general as possible, in order to be used and extended to any urban area of the world at the city level. The extracted street networks can be used for several purposes, such as urban index calculation, digital twins, navigation applications, network analysis, or other research topics, supporting sustainable development and active mobility. Code and examples are available in a GitHub repository.

An additional advantage of this methodology is that it may be used to assess inconsistencies of pedestrian street networks in OSM. The evaluation is achieved by flagging streets with potential issues rather than eliminating them from the network. The results of an assessment may be used to support mapping campaigns seeking to improve the overall accuracy of OSM.

The paper is composed by the theoretical framework in section 2, describing how street networks are modelled, general aspects of OSM, and related work found in academic literature. Second, the methodology for extracting realistic driving, cycling, and pedestrian networks is presented and explained in section 3, including remarks on how it is applied to the assessment of pedestrian networks. Finally, examples are provided and discussed as use cases in section 4, showing the methodology in action for both the extraction of street networks and the pedestrian network assessment.

2. Theoretical Framework

2.1 Modelling of Street Networks

Street networks represent the roads of an area as a set of interconnected points and lines. In this paper, the term street network is used to describe a network of any mode of transportation, i.e., driving, cycling, and pedestrian. Street networks are usually modelled using graphs. A graph $G = \{N, E\}$ is a data structure composed by a set of nodes N (also called vertices), and a set of edges E (also called arcs). Edges connect the nodes of the graph, forming a network.

There are two popular approaches when modelling street networks with graphs: i) primal graphs, where street segments are represented by edges and intersections by nodes; and ii) dual or line graphs, where street segments are represented by nodes and intersections are represented as edges.

Representing the street network as a graph is particularly useful for a wide range of applications, including routing, network analysis, traffic simulations, space syntax, and urban planning. Our methodology features a primal graph approach to represent the resulting street networks.

2.2 OpenStreetMap Characteristics

OpenStreetMap (OSM) is a global, crowdsourced database containing information about street networks, points of interest, administrative boundaries, buildings, public transport, and much more, with a global coverage. In total, as of May 2025, the OSM database is over 150GB big.

Street networks in OSM are stored in a database. Each element in the OSM database can be a point, a line, a polygon, or a relationship. Normally, the nodes of the network are represented as points, while the edges are represented as lines. In addition to the geometry of each feature, a set of properties are stored, called tags. As of 2025, there exist over 100.000 distinct tags in OSM. Tags enrich the elements of the map, allowing to input properties for each type of map feature. In the context of street networks, some tags are fundamental for categorising and filtering. For example, adding the tag ”highway” to a line or point feature allows to specify the street type it represents, the ”access” tag defines if it is possible or not to circulate over a street, or the tags ”footway”, ”sidewalk”, and ”cycleway” give information about the pedestrian and cycling usage of certain street segments.

There are multiple ways to extract information from the OSM database. The OSM foundation provides a public API to query the database, called Overpass API (OpenStreetMap, 2025a). Moreover, a popular alternative for extracting street networks in python environments is OSMnx (Boeing, 2025). This library allows to download street networks to be used in the python graph manipulation library networkx, and to extract points of interest and map features as data frames. The OSMnx library has been extensively used in academic literature and was used as the base line for the evaluation of our methodology.

2.3 Related Work

OSM is widely recognized as a valuable source of geospatial data and is frequently used for generating street network datasets for various applications. In academic literature, it has been previously used for building street networks in various contexts, as it provides flexible and accessible data with respect to other sources (Čerba, 2025), and it is a source of constantly updated, crowdsourced information (Franzini et al., 2020).

Methodologies for the extraction of street networks have been proposed all over the academic literature. In the context of the extraction of driving street networks, examples include extraction methodologies for assessing OSM with authoritative datasets (Fan et al., 2016), data integration of OSM data for the assessment of territorial innovation (Čerba, 2025), and for the calculation of street network indices (Boeing, 2019a, Boeing, 2019b). Examples of street network extraction of cycling street networks encompass the identification of gaps and missing links in cycling networks using OSM data (Vybornova et al., 2023), the extraction of street networks for simulations of active transportation scenarios (Jafari et al., 2022), or the quantification of missing cycling infrastructure (Buczyński and Chavez-Pacheco, 2025). Multiple efforts has also been made for building pedestrian networks, usually contextualized based on their application. Methodologies for building pedestrian networks have been proposed for extracting inclusive and realistic roads for outdoor navigation (Hosseini et al., 2024), for analysing geometrical properties of networks (Hosseini et al., 2023), for the assessment of road hierarchy (Thottolil et al., 2024), for pedestrian urban analysis (Simons, 2023, Bartzokas-Tsiompras, 2022), among others. Machine and deep learning approaches, in conjunction with the street network, have also been explored for the assessment and improvement of OSM data in an automatic way (Kaur and Singh, 2019, Miloudi and Meguenni, 2023). Finally, the quality of OSM has been subject of study in multiple academic articles with respect to the street network (Kaur and Singh, 2019, Franzini et al., 2020, Elias et al., 2020) and to the map features such as points of interest (Klinkhardt et al., 2023). All of this studies highlight the importance of having generalized and realistic methodologies for the extraction of OSM data.

Other contributions towards the extraction of street networks from OSM and their refinement include open source software such as Neatnet (Gaboardi and Fleischmann, 2025), which enables graph simplification from OSM data; OSM SidewalKreator (de Moraes Vestena et al., 2023), which is a tool for extracting and creating sidewalks in OSM; or even OSMnx, which is used not only for data extraction, but also for network simplification (Boeing, 2025).

Our methodology attempts to build on top of existing work and improve the refinement procedures by featuring detailed filters and network-specific post-processing algorithms. The aggregated value lies in its globally-applicable nature and its focus on multiple modes of transportation rather than a single one.

3. Methodology

The methodology presented is focused on the extraction of realistic and usable street networks for driving, cycling, and pedestrian modes of transportation using open data from OpenStreetMap (OSM). This methodology is intended to be as general as possible, in order to be used for extracting street networks anywhere in the world using only free, open source data, and at the city level.

Given the wide variability of mapping accuracy between cities, specially between developed and developing countries, creating a one-fits-all methodology for extracting realistic street networks is a complex task. To generalise as much as possible our methodology and to create street networks both realistic (as close to reality as possible), and usable (enabled for routing, usage in GIS software, etc.), we propose a set of filters and post-processing algorithms for each network type, addressing the particularities of each mode of transportation. With that in mind, the general workflow of the methodology is depicted in figure 1.

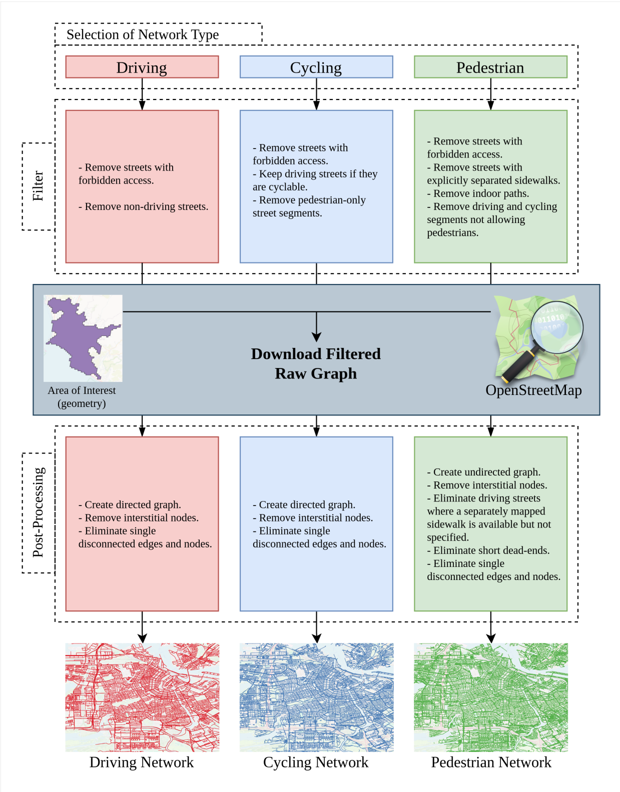

Figure 1. General workflow of the extraction of driving, cycling, and pedestrian street networks.

On the workflow, the first step is to define the type of network to extract, which can be driving, cycling, or pedestrian. Based on the network type, a different set of filters and post-processing steps are selected. A preliminary graph with raw OSM data is downloaded within a specified area of interest (defined by a polygon geometry) and using the specific filters of the network type. The downloaded graph contains each straight segment of the network, meaning that curves and roundabouts are represented using many connected edges. This raw graph is then post-processed using a set of algorithms defined for each network type. Post-processing includes building a directed or undirected graph, depending on the network type, the elimination of interstitial nodes, of single isolated nodes and edges, and of certain inconsistencies. These post-processing steps are devoted to make the resulting graph usable and, if possible, to reduce its size by removing inconsistencies that may create redundant information within the network. The final result is a graph, either directed or undirected, that best fits the selected network type.

The following subsections explain in-depth the filters and post-processing steps for each of the network types.

3.1 Driving Network Extraction

As OSM is primarily designed for driving street networks, extracting this kind of network is not a complex issue. However, there are some considerations to be taken into account.

As the direction in which traffic flows is important, the driving network is represented as a directed graph. This means that 2-way streets are represented by two different edges. This increases the size of the network, but improves the usability, as accurate routing is fundamental for these kind of networks.

3.1.1 Filters: The filters for the driving street network are mostly focused on the street access level, i.e., if vehicles are allowed or not to transit. For this purpose, we use the same filter used by the OSMnx library for driving networks. This filter removes edges based on the access level (e.g., private, forbidden, etc.), which is determined by the OSM tag ”access”. It also removes edges that do not belong to the driving network based on their ”highway” tag, such as certain service streets, parking alleys, pedestrian-only segments (e.g., sidewalks, paths), and cycling infrastructure (e.g., cycleways).

3.1.2 Post-processing: Initially, in order to render the network more usable, the post-processing ensures that the graph is directed. With the initial directed graph, the further post-processing steps of the driving street network are focused on reducing the size of the network by eliminating redundant nodes and edges. First, a procedure to remove interstitial nodes is carried out, where redundant nodes are eliminated. Then, single edges and nodes that are isolated from the graph are eliminated.

3.2 Cycling Network Extraction

Given the increased freedom of movement that bicycles have with respect to cars, the analysis and extraction of cycling networks is increasingly complex. Bicycles usually share space with traffic in urban areas, but may also possess separate infrastructure, such as cycleways. This means that the cycling network is mostly composed of a subset of the driving street networks where bikes are allowed, plus cycling-enabled infrastructure, which includes not only cycleways, but also shared pedestrian paths like sidewalks or paths.

Cycling streets are represented as directed graphs, as they share a great portion of the street network in which the flow direction is fundamental.

3.2.1 Filters: The filters for the cycling network maintain each segment of the driving network unless it is not explicitly stated that cycling is not allowed. This is done using cycling-specific OSM tags such as ”bicycle” and ”cycleway”. For example, a driving street with the tag ”bicycle” equals to ”yes” explicitly specify that the street is cyclable. However, if the tag equals to ”no” or ”forbidden”, it means bicycles are not allowed, and that segment should be removed. To account for cities with low mapping accuracy, in cases where the ”bicycle” tag is missing the street is assumed to allow bicycles anyway. Nevertheless, in certain street types, such as motorways, it is implied that bicycles are not permitted. Due to this, motorways are excluded from the cycling network, except it is explicitly specified that it allows bicycles.

Another filter that is applied to the cycling network is with respect to pedestrian infrastructure, as some pedestrian paths are shared between pedestrians and bicycles. The filter aims to maintain any pedestrian segments that contain the tag ”bicycle” equals to ”yes” or ”designated”. This filter will include in the graph any paths that are originally pedestrian, such as sidewalks or hiking paths, that specify to be shared with bicycles.

3.2.2 Post-processing: Similarly to the driving street networks, cycling networks do not require extensive post-processing, as filtering is enough to refine the network to a desirable extent.

Consequently, post-processing is centred on ensuring that the graph is directed, on removing interstitial nodes, and on removing isolated edges and nodes which might be present after applying the filters.

3.3 Pedestrian Network Extraction

The case of pedestrian networks is particular. Given the wide freedom of movement that pedestrians enjoy, it is complex to decide when a street segment should be removed or not. This is increasingly complex in small urban centres and cities in developing countries, as laws are less restrictive and there is usually lower mapping quality.

3.3.1 Filters: In order to create a filter that is as general as possible for the pedestrian network, we part from the assumption that all driving streets are walkable, unless stated otherwise. This approach is consistent with the OSM pedestrian guidelines (OpenStreetMap, 2024).

Driving and cycling street segments are filtered out if pedestrians are not allowed. This is determined by the OSM tag ”foot”, when its value is either ”yes”, ”designated”, ”permissive”, or ”official”. When the tag ”foot” is not present it is assumed that the street is walkable for most street types. Motorways, similarly to cycling networks, are excluded when the tag ”foot” is not present. This means that certain cycling segments that share space with pedestrians are maintained in the network.

Other filters applied to driving and cycling streets include: the removal of inaccessible streets based on the ”access” tag; the removal of trunks, primary, and secondary streets that do not have sidewalks (specified by the tag ”sidewalk” with value equals to ”no”); and finally the removal of segments that specify that the pedestrian portion of the street is mapped separately, based on the tags ”sidewalk:left”, ”sidewalk:right”, and ”sidewalk:both”.

Additionally to the driving streets where walking is permitted, the pedestrian street network is also composed by segments with tag ”highway” equals to ”footway”, ”pedestrian”, ”living_street”, ”path”, ”track”, and ”steps”. Indoor segments are also removed, which are out of scope of this methodology as we are interested only on the outdoor network. Indoor paths are determined by the presence of the tag ”level”, consistently with OSM simple indoor mapping guidelines (OpenStreetMap, 2025b).

3.3.2 Post-processing: Although the pedestrian filters do much of the heavy lifting, post-processing is still useful for correcting inconsistencies on the resulting network. In particular, the pedestrian post-processing is focused on the elimination of duplicated driving streets where a separately mapped sidewalk is available but is not explicitly specified.

This problem arises from the fact that, in OSM, it is possible to map the main carriageway and the pedestrian sidewalk separately. That means, one segment is used to represent the part where vehicles can circulate, and another segment to represent where pedestrians can walk. On these cases, the driving street segment should specify that a separate geometry is used to represent the sidewalk by using the tag ”sidewalk:left”, ”sidewalk:right”, or ”sidewalk:both” with the value ”separated”. This division makes it easier for data consumers, as the data to be downloaded is less, and pedestrian infrastructure is more accurate. However, streets that have a separately mapped sidewalk do not always specify the separation. This generates more complexity on the resulting graph, an inaccurate map, and redundant information for data consumers.

To overcome this, an algorithm was designed to detect those cases based on geometrical properties, i.e, incident angle and proximity. The algorithm iterates over all the edges of the pedestrian graph. When it finds a driving street (based on its ”highway” tag) it creates an eroded buffer (see figure 2) from the geometry of the street segment. This is done to avoid capturing adjacent edges, and capture only segments that lie on the sides of the road. Then, each segment of the graph that intersects the buffer is selected as a candidate. To speed up the computation of the intersecting geometries, a spatial index was used. Finally, with the selected nearby edges, two checks are made: i) if the edge is a sidewalk (i.e., a sidewalk, a crossing, or a pedestrian footway/cycleway); and ii) if the edge has a similar incident angle, up to a threshold of 15 degrees, with respect to the street. If both proximity and angle matches, the driving street segment is removed from the graph and the sidewalk is maintained. A graphical example of the eroded buffer and selected edge candidates is presented in figure 2.

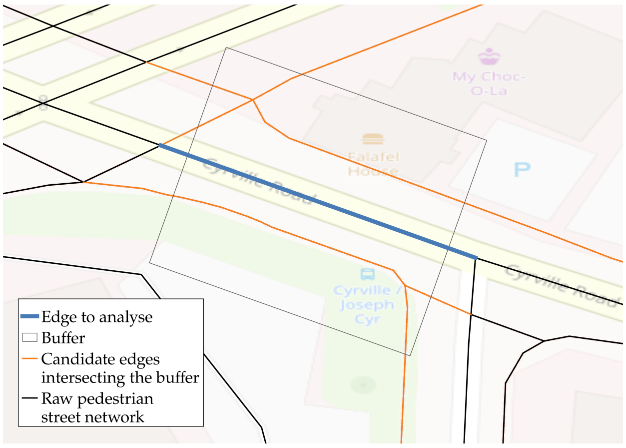

Figure 2. Example of eroded buffer and selected edge candidates for analysis of driving street segments.

3.4 Assessment of Pedestrian Networks

The proposed post-processing of pedestrian networks can be applied to not only refine the network, but also to identify OSM inconsistencies. By flagging inconsistent street segments instead of removing them, it is possible to do an assessment of the pedestrian street network.

Such an assessment can be used for mapping campaigns, as support for automatically identifying issues in the pedestrian network and helping map-makers to focus only on problematic sections of the map. Combining this approach with other methodologies for identifying sidewalks, is a step forward in improving the overall quality of OSM and of pedestrian maps.

4. Results

The results of the study illustrate the usage of the methodology by showing how the refinement of the network reduces or increases the graph’s number of edges and nodes. The results are then discussed to explain why the graph increases or decreases in size.

As the key feature of the methodology is its generalization and global behaviour, multiple capital cities from around the world were processed. The number of nodes and edges from the resulting graphs of five capital cities are reported in table 1. The geometries of the areas of interest for each city were extracted from the GHS Urban Centre Database 2025: GHS-UCDB

R2024A, produced by the European Comission's Joint Research Centre (JRC) (Rivero et al., 2024). This dataset contains urban areas, rather than complete administrative boundaries, avoiding having areas of interest that are too big and empty, and focusing more on urban cores. The selected cities for extracting their street networks are Addis Ababa (Ethiopia), Amsterdam (Netherlands), Montevideo (Uruguay), Kampala (Uganda), and Ottawa (Canada). This selection features variability in continents, income levels, and map quality. Additionally, for reference, the networks for the same five cities were extracted using default OSMnx filters. Both the processed and reference street networks were downloaded in May of 2025. The change of the number of nodes and edges of the resulting graph with respect to the reference network is also reported in table 1. In the table, positive numbers indicate an increase in the number of nodes or edges, while negative numbers indicate a reduction.

City	Resulting Graphs	Change
Addis Ababa, Ethiopia	Walking: 75933 Nodes 106541 Edges Cycling: 74923 Nodes 198553 Edges Driving: 67191 Nodes 180511 Edges	 -0.7% Nodes -0.6% Edges -0.4% Nodes -0.3% Edges -0.4% Nodes -0.3% Edges
Amsterdam, Netherlands	Walking: 123110 Nodes 169317 Edges Cycling: 59572 Nodes 139988 Edges Driving: 25863 Nodes 60151 Edges	 11.8% Nodes 13.8% Edges -9.3% Nodes -9.9% Edges 0.2% Nodes 2.0% Edges
Montevideo, Uruguay	Walking: 24204 Nodes 37341 Edges Cycling: 19627 Nodes 52218 Edges Driving: 16075 Nodes 43767 Edges	 1.0% Nodes -0.3% Edges -0.8% Nodes -0.9% Edges -0.2% Nodes 0.1% Edges
Kampala, Uganda	Walking: 86840 Nodes 100720 Edges Cycling: 85998 Nodes 196126 Edges Driving: 53356 Nodes 123361 Edges	 -2.5% Nodes -1.6% Edges -0.5% Nodes 0.2% Edges -0.4% Nodes -0.1% Edges
Ottawa, Canada	Walking: 88290 Nodes 119958 Edges Cycling: 49675 Nodes 121995 Edges Driving: 10771 Nodes 27139 Edges	 -6.5% Nodes -15.4% Edges 2.2% Nodes 1.9% Edges -0.4% Nodes -0.1% Edges

Table 1. Number of nodes and edges of the walking, cycling, and driving street networks after filtering and post processing for the selected cities, along with the change in the number of nodes and edges with respect to the reference network downloaded with default OSMnx settings.

At first glance, a notable reduction is observed in the pedestrian network of Ottawa, Canada. After inspecting the original OSM data, it was determined that there were many sidewalks mapped separately from the main driving street, including residential streets, which is unusual in OSM even for highly developed cities. However, given the number of removed segments, we can conclude that not all of their respective main driving streets report the sidewalks as separate features. The reduction of nodes and edges in the pedestrian network of Ottawa was of 15.4% and 6.5%, respectively, effectively reducing the resulting graph size.

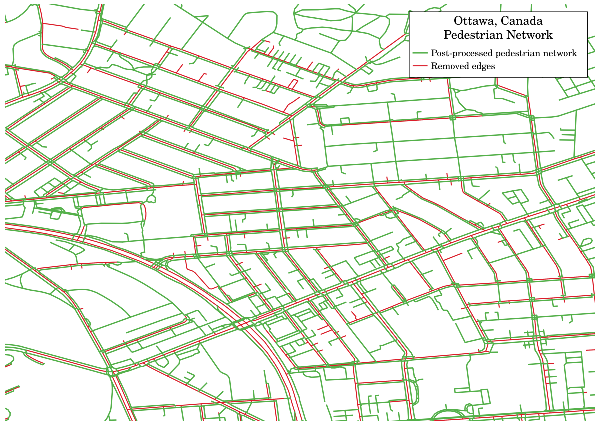

Figure 3. Portion of post-processed pedestrian street network (green) and the removed edges (red) from the original, raw network in the city of Ottawa, Canada.

Figure 3 illustrates the result of the post-processing for a portion of the city of Ottawa. In the figure, green edges represent the final network, while the red ones represent the deleted edges after filtering and post-processing. It can be observed that most of the deleted edges are located within two pedestrian sidewalks, which highlight the effect of the post-processing. After further assessment, we determined that more than 20.000 driving street segments were detected during post-processing having a separately mapped sidewalk but not reporting it as a separate feature.

However, the refinement of the network not only means that the resulting graph is a reduced subset of the raw data. For example, the resulting pedestrian street network of Amsterdam, Netherlands, shows an increase of 13.8% in the number of edges with respect to the reference. After inspection, it was observed that the increase was due to the addition of multiple cycling shared paths that are shared with pedestrians which were not initially considered in the reference network.

These two examples, showing cities in high-income economies and with high mapping quality, show how the methodology behaves in such kind of cities. Presenting reductions of more than 10% in some of the networks, it is a useful tool for constructing such networks with less redundancy. In addition, the resulting networks show paths that are actually located in the pedestrian infrastructure, specially when sidewalks are mapped separately.

On the other hand, the case for cities with less mapping quality shows less (or non-existing) reductions or expansions of the networks. According to the results, in the city of Addis Ababa, capital city of Ethiopia, the changes of the network with respect to the reference are less than 1% in all scenarios. As the edge count of the walking, cycling, and driving networks ranges from 106.541 to 198.553, a 0.5% change represents around 500 edges. This minor change is related to the fact that there are few separately mapped infrastructure for cycling and pedestrian

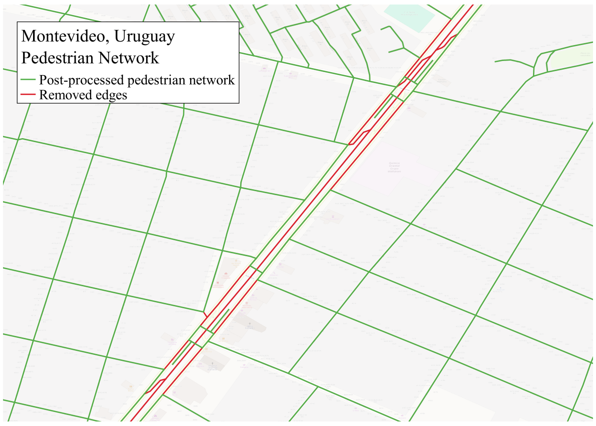

Figure 4. Portion of post-processed pedestrian street network (green) and the removed edges (red) from the original, raw network in the city of Montevideo, Uruguay. Note that, after the removal of edges, disconnections appear on the network.

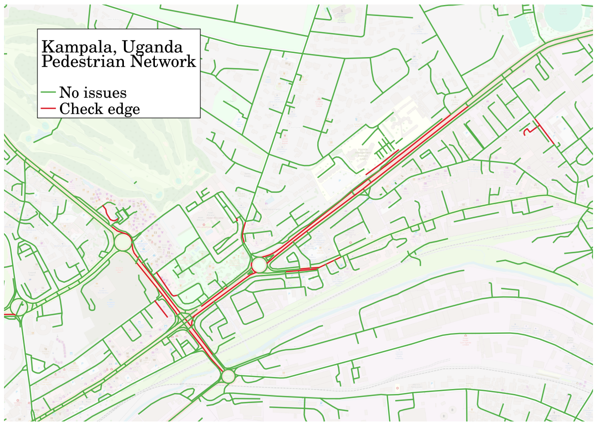

Figure 5. Portion of the pedestrian network of Kampala, Uganda after assessment. Red edges indicate that attention is required from mappers.

paths, meaning that most of the mobility is mapped by the driving network. Given the low amount of additional information provided by such networks, the most realistic scenario is to assume that most driving streets can be used by pedestrians or bikes. Another example of this can be seen in the city of Montevideo, Uruguay, where the maximum change in the number of nodes and edges is of 1%. In figure 4 it is shown how driving segments are removed where sidewalks are present.

When closely looking at the figure, a limitation of the post-processing algorithm arises. When the sidewalks and driving networks are incorrectly mapped, the graph may be disconnected due to the removal of driving street segments that serve as connections between pedestrian paths. This creates longer paths for reaching certain places and disconnections over the network.

Finally, to showcase the usage of the assessment functionality, figure 5 shows the city of Kampala, Uganda. The figure shows a zone of the city where there are many sidewalks separately mapped with respect to the main driving street. Each segment colored in red shows a driving street that is nearby a pedestrian sidewalk and has a similar orientation, and consequently, should specify that the sidewalk of the street is mapped separately using the tags "sidewalk:left", "sidewalk:right" or "sidewalk:both". However, edges in red do not specify such separation, and are marked for assessment. Such information is useful for pinpointing the specific place of the city portrayed in the figure as a location that requires attention for mapping campaigns. It can also highlight, for the case of Montevideo, places where improvements are necessary to maintain connectivity of the pedestrian network, addressing the previously mentioned limitation and supporting the improvement of the overall quality of OSM.

5. Conclusion and Future Work

The extraction of realistic and usable driving, cycling, and pedestrian street networks is a necessity for accurate urban analysis in multiple applications. Using OpenStreetMap (OSM), a global, crowdsourced, open-sourced database containing street networks and points of interest information, we proposed a methodology for the extraction of such networks, taking into account a set of filters and post-processing algorithms based on the network type and characteristics. Particular attention was paid to pedestrian networks, as an inconsistency of OSM is that sidewalks mapped separately from driving street segments do not specify the separation, as should be according to OSM mapping guidelines, thus, generating redundant information.

The methodology was tested by downloading and processing the driving, cycling, and pedestrian street networks of five capital cities from around the world, and comparing them with respect to networks downloaded using the OSMnx python library with default settings. The results show that cities with a higher level of mapping saw larger modifications, both for the removal and addition of information. This was due to the fact that these cities contain more pedestrian and/or cycling paths that are mapped separately from the main driving segments, generating inconsistencies. On the other side, cities with less mapping quality show less variability. This is consistent both for high-level mapping cities, as it allows to construct more accurate and specific navigation paths using only the separated sidewalks and cycleways, and for low-level mapping cities, as it is assumed that the driving network represents both the pedestrian and cycling networks.

An additional functionality of the post-processing algorithm for the refinement of the pedestrian street network is the network assessment or identification of inconsistencies. This is done by flagging, instead of removing, driving network segments that present inconsistencies, particularly when sidewalks are mapped separately from the main carriageway, but the separation is no specified on the driving street segment.

Future work includes the refining of the post-processing algorithm for pedestrian networks; adding more conditions based on local regulations and other cases of inconsistencies; as well as the improvement of the assessment functionality by adding more cases of inconsistencies such as network disconnections.

All the code and examples are available free and open sourced in a GitHub repository (Duque, 2025).

References

Bartzokas-Tsiompras, A., 2022. Utilizing OpenStreetMap data to measure and compare pedestrian street lengths in 992 cities around the world. *European Journal of Geography*, 13, 127-141. https://doi.org/10.48088/ejg.a.bar.13.2.127.138.

Boeing, G., 2019a. The Morphology and Circuity of Walkable and Drivable Street Networks. *Modeling and Simulation in Science, Engineering and Technology*, 271-287. https://doi.org/10.1007/978-3-030-12381-9$_1$2.

Boeing, G., 2019b. Urban spatial order: street network orientation, configuration, and entropy. *Applied Network Science*, 4, 67. https://appliednetsci.springeropen.com/articles/10.1007/s41109-019-0189-1.

Boeing, G., 2025. Modeling and Analyzing Urban Networks and Amenities With OSMnx. *Geographical Analysis*.

Buczyński, A., Chavez-Pacheco, A., 2025. *Quantifying Cycling Infrastructure Investment Needs Across Europe Using OpenStreetMap Data*. 154–159.

de Moraes Vestena, K., Camboim, S. P., dos Santos, D. R., 2023. OSM Sidewalkreator: A QGIS plugin for an automated drawing of sidewalk networks for OpenStreetMap. *European Journal of Geography*, 14, 66-84. https://doi.org/10.48088/ejg.k.ves.14.4.066.084.

Duque, J. P., 2025. Diuke/street-network-indices: First alpha release. https://doi.org/10.5281/zenodo.15498484.

Elias, E. N. N., de Oliveira Fernandes, V., Junior, M. J. A., Schmidt, M. A. R., 2020. The quality of OpenStreetMap in a large metropolis in northeast Brazil: Preliminary assessment of geospatial data for road axes. *Boletim de Ciencias Geodesicas*, 26, e2020012. https://doi.org/10.1590/S1982-21702020000300012.

Fan, H., Yang, B., Zipf, A., Rousell, A., 2016. A polygon-based approach for matching OpenStreetMap road networks with regional transit authority data. *International Journal of Geographical Information Science*, 30, 748-764. https://doi.org/10.1080/13658816.2015.1100732.

Franzini, M., Annovazzi-Lodi, L., Casella, V., 2020. Assessment of the Completeness of OpenStreetMap and Google Maps for the Province of Pavia (Italy). *GISTAM 2020 - Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management*, 270 - 277. https://doi.org/10.5220/0009564302700277.

Gaboardi, J., Fleischmann, M., 2025. uscuni/neatnet: v0.1.1. https://doi.org/10.5281/zenodo.15333394.

Hosseini, R., Lim, S., Tong, D., Sohn, G., Seyedabrishami, S., 2024. A specialized inclusive road dataset with elevation profiles for realistic pedestrian navigation using open geospatial data and deep learning. *Computers, Environment and Urban Systems*, 114, 102199. https://doi.org/10.1016/j.compenvurbsys.2024.102199.

Hosseini, R., Tong, D., Lim, S., Sun, Q. C., Sohn, G., Gidófalvi, G., Alimohammadi, A., Seyedabrishami, S., 2023. A Novel Method for Extracting and Analyzing the Geometry Properties of the Shortest Pedestrian Paths Focusing on Open Geospatial Data. *ISPRS International Journal of Geo-Information 2023, Vol. 12, Page 288*, 12, 288. https://doi.org/10.3390/IJGI12070288.

Jafari, A., Both, A., Singh, D., Gunn, L., Giles-Corti, B., 2022. Building the road network for city-scale active transport simulation models. *Simulation Modelling Practice and Theory*, 114, 102398. https://doi.org/10.1016/j.simpat.2021.102398.

Kaur, J., Singh, J., 2019. An automated approach for quality assessment of openstreetmap data. *2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018*, 707-712. https://doi.org/10.1109/GUCON.2018.8674899.

Klinkhardt, C., Kühnel, F., Heilig, M., Lautenbach, S., Wörle, T., Vortisch, P., Kuhnimhof, T., 2023. Quality Assessment of OpenStreetMap's Points of Interest with Large-Scale Real Data. *Transportation Research Record*, 2677, 661 - 674. https://doi.org/10.1177/03611981231169280.

Miloudi, S., Meguenni, B., 2023. Exploring the Potential of Machine and Deep Learning Models for OpenStreetMap Data Quality Assessment and Improvement. *Leibniz International Proceedings in Informatics, LIPIcs*, 277, 53. https://doi.org/10.4230/LIPIcs.GIScience.2023.53.

OpenStreetMap, W., 2024. Guidelines for pedestrian navigation — openstreetmap wiki,. [Online; accessed 22-May-2025].

OpenStreetMap, W., 2025a. Overpass api — openstreetmap wiki,. [Online; accessed 23-May-2025].

OpenStreetMap, W., 2025b. Simple indoor tagging — openstreetmap wiki,. [Online; accessed 22-May-2025].

Rivero, I. M., Melchiorri, M., Florio, P., Schiavina, M., Goch, K., 2024. Joint research centre data catalogue - ghs-ucdb r2024a - ghs urban centre database 2025 - european commission.

Simons, G., 2023. The cityseer Python package for pedestrian-scale network-based urban analysis. *Environment and Planning B: Urban Analytics and City Science*, 50, 1328-1344. https://doi.org/10.1177/23998083221133827.

Thottolil, R., Kumar, U., Mittal, Y., 2024. Quantitative assessment of urban road network hierarchy, topology, and walkable access using open-source gis tools. 243–258.

Vybornova, A., Cunha, T., Gühnemann, A., Szell, M., 2023. Automated Detection of Missing Links in Bicycle Networks. *Geographical Analysis*, 55, 239-267. https://doi.org/10.1111/gean.12324.

Čerba, O., 2025. OpenStreetMap as the Data Source for Territorial Innovation Potential Assessment. *ISPRS International Journal of Geo-Information*, 14, 127. https://doi.org/10.3390/ijgi14030127.