Master Thesis

submitted within the UNIGIS MSc programme

Interfaculty Department of Geoinformatics - Z_GIS

University of Salzburg

Crossing streets between intersections

Integration of non-dedicated mid-block crossings into a pedestrian network graph

by

Dipl.-Ing. Johannes Bouchain, BSc

12146351

A thesis submitted in partial fulfilment of the requirements of

the degree of

Master of Science – MSc

Advisor:

Dr. Martin Loidl

Hamburg, 11.03.2024

Contents

Summary

In this thesis, the effect of integrating non-dedicated mid-block crossings along residential roads on pedestrian route lengths is examined. This is motivated by the growing importance of walking in the socio-political debate and by previous work that emphasizes the need of comprehensive pedestrian network data. The study uses comprehensive test data from the OpenStreetMap project. Technically it relies on QGIS graphical modeling including comprehensive geometric and attributive expressions, and on R programming language. The design includes three steps: 1) Graph generation of the pedestrian networks, including perpendicular informal crossings, and with 45-degree informal crossings; 2) test routes generation using a newly developed method; 3) model for intersection crossing count indices as selected indicator check. The model is then applied to three test areas in Neukölln and Tempelhof districts, Berlin. The resulting length ratios and crossing counts show that for small distances, the routes are remarkably shorter with non-dedicated mid-block crossings included, when routes start next to a directly crossable street, but for longer routes the length reductions are only small. With 45-degree crossings, overall length reduction is generally higher, but still low for longer routes. Missing crossings at intersections, especially at T-shaped intersections, are an important factor that leads to irregularities in the results. In a concluding discussion and outlook, limitations of the selected approach are explored and the need for further research is identified.

Zusammenfassung

In dieser Arbeit wird untersucht, wie sich die Einbeziehung von freien Querungen entlang von Wohnstraßen auf die Länge von Fußgängerrouten auswirkt. Dies ist motiviert durch einen Bedeutungsgewinn des Fußverkehrs in der gesellschaftspolitischen Diskussion sowie durch vorhergehende Forschungsarbeiten, die den Bedarf an umfassenden Fußwege-Netzdaten hervorheben. Die Studie nutzt umfassende Testdaten des OpenStreetMap-Projekts. Die technische Umsetzung erfolgt anhand von grafischer Modellierung mit QGIS, einschließlich umfassender geometrischer und attributiver Ausdrücke, und der Programmiersprache R. Die Implementierung erfolgt in drei Schritten: 1) Graphengenerierung der Fußwege-Netze ohne freie Querungen, mit rechtwinkligen freien Querungen sowie mit kombinierten 45-Grad- und rechtwinkligen freien Querungen; 2) Generierung von Testrouten mit einer neu entwickelten Methode; 3) Modell für Kreuzungs-Querungs-Indizes als ausgewählte Indikatorprüfung. Das Modell wird dann auf drei Testgebiete in den Berliner Stadtteilen Neukölln und Tempelhof angewendet. Die ermittelten Längenverhältnisse und Querungszahlen zeigen, dass die Routen bei kleinen Entfernungen deutlich kürzer sind, wenn sie neben einer direkt überquerbaren Straße beginnen und freie Querungen möglich sind. 45-Grad-Querungen führen durchgehend zu einer etwas größeren Längenreduktion, die allerdings bei längeren Routen weiterhin relativ gering ist. Fehlende Querungen an Kreuzungen, insbesondere an T-Kreuzungen, sind ein wichtiger Faktor, der zu Unregelmäßigkeiten in den Ergebnissen führt. In einer abschließenden Diskussion und einem Ausblick werden die Grenzen des gewählten Ansatzes erörtert und der Bedarf an weiterer Forschung aufgezeigt.

Preface

Walking is an important means of transportation, not only with regards to sustainable and health-promoting mobility, but also with its role as the most natural way of human locomotion, being part of practically every route from A to B.

At the same time, walking, besides cycling, is the most important way of getting around in my personal life, as practically all my destinations, including several underground and suburban railway stations, are within walking distance. As an urban planner and geoinformatician-to-be, I have a habit of always observing urban space in a planning-analytical way. While discovering far more misconfigurations than good solutions in the urban pedestrian network[1], I asked myself: Why does a city like Hamburg, which wants to become a serious cycling city, often promote cycling at the expense of pedestrian infrastructure rather than at the cost of over-dominant car traffic? In my point of view, becoming as pedestrian-friendly as possible must be as important as promoting urban bicycle traffic against the backdrop of climate change and the urgently needed mobility transition. This is why I decided quite early to dedicate the final thesis of my geoinformatics studies to a relevant scientific question on the topic of pedestrian traffic.

While the geoinformatics discipline offers a large potential to directly address the urban planning level when it comes to sustainable mobility solutions, it seemed more appropriate to work on the analytic part of the pedestrian network. A detailed understanding of the current configuration of urban infrastructure dedicated to walking is a crucial basis for identifying the greatest need for improvement and developing suitable measures for a sustainable pedestrian network.

Many tests of routing engines led to unsatisfying results with regards to shortest and fastest options, especially for short distances at neighborhood level. Of course, this heavily depends on the data source used for routing. But certain aspects seemed to be common in every environment. One of them is the missing inclusion of unmarked or informal crossings at mid-block locations. In many cases, it is legally and practically possible—with regards for example to volume of motorized traffic, on-street parking, and street furniture—to cross a street in between formal crossings. Not including these free crossings seems to generally lead to large and unnecessary detours for pedestrians proposed by the routing engine. I started a research on this topic and found it relevant, with the results now presented here.

I want to thank my supervisor Dr. Martin Loidl for his support and inspirations for this work. Additionally I am thankful for numerous fruitful conversations during the time of this work with people who either have or do not have a geoinformatics background—colleagues, professionals, fellow students, friends and acquaintances. Last but not least, I would like to thank my family for supporting me throughout my intensive commitment to the part-time geoinformatics studies in Salzburg, including this thesis. The many walks with you, including various non-dedicated mid-block crossings, have always been a beautiful balance to studying, among many other things.

Hamburg, in March 2024
Johannes Bouchain

[1] In Hamburg and other cities, especially in the densely populated districts.

1 Introduction

In this chapter, the scientific motivation for this work is presented. Then follow a problem statement, solution approaches, and the research objective with the hypotheses for the study. Afterwards, the structure of the thesis is outlined.

1.1 Motivation

Walking is an original and common form of human locomotion with diverse characteristics, (Hillnhütter 2016 p. 9), depending heavily on physical conditions and preferences. Promoting walking has hugely positive effects (Dinu et al. 2019; Ek et al. 2021), e.g. in terms of mitigating climate change, improving urban quality of life, encouraging social interaction and achieving health benefits. The social cost of walking is low, especially when compared to motorized traffic (Gössling et al. 2019).

With walking as an important and serious means of transportation, digital solutions that support adequate pedestrian navigation and wayfinding are necessary. During a long period, digital routing systems were almost completely car-oriented, not taking into account route possibilities of pedestrians e.g. to walk outside of predefined lanes (Gaisbauer & Frank 2008). With the spread of portable mobile devices that allow for using routing software while walking, an increased focus on improved pedestrian wayfinding can be observed (Bauer et al. 2014). Recent research related to GIS (Geographic Information systems) and for pedestrian networks and routing often focuses on shortest-path options, but also includes quality-aware alternatives to the shortest option (Novack et al. 2018; Siriaraya et al. 2020), GIS-based methods to measure walkability (Fonseca et al. 2022), pedestrian accessibility and personalized pedestrian network profiles (Bolten & Caspi 2021), and also network generation based on pedestrian tracking data (Yang et al. 2020).

For walking, like for other transport modes, finding the shortest path within a given network is a fundamental requirement that such a model must take into account. The most common form of abstracting the existing real world network into a digital model is graphs that consist of edges representing connections between nodes.

The geometric representation of the network determines the quality and level of detail of the possible calculations and analyses using established graph-based routing algorithms such as Dijkstra (Dijksta 1959) or A* (Goyal et al. 2014). Thus, to allow for computing the optimal path as defined by the navigation or research task, the graph used for wayfinding or network analysis must provide an adequate number of nodes as well as all edges at the required sharpness of detail. Additionally, nodes and edges usually provide different attributes in terms of key-value pairs which can be included when identifying the optimal route.

Alternatives exist to find shortest paths via cost surfaces, i.e. the so called Least Cost Path algorithms, applied e.g. for infrastructure planning (Sari & Şen 2022) or using new bottom-up lane-based approaches developed for autonomous driving (Poggenhans et al. 2018; Poggenhans & Janosovits 2020). However, as these approaches require a specific data structure such as raster data or areal vector representations of the infrastructure and thus are hard to use in combination with common routing tools, they are not considered for this work.

Many common representations of street networks are based on a level of detail that targets motorized traffic with street space reduced to one linear geometry or eventually geometrically separated into directional roadways. Sidewalks are not present as separate geometries in such cases. Information on the existence, dimensions and types of sidewalks of the respective street section, if available at all, only exists as attributes of the street centerline. In such cases, common routing algorithms make no distinction as to whether the pedestrian is walking on the left or right-hand sidewalk. Destinations on both sides of the street are therefore always assumed to be equally accessible. This form is referred to below as the attributive solution.

Another form of representing the pedestrian network is using separate geometries for sidewalks and crossings, resulting in a division of the street space into several roughly parallel edges, connected at given positions by streets crossings as approximately perpendicular lines. Compared to the attributive approach, the geometry-based variant is better suited for taking into account street-side accessibility (Rhoads et al. 2023) when using conventional graph-based routing algorithms. This form is referred to below as the geometric solution for comprehensive pedestrian routing graphs.

Street space then is treated as non-accessible and non-crossable, like any other space enclosed by network edges and the smallest enclosing rings consist of several sidewalk and crossing edges in such cases. For low-traffic streets, this pedestrian inaccessibility of the street space between formal crossings and the two opposite sidewalks prevents realistic wayfinding, especially for small-scale pedestrian routes. While walking from A to B, pedestrians might want to cross streets between formal crossing locations, e.g. to shorten the route, to avoid traffic lights and thus waiting time at intersections or to switch to the opposite sidewalk because it seems more comfortable for walking. Such informal crossings usually are possible where the volume of vehicle traffic along the roadway to be crossed is low enough and the spatial configuration of the respective street section enables switching to the opposite side of the street, which must be possible in a safe manner (Naumann et al. 2019).

On the other hand, street space cannot be treated as completely open space. Squares and open green spaces can often even legally be traversed by pedestrians in any possible direction and position (Graser 2016; Hahmann et al. 2018), taking into account the specific abilities of pedestrians to move away from predefined lines. Crossing streets informally at mid-block positions, however, is restricted regarding direction and position for several reasons and requires special consideration.

OpenStreetMap[2] (OSM) is the most popular crowdsourced geospatial database e.g. for streets and paths. A wide range of research works nowadays use and analyze data from the OSM project (Wu et al. 2021), also specifically for pedestrian network data (Bartzokas-Tsiompras 2022) or pedestrian-oriented routing approaches (Rousell & Zipf 2017), and examine the completeness of OSM sidewalk data (Mobasheri et al. 2015; Omar et al. 2022). Prior to these works, different studies focused on the high quality or at least the high-quality potential of Volunteered Geo-

[2] https://wiki.openstreetmap.org/wiki/Guidelines_for_pedestrian_navigation#Sidewalks_and_crossings, last accessed: 2024-01-06.

graphical Information (VGI) from the OSM project from a more general point of view (Corcoran et al. 2013; Goodchild & Li 2012; Koukoletsos et al. 2011). The number of scientifically sound studies that use and analyze OSM data and the large number of often official applications based on OSM are an indicator of the quality generally attributed to this project.

The OSM data structure provides a mixture of both the attributive and the geometric approach for providing detailed sidewalk data. A common mapping standard here is the attributive integration of sidewalk data for smaller roads, like e.g. residential streets, and the provision of separate sidewalk and crossing geometries for larger roads. Using OSM data for pedestrian wayfinding thus often allows for street-side specific accessibility modeling along larger roads, when this can be considered an important information and often critical regarding path length and realistic representation. Along smaller roads, however, street-side specific accessibility is then not taken into account, which often is an adequate representation, as many smaller streets allow for informal crossings at many different positions. The question if an attributive integration of sidewalk parameters to the street centerline or separate sidewalk geometries is the favorable approach in OSM is subject to intense ongoing discussion within the community and will not be treated here.

However, when defining a network with complete separate sidewalk geometries as a goal for OSM data, the question arises as to whether non-dedicated mid-block crossings can be integrated into the applied mechanisms for wayfinding, walking analysis and walking simulation, how and under which circumstances this integration can or should be done and what the effect of their integration into such a detailed network graph is in terms of path length and structure.

While routing through open spaces and sidewalk graph analysis and creation have been subject within recent research projects, street space itself as a special type of partly open space has rarely been taken into account in a detailed, GIS-related form. The motivation behind this work is to make an important contribution to filling this gap.

1.2 Problem statement and solution approaches

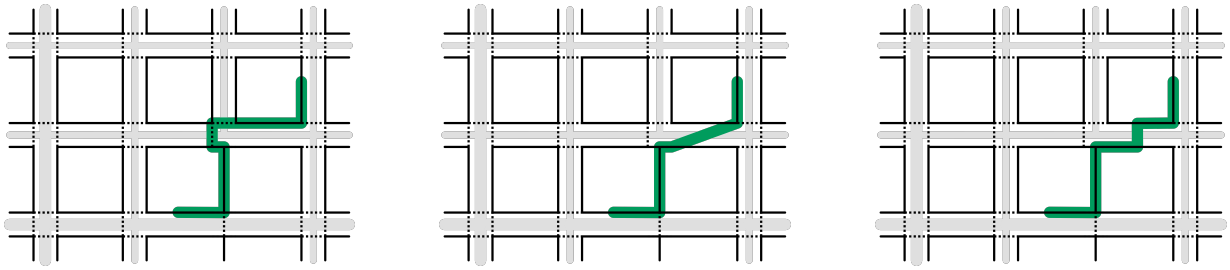
The type of pedestrian network representation within digital routing graphs determines how street space is modeled, and how it can be treated during routing computation. An attribute-based approach with sidewalk data added to the street centerline ignores the existence of street space. Destinations at both street sides are considered to be equally accessible from any point along the street. The travel cost for any crossing along such street sections is zero. On the other hand, network representations with separate geometries for sidewalks and formal crossings result in a travel cost for non-dedicated mid-block crossings that is infinitely high as they are not possible when only considering existing edges of the network.

Based on the assumption that a routing graph with separate sidewalk geometries is in many respects more reasonable than the attribute-based counterpart, the question arises as to how possible non-dedicated mid-block crossings can be sensibly handled within such a structure.

While this question generally covers a wide range of legal, technical and spatial aspects, simplifications and assumptions have been made for this work. The proposed solution comprises

automated generation of an extended graph that includes additional edges at a given interval and angle, representing an approximation to actual non-dedicated mid-block crossing positions between intersections and other formal crossing locations, allowing for the usage of common graph-based routing algorithms. Within the model, all residential roads, regardless of the actual spatial and functional configuration, are provided with such additional crossing edges. It is assumed that with this approach the effect of such an extended graph on shortest-path tasks can be measured in a sufficiently realistic way for the essential aspects that are relevant for answering the research question.

Figure 1.1 illustrates the motivation of this work.



(a) Usual routing based on sidewalks and formal crossings

(b) Ideal shortcut, crossing street space diagonally

(c) Legally correct mid-block crossing

Figure 1.1: Schematic street network with example routes illustrating the motivation of this study

1.3 Research objective and hypotheses

The research objective of this work is to determine the length reduction effect of pedestrian routes including non-dedicated mid-block crossings in relation to the route without such crossings included, depending on different starting points and urban road network configurations.

This research goal is to be validated on the one hand based on legal crossing behavior, including only perpendicular mid-block crossings. On the other hand, it is assumed that actual human walking behavior includes more "natural" but legally incorrect diagonal crossings, so that the research objective must also be validated against a network graph including this type of non-dedicated mid-block crossings.

In addition to the analysis of the path length differences, crossing numbers are analyzed, as it can be assumed that the number of street crossings for routes including non-dedicated mid-block crossings is higher than for the original routes, especially with path lengths as the only cost factor.

In order to partly explain expected differences in the results depending on the network type tested, an example indicator is developed out of several network parameters that are likely to be relevant.

Based on the path length comparison of original routes and the variants including non-dedicated mid-block crossings, the following hypotheses are formulated as a starting point for this work:

- The highest average length reductions occur for short routes with approximately typical street block lengths, when routes start at mid-block positions next to a freely crossable street section.
- Length reductions are generally higher with diagonal non-dedicated mid-block crossings included compared to routes only containing such crossings in perpendicular form.
- With path length as the only cost factor for this study, the selection of crossing positions along freely crossable road sections is quasi-random, as there are only minor length differences between neighboring crossing edges when sidewalks run approximately parallel. For perpendicular crossings, dedicated crossings from the original graph are included into this quasi-random selection.
- In many cases, the number of street crossings per route increases when non-dedicated mid-block crossings are included, especially with integrated diagonal crossing directions.
- Irregularities with regards to network configuration, for example unusually large distances between dedicated crossings at certain points, lead to noticeable length reductions even for long walking distances.

In the context of this work, high length reductions are those that save more than 10% compared to the original route lengths.

This work focuses on a new methodological approach to test routes generation and route comparison and proposes different analysis steps with regards to starting points, route lengths, and network structure. A full statistical analysis including t-tests for significance is not part of this study.

1.4 Structure of this work

Chapter 2 introduces basic concepts and approaches from previous works that are relevant for understanding the background of this work. Chapter 3 describes the design and implementation of the study, including model design assumptions, applied methods, study area selection, and the actual implementation. In chapter 4, the results are presented and analyzed. In chapter 5 the previously analyzed results are evaluated and discussed, including a reference back to the research objective and the working hypotheses. This also includes an outlook with open questions for future work, followed by a conclusion in chapter 6.

2 Fundamentals of realistic pedestrian routing

This work is intended to contribute to realistic and comprehensive pedestrian routing, taking into account the specific capabilities and constraints of pedestrians when walking from A to B. The underlying network graph for such comprehensive routing solutions should include all formal and informal features that pedestrians are allowed to use within their itineraries.

These different types of pedestrian network elements that are necessary for comprehensive routing solutions have been subject to scientific research that is referred to within this chapter, revealing the starting point of this study. Starting with a general look on walkable space and its extraction using GIS-based approaches, current approaches related to the integration of sidewalks and street crossings as separate geometries for a comprehensive network graph are examined, followed by proposed algorithms to integrate open spaces. In the next step, recent research related to analysis and graph integration is examined. The OSM project and its suitability for comprehensive pedestrian routing is then analyzed using a review of recent research. Finally, the importance of the shortest routes for pedestrians and methods for selecting test routes are presented based on selected research works.

2.1 Walkable space

Digital solutions for pedestrian navigation must take into account the special abilities, but also the specific constraints of people who walk. These differ from those of car drivers, which have long been the benchmark for navigation solutions.

A growing interest in pedestrian-friendly routing solutions can be observed since the second half of the 2000s. Derived from the cognitive wayfinding process of pedestrians and taking into account the early and still valid considerations of Kevin Lynch (Lynch 1960) in this regard, Gaisbauer & Frank (2008) propose a model for pedestrian routing that relies on all components of walkable space. Due to the unavailability of comprehensive linear pedestrian network data at the time, their approach builds upon skeletonizing methods, i.e. algorithms for reasonable centerline extraction, of high-resolution raster images (Walter et al. 2006) and vector polygon data (Elias 2007). Walkable space can thus be derived directly from these sources or by subtracting blocked areas such as buildings and traffic lanes. Gaisbauer & Frank (2008) also highlight that with this model, street space remains fully blocked and that street crossings must be added to this type of skeleton model. While they propose a method to solve this for crossings at intersections by connecting the skeleton edges of opposite sidewalk corners, the question of how to detect and integrate dedicated street crossings at mid-block positions still remains unanswered. Also, remaining pedestrian space also includes larger areal features, within which possible pedestrian routes can only be modeled very inadequately and unrealistically using the skeleton method. However, it is already largely clear from this work which formal elements must be integrated into a comprehensive graph for pedestrian navigation.

2.2 Integrating sidewalks and formal street crossings

Sidewalks and street crossings are important formal and linear pedestrian network elements and their presence as separate geometries is crucial for comprehensive and realistic routing solutions.

To integrate sidewalks and street crossings as separate geometries, either comprehensive network data sources containing sidewalks as separate edges or adequate graph generation techniques are necessary. As the availability of sidewalk data was a major issue for a long time—and often still is today—, several previous works focused on sidewalk graph generation techniques. In addition to remote sensing as an obvious source, Karimi & Kasemsuppakorn (2013) find network buffering as a suitable and fast method. Additionally, they examine collaborative mapping based on GPS (*Global Positioning System*) traces of a large number of participants and image processing. All three techniques are found suitable for integrating sidewalks, except that no bridges and tunnels can be detected using GPS traces and image processing techniques may encounter performance problems in the case of shadows and dense features.

Especially the GPS approach seems promising, given the fact that nowadays pedestrians usually carry smartphones suitable for tracking. Mobasheri et al. (2018) present an approach to add sidewalks based on GPS traces of wheelchair users. However, complex algorithms and resource-intensive processing are necessary to create the most suitable sidewalk centerline from these data.

Another approach to integrate sidewalks is to use graph generation techniques relying on geometric or attribute-based rules to derive the necessary edges from existing features, usually from street centerline representations. However, more in-depth approaches are necessary to also take varying widths of sidewalks and roadway-sidewalk distances into account. Li et al. (2018) propose a semi-automated method to generate sidewalk data derived on the one hand from street centerline data and on the other hand from parcel polygons, resulting in sidewalk edges that do not have a uniform distance to roadways, but take varying widths of public street space into account.

Rhoads et al. (2023) provide an in-depth review of digital sidewalk networks. According to their study, sidewalk centerlines are directly available only for few cities and do not always contain attributive information on sidewalk characteristics. They find curb lines as the most frequently available source for sidewalk centerline derivation, but these datasets do not always have a corresponding sidewalk and do not provide any sidewalk attributes. Sidewalk polygon data as another alternative implicitly includes attribute data such as area and width, but often contains digitization errors and requires significant processing for sidewalk line generation.

Recent research suggests that extracting pedestrian network data from its polygon representation is still a valuable approach for specific scenarios. Valls & Clua (2023) propose a high-resolution method to skeletonize sidewalk polygons using an approach based on Voronoi diagrams.

However, Rhoads et al. (2023) additionally highlight that research on pedestrian networks often relies on constructing and inserting the necessary geometries and associated attributive data by hand for the selected study area at a high cost, taking important resources away from the actual core of research.

Machine learning techniques may help to partly overcome this issue and recent research has examined related approaches. Wanyan et al. (2023) introduce an approach for network generation from remote sensing imagery using self-supervised learning, highlighting problems such as

expensive and time-consuming annotation and obscuration of target detection objects. Verma & Ukkusuri (2023) used Google Satellite imagery from different North-American cities to detect street crossings using a deep learning model based on object detection.

It can be concluded that still today, more or less automatized and expensive graph generation techniques are required if a specific urban area is to be used as study area for comprehensive pedestrian routing including separate sidewalks and street crossings. On the other hand, only few urban areas can be selected as data source for comprehensive pedestrian network graphs directly suitable for routing.

2.3 Integrating open spaces

Besides mainly linear pedestrian structures, there are areal pedestrian network features such as squares and other open spaces that pedestrians are allowed to cross freely. The integration of these features into a graph for realistic pedestrian routing is more challenging, as usually the itineraries representing the numerous possibilities of traversing such areas in a natural way initially are not present within the underlying network graph.

Besides the need for customized pedestrian routing solutions and pedestrian-specific treatment of linear network features, Bauer et al. (2014) provide an early approach for the integration of open spaces into a network graph. They use a raycasting algorithm to detect if all parts of direct connections between nodes of a polygon representing a freely traversable square are actually located within the polygon. They then add the fully included lines as additional shortest-path routing edges to the respective open space.

As an alternative but at the same time an extension of this work, Dzafic et al. (2015) propose an algorithm based on a spiderweb graph, i.e. raster lines with added diagonal lines and common nodes at all line intersections. They conclude that the spiderweb graph already leads to quite suitable results, but combining their approach with that of Bauer et al. (2014) leads to even better paths via open spaces. Andreev et al. (2015) extended previously elaborated algorithms to include cases where the start or end point of the route is located at arbitrary positions within the square.

Graser (2016) introduces other algorithms such as skeleton, grid and visibility graph. Additionally, she examines least-cost path solutions based on polygon features representing a cost surface with different values. With this latter solution rejected as it does not result in a routable vector-based network graph, she finds the visibility graph the most suitable solution in terms of implementation complexity and graph properties.

Hahmann et al. (2018) reviewed a large number of different algorithms of prior works that focused on pedestrian routing through open spaces, including two different versions of visibility graph: one that connects all polygon nodes to all others, and one that only connects nodes that themselves are connected to adjacent network edges and not only to the open space polygon. They find the most suitable model to be located in between these two poles of visibility graph complexity. However, a visibility graph is not suitable for the inclusion of arbitrary start or end points of the route located at the respective open space polygon, which actually also counts

for arbitrary positions at the polygon edges, normally the closest points for surrounding street addresses.

Obviously, street space that can be traversed with low risk and generally at low cost cannot be converted into edges of a routing graph using the same algorithms as for fully open spaces like squares. Legal and physical restrictions reduce allowed and practically feasible directions and positions to cross streets at non-dedicated mid-block positions. However, the integration of the additional edges into a routing graph is almost equally challenging. Statically adding all relevant nodes and edges to the routing graph for a large area is a critical performance issue. On the other hand, post-processing of routes based on the original graph to include additional edges only where necessary and to generate a corresponding corrected route would require approaches for graph-based routing that differ in many respects from the routing services commonly used today. This is an obvious reason why current routing services do not yet include free crossings of squares[3] and streets.

2.4 Integrating non-dedicated crossings

Besides formal linear network features and independent itineraries across freely walkable open spaces, street crossings at non-dedicated positions must be integrated for a comprehensive and realistic pedestrian routing approach. Generally, there are some similarities here to open-space routing approaches in terms of technical implementation options. However, it is obvious that street space is not as freely crossable as squares or other open areas and subject to legal and physical restrictions.

Streets may be crossed informally at intersections where formal street crossings are missing across the respective merging street or at arbitrary mid-block positions. Informal crossings at intersections have been examined as part of a comprehensive pedestrian network typology by Cambra et al. (2019). They propose to add such crossings to the routing graph at low-traffic single lane streets. However, their use-case does not address situations where even informal crossings are not possible because they are blocked, for example by parked cars. Also, the distinction between formal and informal crossings at an intersection often is not obvious, because in many cases crossings at intersections are considered formal only because they are the logical and straight extension of sidewalks via merging streets. Thus, such crossings must be considered apart from non-dedicated mid-block crossings and require further dedicated research.

Non-dedicated mid-block crossings have generally been subject to scientific investigation since the early 2000s (Baltes & Chu 2002), but the focus is on major urban roads and on behavioral issues here, which is also the case for recent works (Govinda & Ravishankar 2023; Guío-Burgos et al. 2022). The major part of theses studies has been conducted in Asian or Latin-American cities, with only few studies investigating mid-block crossing behavior in European cities (Lassarre et al. 2012; Yannis et al. 2013).

[3]Hahmann et al. (2018) mention that a solution for routing through open spaces had recently been integrated into the OpenRouteService routing application. However, this could not be confirmed by testing selected open spaces in Heidelberg, Germany, while working on this study (https://maps.openrouteservice.org/, last accessed: 2024-03-01).

Crossing behavior at non-dedicated mid-block positions specifically on low-traffic residential streets has rarely been addressed as a specific issue. This is obvious, because due to lower vehicular speeds and lower traffic volumes, pedestrian safety and risk-taking behavior as the triggering factors in many studies are much less critical here. Research results exist solely as a by-product of comprehensive studies on pedestrian crossing behavior. Abdullah et al. (2022) conducted an online survey on the behavior of pedestrians at unmarked mid-block locations. 79% of the 220 respondents stated that they cross at unmarked mid-block locations along local/residential roads sometimes to always. About two thirds of the participants belonged to developing countries, which indicates that the validity of the numbers for Western European cities is severely limited.

The previously presented studies usually are not intended to improve pedestrian routing. Such an approach, however, has been introduced by Naumann et al. (2015), who developed a safety index for road crossings. Their findings indicate that marked crossings close to intersections have the highest safety values compared to other crossing types, followed by non-dedicated mid-block crossings along minor roads with low traffic volume. Building on this work, they propose the integration of perpendicular non-dedicated mid-block crossings at 10 m intervals (Naumann 2018; Naumann et al. 2019; Naumann & Kovalyov 2017). They highlight the importance of such an integration for the outskirts due to the high density of formal crossings in central urban areas. However, they point out that a large number of additional nodes and edges are generated in this way, with negative effects on the processing time.

This thesis, in contrast, is based on the assumption that in many cases it is reasonable to integrate non-dedicated mid-block crossings even in high-density neighborhoods.

2.5 Using OSM for realistic pedestrian routing

In recent years, the suitability of OSM data for pedestrian navigation and wayfinding has been examined in detail. Due to still missing comprehensive data including separately mapped sidewalk geometries at the time, several works highlight approaches of generating missing geometric sidewalk data. Naumann & Kovalyov (2017) find that sidewalk-oriented routing based on street centerline geometries with pedestrian-relevant attributes is hard to achieve. Thus, they propose to use the relevant attributes to generate sidewalk geometries from OSM road geometries based on a standardized geometric approach. The suggested workflow contains the construction of sidewalk edges, the connection of these separate edges, and the addition of dedicated road crossings. As a last step, they even provide for the integration of non-dedicated road crossings (see section 2.4), which is a profound difference to comparable works. However, such an approach neglects the irregularities that usually exist within a sidewalk network such as varying sidewalk-roadway distances. Also, their approach does take into account that already at the time separately mapped sidewalks existed for several roads in well-mapped areas, especially for larger roads.

Approaches to generate sidewalks from street centerline data requires comprehensive attributive data on which street section actually provide left-hand, right-hand, or on both sides. Mobasheri et al. (2015) examined attributive sidewalk information in detail and found that at the time,

only 5.6% of streets in Berlin provided a `sidewalk` tag. However, already two years later the number had been grown to 8.2%, according to a subsequent study that compared OSM sidewalk data completeness in several larger German cities (Mobasheri et al. 2017). This suggests that the amount of suitable sidewalk information within the OSM database is rapidly growing.

With the continued growth of the amount of sidewalk information in OSM and the emerged suitable and comprehensive sidewalk scheme, research shifted to explicitly examine the existence of separately mapped sidewalk geometries. The OpenSidewalks project launched in 2016 (Bolten et al. 2017) made an important contribution to the expansion of the OSM scheme for comprehensive sidewalk mapping. It also proposed effective methods for transferring municipal sidewalk data to the OSM database. Recently, a number of local mapping communities began mapping sidewalks separately for all street types in their city. For example, Cambra et al. (2019) mention the efforts of the Warsaw OSM community to complete geometric sidewalk data.

Omar et al. (2022) analyzed sidewalk completeness of Chicago, Seattle and New York City, focusing on both attributive and geometric sidewalk data. They showed that large areas of the cities studied already provided complete geometric sidewalk data in OSM. Their work is also an important contribution to analyzing the trustworthiness of OSM sidewalk data, as they went beyond simply analyzing coverage.

However, in-depth research to analyze the tagging completeness of sidewalk geometries, street crossing geometries and street centerlines is still pending, given the fact that comprehensive routing must include sidewalks and street crossings not only based on geometry, but as weigthed and filtered edges that depend on numerous attributes. Bartzokas-Tsiompras (2022) shows the different levels of attributive sidewalk completeness for selected cases as part of a study that has a different focus, but still as an indicator for pending in-depth examination. Additionally to the attributive completeness of sidewalk geometries, street centerlines must include the information that the respective street section provides separately mapped sidewalk geometries and that routing services must exclude the street centerline from the pedestrian network graph. Dedicated research for this part of data completeness is necessary.

Detailed and realistic pedestrian routing not only depends and the provision of actual roadway, sidewalk and street crossing information, but also on external elements as part of urban street space. Seidel (2022a) presented the "Straßenraumkarte Neukölln" (street-space map of Neukölln, Berlin) as a comprehensive approach to integrating street-space elements in detail. The map emerged due to the exceptional efforts of members of the *Verkehrswende* (mobility transition) OSM group in Berlin, founded in 2019, to record on-street parking in detail, with the approach quickly extended to cover all relevant street-space features. The map itself was published in 2022 and has gained attention in the FOSSGIS (Lingner et al. 2023; Seidel 2022b) and OSM communities. It can be concluded, with subject to analytical verification, that this level of detail is indeed an exception within the current state of the OSM map and a promising test case to use a ready-made pedestrian network graph that includes all formal pedestrian features.

Another aspect that is relevant here is the quality assessment of OSM data. As the database is the product of mapping efforts of millions of people[4] around the globe, it is also subject to inconsistencies and a variety of mapping errors. Extrinsic quality assessment using ground truthing methods or quality-proven authoritative data is expensive. But internal assessment methods should be applied, for example when checking potential test data for consistency. Recent works have often focused on the road network for quality assessment, as it is the most commonly used part of the OSM database, but the proposed methods are generally also transferable to the pedestrian network. Alghanim et al. (2021) propose a comprehensive intrinsic quality assessment approach for OSM road data including machine-learning techniques. Zhang et al. (2021) present an approach for parallel roadway detection using e.g. the rules of symmetry and continuity that appears to be transferable for detecting inconsistencies of roadways and parallel sidewalks. Additionally, approaches to automatically correct missing network data, for example by closing obvious gaps, have been proposed, e.g. for road data (Funke et al. 2015).

However, all these approaches require complex algorithms or tools that are not always commonly available and simple built-in network analysis tools of the software used may be sufficient to check for the suitability of selected OSM pedestrian network data for comprehensive and realistic pedestrian routing.

2.6 Shortest paths for pedestrians

It is obvious that especially for pedestrians, the shortest route between A and B does not necessarily represent the preferred, i.e. the most realistic one. In previous work, this has often been highlighted, also quite recently by Koritsoglou et al. (2022). They emphasize that pedestrian navigation systems must provide several alternative routes, usually selected using a k-shortest path algorithm or by adding penalties to network edges. In pedestrian navigation applications, these are often customizable. Hosseini et al. (2023) also highlight that shortest paths might not be selected in order to walk more accessible, convenient or comfortable routes. They present a method to analyze the geometric properties of shortest pedestrian paths based on similarity, route curviness, road turns and intersections, and road gradients.

This shows that selecting realistic routes for pedestrians within a comprehensive network graph goes far beyond the simple shortest route approach. However, shortest route studies are a widely used approach in research and a legitimate starting point, also because often the shortest path is actually desired. Detecting the shortest route option is supposed to be especially relevant when integrating informal route segments, because it is obvious that these are often selected with the aim of shortening the route in relation to the formal alternative, at least in the section concerned.

[4]According the OSM statistics page, there are currently over ten million registered users with 1.75 million of them having contributed at least once (data from 2021) (https://wiki.openstreetmap.org/wiki/Stats last accessed: 2024-03-10).

2.7 Selection of test routes

The intention of this work is not only to examine the effects of non-dedicated mid-block crossings on pedestrian route lengths, but also to examine these based on a large set of test routes within strategically selected parts of a larger network. The question arises as to whether a suitable method already has been developed in previous work for comparing an original network to the same network with additional nodes and edges at certain positions.

Network length comparison is often performed with the goal of checking the test network for completeness relative to a selected reference network. Graser et al. (2014) give a comprehensive overview on network comparison methods. Besides comparing the full network lengths, they highlight network cells comparison as applied by Haklay (2010). Such methods have been applied by other researchers, often to measure OSM network completeness compared to an official reference graph, such as GIP[5] in Austria, Ordnance Survey data in the United Kingdom, or TIGER[6] data in the United States. Ahmed et al. (2015) conducted a network comparison of OSM and Tele Atlas[7] data with the extended goal to not only compare path lengths, but also similarities of shapes and similarities of connectivity.

Zielstra & Hochmair (2012) conducted a comparison of pedestrian networks by generating 1000 origin-destination point pairs, with the resulting route lengths according to typical walking distances. This approach has many advantages as it is fully random and thus statistically valid, but does not allow to statistically test for network characteristics around a given start point.

All presented network comparison approaches have the goal to compare data from different sources, with resulting paths that may be quite similar, but already differ at their start and end points depending on the ground truthing accuracy of the selected data set.

Another approach is single route comparison, the most common method used to test the open space routing algorithms mentioned in section 2.3. Start and destination points are strategically selected, so that the resulting routes contain sections that are relevant for comparison, e.g. a square. The compared graphs are not two different data sources, but an original graph and an extended graph, matching the original one except for additional nodes and edges included using a selected algorithm. The compared routes may fully match, but they can also differ partly or completely except for start and end points, depending of the cost factors selected and the geometric properties of the additional nodes and edges.

Novack et al. (2018) use an almost similar approach for single pedestrian routes, with varying cost factors in order to generate routes with different qualities such as green, social, or quiet routes. This allows for in in-depth comparison of single routes. However, the work did not

[5] Graphenintegrations-Plattform, a standardized authoritative reference graph for the Austrian transportation network.

[6] *T*opologically *I*ntegrated *G*eographic *E*ncoding and *R*eferencing system, a public domain data source provided by the United States Census Bureau.

[7] Tele Atlas is a Netherlands-based digital mapping company, today a subsidiary of TomTom, a manufacturer for automotive navigation systems.

incorporate a way to statistically check a large set of different routes with regards to the relevant properties.

In relation to real-life situations, such as a single address that is an everyday start point for routes in different directions, a combined methodological approach of strategical selecting only the start point, but generating many more or less random routes from this start point seems valuable, especially for this study. Route lengths can be either random with e.g. given minimum and maximum values or an approximate range, or they can be set to one fixed value. A comparable approach has not been applied within the previous work examined for this study.

3 Study design and implementation

In this section, the design and implementation of the study to be conducted as core element of this research work is introduced. The level of abstraction with the associated assumptions is presented, followed by study area data requirements and selection. In a next step, the technical approach is described, followed by a presentation of the actual set-up for conducting the study.

3.1 Model design assumptions

This study is limited to analyzing route lengths and the number of road crossings of the shortest possible pedestrian routes, not taking into account personal preferences and environmental constraints that would lead to choosing other path options in reality. In addition, the simplifying assumption is made that all and exclusively residential streets with sidewalks on both sides are continuously crossable. With these assumptions, the aim is to provide initial components for a future comprehensive investigation of non-dedicated mid-block crossings. The following sub-sections explain these choices and other assumptions that were made when designing this study.

3.1.1 Perpendicular and 45-degree crossing directions

According to traffic regulations in Germany[8] it is mandatory to cross streets in the shortest possible form perpendicular to the roadway direction, while e.g. the Austrian[9] regulation is less specific, only prescribing to cross in a straight line. With the German regulation as reference, only the perpendicular direction is legal option for crossing streets at non-dedicated positions. However, observations and a sensible approach suggest that streets are often crossed diagonally in a form that reduces turning angles and allows for route shortening. The inclusion of diagonal crossing options into a network graph is especially important when the task is to predict or analyze actual trajectories and not to propose routes the can be used in a completely legal manner.

With this in mind, two types of non-dedicated mid-block crossings are examined: exclusively perpendicular crossings and a combination of perpendicular and 45-degree crossings, with the latter integrated to the extended network graph in both directions. This offers the possibility to not only compare the original routes to one version of modified routes, but also to compare the two different types of modification.

[8] "Wer zu Fuß geht, hat Fahrbahnen unter Beachtung des Fahrzeugverkehrs zügig auf dem kürzesten Weg quer zur Fahrtrichtung zu überschreiten. […]" (StVO Straßenverkehrs-Ordnung vom 6. März 2013, zuletzt geändert 28. August 2023, §25 (3)). Online available at https://www.gesetze-im-internet.de/stvo_2013/BJNR036710013.html, last accessed: 2024-03-09. Translated version: *Pedestrians must cross roadways quickly by the shortest route at right angles to the direction of travel, taking into account vehicle traffic.*

[9] "An Stellen, wo der Verkehr weder durch Arm- noch durch Lichtzeichen geregelt wird, dürfen Fußgänger die Fahrbahn unter Bedachtnahme auf das Verkehrsaufkommen auf geradem Weg überqueren. […]" (Gesamte Rechtsvorschrift für Straßenverkehrsordnung 1960, Fassung vom 10.03.2024, § 76 (4) (a)). Online available at https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10011336, last accessed: 2024-03-10. Translated version: *In places where traffic is not controlled by arm or light signals, pedestrians may cross the road in a straight line, taking into account the volume of traffic.*

3.1.2 Continuous crossability of residential streets

Traffic speed, traffic volumes and environmental settings considerably vary even within the same road category. Even small streets might be uncrossable due to physical constraints of the built environment, while larger roads might on the other hand be crossable when they do not provide such constraints and traffic volume is low. However, it can be assumed that most residential streets can be crossed safely at many positions and the number of larger urban roads suitable for mid-block crossing is quite low.

With regards to data availability and research scope, the integration of non-dedicated mid-block crossings only for residential streets is a sensible compromise. It is accepted here that this overstates the realistically possible crossings within residential streets, especially for densely populated areas e.g. with continuous street-side parking, while at the same time crossing options of some higher-level roads are understated.

When crossing streets between intersections, pedestrians must choose an appropriate position to change to the opposite sidewalk. In a realistic setting, this choice is determined by numerous factors such as distance to formal crossings, street furniture, trees, parked cars, curb height, traffic volume and observation by other passers-by. A digital representation of these factors would require a data set that is much more granular than the highest level of detail currently available in the OSM database or would require the inclusion of additional data sources. The safety index for road crossings, as presented in previous research, would have to be included and extended to a comprehensive feasibility index.

Thus, for this work the simplified assumption is made that residential streets are equally crossable at all mid-block positions, while at the same time a short uncrossable section next to formal crossings is preserved[10]. With this unweighted approach, a virtually random selection of crossing positions is expected, determined only by the lengths of the respective crossing edges with only minimal variances along a street that provides parallel sidewalks. For perpendicular crossings, this selection includes formal crossings along the respective street section, as both can have almost the same lengths. The 45-degree crossings, however, are assumed to be always preferred against the formal crossings, except where formal crossings lie within the actual travel direction and allow to continue without having to turn.

3.1.3 Relevance of the numbers of street crossings

It can be assumed that non-dedicated mid-block crossings that are very close to formal crossing locations or only save a small amount of distance or time are only practiced by a small number of pedestrians. The same counts for multiple crossings of winding roads in order to shortcut routes.

[10] German traffic regulations no longer stipulate a minimum distance to dedicated crossings. The Austrian road traffic regulations, on the other hand, contain a specification of 25 m (§ 76 (5)), but this is also not mandatory in certain traffic conditions.

Against this background, the analysis of crossing counts by comparing routes based on the original network and the network with additional mid-block crossings is worth a closer look, as the shortest-path approach does not include other decisive factors than path length and may result in partly unrealistic routes. On the other hand, as described in the previous section, a short uncrossable section next to dedicated crossings is preserved, assuming that people who are used to cross mid-block tend to not do so very close to existing formal crossings.

3.1.4 Dense and regular urban fabric

The type of urban fabric is another important factor that determines the motivation to cross streets at mid-block positions. It can be assumed that the amount of mid-block crossings per route depends e.g. on block size lengths, regularity of the street grid, straightness of streets and street visibility.

The model developed here is especially adapted to a dense and more or less regular urban fabric, offering route options in many directions and providing streets that can be crossed at mid-block positions area-wide, but with the network gaps that exist even in densely populated areas. Including areas with a rectangular street grid seems valuable, as it can be assumed that mid-block crossings are especially important for routes that run diagonally to the main directions of the street grid. However, including some neighborhoods with a curved street layout is a crucial part of the selected setting as well to include possible "double crossings" in order to use the shorter sidewalk at the inner side of the curve.

3.1.5 Real-world data

It is assumed that the effect of non-dedicated mid-block crossings depends on several parameters of the tested network. Analyzing typical network types could therefore be conducted based on fictional path networks. This is a valuable approach and often practiced for specific network analysis purposes. In this case, it could be based on a simple street grid with automatically added sidewalks and crossings. However, this prevents the investigation of special cases and network configurations that are not included in simplified fictional model networks.

Highly detailed real-world test data already providing all geometries and attributes except the additional crossing geometries therefore was chosen to show how the results are influenced by special cases within the network. It is precisely the irregularities within a supposedly regular structure that are of interest here.

3.2 Methods of this work

This study introduces a novel method, parts of which are based on elements from previous studies. The new method is developed due to the lack of prior studies that compare an original network graph to a graph that extends the original graph by additional nodes and edges. In particular, the method for generating the test routes represents a specialized approach that can be classified between the testing of individual selected routes, for example across a square or within a district, and the generation of a certain number of routes with random start and end points within a selected larger area.

3.2.1 Extended graph generation based on geometric and attributive rules

As presented within the literature review, automatized graph generation has been the subject of many previous studies. If additional elements are to be added to existing graphs for the network analysis, geometric rules are usually set up to determine the positions, angles and lengths of the new nodes and edges based on the configuration of existing network elements. These rules may be conditional with regards to the presence or values of relevant attributes. Furthermore, they can vary based on conditions that check the presence or values of relevant attributes.

For this study, the graph generation method is a mixture of a geometric and attributive approach to define the rules for the network elements to be added. With the decision taken to use OSM data, the level of complexity of the necessary attribute-related expressions is comparably high, as OSM data contains numerous cases where several different tags can be used for the same purposes. Geometric expressions, on the other hand, must be configured in such a way that they take into account the special features of the networks to be tested. Therefore, an important part of the graph generation method chosen here is to make the decisive values within the model dynamically adaptable. This enables an iterative approach to an optimized result for the selected network.

3.2.2 Route generation using *Concentric Circular Destination Points* (CCDP)

The selected design for this study contains a comparison of a large set of test routes with different lengths with a large variety of network conditions at the start or the end of the route. The manual selection of a small number of test routes is not suitable here because of the small sample size, while a fully randomized selection of start and end points for a large number of test routes does not allow for a comparative check of the start or end points.

Based on these requirements, the *Concentric Circular Destination Points* (CCDP) method is developed. A defined number of test points is generated on an imaginary circle around a strategically selected start point at equidistant angles with a given initial radius. This is repeated with the step-wise increment of the radius using a given step size until the predefined maximum radius is reached. Now these regularly distributed points are moved to the closest point on an edge of the underlying network, resulting in a quasi-random distribution of points, with a higher density close to the manually selected start point. These points are then used as end points for test routes using the original network and versions using the graph generation model. The destination point generation steps are repeated for start points at other strategically selected positions of the network to be able to statistically compare different types of network layouts.

There are a few downsides of this method that need to be discussed, namely the lack of control of resulting test route lengths due to irregular network configurations, the dependency of a dense street grid without large gaps, and congruent resulting routes. However, these are accepted, as the results are still expected to be sufficiently meaningful for the purposes of this study.

3.2.3 Selective indicator check

Due to the complexity of comprehensive pedestrian networks, in particular in the case of the explicit inclusion of sidewalks and crossings, numerous possible factors for diverging results are expected when analyzing route lengths and crossing numbers of the different route types. The design of the study includes an examination of the correlation between the results and one selected explanatory indicator.

The selection of the indicator and the development of the corresponding model take place after the graph generation and test routes generation steps and are part of the analysis. The results presented in section 4 suggest that the selected route measurements vary depending on the "crossing completeness" of intersections, i.e. the number of merging streets at intersections that do not provide a crossing geometry within the underlying network graph. Accordingly, a model is developed to

- extract all relevant intersections,
- count the number of merging streets per intersection,
- split the result into T-shaped (3 merging streets) and X-shaped ($\geq$ 4 merging streets) intersections and to
- calculate a crossing index for each intersection (0 = all merging streets provide crossing geometries, -1 = all except 1 merging streets provide crossing geometries, etc.).

3.3 Study area: Neukölln and Tempelhof districts, Berlin

An analysis using Overpass queries[11] reveals that within the OSM database, currently no other urban areas than the Berlin use case are suitable as test areas for the chosen setup, i.e. using existing sidewalk data with comprehensive tagging, allowing for graph generation to be reduced to the additional crossing lines. Subsequent in-depth testing reveals the full suitability of the Neukölln district data as test area for this work.

The district of Neukölln is located about five kilometers south of the city center of Berlin, the German capital. The urbanization of this area took place from the end of the 19th century and thus somewhat later than that of the more centrally located urban expansion areas. However, the street network is seamlessly connected to that of the city areas further north and also adopts its structure of predominantly rectangular street blocks with concentric alignment to the city center. An exception with a rather irregular but nevertheless dense street network is the central area of the former village of Rixdorf, from which today's Neukölln district emerged. The area is densely populated and provides a mostly closed building development. Street space within residential roads is dominated by on-street parking and at many intersections there are no lowered curbs nor any other facilities for barrier-free crossing, reducing an adequate usability of the sidewalks to "pedestrians without wheels".

[11]The analysis is conducted using adequate Overpass queries, see Example Overpass queries in appendix.

Figure 3.2 (p. 26) shows example situations in Neukölln that lie outside the finally selected parts of the street network but still give an impression on typical street layout at intersections and mid-block[12].

Following the positive test of the Neukölln area for usability in this study, other Berlin districts are analyzed for comprehensive OSM mapping of sidewalks and crossings, particularly to find an area with curved residential streets that can be compared to the predominantly straight streets in Neukölln. This leads to the detection of the *Gartenstadt Neu-Tempelhof* (New-Tempelhof Garden City) area in the north of Tempelhof district, about 4 kilometers west of Neukölln. Here, the geometric and attributive structure fulfills all requirements as an additional test area. Development of the area started a few years prior to World War I, was interrupted during the war and continued during the 1920s, was some later additions during the 1950s. Block sizes are larger here than in Neukölln district, with a partly regular structure containing both curved streets and rectangular street blocks.

Figure 3.1 shows the location of Neukölln and Tempelhof districts in Berlin.

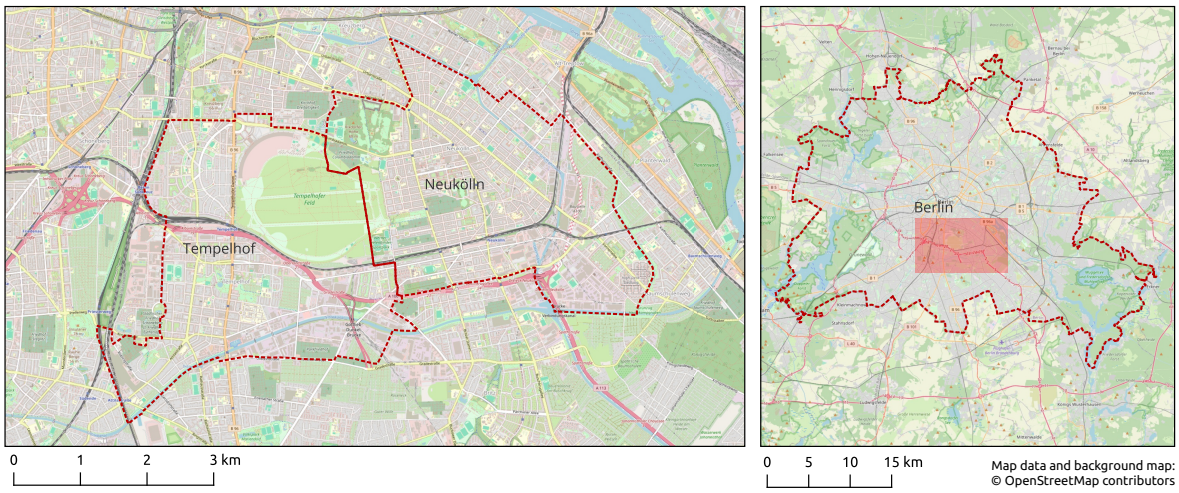


Figure 3.1: Location of Neukölln and Tempelhof districts within the city of Berlin

3.4 Required geometric and attributive representation

In order to reduce the attributive requirements, it is decided to integrate the road type detection for sidewalks based on a geometric approach using buffering, but to extract roadways that provide separate sidewalks on both sides based on attributes added to the roadway geometry.

The required level of detail for the network to be used thus includes

- a comprehensive geometric representation of the footpath network consisting of
  - a complete geometric representation of sidewalks for all road types,

[12]The photos were taken on 2023-10-10.

(a) Intersection with marked crossings and lowered curbs (Jonasstraße/Ilsestraße)

(b) Partly marked crossing with lowered curbs (Altenbraker Straße/Thomasstraße)

(c) Unmarked crossing without lowered curbs (Altenbraker Straße/Jonasstraße)

(d) T-shaped intersection with missing crossings via through street (Ilsestraße/Ilsenhof)

(e) Mid-block situation with street-side parking (Schierker Straße)

(f) Mid-block situation with lowered curbs for property access (Jonasstraße)

Figure 3.2: Example Neukölln situations at intersections and mid-block

  - all formal crossings, either structurally recognizable or integrated because of their obvious location, e.g. at unmarked positions at intersections where crossing is generally possible,
  - a full geometric representation of all other linear and formally walkable network elements, and
  - ideally also a linear geometric representation of the main itineraries via open spaces as well as

- attributes that allow for the detection of roadways that provide separate sidewalk geometries.

Different pre-tests and the analysis of the OSM sidewalk mapping approach[13] (including sidewalk schema proposals[14]) have shown that all necessary tags already exist either in a fully approved or in a proposed form.

With the decision taken to use the Berlin data as use case, the aforementioned general requirements for detailed mapping are examined, resulting in a translation of these requirements into specific tags to be used within the graph generation model.

The chosen approach requires a solution to detect which sidewalks belong to residential roads (`highway=residential`). An approach to solve this based on attributes is not selected as the required tags are still in proposal state and only exist in Berlin so far (e.g. `is_sidepath:of`[15] as a road type reference tag for sidewalks). This would prevent using the model in other areas where sidewalk geometries have already been added for all road types without such experimental tags. Thus, a geometric approach is decided upon.

On the other hand, the detection of roads that provide separate sidewalk geometries on both sides is possible using the accepted tags `sidewalk=separate` or `sidewalk:both=separate`, allowing for an attributive approach to filter out the respective residential roads.

As the approach only needs network edges, all tags have to be extracted only as ways[16]. Nodes and relations can be excluded.

The required tags and their respective purpose are:

- **`highway=*`**
  The approach includes an extraction of the full network of paths and roads in the target area.
- **`footway`**

[13] https://wiki.openstreetmap.org/wiki/Sidewalks, last accessed: 2024-02-21.
[14] https://wiki.openstreetmap.org/wiki/Proposal:Sidewalk_schema, last accessed: 2024-02-21.
[15] https://wiki.openstreetmap.org/wiki/Proposal:Key:is_sidepath:of, last accessed: 2024-02-21.
[16] The OSM data model contains *nodes* (points), *ways* (polylines and polygons), and *relations* (combination of points and ways into one grouped feature).

  - `footway=sidewalk`
    All footpaths that are sidewalks need to be present as geometries and tagged as such, regardless of the type of the associated road.
  - `footway=crossing`
    All existing crossings need to be present and tagged as such. These are usually marked crossings and unmarked crossings at intersections.

- `sidewalk=separate, sidewalk:both=separate`
  All roadways that have sidewalks on both sides need a clear attribute indication in this regard.
- `foot=use_sidepath`
  All roadways that are not suitable for walking and have sidewalks either on one or both sides need to provide this tag. Extracting the required information from the different sidewalk tags would be possible. Also it could be defined generally that all roadways except pedestrian streets and living streets can not be used by pedestrians. But this would make the attributive approach much more complex. The model uses this tag for roadway—sidewalk reference.

Areal features, bridges, and tunnels are not taken into account due to lack of relevance for the test area.

3.5 Open-source software

Several web-based or locally installed open-source tools are applied to conduct this study.

Overpass QL

The *Overpass QL*, an OSM-specific query language for precise extraction of spatially or attributively selected data from the OSM database is used to detect and analyze areas with separately mapped sidewalks for all street categories. It is also used to download the required data for the test areas.

QGIS, Graphical Modeler, Geometry Expressions

The open-source desktop GIS software *QGIS* is used to work with the downloaded data, namely for turning the data into a routable graph. The extended graphs and the test routes generation is also done using QGIS, with the required models created using the built-in *Graphical Modeler*, including numerous and complex *QGIS Geometry Expressions*. The spatial analysis of the results is also conducted using QGIS.

QGIS built-in network analysis tools

QGIS offers five built-in network analysis tools that are suitable for this study. These tools can be accessed via the *Processing Toolbox* and offer sufficient individual parameterization to carry out the planned tests. “Shortest path (point to layer)” and “Service area (from layer)” are the two algorithms used within the scope of this study.

While the resulting routes are always created without any errors, the service area tools unfortunately do not output a clean vector layer but may, depending of the complexity of the selected routing graphs, contain overlaps and "hidden dangles", i.e. line overlaps that are not visible and lead to false results when analyzing contained route lengths. These errors need to be discussed.

R programming language

The *R* programming language is used for the statistical analysis of the data using an attribute-only CSV export of the test-routes data, with the attributes that are geometry-dependent pre-calculated using QGIS and additional columns derived from the pre-calculated data added using R.

3.6 Implementation

This section describes how the previously introduced setup is implemented.

3.6.1 Study area data extraction

The study area is downloaded using an appropriate Overpass query[17] and saved to disk[18]. The download contains all ways tagged as `highway=*` and the related district boundary relations. First tests showed that there are only a few errors and that manual editing of the data in QGIS is only necessary for very few (< 10) relevant features.

3.6.2 Graph-generation

The additional crossing geometries are then created using the model shown in figure 3.3 (p. 30), which outputs the necessary layers for test routes generation: original routable graph, graph with added 90-degree crossings and graph with combined 45-degree and 90-degree crossings, each as combined layer and as separate layers for original network and additional crossing geometries[19]. The crossing interval is set to 10 m, the minimum distance to existing crossing geometries to 15 m, the initial length of the crossing lines per side is set to 15 m, and the maximum distance between sidewalk and roadway—used for the buffer to detect all sidewalks of residential streets—to 20 m.

A section of the resulting graph is shown in figure 3.4 (p. 31).

3.6.3 Network service areas

The generated network graphs include isolated sections not connected to the main part of the network. To remove these unconnected parts, service areas from the selected start points are created for each of the graphs to be used as cleaned network graphs for test routes generation. The distance is set large enough to cover the expected extent of all test routes.

The additional model used for this intermediate step is shown in figure 3.5 (p. 32).

[17] See Example Overpass queries in appendix.
[18] Download dates: Neukölln data 2023-10-21, Tempelhof data: 2024-01-20.
[19] The most important attributive and geometric expressions used within the model are described in section Example geometric and attributive expressions in the appendix.

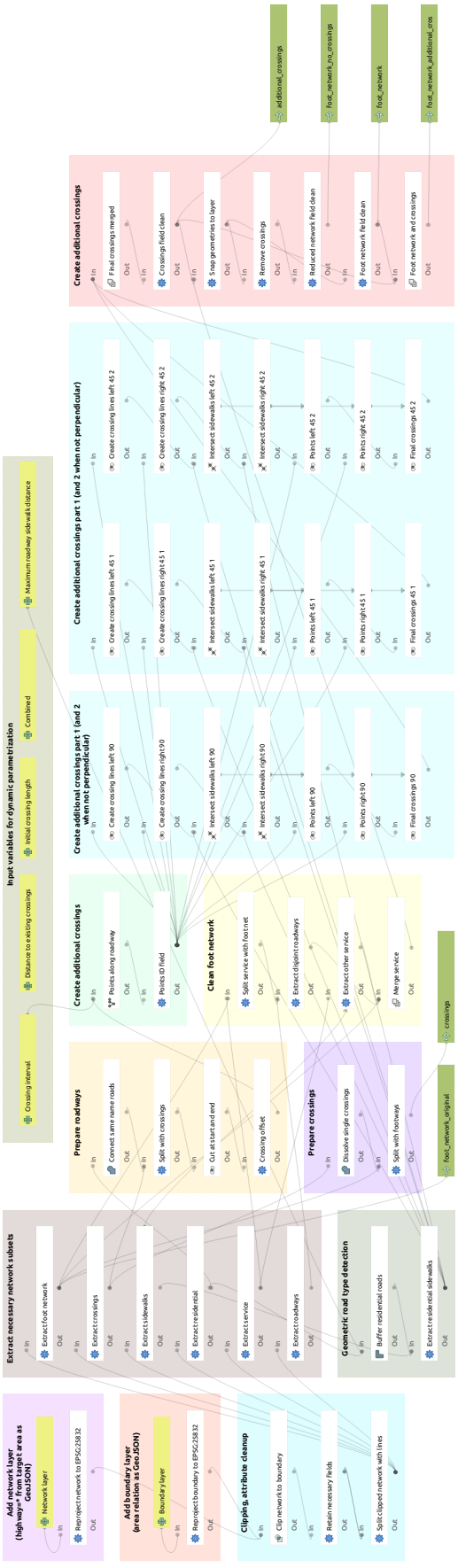


Figure 3.3: Graphical model for generation of additional crossing edges

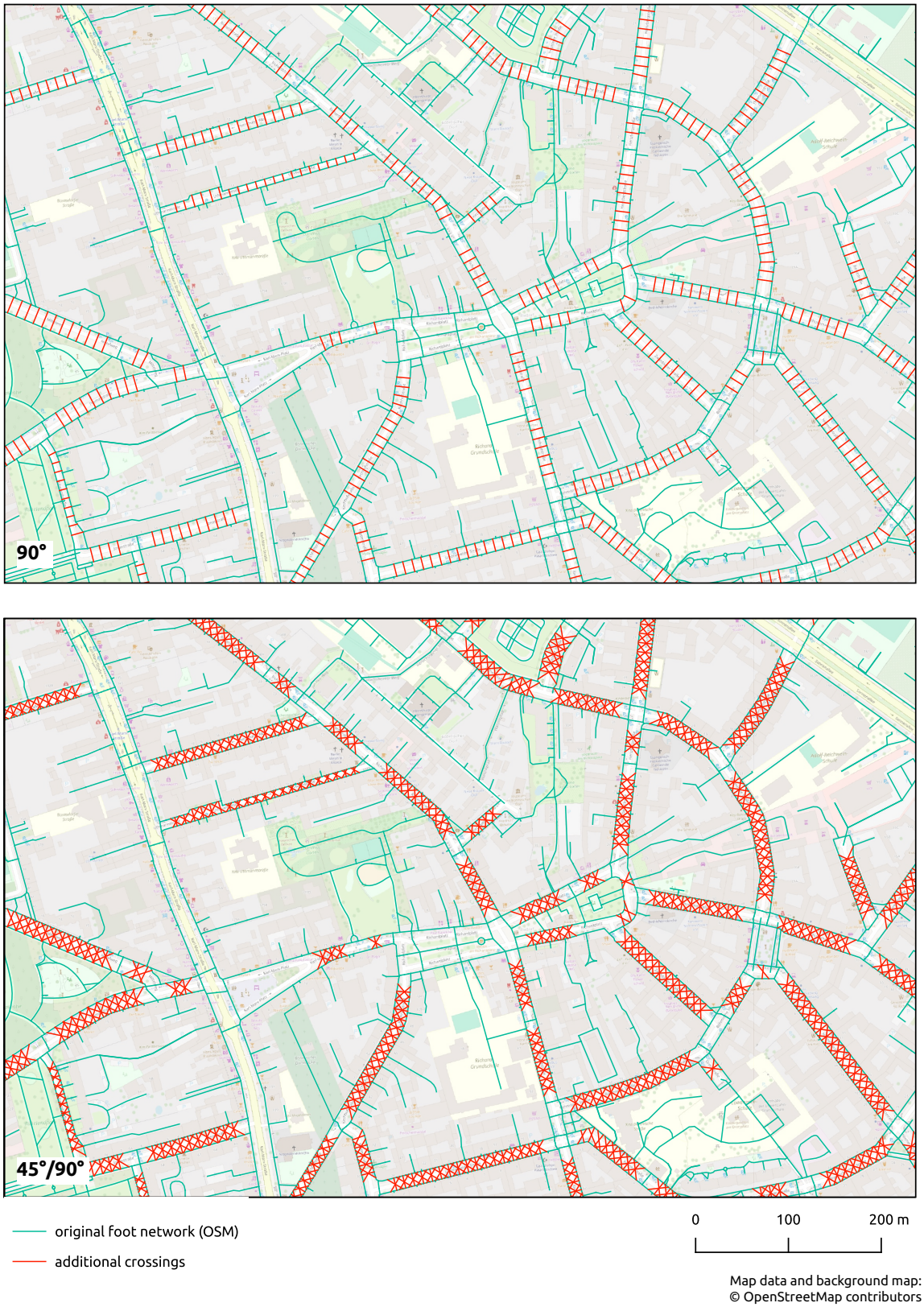


Figure 3.4: A section of the resulting network graphs

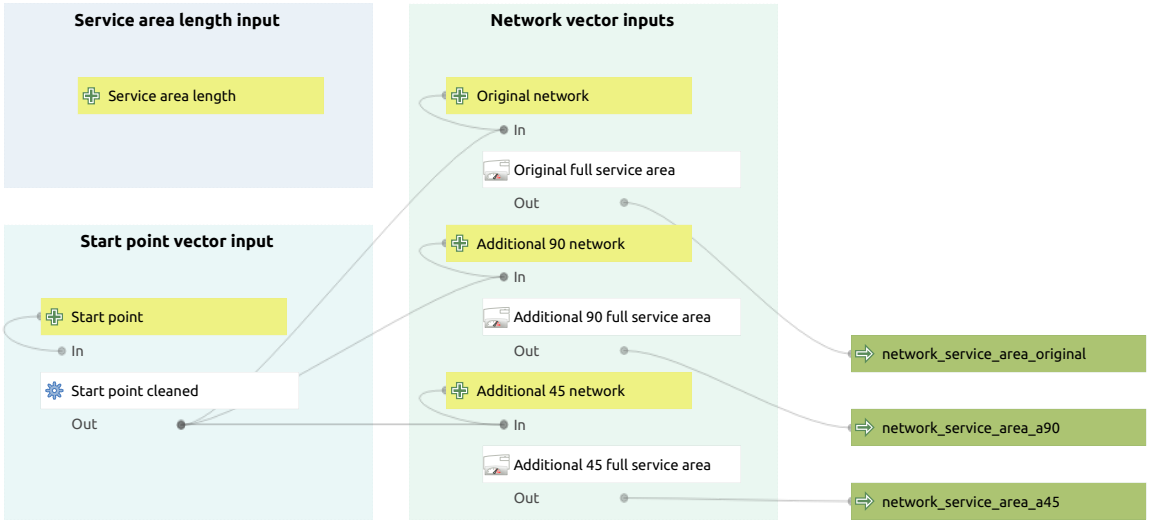


Figure 3.5: Graphical model for network service area generation

3.6.4 Start point selection

As part of the CCDP method introduced in section 3.2.2, three start points are selected:

- **Start point A** (Neukölln) at the sidewalk in front of Richardplatz 18 residential building, selected to gain results within an irregular network, starting next to a directly crossable residential street and surrounded by residential streets to each direction.
- **Start point B** (Neukölln) at the southern entrance of Boddinstraße metro station at a traffic island between the directional roadways of Hermannstraße main street, selected to gain results within a more regular network, starting next to a street the is not freely crossable.
- **Start point C** (Tempelhof) at the sidewalk in front of Hessenring 46 residential building, selected to gain results within an irregular network with curved streets, starting next to a directly crossable and curved residential street, but close to an intersection with a major road that is not freely crossable.

3.6.5 Test-routes generation

The test routes are then created by applying the model shown in figure 3.6 (p. 34) for each of the three selected start points, based on the CCDP method introduced in section 3.2.2. The model includes the calculation of all necessary attribute fields that rely on the geometries of the related routes.

The number of concentric test points is set to 32, minimum distance to 25 m and maximum distance to 500 m using a 25 m increment, i.e. 20 distance steps. With these settings, 1920 test routes are created for each of the three types (640 routes per start point).

The different steps of the implemented CCDP method are visualized in figure 3.7 (p. 35).

The main geometric expression used within the model is described at the end of section Example geometric and attributive expressions in the appendix.

3.6.6 Preparation of analysis in R

All resulting routes are stored as vector layers in either Geopackage or GeoJSON formats. A merged file of all routes is exported without geometry in CSV format to to perform the necessary calculations in R. Within the main R script, the data stored in a single data frame, with all additionally calculated fields added as columns.

3.6.7 Calculation of intersection crossing indices

As part of the analysis, the intersection crossing completeness indicator is selected and integrated using the method described in section 3.2.3. The graphical model to generate the necessary crossing indices for each intersection of residential streets or residential and living streets is shown in figure 3.8 (p. 36).

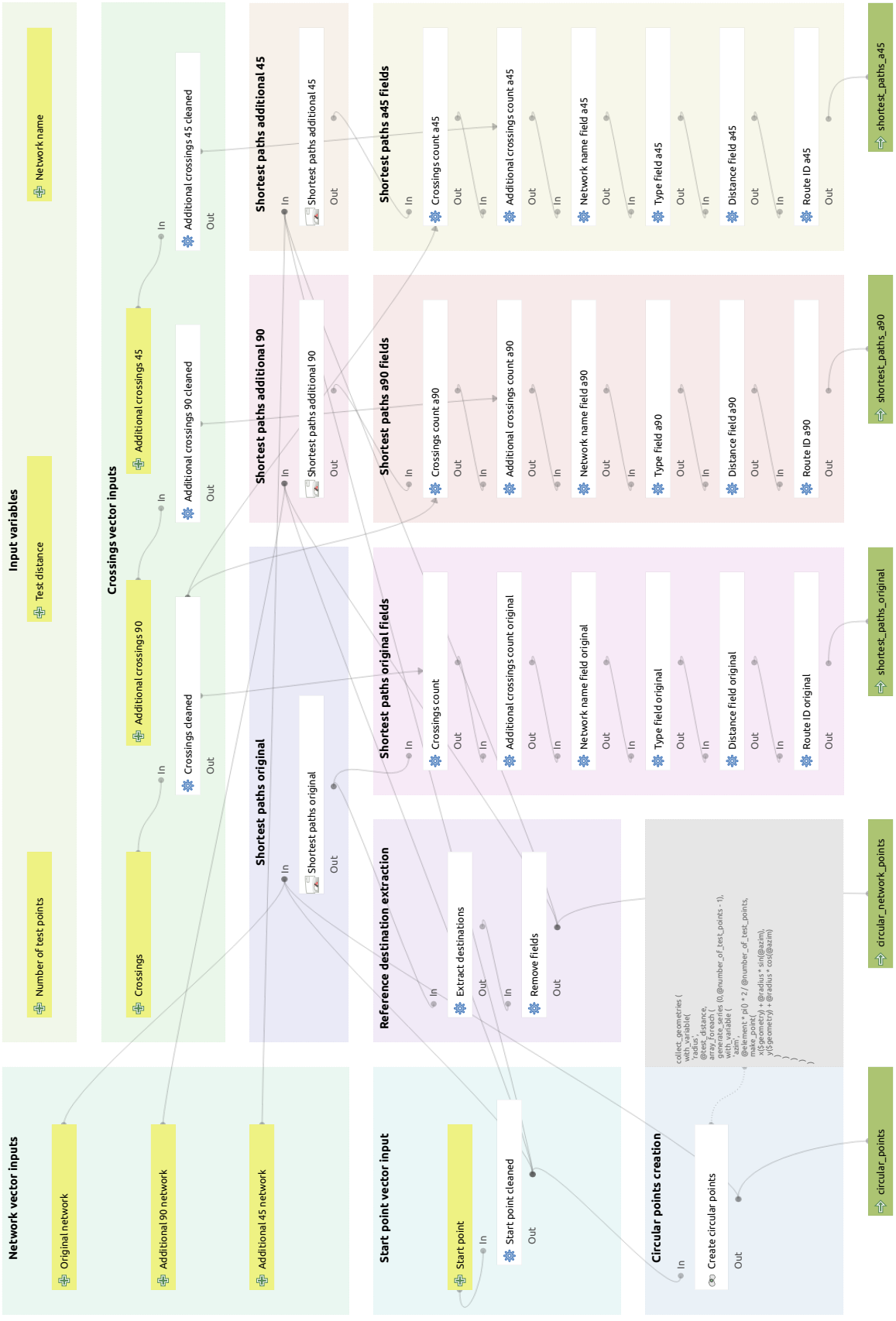


Figure 3.6: Graphical model for test routes generation using CCDP method

Figure 3.7: Implemented steps of test routes generation using CCDP method

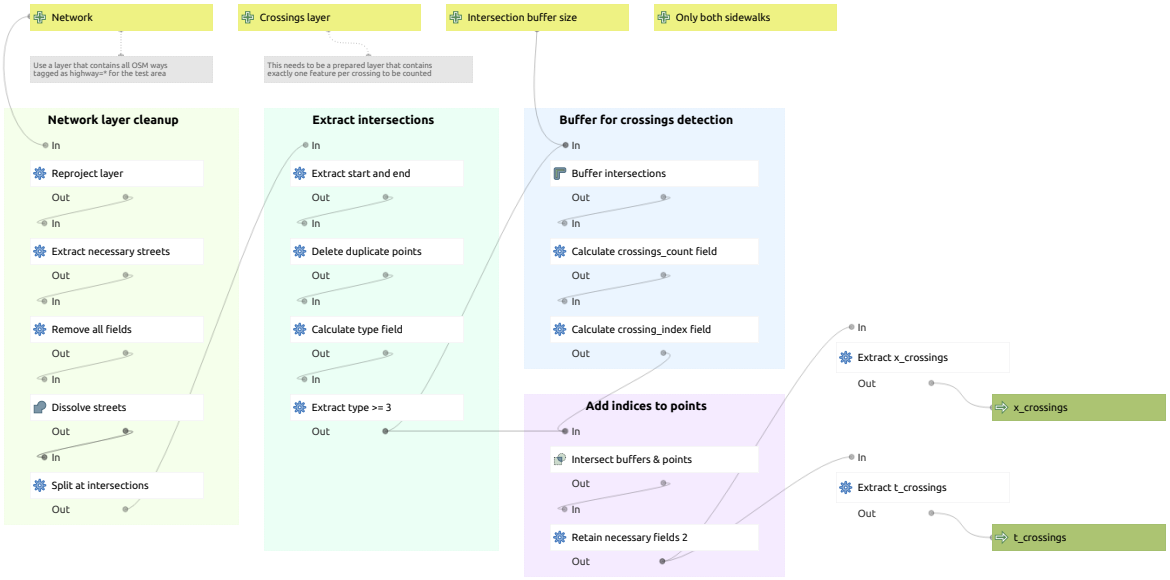


Figure 3.8: Graphical model intersection crossing indices calculation

4 Results and analysis

In this section the results of the model implementation are presented and analyzed. First, the summarized metrics for lengths ratios and crossing counts are examined for the full data set and for each of the three single test networks. The focus is then shifted to the path length ratios, including a path length distribution check and the presentation of the average path length ratios per 25 m path length group. This is then followed by an intersection crossing index analysis as well as an analysis of street grid direction dependency. Finally, all analysis steps are summarized.

In order to simplify the description of the observations within this chapter, the routes are referred to as *original* routes, as *a90* routes for the network graph that includes 90-degree non-dedicated mid-block crossings, and as *a45* routes for the network graph that contains both 45-degree and 90-degree non-dedicated mid-block crossings.

4.1 Overall results for full data set

Figure 4.1 shows an overview map of the resulting routes from the three selected start points.

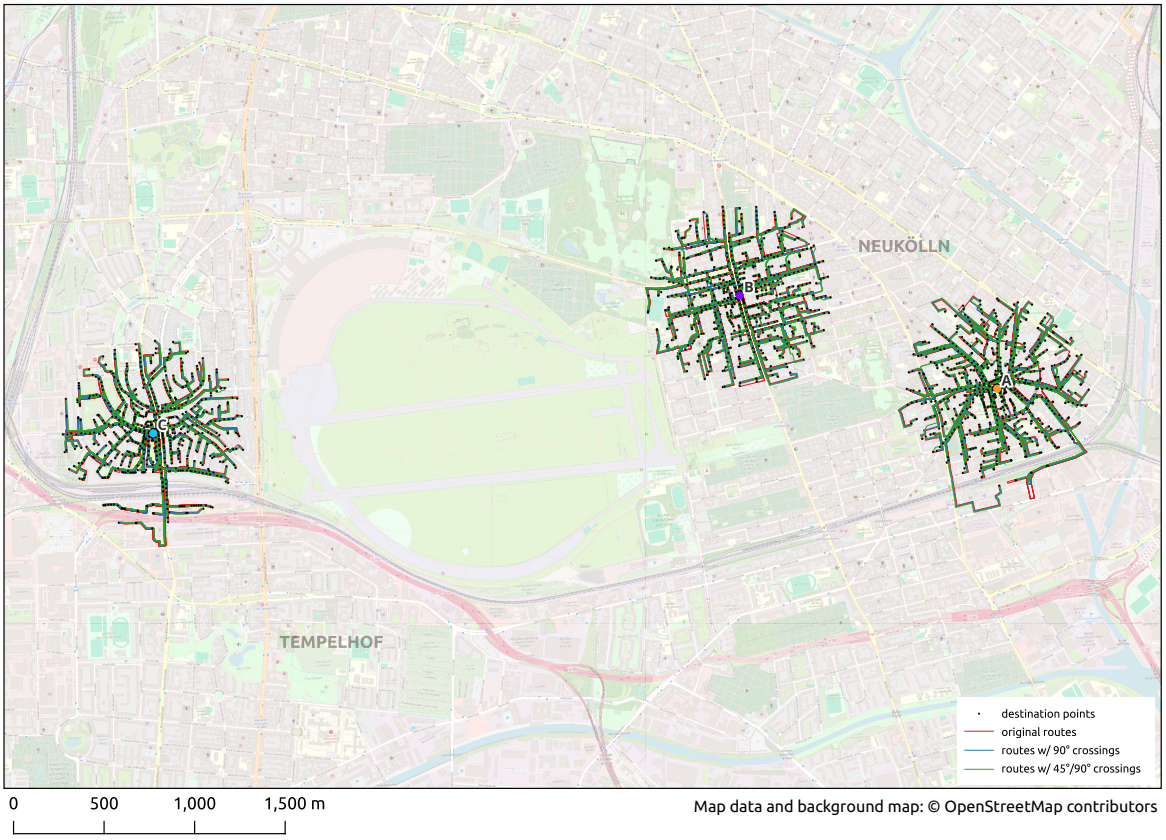


Figure 4.1: Results overview map for networks a, b, and c

Table 4.1 shows that *original* route lengths vary from about 20 m to just under 1.5 km, with a mean of about 357 m. For the *a90* and *a45* extended graphs, the mean length value is 5 respectively 11 m lower than for the original routes. Median values are a bit lower than mean values, showing that there are more shorter routes.

Table 4.1: Length metrics (meters) of full test routes data set

type	mean	median	min	max
original	357.56	352.18	20.36	1374.91
a90	352.47	349.48	20.36	1236.84
a45	346.64	346.36	20.36	1236.84

The numbers reveal that at least for the longest route, the path-length reduction effect of the additional non-dedicated crossings is highly noticeable. For the shortest route there is no path-length reduction, but the overall numbers do not yet reveal the length-related distribution of length reductions.

Table 4.2 shows that the average length reduction for the *a90* type is just over 2% and just under 4% for the *a45* type. The higher median values show that for most cases the length reduction is even lower. Additionally, the last three columns show the ratios of routes providing a length ratio lower than a given value, revealing e.g. that only 3.5% of all routes "save" more than 20% of the route length for the *a90* type, compared to 4.3% for the *a45* type. For routes that save 10% of route length, referred to as high length reduction for this work, these values increase to 6.3% for the *a90* type and 8.2% for the *a45* type.

Table 4.2: Length ratio metrics of full test routes data set

type	mean	median	min	max	ratio of rts. w/ lratio < 0.8	ratio of rts. w/ lratio < 0.9	ratio of rts. w/ lratio < 0.98
a90	0.979	1.000	0.305	1	0.035	0.063	0.109
a45	0.961	0.984	0.230	1	0.043	0.082	0.425

This shows that, although the average length reduction for the *a90* and *a45* routes is quite low, there are cases where the *a90* and *a45* routes are two thirds to three quarters shorter than the *original* route, as the low minimum values of around 30.5% and 23% indicate. Also, as the median values show, at least 50% of all *a90* routes have the same lengths as their original counterparts, whereas more than 50% of the *a45* routes are shorter compared to the respective original route.

Table 4.3 shows the main metrics for the number of streets crossings per route. All three network types include a maximum of 9 street crossings at original crossings or, for the *a90* and *a45* types, also at non-dedicated mid-block positions. The median number of crossings is 2 for each case and the mean of around is around 2.47 for the *original* routes, a bit higher for the *a90* type and once again higher for the *a45* type.

Table 4.3: Crossing count metrics of full test routes data set

type	mean	median	min	max
original	2.47	2	0	9
a90	2.50	2	0	9
a45	2.57	2	0	9

Although minimum, maximum, and median values are the same for all types, higher mean values indicate that there are cases where the *a90* route and especially the *a45* route contain more street crossings than the respective *original* route.

Table 4.4 shows the metrics for crossing count differences compared to the *original* routes that vary between -1 and 2 for both the *a90* and the *a45* type, each with a median of 0. Mean values are about 0.03 for the *a90* type and around 0.1 for the *a45* type. 92.6% of all *a90* routes and 84.9% of all *a45* routes contain the same number of crossings as the respective *original* routes.

Table 4.4: Crossing count difference metrics of full test routes data set

type	mean	median	min	max	ratio of rts. w/ 0 crossing count difference
a90	0.03	0	-1	2	0.93
a45	0.10	0	-1	2	0.85

This shows that for both the *a90* and *a45* type there are cases where the number of crossings is reduced, but more routes that provide more crossings than the respective *original* route.

4.2 Overall results for single networks A, B, and C

The route comparison metrics are now examined for each of the three networks separately. The analysis comprises the comparison related to the full network metrics and across the different networks, revealing discrepancies caused by the different characteristics of the starting points and the varying network layouts.

Network A

Figure 4.2 (p. 40) shows the resulting routes for all three types from selected start point *A*.

Figure 4.3 (p. 41) shows the routes separated into four length ratio groups for the *a45* type and figure 4.4 (p. 42) shows the crossing count differences separated into four groups for the *a45* type, both in a cumulative view to provide a basic orientation.

The length ratio maps reveal that there is a noticeable compass direction dependency for *network A*, with visibly reduced direct accessibility for the northern and northeastern direction for the *original* routes. Crossing count differences vary to a maximum of 2. The respective map (marked "+2") already reveals the "double street crossings" because of sharp turning angles of the route as one of the reasons for additional crossings.

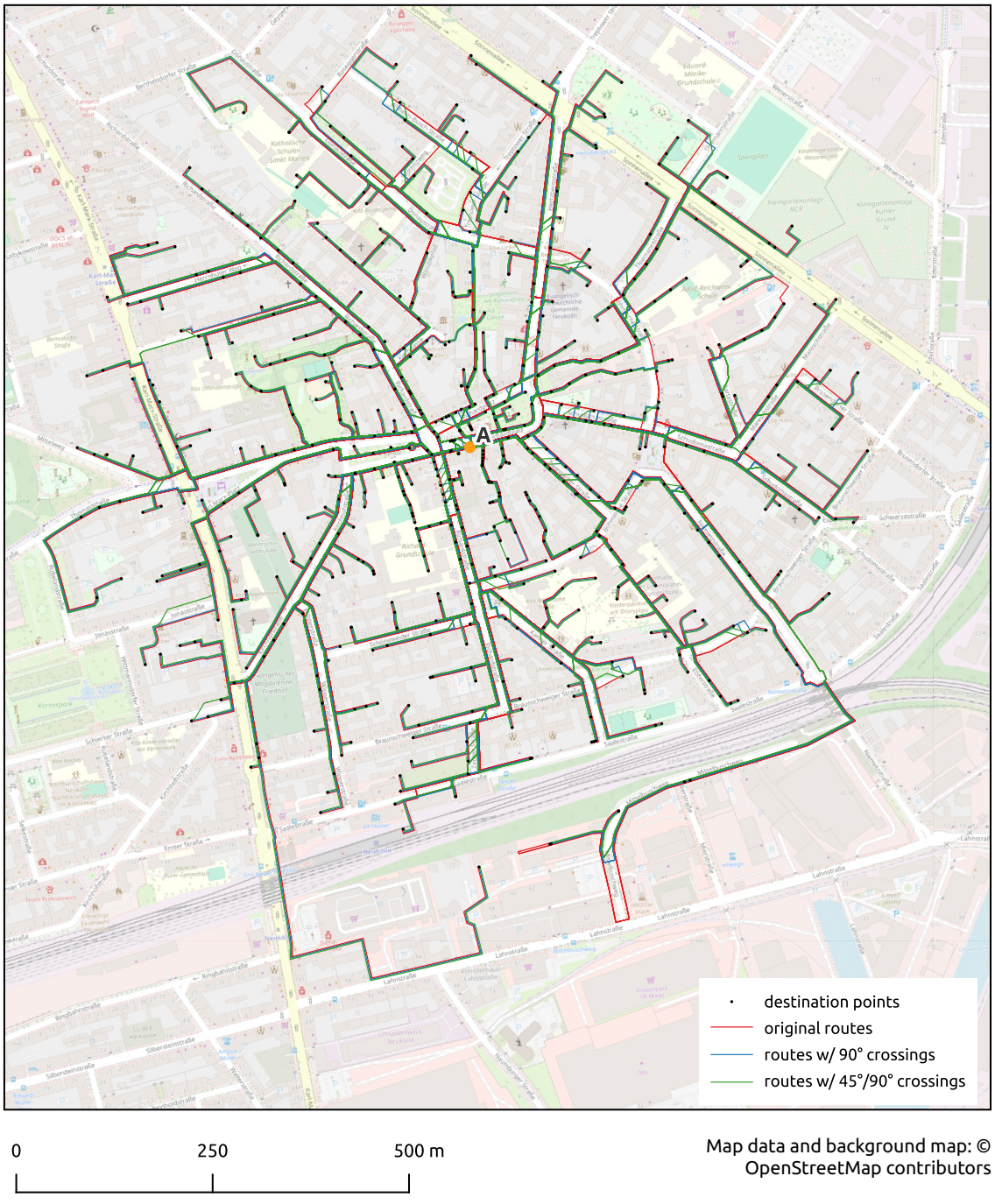


Figure 4.2: Overview map of resulting routes for start point A

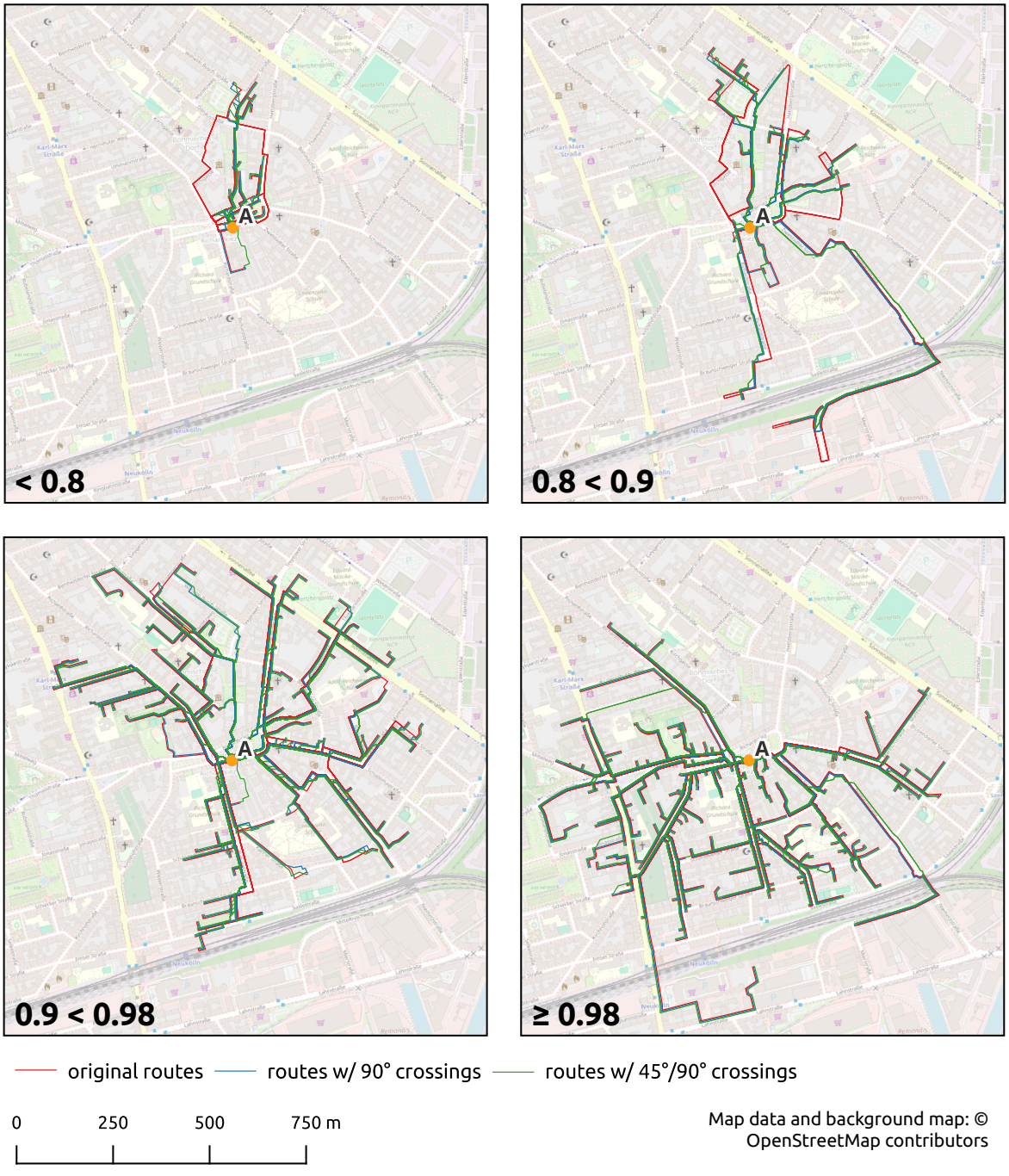


Figure 4.3: Map of resulting routes for start point *A* grouped by length ratios for the a45 type

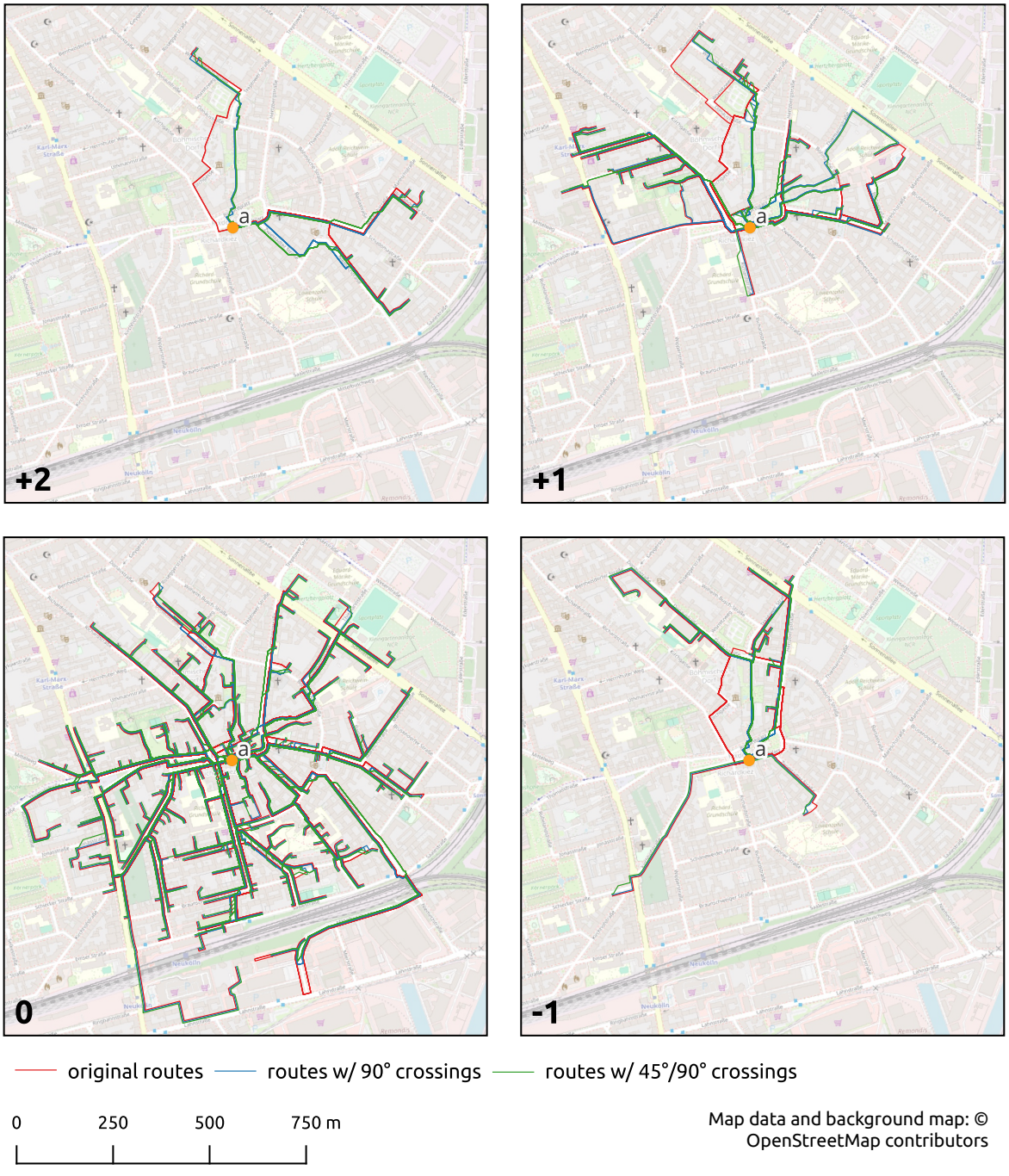


Figure 4.4: Map of resulting routes for start point *A* grouped by a45 type crossing count differences

Table 4.5 shows that the routes with the highest overall relative length reduction of the full data set for both the *a90* and *a45* type are contained in *network A*, because the numbers are the same as in the respective table presented in the previous section. However, mean values are remarkably lower for both types with 95.6 and 93.4 respectively, as is the median for the *a45* type with 97.8 compared to 98.4 for the full data set. 7.8% of the *a90* and 9.8% of the *a45* routes are more than 20% shorter than their *original* counterparts.

Table 4.5: Length ratio metrics of network A

type	mean	median	min	max	ratio of rts. w/ lratio < 0.8	ratio of rts. w/ lratio < 0.9	ratio of rts. w/ lratio < 0.98
a90	0.956	1.000	0.305	1	0.078	0.134	0.209
a45	0.934	0.978	0.230	1	0.098	0.169	0.522

The values for *network A* reveal that there are more routes with large path length reductions within this part of the data set than compared to the full data, which is especially true for the *a45* type.

Table 4.6 shows that mean values are again both visibly higher than their counterparts for the full data set. Median, minimum and maximum values, however, are the same as for the complete data, with 0, -1, and 2 respectively. The last column shows that 86.4% of the *a90* and 80.5% of the *a45* routes have the same number of crossings as the respective *original* routes.

Table 4.6: Crossing count difference metrics of network A

type	mean	median	min	max	ratio of rts. w/ 0 crossing count difference
a90	0.05	0	-1	2	0.86
a45	0.12	0	-1	2	0.80

The values reveal that in comparison with the full data set, more *a45* and *a90* routes within *network A* contain additional street crossings than the respective *original* routes.

Network B

Figure 4.5 (p. 44) shows the resulting routes for all three types from selected start point *B*.

Figure 4.6 (p. 45) shows the routes separated into four length ratio groups for the *a45* type and figure 4.7 (p. 46) shows the crossing count differences separated into four groups for the *a45* type, both in a cumulative view to provide a basic orientation.

The length ratio maps show that for the original paths there is an important barrier to the east for the original routes. To the other directions, direct accessibility is either not or partly reduced for the original paths compared to the *a90* and *a45* types, almost clearly divided into directional sectors. Crossing count differences show a mixed pattern with -1 difference routes reduced to the east. However, the higher-level relevance can be put into question here, as for the +1 and -1

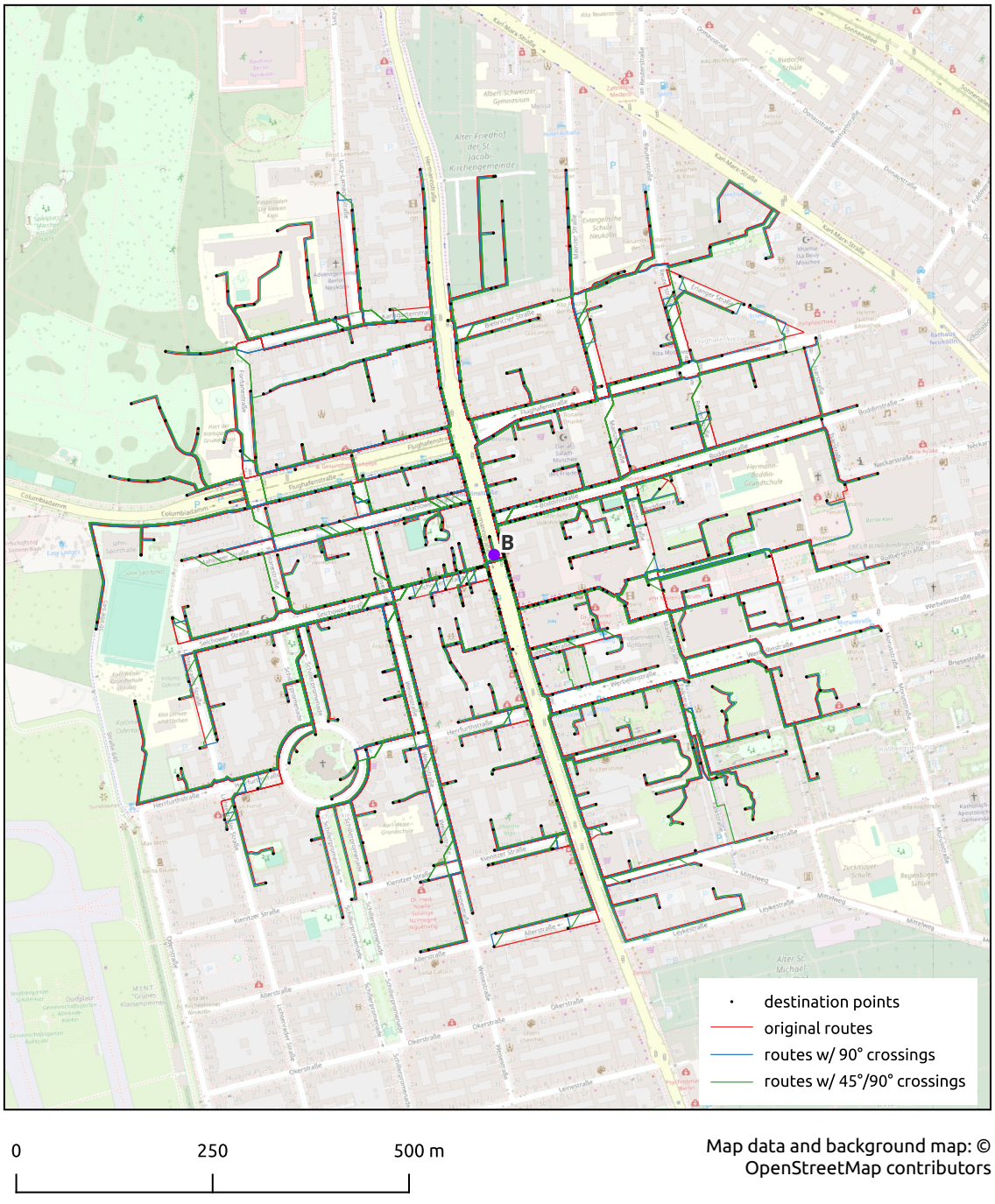


Figure 4.5: Overview map of resulting routes for start point *B*

Figure 4.6: Map of resulting routes for start point *B* grouped by length ratios for the a45 type

Figure 4.7: Map of resulting routes for start point *B* grouped by a45 type crossing count differences

routes the reason for the difference often seems to be T-shaped crossings which lead to different numbers of crossings depending on the street side used.

Table 4.7 shows that for *network B*, the maximum path-length reduction is only about 17 percent. Mean ratio values are remarkably closer to 1 than for the full data set, showing that the overall path-length reduction effect of the additional crossings is substantially lower for this network than for *network A*. No routes of this network "save" more than 20% of the path length for the *a45* and *a90* types, but still there is a 10% length reduction for 2.2% resp. 3.3% of the test routes.

Table 4.7: Length ratio metrics of network B

type	mean	median	min	max	ratio of rts. w/ lratio < 0.8	ratio of rts. w/ lratio < 0.9	ratio of rts. w/ lratio < 0.98
a90	0.996	1	0.83	1	0	0.022	0.044
a45	0.982	1	0.83	1	0	0.033	0.320

For *network B*, the more regular street grid and intersection layout compared to *network A* appears to be an important factor for the smaller length reductions. Also, non-dedicated mid-block crossings are not possible at very short distances, as the route does not start next to a residential street.

Table 4.8 shows that average crossing count differences are positive here as well, with the *a90* value even a little higher than the one for *network A*, but with a remarkably lower *a45* value (0.06 compared to 0.12 for *network A*). The maximum difference for both the *a90* and the *a45* type is 1 compared to 2 for the full data set and for *network A*. Here, with 94.4% resp. 90.8%, even more routes provide the same number of crossings for the *a45* and *a90* types as for the *original* route.

Table 4.8: Crossing count difference metrics of network B

type	mean	median	min	max	ratio of rts. w/ 0 crossing count difference
a90	0.05	0	-1	1	0.94
a45	0.06	0	-1	1	0.91

It can be concluded that also here, the more regular street grid and intersection design layout of *network B* compared to *network A* leads to a reduced length-reduction effect of the *a45 type*.

Network C

Figure 4.8 (p. 48) shows the resulting routes for all three types from selected start point *C*.

Figure 4.9 (p. 49) shows the routes separated into four length ratio groups for the *a45* type and figure 4.10 (p. 50) shows the crossing count differences separated into four groups for the *a45* type, both in a cumulative view to provide a basic orientation.

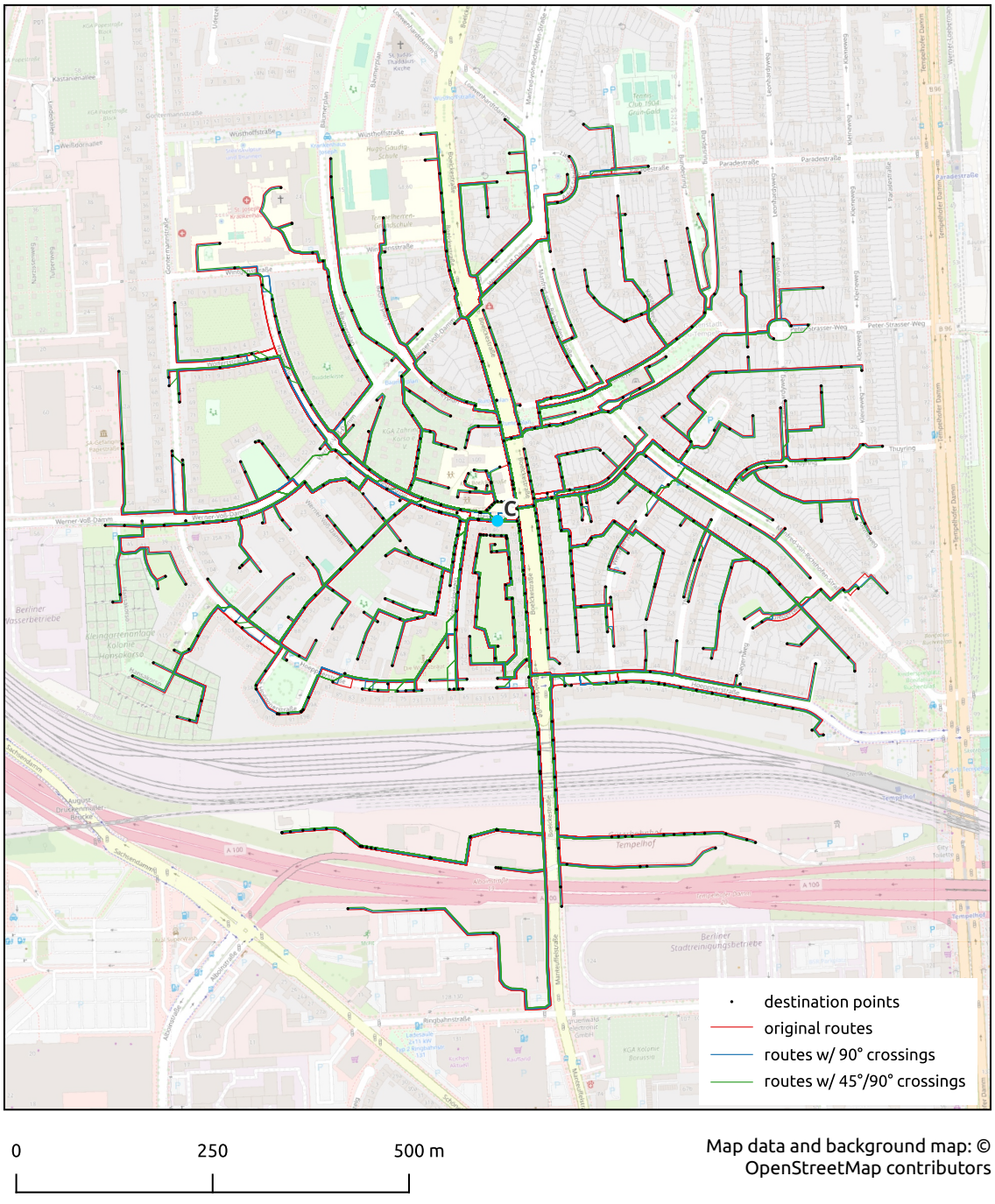


Figure 4.8: Overview map of resulting routes for start point C

Figure 4.9: Map of resulting routes for start point *C* grouped by length ratios for the a45 type

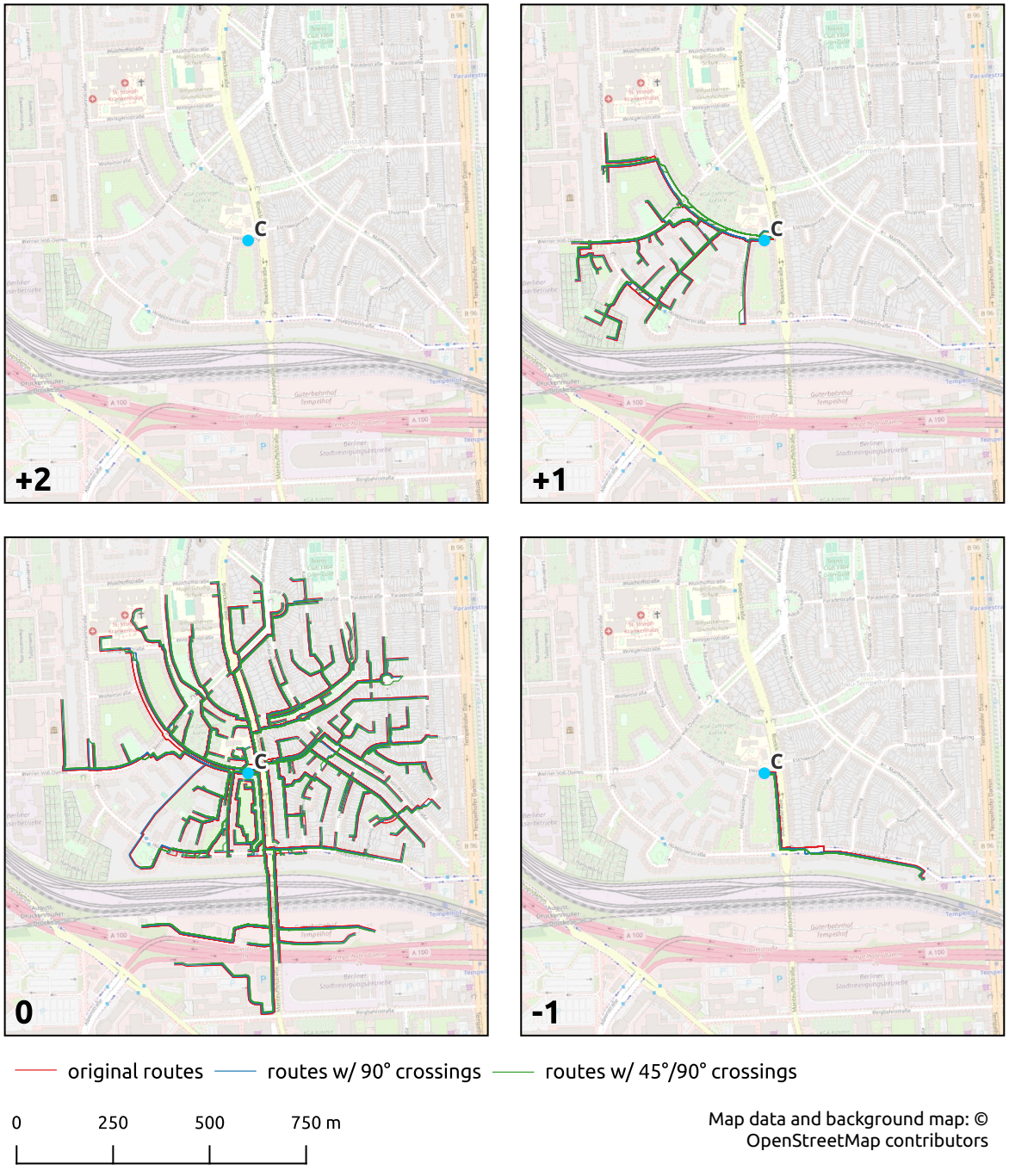


Figure 4.10: Map of resulting routes for start point *C* grouped by a45 type crossing count differences

As the length ratio maps show, far less and only very short routes for *network C* have a ratio value lower than 0.8 compared to *network A*. Length reduction is still higher for routes to the south-western direction than for other directions. This also counts for the crossing counts difference, which is positive only for several routes to the west and south-west.

Table 4.9 shows that the minimum length ratio for *network C* is about 43.5 percent for type *a90* and approximately 35 percent for type *a45*. Mean values lie between those of *network A* and *network B*. The median value for the *a90* type is minimally lower than 1, compared to exactly 1 in all other cases. 2.8% of the *a90* routes and 3.1% of the *a45* routes save more than 20% of the *original* route length.

Table 4.9: Length ratio metrics of network C

type	mean	median	min	max	ratio of rts. w/ lratio < 0.8	ratio of rts. w/ lratio < 0.9	ratio of rts. w/ lratio < 0.98
a90	0.986	0.997	0.436	1	0.028	0.033	0.075
a45	0.967	0.982	0.351	1	0.031	0.045	0.433

Network C, like *network A*, thus contains routes that are remarkably shorter for for the *a90* and *a45* types compared to the original routes, but the overall length reduction effect that is not as high as for *network A*.

Table 4.10 shows that mean values for the *network C* crossing count difference are close to the values for the complete test routes, minimally negative for the *a90* type here and positive for the *a45* type, with a value close to the one of *network A*. Minimum crossing count differences are -1 and 1 for both types, compared to -1 and 2 for the full data set and for *network A*. 96.9% of the *a90* routes and 83.4% of the *a45* routes provide the same number of street crossings as their *original* counterpart.

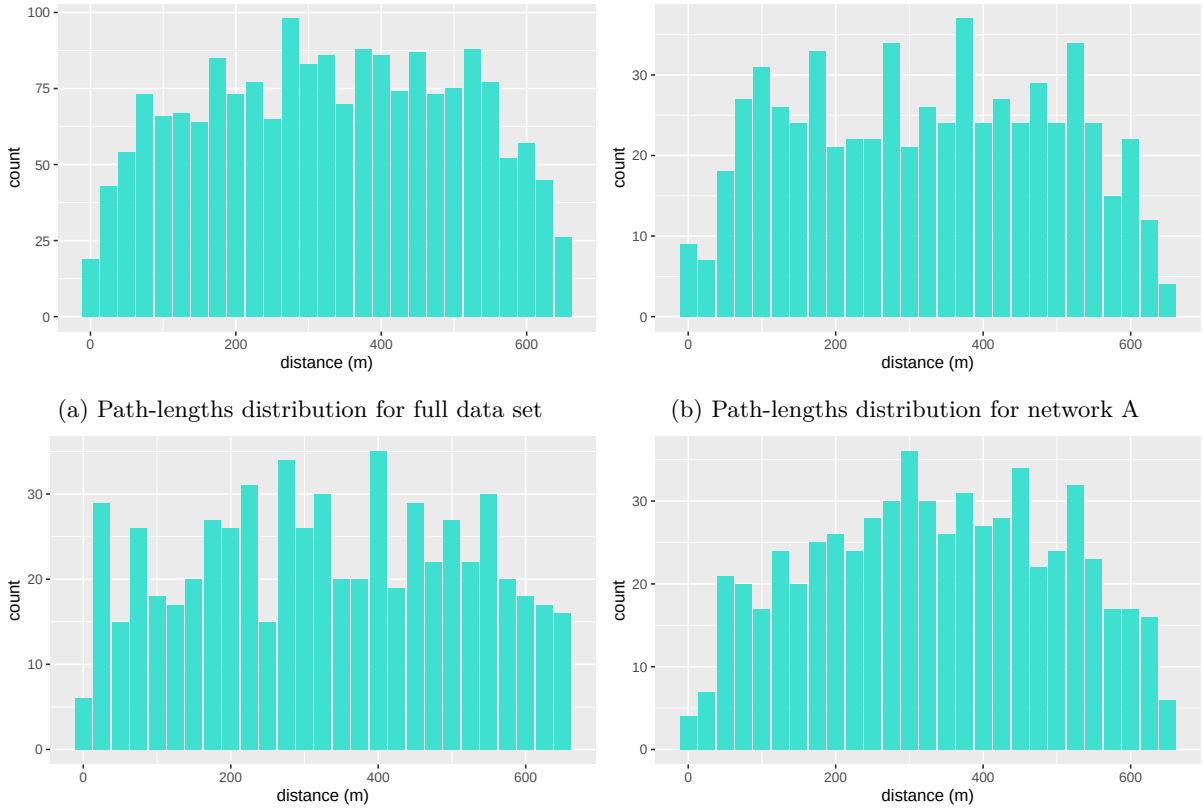
Table 4.10: Crossing count difference metrics of network C

type	mean	median	min	max	ratio of rts. w/ 0 crossing count difference
a90	-0.01	0	-1	1	0.97
a45	0.12	0	-1	1	0.83

The most important point here is that the *a45* type causes additional crossings for several routes, while for very few cases the *a90* routes provide less crossings than their original counterpart.

4.3 Path-length distribution

Figure 4.11 shows the path-length distributions for the full network and for networks *A*, *B*, and *C* separately. The path-length distribution is presented as a pre-test to check for sufficient routes in each of the 25 m path length slots. The sub-figures show that for the major part of the distance slots there are at least 50 test routes, whereas the single test cases *A*, *B*, and *C* provide at least 15 test routes for almost all test distances.



(a) Path-lengths distribution for full data set

(b) Path-lengths distribution for network A

(c) Path-lengths distribution for network B

(d) Path-lengths distribution for network C

Figure 4.11: Path-lengths distributions

4.4 Path-length ratios grouped by 25 m slots

To visualize the results for specific path lengths, the test routes are grouped by 25 m slots to output the path-lengths ratios for each step. First, the path-length ratios are presented for the full data set, followed by a visualization and description for each of the three single networks.

Figure 4.12 shows that length reductions are highest for the full data set between 50 and 150 m route length. For routes longer than 200 m the path reductions are around 1.25 percent for the *a90* type and 2.5 percent for the *a45* type, with some slight variances.

Network A

Figure 4.13 shows the mean path-length ratios per 25 m distance slot for *network A*. Here, the average length reductions are even larger at 100 m than for the full data set. At the same time,

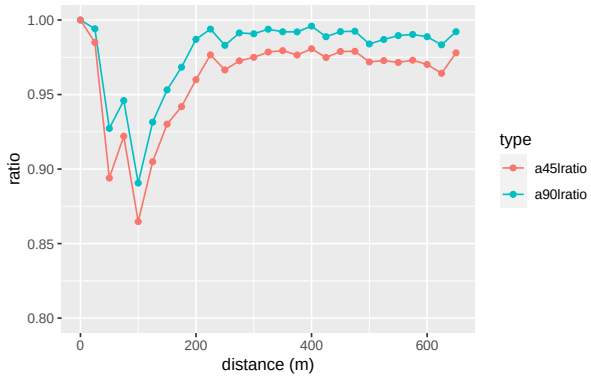


Figure 4.12: Path-length ratio distribution for full data set

length ratios are much smaller for longer distances than compared to the overall numbers, with more variance and a noticeable valley at 625 m.

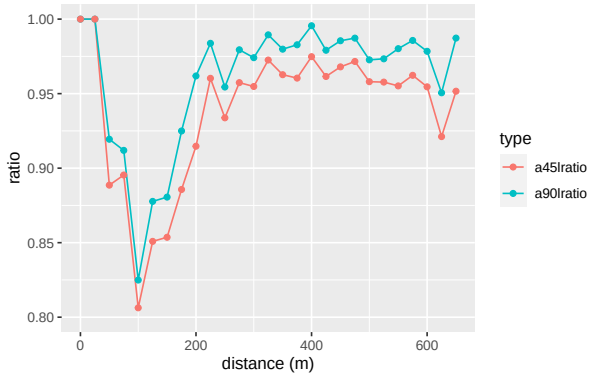


Figure 4.13: Path-length ratio distribution for network A

Network B

Figure 4.14 (p. 54) shows that there are no path length reductions for the *a90 type* within *network B* for route lengths up to 300 m, while here the *a45* route length reduction varies between 0 and 2.5 percent. For longer routes, ratios vary between 0.98 and 1 for the *a90* type and between 0.96 and 0.99 for the *a45* type.

Network C

Figure 4.15 (p. 54) shows that *network C*, like *network A*, has large path-length reductions for short routes, shifted to 25 m to 125 m in this case. Average length reductions are a bit smaller than for *network A*. For larger distances, average values vary between 0.99 and 1 for the *a90 type* and 0.975 to 0.99 for the *a45 type*, with *a45* values constantly reducing whit larger distances. There are far less irregularities compared to *network A*.

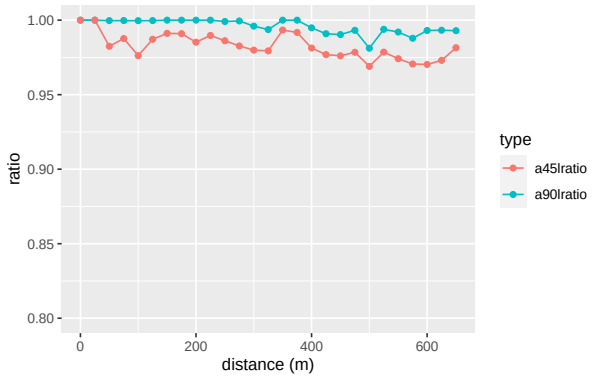


Figure 4.14: Path-length ratio distribution for network B

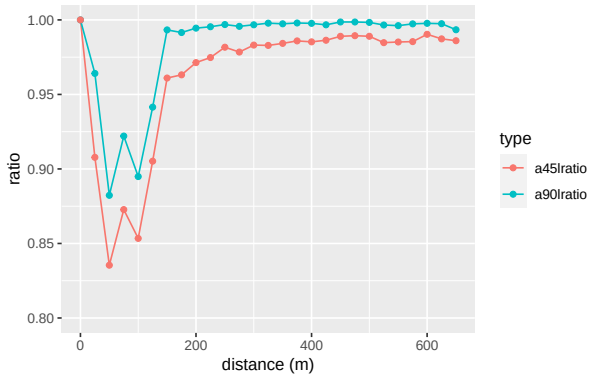


Figure 4.15: Path-length ratio distribution for network C

4.5 Single route examples for length ratios and crossing count differences

In this section, example routes are presented which show certain correlations between characteristics of the footpath network and the length reductions for *a45* and *a90* types. First, four routes are selected for each network, each with the lowest length ratio per distance group ($<$ 200 m, 200 $<$ 400 m, 400 $<$ 600 m, and $\geq$ 600 m). For each group, the route with the lowest length ratio for the *a45* type is selected. This is followed by a closer look at four example routes, two for the largest crossing count difference (+2) and two for the lowest one (-1).

Figure 4.16 (p. 56) shows the four routes for *network A* with the lowest *a45* length ratios per distance group. Route 1) shows a typical case for a very short route that starts directly at a freely crossable road section and ends at a close location at the opposite side of the street. Length reduction for routes including non-dedicated mid-block crossings is especially remarkable here when the next dedicated crossing causes a relatively large detour. This counts for both the *a45* and the *a90* cases, with a less smaller length reduction effect for the *a90* type—as is the case for all other example routes for this network. Route 2) includes the same direct street crossing, followed by another one that is used to reach an inner-block footpath not providing any dedicated crossings at the point where it is connected to the street network. Route 3) includes the same non-dedicated crossings as route 2), but includes a third directly before the end of the route at a T-shaped intersection with missing crossings across the through-street. This leads to a completely different original route. 4) shows a route to the southern direction where, once again, a T-shaped intersection with missing crossings causes a detour for the original route where a dedicated mid-block crossing at a certain distance has to be used.

In figure 4.17 (p. 57), four routes for *network B* providing the lowest *a45* length ratios per distance group are presented. Route 1) is a relatively short route including one non-dedicated mid-block crossing before reaching the destination. The length-reduction effect is only remarkable for the *a45* type here, as the non-dedicated crossing for the *a90* type only leads to a minor reduction compared to the dedicated crossing that is part of the *original* route. Route 2) uses an inner-block footway through a shopping mall as shortcut for the *a45* and *a90* types. The length reduction for both relevant types is caused by a T-shaped intersection close to the destination that does not provide dedicated crossings via the through-street, leading to a completely different *original* route. The same T-shaped intersection is also causing the length reduction for routes 3) and 4).

Figure 4.18 (p. 58) shows the four routes for *network C* with the lowest *a45* length ratios per distance group. Route 1) is another case for a very short route crossing a residential street that can be freely crossed, causing a comparably large detour for the *original* route. Route 2) shows a typical case for length reduction that is caused by the 45-degree crossing angles for the *a45* type and by slightly longer dedicated crossings compared to non-dedicated crossings for the *a90* type. Length reductions normally are quite small here but can become remarkably larger when the layout of dedicated crossings causes detours. Routes 3) and 4) contain the same cases but also provide a "double crossing" of the *a45* type across a curved street, as the inner side of the curve is shorter compared to the outer side, additionally providing a street crossing that causes a small detour for the other route types.

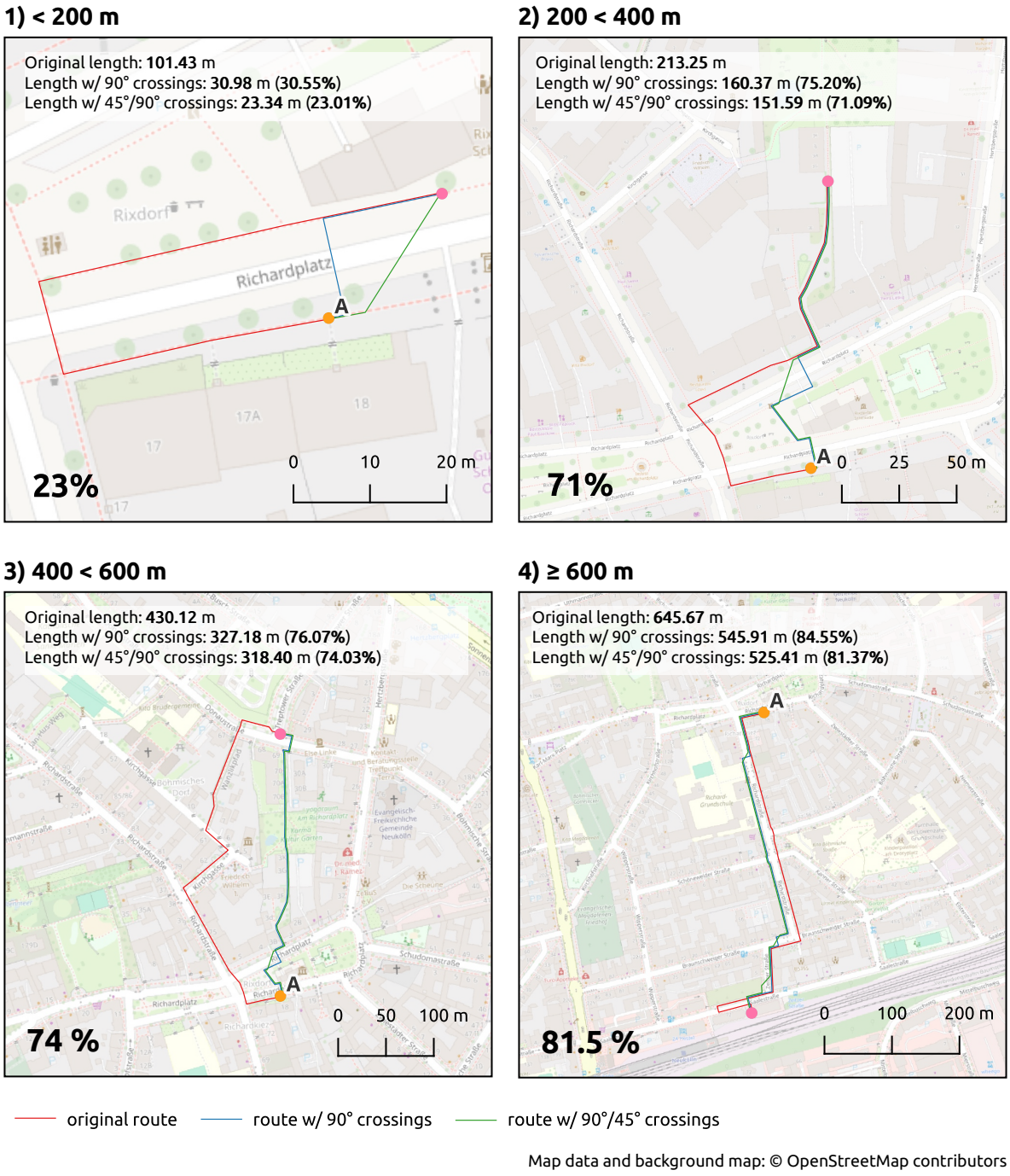


Figure 4.16: Example routes with lowest length ratios per 200 m distance group for *network A*

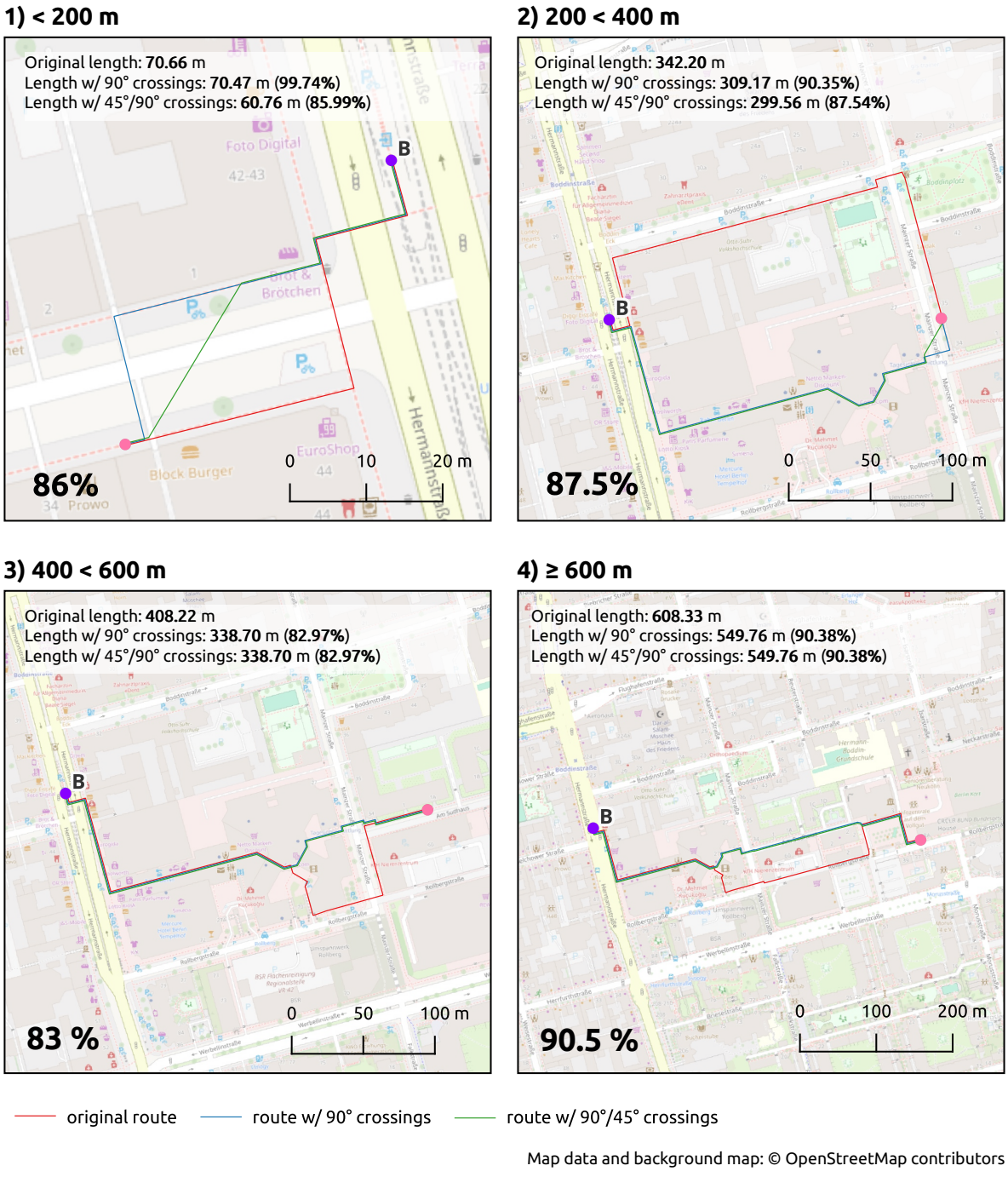


Figure 4.17: Example routes with lowest length ratios per 200m distance group for *network B*

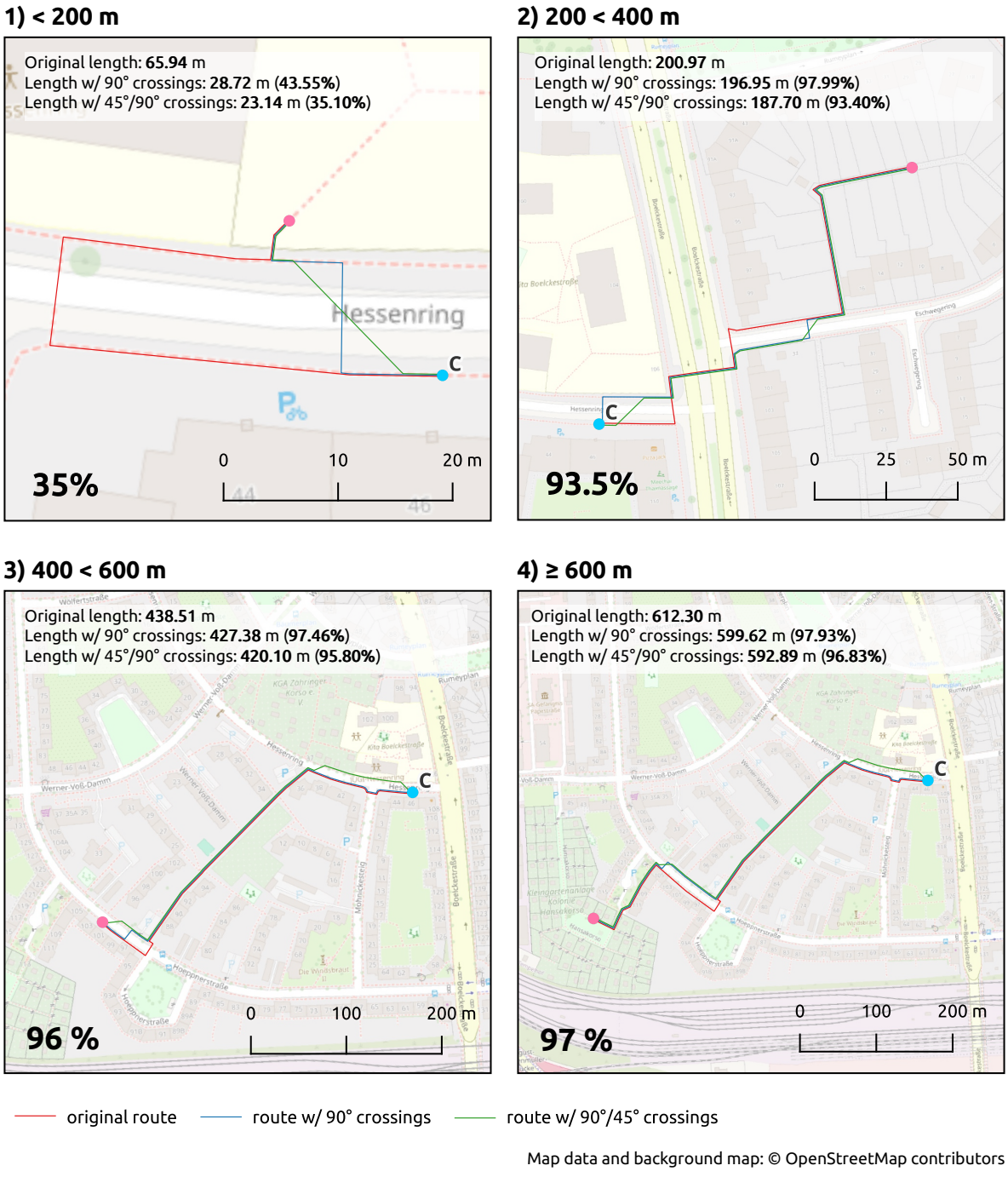


Figure 4.18: Example routes with lowest length ratios per 200m distance group for *network C*

The four route examples illustrated in figure 4.19 (p. 60) show typical characteristics of routes with additional or less street crossings for the *a45* and partly also the *a90* type compared to the *original* routes. Route 1) shows a case where two street crossings are added for the *a45* type, leading to the sectional usage of the opposite street side to the respective *original* and *a90* routes. This section contains an additional T-shaped crossing, otherwise this case would only add one street crossing. Route 2) illustrates a case where three non-dedicated mid-block crossings and no dedicated crossings are used for the *a45* type. The non-dedicated crossings are located very close to each other, caused by the closely consecutive sharp turning angles of the route. The *a90* type contains one dedicated and two non-dedicated crossings here, which—like for the *a45* case—leads to a resulting route that may be a bit shorter but looks far more complicated than the *original* route. Even though many other route example show cases for the *a45* and *a90* types that people who are used to cross mid-block might use, it is likely that also this group would rather take the much more comfortable *original* route in this case. Route 3) shows a case where the *a45* and *a90* types contain one crossing less than the *original* route, caused by an extra T-shaped crossing within the *original* route that runs parallel. The case illustrated with route 4) appears several times within the selected routes. Here, the additional crossing of the *original* route is only caused by crossing at two sides of a T-shaped intersection at the end of the route, whereas the *a45* and *a90* routes cross the through-street prior to the crossing and do not use any of the crossings at the respective T-shaped intersection. In general, the reduction of crossings by the *a45* and *a90* routes is rather quasi-random and of low relevance.

4.6 Intersection crossing indices

As pointed out when analyzing the example routes within the previous section, missing dedicated crossings at T-shaped intersections are causing detours for the *original* routes in several cases. The completeness of intersections with regards to pedestrian crossings is therefore examined in order to determine whether this is a meaningful indicator for path length reductions due to non-dedicated mid-block crossings.

The incomplete provision of T-shaped intersections with crossing geometries in the OSM database even for the comprehensive Neukölln case leads to a profound analysis. The main reason for why the local OSM community decided to not add geometries for unmarked crossings here is based on the regulations of the German Road Traffic Act . Intersection areas must be kept clear up to a distance of 5 m from the points where the edges of the roadway meet[20]. However, this does not include the corresponding areas at the opposite side of the merging streets at T-shaped intersections. Thus, crossing the through-street usually is physically blocked by parked cars here. Figure 4.20 (p. 61) illustrates this situation.

The model for intersection crossing index calculation described in section 3.6.7 is used to analyze all X-shaped and T-shaped intersections of residential roads including living streets that are contained within the three test networks.

[20]StVO, Straßenverkehrs-Ordnung vom 6. März 2013, zuletzt geändert 28. August 2023, §12 (3) 1. Online available at https://www.gesetze-im-internet.de/stvo_2013/BJNR036710013.html, last accessed: 2024-03-09.

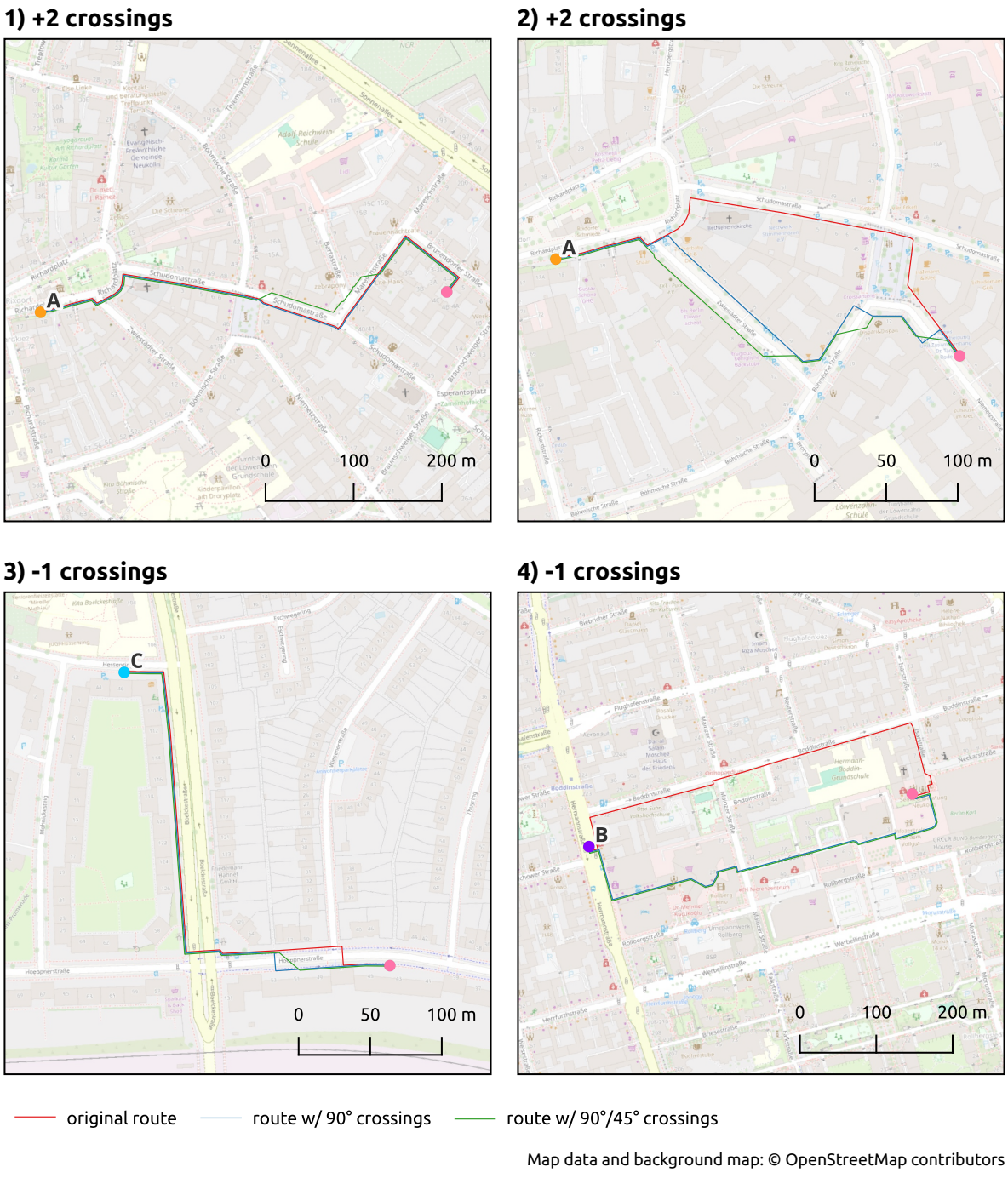


Figure 4.19: Example routes with maximum and minimum crossing count differences

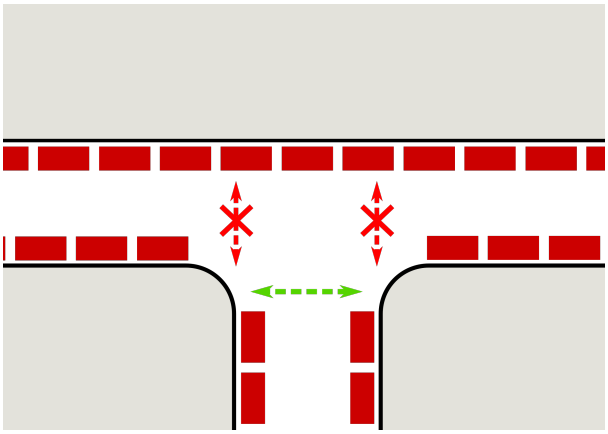


Figure 4.20: Scheme of a T-shaped intersection showing the crossings via the through-street blocked by parked cars

Table 4.11 shows the summarized results for the intersection crossing type and index calculation.

Table 4.11: Crossing counts and mean crossing index per network

network	intersections count	X intersections count	T intersections count	T inersections %	intersection with index < -1	mean index
A	49	8	41	83.67	15	-0.959
B	29	11	18	62.07	4	-0.483
C	41	7	34	82.93	2	-0.293

Network A has an average intersection crossing index of around -0.96, which is remarkably lower than that of networks B (-0.48) and C (-0.29). *Network B* includes the lowest number of relevant intersections. Here, the percentage of T-shaped intersections is 62%, which is lower than that of *network A* (83.7%) and *network C* (82.9%). *Network C* has the lowest average intersection crossing index and also the lowest number of T-shaped intersections with an index < -1 with a value of 2 compared 4 for *network B* and the remarkably higher value of 15 for *network A*.

Figure 4.21 (p. 63) shows the distribution of intersection crossing indices within *network A*. Several intersections with a low index (orange and red) are indeed located at crucial positions relatively to the selected start point, with some of them obviously relevant for path-length reductions of the *a45* and *a90* routes.

Figure 4.22 (p. 64) illustrates the distribution of intersection crossing indices within *network B*. Within the south-western part of the network, crossings are complete at almost all intersections. Only within the eastern part of this test network there are a few intersections with relevantly low indices. Also, the eastern part provides only very few relevant crossings. Especially the T-shaped intersection marked in red (due to the only crossing not marked as such within the test data set) causes relevant detours of the *original* route in some cases, also part of the example route analysis above.

Figure 4.23 (p. 65) shows the distribution of intersection crossing indices within *network C*. Within this network, there are practically no T-shaped intersections with a relevantly low index of -2 or less, indicating that these intersections are "complete" with regards to crossing geometries within the used OSM data set. One X-shaped intersection (which actually is a T-shaped intersection, as it contains one street with separately mapped directional roadways) has a very low value and causes detours in some cases. Two T-shaped intersections with low values in the south-east are of minor relevance here but might be important for other start points. Another important factor here is the generally lower crossing density of *network C* compared e.g. to *network A*.

4.7 Dependency of street grid direction

As *network B* is based on a much more regular street grid than the other two networks, it is worth examining the dependency between route direction and street grid direction. The results in table 4.12 and table 4.13 show that for routes that run mainly diagonal to the main directions of the street grid, path lengths and non-dedicated crossing counts are remarkably lower for routes based on the *a45* graph compared to routes that run parallel to the street grid. For the *a90* graph, there is no remarkable effect on path lengths, but the mean crossing count difference still is substantial. However, the quasi-random integration of 90 degree crossings already pointed out earlier has to be taken into account here.

Table 4.12: Network B direction comparison for a90 type

type	mean length ratio	mean non-ded. mid-bl. crossings count
parallel	0.994	0.138
diagonal	0.999	0.206

Table 4.13: Network B direction comparison for a45 type

type	mean length ratio	mean non-ded. mid-bl. crossings count
parallel	0.989	0.338
diagonal	0.979	1.075

Figure 4.24 (p. 66) shows the routes for network B that run approximately parallel to the main street grid directions and figure 4.25 (p. 67) shows the mainly diagonal routes. Although the figures described above indicate that the amount of non-dedicated mid-block crossings is remarkably higher for the diagonal routes, this does not become clear from a visual point of view when comparing the two maps, as many parts of the included network still provide irregularities and especially in the eastern part there are only few fully crossable residential streets included.

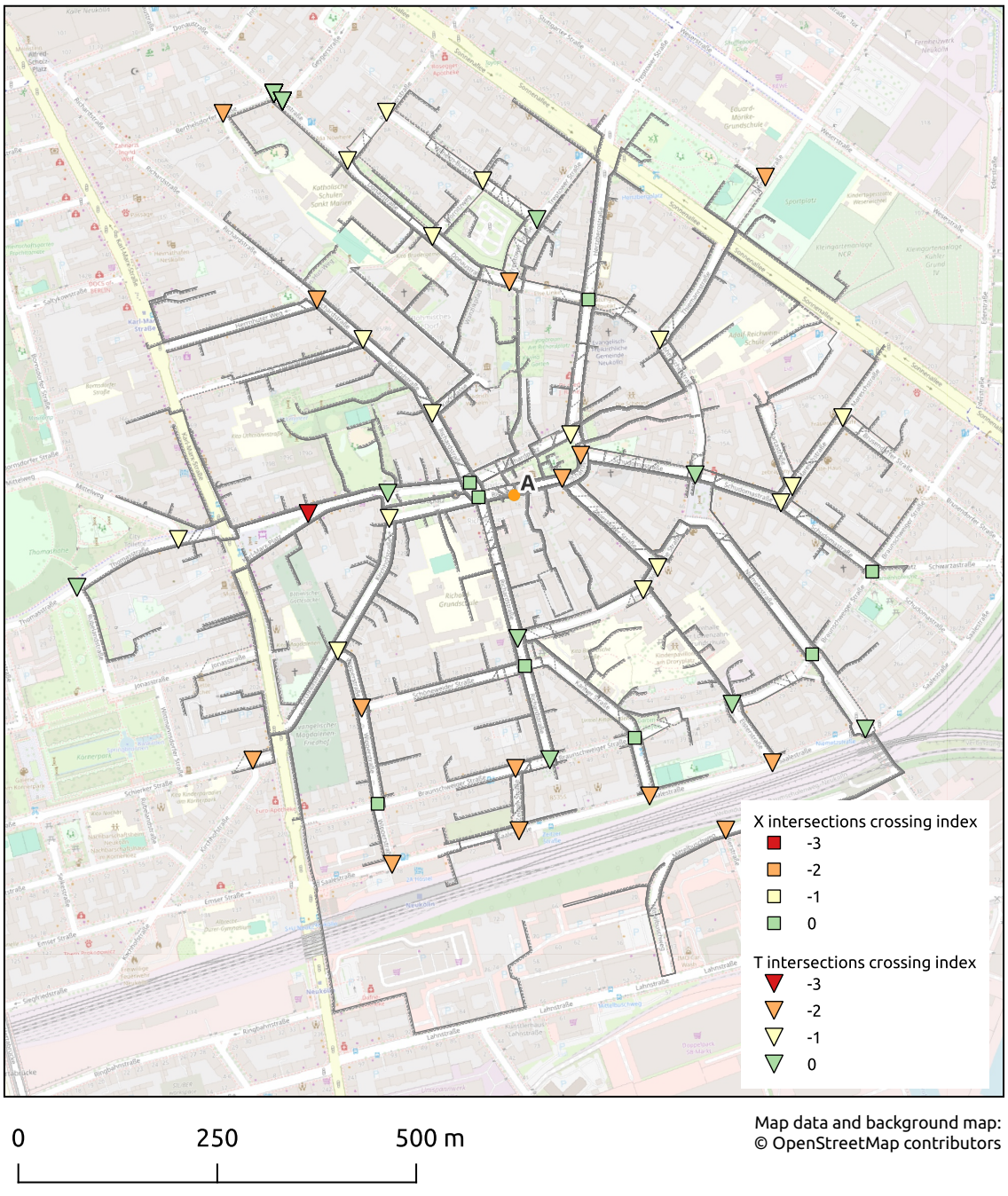


Figure 4.21: *Network A* routes with intersection crossing indices

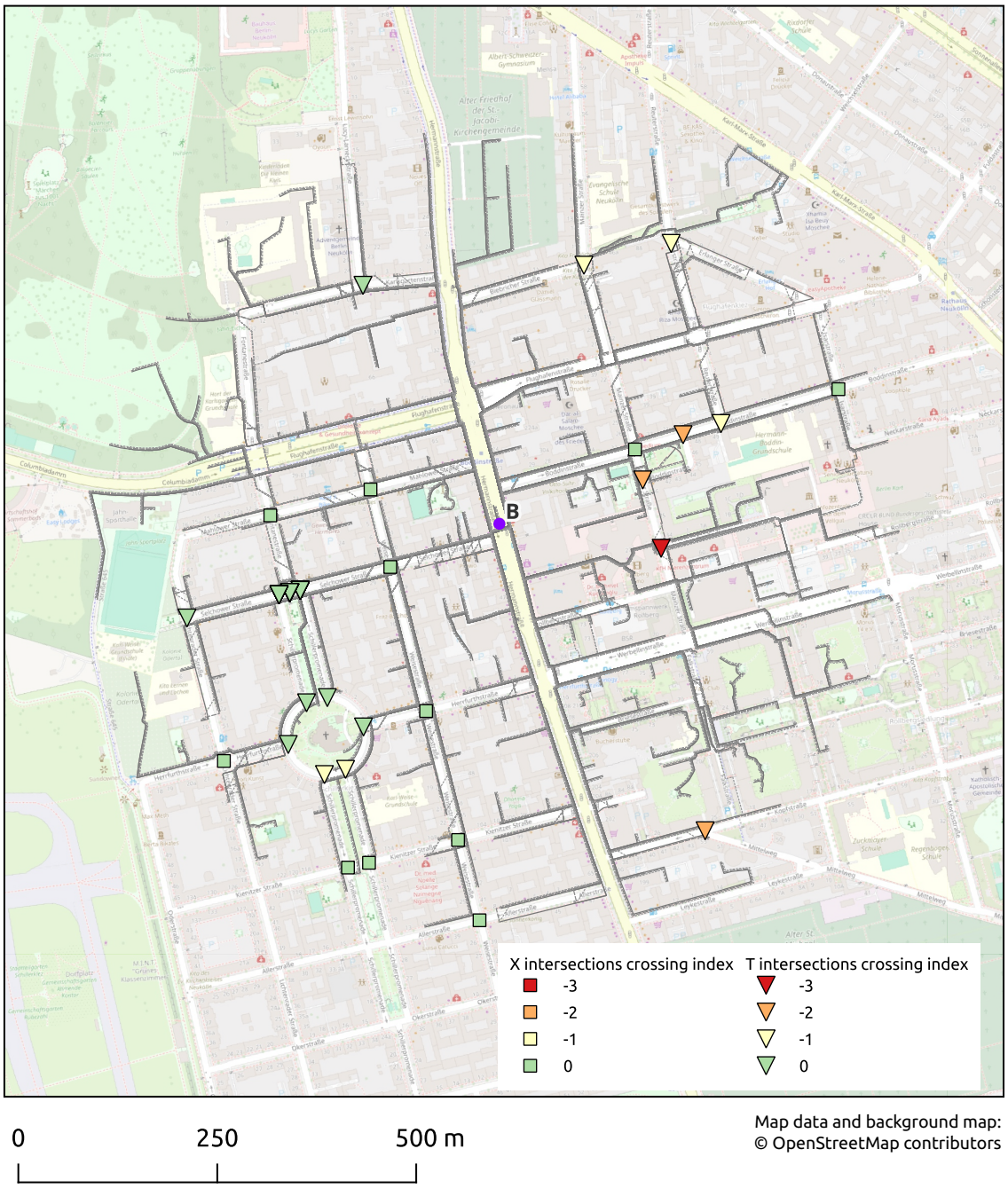


Figure 4.22: *Network B* routes with intersection crossing indices

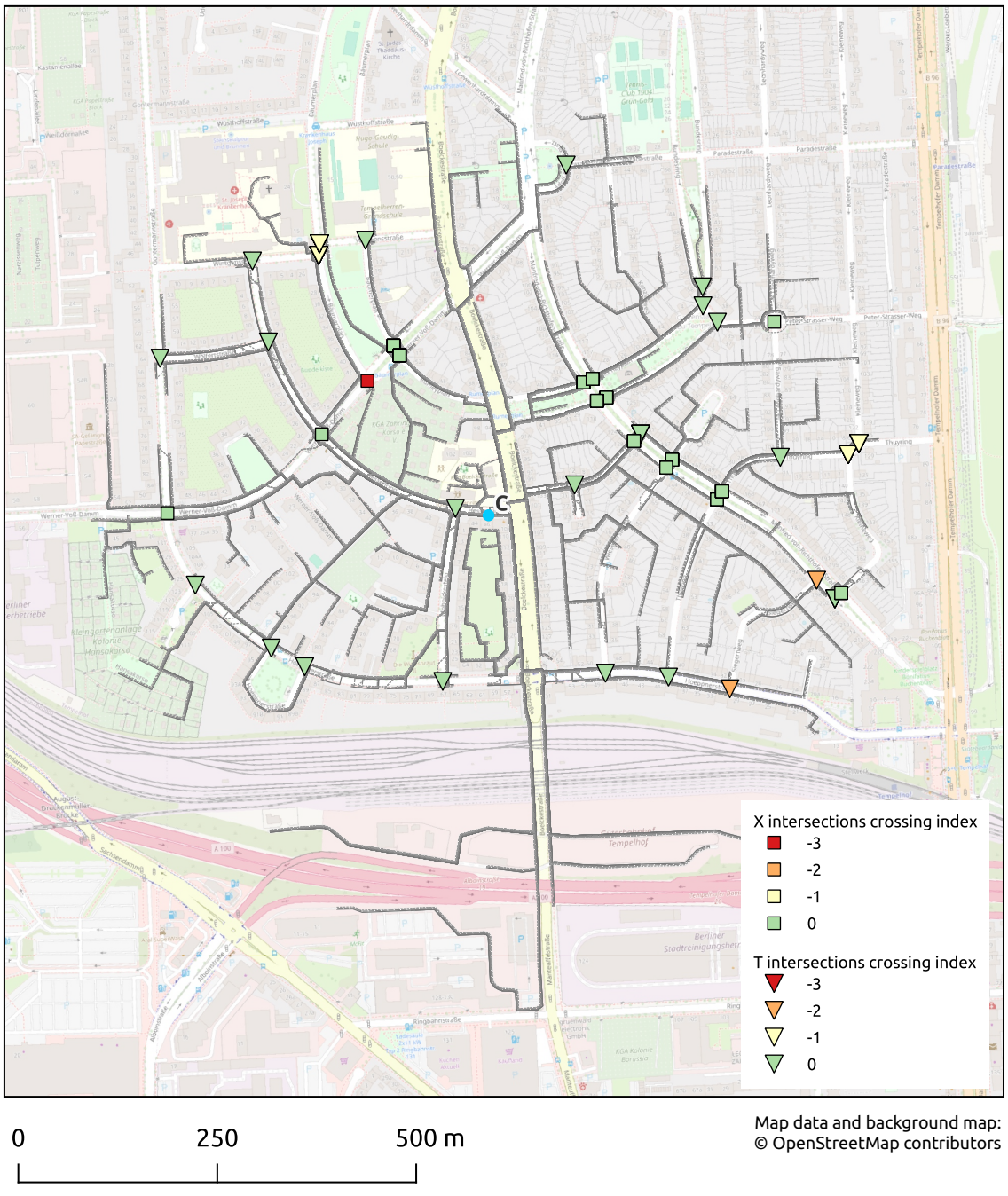


Figure 4.23: Network C routes with intersection crossing indices

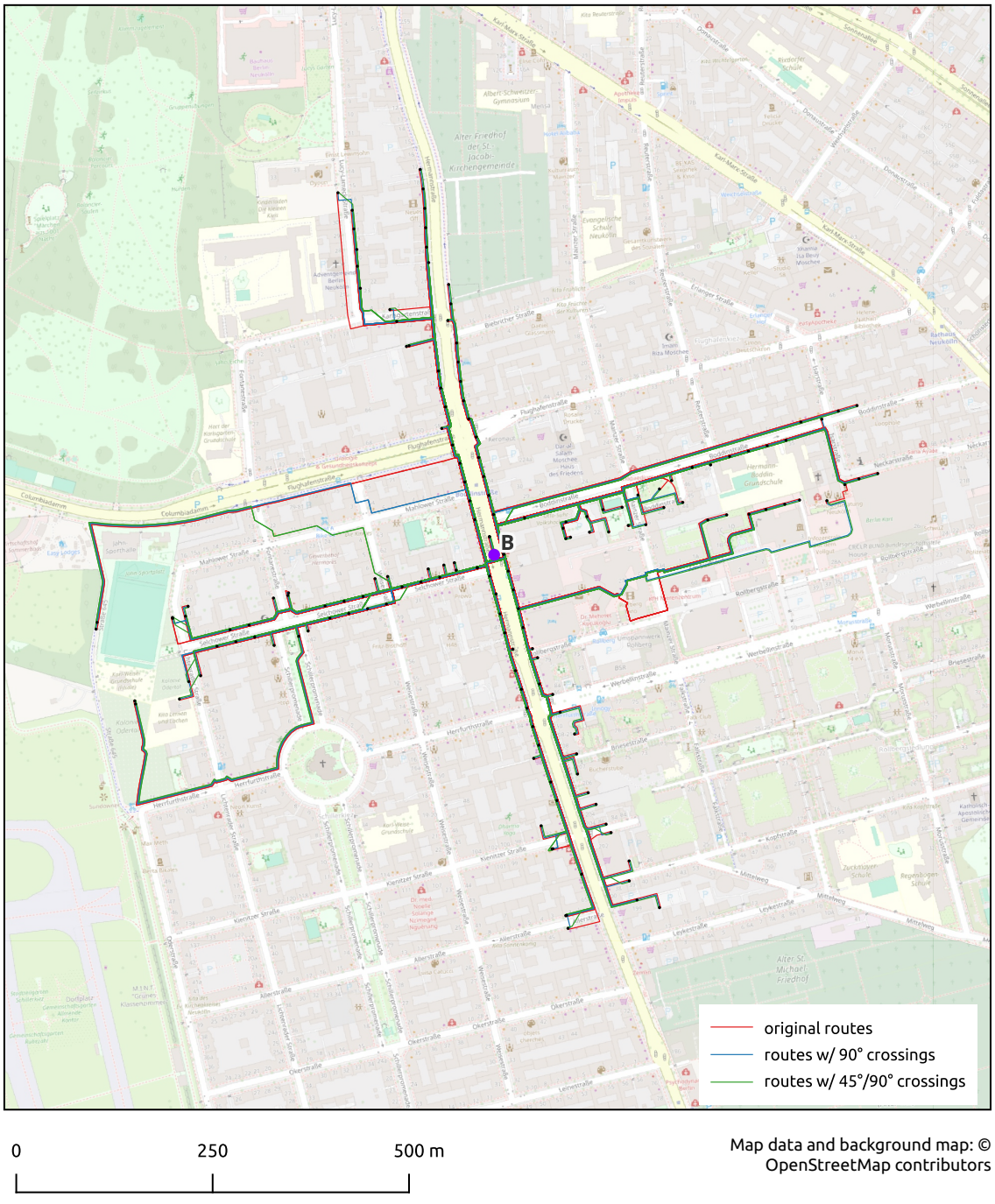


Figure 4.24: Network B routes parallel to main street grid directions

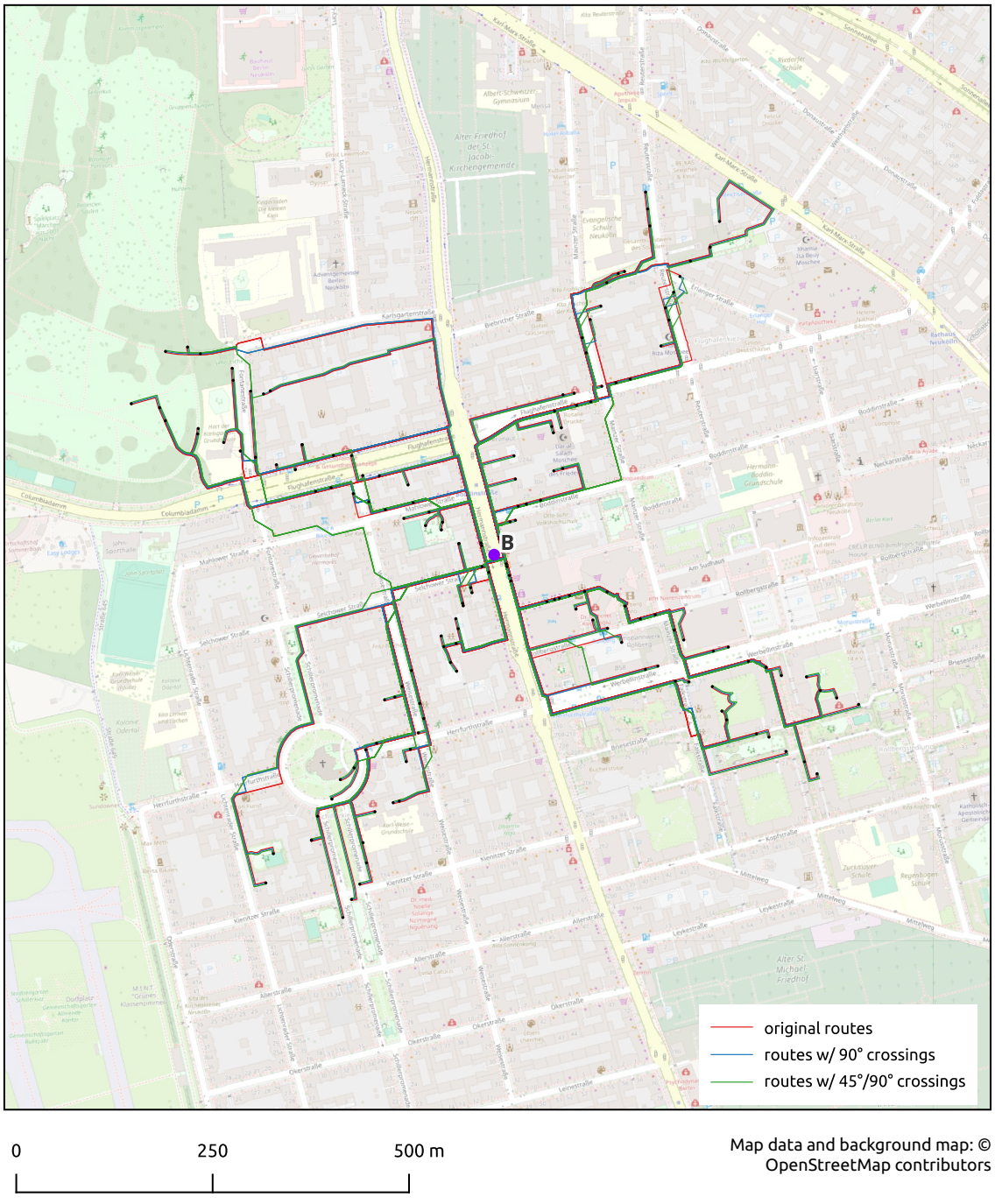


Figure 4.25: Network B routes diagonal to main street grid directions

4.8 Summarized analysis

The results of the Neukölln and Tempelhof OSM use cases show that with non-dedicated mid-block crossings included, routes become remarkably shorter only in specific cases. The average length reduction is about 2% for the *a90* type and 4% for the *a45* type, which still may be relevant for pedestrians who tend to cross mid-block where possible at a suitable cost. If the selected routes start next to a freely crossable residential street, route lengths get shorter for small distances (< 200 m for the tested networks) for both the *a90* and the *a45* type, with some resulting routes that have only about 25% of the length compared to the *original* route.

The additional analysis of the crossing counts has shown that for the tested distances (mainly up to 600 m with some longer routes) there is a maximum of 2 additional crossings in some cases for the *a45* type and in even less cases also for the *a90* type. In most cases, crossing count differences vary between -1 and 1, where the analysis of selected routes shows that this is mainly due to T-shaped intersections within sections where the different route types use opposite sidewalks along a residential street. A second reason for an increased number of crossings for the *a45* and *a90* types are sharp crossing angles of residential streets or curved residential streets within the routes, adding one crossing in most cases for the tested route lengths and two crossings in very few cases.

An analysis of the included intersections of residential streets has shown that these values depend on the intersection crossing index that represents the "crossing completeness" of the intersections (see section 3.6.7), which is relatively low for *network A*. For the other networks, "crossing completeness" is high and average path-lengths reductions are low, suggesting that this is indeed an important indicator to explain the much more remarkable average path-length reductions for the irregular *network A*, especially with regards to T-shaped intersections that often do not provide crossing geometries via the through-street within the respective part of the tested OSM network data set.

An additional test has shown that within *network B*, which provides the most regular test network, the mean number of non-dedicated mid-block crossings is remarkably higher for the *a45* type when extracting routes that run diagonal to the street grid than when only parallel routes are taken into account. This also counts for average path-length reductions of the *a45* type. For the *a90* this only counts partly for the average non-dedicated crossing numbers, but not for average length ratios.

5 Discussion

This work has examined the effects of two types of additional non-dedicated mid-block crossings added to an original OSM graph with regards to route lengths and crossing numbers. The results presented and analyzed in the previous chapter suggest that the overall effect of the additional crossings both for length reduction and number of street crossings is not highly remarkable, but becomes important in a large number of cases when specific start points, route distances, network layouts and route directions are considered.

In this chapter, the previously analyzed results are discussed in a broader context in order to cast the light on limitations of the selected approach and to extract relevant aspects and approaches for future research. First, the focus is on the parameters analyzed in the previous chapter: route length, crossing numbers, network layout, and route direction. Second, a closer look is taken on the relevance of the two selected types of non-dedicated mid-block crossings. Third, different aspects for each of the developed models and the underlying approaches are reflected. Finally, an outlook is given.

5.1 Selected measurements

This section discusses the results for the selected measurements with regard to analyzed effects and limitations. It is important to note here that the results are always considered relative to the extracted digital routing graph and a comparison with the actual physical reality is not part of the study.

The hypothesis that length reductions for routes including non-dedicated mid-block crossings are highest for short routes approximately at average street block length is confirmed. However, there are several cases where even for the longer tested routes the length reduction is high ($>$ 10%), especially for network A, providing the most irregular structure. Also, as expected, there is a tendency to more street crossings for routes with non-dedicated mid-block crossings included. However, validation for longer pedestrian routes than tested within this work is still pending and should be included in future work. The hypothesis that network irregularities lead to high path length reductions also for several longer routes is confirmed, with incomplete crossing provision at T-shaped intersections as an important factor revealed when analysing the results in detail.

5.1.1 Route length

The results show that the effect of route length is remarkably different for added 90-degree crossings or combined 45-degree and 90-degree crossings. The integration of 90 degree mid-block crossings already results in different routes in some cases, compared to the original ones. In other cases, routes stay the same, except that perpendicular crossings are included at mid-block positions instead at intersections in some cases. However, route lengths are not or only minimally altered in most situations, with perpendicular mid-block crossings included that result in routes that often are only a few centimeters or millimeters shorter than the original ones. Remarkable shortcuts enabled by the perpendicular crossings can be observed especially at the start or end of the route, if located at a residential street that provides non-dedicated mid-block crossings. In contrast, the addition of 45-degree crossings already saves a higher amount of distance because

crossing streets in such an angle almost always is a shortcut, except when points directly opposite on the other side of the street need to be reached. With 45-degree crossings, non-dedicated mid-block crossings are always preferred against perpendicular crossings at intersections when the shortest route runs along a freely traversable street section that needs to be crossed to reach the destination.

In summary, the results have shown that many very short routes relying only on original crossing geometries in the area of residential streets become remarkably shorter when non-dedicated mid-block crossings are added, especially when the routes start next to a freely crossable street section. For longer routes (> 200m) that usually are more relevant for typical routing and wayfinding cases, the shortening effect of the added street crossings is not substantial in most cases. However, several factors may lead to much shorter routes with high length reductions even for larger distances, and the results suggest that missing crossing geometries at intersections are an important indicator here. Within the test area, crossing geometries via the through-street at T-shaped intersections are missing in numerous cases.

Parameterization of the selected model partly influences the exact route lengths and especially route-length ratios of compared original and extended graphs. Besides exclusively defining all residential streets as freely crossable, adding non-dedicated crossings at 10 m intervals and not closer than 15 m to original crossing geometries within the model partly influences the results. It can be assumed that a reduction of these numeric parameters, especially the reduction of the distance value to original crossing geometries, would minimally increase mean path-length reductions for the graphs including additional crossings, but the fundamental findings would not change.

Maximum considered route lengths and the consideration of only certain road network types in densely populated urban areas are limitations that future studies should overcome. To address the issue of incomplete intersections in terms of crossing geometries, adding informal crossings on all intersection legs that do not provide dedicated crossings as proposed by Cambra et al. (2019) might be worth examining. Also, Naumann et al. (2019) propose to build a graph exclusively based on sidewalk geometries and geometrically added crossings and route length effects of such an approach seem worth examining.

5.1.2 Number of crossings per route

The results show that within the tested networks, the number of crossings per route varies between -1 and +2 when comparing the original route to the one with perpendicular or combined 45-degree and perpendicular crossings added. However, more than 92% of the routes based on added perpendicular crossings and almost 85% based on the combined crossing directions provide the same crossing numbers as their counterparts based on the original graph. The tested distances of only up to 600 m for the largest part of the routes are an important limitation here and it would be worth examining the effect for longer pedestrian itineraries. Also, with the analysis of selected routes, it has been shown that the values are highly influenced by crossings with minor importance where pedestrians have right of way, mostly via roads merging at T-shaped intersections not contained in all three versions of the route. This reduces the relevance

of the results, but still reveals a tendency towards more street crossings, as there are far more resulting routes with a positive crossing count difference than with a negative one.

Apart from testing with longer distances, a more fine-grained examination seems highly valuable for future work. The proportion of dedicated and non-dedicated crossings within the routes is worth a closer look here. But this counts also for the distinction between primary and secondary road crossings within the tested routes. The distinction between these two crossing types has been proposed by Lassarre et al. (2012). Primary crossings are those that occur perpendicular to the trip roadway by crossing it, at intersections or mid-block, while crossings parallel to the trip direction via merging streets are called secondary crossings. It is likely that additional primary crossings at non-dedicated mid-block positions have a greater effect on the speed, risk and quality of the route than secondary crossings, but this should be verified as part of future work.

5.1.3 Crossing completeness of intersections

The completeness of crossings at intersections has been selected as an indicator for length differences between the routes based on the original graph and those that include additional street crossings. The results have shown that the density and number of intersections that do not have crossing geometries on all sides is very likely to be a decisive factor in influencing route lengths when additional crossings are added. Length reductions for the routes including additional crossings are lower, especially for large tested distances, when less intersections with incomplete crossing geometries are included, as shown with the relevant comparison of the three test networks. This negative correlation between length reduction and intersection crossing indices—0 for "complete" intersections, negative values according to the number of intersection legs that do not provide a crossing geometry—was shown using a simplified model that could be further refined in future work.

Another point that deserves in-depth consideration in future work is that route-length comparison of an original graph and a graph with included non-dedicated mid-block crossings can be used, with a slightly adapted method and model, to measure connectivity of a pedestrian network with separate sidewalk geometries. With such an approach, the role of street space as barrier for pedestrian connectivity could be examined, especially when also adding informal crossings at all intersection legs with missing formal crossings to compare with the theoretical state of maximum connectivity.

The effect of other indicators on route-length ratios is also worth testing, such as street block sizes, intersection density, mean distances between crossings, number and positions of dedicated mid-block crossings, or the amount of footpaths through block interiors. Additionally, route length in many cases is not the crucial and sole factor for pedestrian route selection. Future work should also take other factors for route choice such as time and comfort into account when further investigating the importance of non-dedicated mid-block crossings, also casting the light on different types of dedicated street crossings.

5.1.4 Route direction

The hypothesis underlying the idea of considering the main route direction in relation to the street grid direction is that the number of informal crossings increases sharply for very regular rectangular road networks when routes run diagonally to the street grid, especially when 45-degree crossings are included. In this way, the greatest possible approximation to the straight line diagonally to the road network is achieved.

This effect could only be demonstrated to a limited extent using the test data set, as even the test network with the greatest regularity is still partly irregular and the tested routes do not contain a large number of "diagonally crossed" street blocks.

Nevertheless, with these limitations in mind, the results suggest that this applies primarily to 45-degree crossings and that the number of informal crossings is actually remarkably higher than for routes that run parallel to the road network. It would be worth to further examine this connection based on a given test area with a fully regular street grid or a fictional regular network.

5.2 Relevance of the examined crossing types

Comparably small reductions in length for the case with perpendicular crossings were an expected result and the reason why a second variant was selected that allows angled crossings, as this represents in many cases a more "natural" form of crossing low-traffic streets mid-block. This section discusses the relevance of the investigated crossing types on the basis of the analyzed results.

The hypothesis that 45-degree intersections lead to a generally higher length reduction is confirmed, as is the hypothesis that quasi-random positions are chosen for non-dedicated mid-block crossings along freely crossable street sections—with the limitations of this study partly restricting general applicability. There are, however, additional findings and considerations to be discussed that go beyond the confirmation of the hypotheses.

5.2.1 Perpendicular crossings

When looking at the length-related results for the 90-degree variant in isolation, it does not seem valuable to integrate corresponding crossings in a continuous form into a pedestrian routing graph used for wayfinding purposes. It is more obvious here to use adequate parameters of the respective network and street space to extract road sections where the addition of corresponding non-dedicated mid-block crossings has a major effect, for example roads with large distances between designated crossings.

If the network contains intersections with mainly complete crossing geometries, the relative length-reduction of the routes for the perpendicular type tends to be particularly low. It can be assumed that this also counts for generally short distances between crossings, e.g. because of small block sizes or a high amount of dedicated mid-block crossings.

If many incomplete intersections are present, a possible reduction in the threshold for adding crossing geometries at intersections, especially across the through street at those in a T-shape, may be a useful step prior to adding non-dedicated mid-block crossings.

The results have shown that within larger parts of the routes, additional 90-degree crossings are selected in a "quasi-random" form, often influenced by minor differences between the length of dedicated crossings and non-dedicated mid-block crossings across the same street. For these portions, additional cost factors are necessary to control the prioritization of these crossings. An integration then only appears to be valuable when the costs of non-dedicated mid-block crossings are significantly lower then those of dedicated crossings. The determination of these cost factors should be target-group dependent and should also be adjustable according to personal preferences within an appropriate routing software.

The results also show that additional perpendicular crossings often lead to shortcuts especially at the beginning and at the end of the routes, depending on the network layout at the respective positions. An alternative approach here would be to check for possible shortcuts especially at both ends of the route and to ignore perpendicular crossings for the section in between. However, such an approach requires alternative technical solutions, such as post-processing after route generation, as it is not possible with the common algorithms to identify which nodes and edges belong to the beginning and ending section of a route prior to calculating it.

It therefore seems more valuable to include cost factors which, based on certain geometric network parameters, also depend on how likely and beneficial non-dedicated mid-block crossings are at the respective positions.

5.2.2 Combined 45-degree and perpendicular crossings

Due to their geometric configuration, 45-degree crossings are a shortcut per se in most cases. Although not legally correct, this way of crossing low-traffic streets at non-dedicated positions mid-block can therefore be assumed to be practiced in many cases.

It is obvious that the integration of such crossings into a graph intended for routing is not useful, as long as crossing a street in other than the perpendicular direction is legally incorrect. However, this type of graph may be used for other purposes, such as analysis of actual pedestrian itineraries. In any case, adequate cost values are necessary here to create realistic routes representing the actual behavior of target groups who tend to cross streets this way. In networks containing many curves and sharp turning angles, the amount of non-dedicated mid-block crossings compared to dedicated crossings tends to be especially high with 45-degree crossings included. The results have shown this with regards to "double crossings" e.g. to avoid the longer outer side of a curved street. This requires adequately adjusted cost factors, as it can be assumed that the threshold value for crossing streets at non-dedicated mid-block positions is not exclusively determined by the path length. This adjustment requires further research dedicated to low-risk non-dedicated mid-block crossing behavior.

The integration of non-dedicated mid-block crossings into publicly used routing graphs is more likely to happen in a perpendicular form. Allowing for an individual setting with specifically

low cost values for perpendicular non-dedicated mid-block crossings is likely to lead to results that are close to those presented here for the type including 45-degree crossings.

5.3 Considerations for graph generation

The QGIS graphical model that has been developed and implemented for generating the different routing graphs as described in chapter 3 contains several basic considerations and technical steps that need to be discussed in terms of consistency and transferability.

5.3.1 Extraction of formally walkable features

Extracting all sections of officially walkable streets and paths from a given OSM data set requires a complex query because there is a large variety of tagging schemes and mapping styles. In several cases more than one tag exists that can be applied for a specific purpose and tag selection depends on individual preferences of the mappers or on local "quasi-standards", i.e. preferred tags of the local mapping community. In some cases, mappers can choose between tags with different levels of implicitness, e.g. `sidewalk=separate` is correct and usually used for roadways where parallel sidewalks have been mapped separately, whereas `sidewalk:both=separate` may be used for the same purpose, but explicitly indicates that separate sidewalk geometries exist on both sides of the roadway.

For the developed model, the query to extract all pedestrian edges from the full `highway=*` polyline layer has been developed in an iterative approach, including visual and technical checks of the resulting graph and step-wise enhancements[21].

The final version still leads to errors within the study area data set. The most important one is *isolated network sections*, i.e. edges not connected to the main part of the network, mainly for the following reasons:

- Intersections that contain streets with separately mapped directional roadways often provide short street sections connecting both directional roadways. Within the test area, these short sections often do not provide a value for the `foot` tag, where `foot=no` or `foot=use_sidepath` would exclude these sections from the resulting graph.
- In some cases, inner-block footpaths, mostly for private use, are only connected to the street network by using the private interior of the building. These sections are usually not mapped as footpath geometries in OSM, resulting in isolated network sections. As these sections usually provide an `access=private` tag, the inclusion of this tag would eliminate the error in most cases, as the corresponding paths would not be included. However, excluding all private network sections would also remove network sections relevant for reaching addresses within the network.
- Access roads and driveways mapped as `highway=service` with different additional tags are usable by pedestrians in most cases. However, these are not always connected to separately

[21] The final attributive expression can be found in section Example geometric and attributive expressions in the appendix.

mapped sidewalk geometries, which is especially relevant when separate sidewalks exist only on one side of the street, opposite to the merging `service` highway. Additionally, the decision to erase all parts of access roads between the sidewalk and the roadway that they merge into leads to isolated network sections in some cases.

Another error appears in some cases that is relevant for crossing counts within the routes. This is caused by actual mapping errors:

- The Neukölln OSM data is of a comparably very high quality for all elements with street-space relevance. Only in very few cases, ways are not correctly tagged because they have not been split at positions where tagging changes, e.g. `footway=crossing` continued for a still connected way that requires `footway=sidewalk` tag.
- In very few cases, crossings are not marked as such, with `crossing` tags completely missing.

This shows a positive side effect of the developed model, namely the identification of mapping errors in relation to the respective aspects.

The resulting graph was considered usable even with the weaknesses mentioned, as these could be either eliminated by subsequent steps or by minor manual corrections. Still, the pedestrian network query is specifically adapted to the Neukölln and Tempelhof use cases and may need to be altered to apply the model within another test area with a suitable OSM data level of detail, where the network layout and mapping styles may significantly vary.

5.3.2 Areal features

The model ignores areal pedestrian features, because in most cases linear features have been mapped in OSM within the test area to represent the main pedestrian itineraries freely crossable open spaces, e.g. across squares. Visual and random routing checks of the resulting graph have shown that the ignorance of areal features does not lead to relevant gaps within the network. No network gaps could be found within the Neukölln test networks and only a few cases could be found within the Tempelhof data set at positions not influencing the resulting test routes.

As the literature review has shown, the integration of areal features, i.e. open spaces, into a pedestrian network graph has been subject to numerous studies. A comprehensive pedestrian network graph should therefore include solutions in terms of realistic path options for both open spaces and fully or partly crossable street spaces and further research should examine this combined approach.

5.3.3 Bridges and tunnels

Bridges and tunnels have not been taken into account, as the test area does not contain any bridges where relevant path or road segments cross each other. The area contains a few pedestrian tunnels at metro stations, resulting in a few cases where planar line crossings at different vertical layers, included as attributive information, have been interpreted as actual intersections. These cases—namely at several metro and suburban railway stations—did not influence the result, as these locations are not part of the tested routes.

However, including bridges and tunnels may be highly relevant in other test areas, particularly where height differences in the terrain or certain building structures lead to crossings of relevant linear infrastructure elements at different vertical levels. The inclusion of bridges and tunnels is therefore a valuable future enhancement of the model.

5.3.4 Crossing types

Existing crossings, i.e. crossings that have been mapped in OSM as actual geometries, have been included into the graph generation model without taken crossing types into account. Crossings are included, simply by checking if the respective network part is tagged as `footway=crossing`, ignoring the `crossing=*` tag. The only selective step here was to exclude street crossings that can be used exclusively by cyclists. This simplified approach was chosen here in order to limit the complexity of the study.

The inclusion of crossing types is in any case a reasonable further enhancement. However, as table 5.1 shows, even the highly detailed Neukölln OSM street space data is still not complete at this point[22], with around 45 percent of all crossings missing the `crossing` tag.

Table 5.1: OSM crossing type distribution for Neukölln

type	count	percent
marked	13	1.01
traffic_signals	213	16.59
uncontrolled	7	0.55
unmarked	464	36.14
zebra	12	0.93
[no crossing tag]	575	44.78

Additionally, 36 percent of the crossings in Neukölln district in Berlin are tagged as `crossing=unmarked`. The OSM Wiki tag page[23] proposes to only use this type of crossing when the position is recognizable by a structural measure. At the same time, it is pointed out that the tag is often used for cases where there are no structural indications of a dedicated crossing, but a geometry has been added merely as an obvious extension of the sidewalk lines across the roadway. This also applies for Neukölln, with a large number of cases not even providing a lowered curb. This might also have been motivated by the fact that street-side parking is not allowed at such positions and crossing is generally possible, at least for "pedestrians without wheels". This does not count for street-side parking along the through-street opposite of the merging street at T-shaped intersections, as this area is not to be kept clear according to the applicable German traffic laws. Accordingly, crossing geometries via the through-street have not been added at T-shaped intersections in many cases. Integrating such crossings as geometries tagged with `crossing=informal` as the "lowest value" (i.e. "highest cost") crossing

[22] Date of issue (download of Neukölln `highway=*` data set used within this work): 2023-10-21.
[23] https://wiki.openstreetmap.org/wiki/Tag:crossing%3Dunmarked, last accessed: 2024-02-06.

type is an option that is discussed by members of the local OSM group in Berlin[24] and would significantly change the results, as the observed irregularities in the length ratios are expected to be leveled out at greater routing distances.

Different OSM mapping strategies and translation of real-world crossing layouts into digital representations are therefore another factor that limits the transferability of the results to other test areas, especially with regards to unmarked and informal crossings. Thus, the results highly depend on the decision if and how crossings have been added as linear geometries. This is especially relevant for crossing positions that are obvious from a geometric point of view, but do not provide any physical structure that indicates that an unmarked or informal crossing exists.

Future work should therefore integrate crossing types, e.g. with different values when calculating with specific cost factors other than path length to analyze fastest or most comfortable route options. This, of course, depends on detailed available data for the respective test area within the OSM database or within additional data sources.

5.3.5 Crossing graph generation

The selected geometric expressions to generate the additional crossing geometries at a given interval along residential streets providing separately mapped sidewalks on both sides, with perpendicular or combined 45-degree and perpendicular crossings every 10 m and a minimal distance of 15 m to existing crossing geometries, led to very few errors. These were not considered relevant for the tests to be conducted with the resulting graph.

One error that could be identified in very few cases is the addition of crossing geometries at the wrong street side at intersections where not all merging streets provide separate sidewalk geometries on both street sides. It would be worth to further investigate this error for a future enhanced version of the graph generation step.

In addition, when generating the crossing geometries, it was decided to first give them a certain length and then to trim the resulting lines down to the relevant sidewalk geometry. The length value is included as a customizable parameter in the model, but required an iterative approach, as too long lines may lead to undesired connections to neighboring streets and too short lines may prevent the generation of crossing geometries along the relating sections. This is therefore another point that must be taken into account when transferring the model to other test areas, as distances between the roadway and the sidewalks may vary to a much larger extent than within the Neukölln and Tempelhof data sets.

One solution, which requires further detailed consideration with regard to the geometric expression, would be to project the required end points directly onto the sidewalk geometries and then create the connection from the corresponding point on the roadway to the projected points.

[24] This has been pointed out in an e-mail discussion in February 2024 with two of the authors of Lingner et al. (2023) (A. Seidel and T. Jordans).

5.3.6 Performance issues

Testing the generated routing graphs for performance was not part of this study. Within the technical environment used for this work, graph generation and test routes calculation could be conducted within reasonable time. However, it is obvious that this depends on the one hand on the size of the network, on the other hand the performance of the technical environment and the actual algorithms used are highly relevant.

Table 5.2 shows that the number of features and vertices increases by a high amount for the routing graphs including additional crossings. The original graph used for comparison here is not the original OSM graph, but a graph prepared for routing with lines split at all intersections, but not split in between.

Table 5.2: Comparison of feature and vertice counts for the three network types

type	features	vertices	features ratio	vertices ratio
original	16117	41564	1.00	1.00
a90	26005	73131	1.61	1.76
a45	34627	107551	2.15	2.59

The numbers suggest that this becomes critical for larger networks. Instead of preparing the full graph for routing with additional crossings, a post-processing approach that adds the additional crossings only where relevant for a pre-calculated route may be a valuable solution.

Additionally, the models used here contain some redundancies, i.e. repeating steps with minor differences. As the QGIS model builder does not offer a possibility to repeat a group of consecutive processing steps for different layers, this results in redundancies in the model layout.

A solution to prevent these redundancies would be to turn the model into a Python-based processing script and to alter the result by adding the respective loops. Another solution would be to split the model into different sub-models. However, the QGIS graphical modeler has recently undergone steady improvements and it is hoped that reusable groups of processing steps will finally be possible in the future.

5.4 Considerations for test routes generation

The QGIS graphical model developed and used for test routes generation has proven to work as expected with no known errors. However, there are a few general considerations for the underlying method that need to be discussed. Also, as the model is based on the intermediate step of turning the network to be used into a single service area, the downsides of this step are worth a closer look.

5.4.1 General methodological considerations

The *Concentric Circular Destination Points* (CCDP) method led to useful results in this case. Generating numerous test routes from one strategically positioned start point helps on the

one hand to examine the effect of selected network configurations close to the beginning of the route. On the other hand, the quasi-random distribution of destination points helps to identify correlations between network configurations and routing results that were not obvious beforehand.

However, there are a few downsides of the method that should be reflected. One important point here is that the method can only be applied in a meaningful way within a dense street network, not containing major gaps in any direction from the selected start point. If the dense network spans only to certain directions, reducing the network point generation to certain angles may help to partly overcome this issue. Still, the problem persists that the method is not suitable for networks containing major holes, as this leads to a accumulation of network points around the edge of the hole. Also, it has been accepted that the method leads to identical routes in several cases and duplicate test routes have not been removed.

Another problem is that the method does not allow for the control of route length. Even for network points with a small Euclidean distance to the start point, the route length can be very long. Also, with some important barriers within the selected area, single routes may be much longer than the selected maximum Euclidean distance, as large detours are required. Altering the method to control route length, such as generating a service area around the start point and using a selection of routes in different directions that provide the selected length is a possible solution here.

The uneven distribution of destination points, i.e. decreasing density of test points with growing test radii, is also not ideal. However, this problem was not considered relevant. Also, higher density of test points for small test radii helped to generate approximately the same number of test routes for short and long distances.

5.4.2 Start point selection

The method for test-routes generation has been developed in order to examine a large number of routes from strategically selected start points that span across different lengths. Technically, the model developed for the method is only based on one start point. However, part of the method is to not only select one starting point, but several points, in order to be able to look specifically at the setting at network positions with selected properties. In this way, one end of the route can be controlled manually, while the other end of the route has quasi-random configurations due to the selected method.

The results show that for the three selected start points, the measured values noticeably vary in different aspects. Examining strategically positioned start points based on other parameters than the ones selected here, e.g. starting mid-block at a long residential street with no dedicated crossings or generally in an area with far larger distances between dedicated crossings, is still worth a closer look.

5.4.3 Service area as routing graph

As described for the graph generation model, the resulting graph for the Neukölln and Tempelhof test areas contains some isolated parts. Thus, the decision has been taken to remove these isolated parts by calculating a large service area with a distance that is large enough to fully contain all test routes. This produces a connected graph and thus removes the problematic isolated parts that could otherwise contain the quasi-randomly selected destination points.

The main question here is: Does a service area polyline layer in QGIS, based on a fully routable graph, result once again in a fully routable graph? Unfortunately, this must be partly denied, at least for the service area algorithms contained in the QGIS version used. Dangles (i.e. dead ends) within the service area that can be reached with a distance smaller than the selected service area distance produce a line overlap in some cases, because a 180 degree turn is added at the end point of the dead end. Routes that contain this overlapping line result in an unnecessary detour via the end of the dangling line in selected cases, although the underlying rule has not yet been identified. This was a known problem when developing the test routes generation method.

Numerous random checks for route length correctness were conducted. Displaying the resulting routes with a an offset also helped to visually check for routing correctness, as this reveals line overlaps. These steps revealed correct cost values, i.e. length values, for the resulting routes, except for two cases. These have been corrected within the attribute table by replacing the wrong cost value with the correct one, but the related geometry has not been altered in order to be able to visualize the error.

Figure 5.1 shows the two relevant cases where routing based on the service area graph leads to wrong results, i.e. unnecessary detours, in this case along a single driveway centerline.

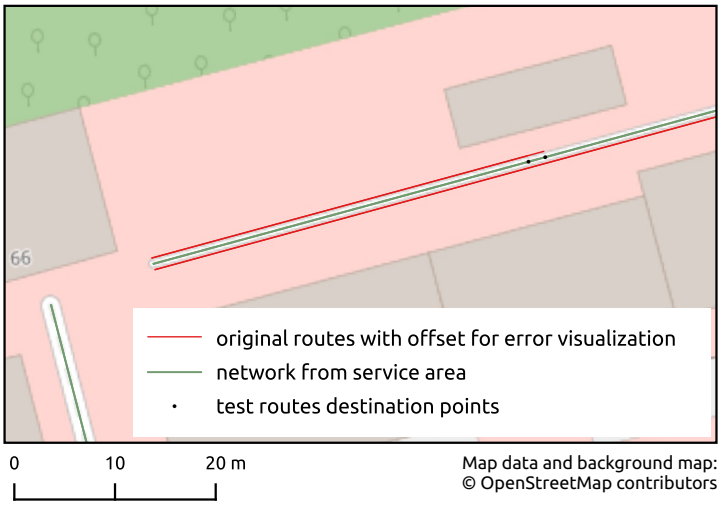


Figure 5.1: Visualization of routing errors from service area graph (unnecessary detours)

A major future enhancement of the selected process would be to either alter the graph generation model in order to fully prevent isolated network edges or to enhance the test routes generation process in order to remove the problematic line overlaps. The problem of possibly overlapping lines within the resulting layer of service area calculation has been added as an issue within the

QGIS Github repository. An alteration of the service area processing tools in order to create a fully routable and correctly measurable graph would be very helpful. The desired output would actually be the original network layer cut to the service area extent, maintaining the original attributes.

Switching to other software solutions such as *PostGIS* and its *pgRouting* extension is also possible, but a technical goal of this study was to examine the current possibilities and limitations of the QGIS environment.

5.4.4 Iteration over distance values with step-wise increment

The applied method for test routes generation contains the positioning of test points on a circle with an adjustable amount of constantly increased azimuth values. A second increment level is a step-wise increase of the radius for test point generation. For testing purposes, i.e. to output each radius as a separate layer, only the azimuth increment has been integrated to the graphical model. To also automatize the radius increment, a batch process has been set up with step-wise increased values (steps of 25 m for the applied use case).

In order to simplify the setup for future use cases, a combination of both increment levels (azimuth and radius) into the model is reasonable. This would require a more complex approach when separate layers are required for each radius. A possible solution is to transfer the model into a Python-based processing script to include the necessary loops. If the inclusion of the test radius as an attribute value within one combined layer is sufficient, the geometry expression can be extended to include an iteration over different test radii and the radius needs to be added as an attribute to the respective points afterwards.

5.5 Considerations for intersection crossing indices calculation

The model to calculate intersection crossing indices has been developed in order to examine one probable indicator for path length reductions with additional crossings included. For the tests to be conducted, the current state of the model was sufficient, but contains a few problems that have to be taken into account.

The model is still in an early test status. Its accuracy was sufficient to show the correlation between path length reductions and the number of "incomplete" intersections, i.e. intersections with missing crossings, in this case.

Known problems of the current state of the model are

- the inability to deal with crossings that are very close to each other
- the unsolved issue of intersections containing separately mapped directional roadways and
- the inability to deal with intersections not providing separate sidewalk geometries on both sides along all merging streets.

Also, the isolated consideration of this indicator has limited significance and must be viewed in the context of other factors such as the number of dedicated mid-block crossings, average block sizes or crossing distances, and intersection density.

5.6 Outlook

The previous sections have discussed specific aspects of the results and model design, also highlighting considerations that could not be integrated to the setup within this study and proposing the integration of these aspects into future studies. Here, a more general outlook is given as a guideline for future research related to the relevance of street space configurations for pedestrian route choice.

5.6.1 Fundamental research on low-risk mid-block crossings

As shown within the literature review, prior work regarding non-dedicated mid-block crossings was focused on behavioral issues and mainly reduced to high-risk mid-block crossings along major urban roads. In order to enable the development of suitable routing solutions including low-risk mid-block crossings for pedestrians, additional fundamental research is needed: Where, how and when do pedestrians cross low-traffic streets mid-block? What are the main motivations for such crossings? Which street-space elements encourage people to cross freely and which prevent them from doing so? What differences exist between specific groups of pedestrians? And especially: Do pedestrians themselves see the need to integrate such crossings into a routing graph?

5.6.2 Fine-grained street traversability calculation

Additional research is also needed for a realistic determination of crossable street sections and for weighted traversability indices, which, of course, must contain other factors than the length of the respective edges. As shown in the literature review, first approaches exist (Naumann et al. 2015), but a more in-depth look is necessary. What different types of barriers—fixed structures or temporary elements—exist and how should they be integrated into a traversability index? How do general layout and geometry of low-traffic streets, e.g. straight versus curved streets or broad versus narrow streets, affect crossability? How relevant are traffic volume and maximum allowed speeds? Answering these questions helps to find separate index values for both side areas of the street on the one hand and for the actual traffic lanes, allowing for a meaningful and fine-grained determination of street space permeability for pedestrians.

5.6.3 Enhanced and comprehensive routing solutions

Obviously, mean pedestrian route lengths are much shorter than those for motorized traffic or, in a less significant form, cyclists. This is also reflected in the network of footpaths, which in urban areas—apart from areas that were built completely car-oriented in the twentieth century—is often much denser than the road network and includes many off-street areas that are not accessible to motorized traffic. If the much-discussed open spaces are also included and sidewalks exist as separate edges in the network, the relative density of paths used by pedestrians is already relatively well taken into account and detailed realistic routing is largely possible.

However, as analyzed and discussed within this study, treating street space as an inaccessible area for pedestrians, except the narrow areas of dedicated crossings, does not fully reflect the actual routes that pedestrians take and are allowed to take. Comprehensive routing solutions for pedestrians should therefore include not only sidewalk-specific geometric data with all necessary

tags relevant for realistic cost values and solutions for shortest or least-cost paths via open spaces, but also allow pedestrians to cross at non-dedicated mid-block crossings where the street space is sufficiently permeable. Applications based on such routing solutions should include individually adjustable parameters, excluding such crossings e.g. for pedestrians with limited mobility who are unable to cross where curbs are not lowered or other relevant barriers exist.

Future work should therefore address the above-mentioned aspects as important elements of enhanced and comprehensive solutions, helping to reach new milestones towards fully suitable and pedestrian-friendly routing.

5.6.4 Towards comprehensive pedestrian network data

Improved and realistic routing solutions of course depend on the availability of comprehensive footpath network data. The case of Neukölln is an exception so far, revealing a large potential for the digital and routing-friendly mapping of all relevant street-space elements. Only a few other urban and suburban areas in OSM worldwide come close to this level of detail. But other areas with a similar fine-grained data structure to Neukölln are sure to emerge in the near future, allowing to select test areas within other types of urban fabric and with different street patterns. Contributing to high-resolution street space data in OSM normally is not a scientific task but that of a very large community of voluntary mappers—and thanks to many of them studies like this one exclusively based on OSM data are possible. Researchers, however, contribute indirectly to the continuously improved quality of the OSM database due to their scientific valorization of this data base—by using it and by pointing out its potential.

As previous research has shown, comprehensive pedestrian network data is not only rare within the OSM data base, but there are also very few official comprehensive data sets, offered by municipalities e.g. within their open data portals. It would be desirable to have access to detailed sidewalk data from municipal or federal sources in order to be able to carry out studies like this one based official data. This requires an ongoing scientific effort in finding solutions for the automatic creation of high-quality foot network data including realistic generation of sidewalks and crossings.

6 Conclusion

Depending on different environmental and behavioral factors, pedestrians might cross streets at non-dedicated mid-block positions within their itineraries when risk is low and the benefit is high. Although route length is only one of numerous motivational factors for such street crossings, a closer look at the path-length reduction effect of almost freely crossable residential streets seemed valuable as an isolated component of dealing with such additional crossings for pedestrian routing and network analysis. Additionally to path length, the differences in crossing counts has been examined.

The approach of this study was to compare shortest paths within an original OSM pedestrian network graph extracted for Neukölln and Tempelhof districts in Berlin, Germany, to those within two extended graphs of the same network including different types of non-dedicated mid-block crossings along residential roads: perpendicular crossings and combined bidirectional 45-degree crossings and perpendicular crossings. A model was developed to generate the extended network graphs based on a mixed attributive and geometric approach. For analysis purposes, a large set of test routes from three manually selected start points with different prerequisites within a dense urban fabric was then generated using a self-developed method. The pedestrian crossing completeness of intersections was selected as one of several indicators explaining the differences regarding the results of the three tested networks.

It was found that path-length reduction is highest for short routes, as the additional crossings e.g. allow for reaching an address at the opposite side of the street by direct street crossing instead of a detour via the closest dedicated crossing, located at the closest intersection in most cases. Knowing that such short distances are of low importance for actual routing scenarios, a closer look was taken at longer routes with remarkable path-length reductions. Intersections, especially T-shaped ones, not providing crossing geometries via all merging streets were found to be an important indicator that influences path-length reduction within the tested networks. However, for most longer routes, perpendicular additional crossings lead to no or only minimal path length reductions, whereas including 45-degree crossings leads to more perceptible length reductions because of the shortcuts they provide in the majority of cases.

The main contribution of this study is therefore a detailed length-based analysis of non-dedicated mid-block crossings automatically added into a network graph in two different forms, including a specially developed methodological and technical approach. It was built upon previous works focusing on or including findings for non-dedicated mid-block crossings along low-traffic streets and on several previous studies covering important subareas of the investigation conducted here. The study has some important limitations, namely the simplified assumption of full and exclusively crossable pedestrian streets, maximum path lengths, focus on densely urbanized areas, ignorance of crossing types and the dependency on a level of detail of OSM data currently only available within the tested area. However, it still is an important contribution to a further improvement of pedestrian navigation and GIS-based analysis of footpath networks in order to give even more scientific weight to this most natural, sustainable, and healthy mode of transport.

References

Abdullah, M.; Dias, C. & Oguchi, T. (2022). Road Crossing at Unmarked Mid-Block Locations: Exploring Pedestrians' Perception and Behavior. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, *46*(2), 1681–1698. https://doi.org/10.1007/s40996-021-00701-z

Ahmed, M.; Fasy, B. T.; Hickmann, K. S. & Wenk, C. (2015). A Path-Based Distance for Street Map Comparison. *ACM Transactions on Spatial Algorithms and Systems*, *1*(1), 1–28. https://doi.org/10.1145/2729977

Alghanim, A.; Jilani, M.; Bertolotto, M. & McArdle, G. (2021). Leveraging Road Characteristics and Contributor Behaviour for Assessing Road Type Quality in OSM. *ISPRS International Journal of Geo-Information*, *10*(7), 436. https://doi.org/10.3390/ijgi10070436

Andreev, S.; Dibbelt, J.; Nöllenburg, M.; Pajor, T. & Wagner, D. (2015). Towards realistic pedestrian route planning. In *15th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2015)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Baltes, M. R. & Chu, X. (2002). Pedestrian Level of Service for Midblock Street Crossings. *Transportation Research Record: Journal of the Transportation Research Board*, *1818*(1), 125–133. https://doi.org/10.3141/1818-19

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*(2), 127–141. https://doi.org/10.48088/ejg.a.bar.13.2.127.138

Bauer, C.; Almer, A.; Ladstätter, S. & Luley, P. M. (2014). Optimierte Wegefindung für Fußgänger basierend auf vorhandenen OpenStreetMap-Daten. *Angewandte Geoinformatik*, *2014*, 408–413.

Bolten, N. & Caspi, A. (2021). Towards routine, city-scale accessibility metrics: Graph theoretic interpretations of pedestrian access using personalized pedestrian network analysis. *PLoS One*, *16*(3), e0248399. https://doi.org/10.1371/journal.pone.0248399

Bolten, N.; Mukherjee, S.; Sipeeva, V.; Tanweer, A. & Caspi, A. (2017). A pedestrian-centered data approach for equitable access to urban infrastructure environments. *IBM Journal of Research and Development*, *61*(6), 10–1.

Cambra, P. J.; Gonçalves, A. & Moura, F. (2019). The digital pedestrian network in complex urban contexts: A primer discussion on typological specifications. *Finisterra*, *54*(110), 155–170. https://doi.org/10.18055/Finis16414

Corcoran, P.; Mooney, P. & Bertolotto, M. (2013). Analysing the growth of OpenStreetMap networks. *Spatial Statistics*, *3*, 21–32. https://doi.org/https://dx.doi.org/10.1016/j.spasta.2013.01.002

Dijksta, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269–271.

Dinu, M.; Pagliai, G.; Macchi, C. & Sofi, F. (2019). Active Commuting and Multiple Health Outcomes: A Systematic Review and Meta-Analysis. *Sports Medicine*, *49*(3), 437–452. https://doi.org/10.1007/s40279-018-1023-0

Dzafic, D.; Klug, S. & Franke, D. (2015). Routing über Flächen mit SpiderWebGraph. *AGIT Journal für Angewandte Geoinformatik*, *1*, 516–525.

Ek, K.; Wårell, L. & Andersson, L. (2021). Motives for walking and cycling when commuting – differences in local contexts and attitudes. *European Transport Research Review*, *13*(1), 46. https://doi.org/10.1186/s12544-021-00502-5

Elias, B. (2007). Pedestrian Navigation - Creating a tailored geodatabase for routing. In *Navigation and Communication 2007 4th Workshop on Positioning* (pp. 41–47). https://doi.org/10.1109/WPNC.2007.353611

Fonseca, F.; Fernandes, E. & Ramos, R. (2022). Walkable cities: Using the smart pedestrian net method for evaluating a pedestrian network in Guimarães, Portugal. *Sustainability*, *14*(16), 10306. https://doi.org/10.3390/su141610306

Funke, S.; Schirrmeister, R. & Storandt, S. (2015). Automatic Extrapolation of Missing Road Network Data in OpenStreetMap. *MUD'15: Proceedings of the 2nd International Conference on Mining Urban Data*, *1392*, 27–35.

Gaisbauer, C. & Frank, A. U. (2008). Wayfinding Model For Pedestrian Navigation. *AGILE 2008 Conference-Taking Geo-Information Science One Step Further, University of Girona, Spain*, *9*.

Goodchild, M. F. & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120. https://doi.org/10.1016/J.SPASTA.2012.03.002

Gössling, S.; Choi, A.; Dekker, K. & Metzler, D. (2019). The Social Cost of Automobility, Cycling and Walking in the European Union. *Ecological Economics*, *158*, 65–74. https://doi.org/10.1016/j.ecolecon.2018.12.016

Govinda, L. & Ravishankar, K. (2023). Pedestrian Gap Acceptance Behavioral Modelling at Midblock and Uncontrolled Intersections. *U.Porto Journal of Engineering*, *9*(1), 150–159. https://doi.org/10.24840/2183-6493_009-001_001113

Goyal, A.; Mogha, P.; Luthra, R. & Sangwan, N. (2014). PATH FINDING: A* OR DIJKSTRA'S? *International Journal in IT & Engineering*.

Graser, A. (2016). Integrating open spaces into OpenStreetMap routing graphs for realistic crossing behaviour in pedestrian navigation. *GI_Forum*, *4*(1), 217–30. https://doi.org/10.1553/GISCIENCE2016_01_S217

Graser, A.; Straub, M. & Dragaschnig, M. (2014). Towards an Open Source Analysis Toolbox for Street Network Comparison: Indicators, Tools and Results of a Comparison of OSM and the Official A ustrian Reference Graph. *Transactions in GIS*, *18*(4), 510–526. https://doi.org/10.1111/tgis.12061

Guío-Burgos, F. A.; Combariza-Pinzón, M. J. & Cerquera Escobar, F. Á. (2022). Pedestrian gaps and walking speed at uncontrolled midblock crosswalks. *Revista Facultad de Ingeniería Universidad de Antioquia*. https://doi.org/10.17533/udea.redin.20220371

Hahmann, S.; Miksch, J.; Resch, B.; Lauer, J. & Zipf, A. (2018). Routing through open spaces–a performance comparison of algorithms. *Geo-Spatial Information Science*, *21*(3), 247–256. https://doi.org/10.1080/10095020.2017.1399675

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, *37*(4), 682–703.

Hillnhütter, H. (2016). *Pedestrian access to public transport.* University of Stavanger, Norway.

Hosseini, R.; Tong, D.; Lim, S.; Sun, Q. C.; Sohn, G.; Gidófalvi, G.; et al. (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*, *12*(7), 288. https://doi.org/10.3390/ijgi12070288

Karimi, H. A. & Kasemsuppakorn, P. (2013). Pedestrian network map generation approaches and recommendation. *International Journal of Geographical Information Science*, *27*(5), 947–962. https://doi.org/10.1080/13658816.2012.730148

Koritsoglou, K.; Tsoumanis, G.; Patras, V. & Fudos, I. (2022). Shortest Path Algorithms for Pedestrian Navigation Systems. *Information*, *13*(6), 269. https://doi.org/10.3390/info13060269

Koukoletsos, T.; Haklay, M. & Ellul, C. (2011). An automated method to assess data completeness and positional accuracy of OpenStreetMap. In *GeoComputation* (Vol. 3, pp. 236–241).

Lassarre, S.; Bonnet, E.; Bodin, F.; Papadimitriou, E.; Yannis, G. & Golias, J. (2012). A GIS-based methodology for identifying pedestrians' crossing patterns. *Computers, Environment and Urban Systems*, *36*(4), 321–330. https://doi.org/10.1016/j.compenvurbsys.2011.12.005

Li, H.; Cebe, J.; Khoeini, S.; Xu, Y. "Ann".; Dyess, C. & Guensler, R. (2018). A Semi-Automated Method to Generate GIS-Based Sidewalk Networks for Asset Management and Pedestrian Accessibility Assessment. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(44), 1–9. https://doi.org/10.1177/0361198118757981

Lingner, L.; Seidel, A. & Jordans, T. (2023). Parkraumanalyse für deine Stadt mit OpenStreetMap. *FOSSGIS-Konferenz 2023: Berlin 15.-18. März 2023.*

Lynch, K. M. (1960). *The image of the city.* Cambridge: The MIT Press.

Mobasheri, A.; Bakillah, M.; Rousell, A.; Hahmann, S. & Zipf, A. (2015). On the completeness of sidewalk information in OpenStreetMap, a case study of Germany.

Mobasheri, A.; Sun, Y.; Loos, L. & Ali, A. L. (2017). Are crowdsourced datasets suitable for specialized routing services? Case study of OpenStreetMap for routing of people with limited mobility. *Sustainability*, *9*(6), 997. https://doi.org/10.3390/SU9060997

Mobasheri, A.; Huang, H.; Degrossi, L. C. & Zipf, A. (2018). Enrichment of OpenStreetMap data completeness with sidewalk geometries using data mining techniques. *Sensors*, *18*(2), 509. https://doi.org/10.3390/s18020509

Naumann, S. (2018). *PERRON - enhanced pedestrian routing and navigation as well as walkability assessment of pedestrian ways. ERA-NET TRANSPORT III: Future travelling. Laufzeit: 01.10.2015-30.09.2017. Schlussbericht zu den Arbeiten des ifak.* Magdeburg: Institut für Automation und Kommunikation e.V. Magdeburg; https://doi.org/10.2314/GBV:1023540614

Naumann, S. & Kovalyov, M. Y. (2017). Pedestrian Route Search Based on OpenStreetMap. In G. Sierpiński (Ed.), *Intelligent Transport Systems and Travel Behaviour* (Vol. 505, pp. 87–96). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-43991-4_8

Naumann, S.; Czogalla, O. & Kühner, F. (2015). A safety index for road crossing. In *FastZero15: Future Active Safety Technology Towards Zero Traffic Accidents Symposium. Gothenborg, Sweden* (pp. 9–11).

Naumann, S.; Kovalyov, M.; Graser, A.; Straub, M.; Entler, B. & Schwarz, S. (2019). Fußwege in der Mobilitätskette: Neue Ansätze für Routing und Navigation im Umweltverbund. *Der Nahverkehr*, *37*(3), 57–65.

Novack, T.; Wang, Z. & Zipf, A. (2018). A system for generating customized pleasant pedestrian routes based on OpenStreetMap data. *Sensors*, *18*(11), 3794. https://doi.org/10.3390/s18113794

Omar, K. S.; Moreira, G.; Hodczak, D.; Hosseini, M. & Miranda, F. (2022). Crowdsourcing and Sidewalk Data: A Preliminary Study on the Trustworthiness of OpenStreetMap Data in the US. https://doi.org/10.48550/ARXIV.2210.02350

Poggenhans, F. & Janosovits, J. (2020). Pathfinding and Routing for Automated Driving in the Lanelet2 Map Framework. *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 1–7. https://doi.org/10.1109/ITSC45102.2020.9294376

Poggenhans, F.; Pauls, J.-H.; Janosovits, J.; Orf, S.; Naumann, M.; Kuhnt, F. & Mayr, M. (2018). Lanelet2: A high-definition map framework for the future of automated driving. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 1672–1679. https://doi.org/10.1109/ITSC.2018.8569929

Rhoads, D.; Rames, C.; Solé-Ribalta, A.; González, M. C.; Szell, M. & Borge-Holthoefer, J. (2023). Sidewalk networks: Review and outlook. *Computers, Environment and Urban Systems*, *106*, 102031. https://doi.org/10.1016/j.compenvurbsys.2023.102031

Rousell, A. & Zipf, A. (2017). Towards a Landmark-Based Pedestrian Navigation Service Using OSM Data. *ISPRS International Journal of Geo-Information*, *6*(3), 64. https://doi.org/10.3390/ijgi6030064

Sari, F. & Şen, M. (2022). Highway Route Planning via Least Cost Path Algorithm and Multi Criteria Decision Analysis Integration, a Comparison of AHP, TOPSIS and VIKOR. *International Journal of Environment and Geoinformatics*, *9*(2), 27–38. https://doi.org/10.30897/ijegeo.900200

Seidel, A. (2022a). Die Neuköllner Straßenraumkarte - Ein detaillierter Plan des öffentlichen Raumes auf Basis freier OpenStreetMap-Geodaten. *Kartografische Nachrichten*, *2022*(3), 10–17.

Seidel, A. (2022b). Die Neuköllner Straßenraumkarte – ein hochaufgelöster OSM-Mikro-Mapping-Kartenstil. *FOSSGIS-Konferenz 2022: Online 9.-12. März 2022.*

Siriaraya, P.; Wang, Y.; Zhang, Y.; Wakamiya, S.; Jeszenszky, P.; Kawai, Y. & Jatowt, A. (2020). Beyond the Shortest Route: A Survey on Quality-Aware Route Navigation for Pedestrians. *IEEE Access*, *8*, 135569–135590. https://doi.org/10.1109/ACCESS.2020.3011924

Valls, F. & Clua, Á. (2023). Modeling Barcelona sidewalks: A high resolution urban scale assessment of the geometric attributes of the walkable network. *PLOS ONE*, *18*(7), e0284630. https://doi.org/10.1371/journal.pone.0284630

Verma, R. & Ukkusuri, S. V. (2023). Crosswalk Detection from Satellite Imagery for Pedestrian Network Completion. *Transportation Research Record: Journal of the Transportation Research Board*, 03611981231210545. https://doi.org/10.1177/03611981231210545

Walter, V.; Kada, M. & Chen, H. (2006). Shortest path analyses in raster maps for pedestrian navigation in location based systems. *International Symposium on "Geospatial Databases for Sustainable Development," Goa, India, ISPRS Technical Commission IV*, *2006.*

Wanyan, X.; Seneviratne, S.; Nice, K.; Thompson, J.; White, M.; Langenheim, N. & Stevenson, M. (2023). Scalable Label-efficient Footpath Network Generation Using Remote Sensing Data and Self-supervised Learning. https://doi.org/10.48550/ARXIV.2309.09446

Wu, H.; Lin, A.; Clarke, K. C.; Shi, W.; Cardenas-Tristan, A. & Tu, Z. (2021). A comprehensive quality assessment framework for linear features from Volunteered Geographic Information. *International Journal of Geographical Information Science*, *35*(9), 1826–1847. https://doi.org/10.1080/13658816.2020.1832228

Yang, X.; Tang, L.; Ren, C.; Chen, Y.; Xie, Z. & Li, Q. (2020). Pedestrian network generation based on crowdsourced tracking data. *International Journal of Geographical Information Science*, *34*(5), 1051–1074. https://doi.org/10.1080/13658816.2019.1702197

Yannis, G.; Papadimitriou, E. & Theofilatos, A. (2013). Pedestrian gap acceptance for mid-block street crossing. *Transportation Planning and Technology*, *36*(5), 450–462. https://doi.org/10.1080/03081060.2013.818274

Zhang, X.; Wang, T.; Jiao, D.; Zhou, Z.; Yu, J. & Cheng, X. (2021). Detecting inconsistent information in crowd-sourced street networks based on parallel carriageways identification and the rule of symmetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, *175*, 386–402. https://doi.org/10.1016/j.isprsjprs.2021.03.014

Zielstra, D. & Hochmair, H. H. (2012). Using free and proprietary data to compare shortest-path lengths for effective pedestrian routing in street networks. *Transportation Research Record*, *2299*(1), 41–47. https://doi.org/10.3141/2299-05

Appendix

Example Overpass queries

In this work, all OSM data has been queried using Overpass query language on the Overpass Turbo website[25]. With this approach, data can be downloaded from the current state of the OSM database instead of using prepared thematic and regional OSM data sets from common providers

Extraction of all ways tagged as `highway` within test area

The full street traffic network of Neukölln was then exported based on the following Overpass query using the Overpass Turbo website:

```
[out:json][timeout:25];
{{geocodeArea:Berlin-Neukölln}};
way[highway](area);
out geom;
```

`{geocodeArea:Berlin-Neukölln};` has been replaced by `{geocodeArea:Berlin-Tempelhof};` to extract the data for the Tempelhof test network.

Extraction of test area boundary

```
[out:json][timeout:25];
relation["admin_level"=10][name="Neukölln"];
out geom;
```

For the productive case used in this work, no other overpass queries were necessary. However, several tests have been conducted to check for direct Overpass filtering options instead of exporting the full `highway` data set.

Check for `is_sidepath:of` all over Germany

The `is_sidepath:of` tag has been used as an indicator for the Berlin-Neukölln level of detail for street-space mapping in OSM[26].

```
[out:json][timeout:25];
{{geocodeArea:Deutschland}};
```

[25] https://overpass-turbo.eu/

[26] The `is_sidepath:of` tag is an important part of the Neukölln use case. However, a continued usage of this tag and the related `is_sidepath:of:name` tag is unlikely, as members of the Berlin OSM group stated during an online meeting on 2024-03-05. Such a multi-level tag structure is too complicated and should better be resolved using relation in OSM. Additional checks for the suitability of the outdated `associatedStreet` relation type (https://wiki.openstreetmap.org/wiki/Relation:associatedStreet, last accessed: 2024-03-10) where also conducted, but no test areas were found with complete provision of this type of relations for associated sidewalks.

```
way["is_sidepath:of"](area);
out geom;
```

As of March 2024, few municipalities or districts provide the same level of detail as Berlin-Neukölln. Interestingly, there seem to be some corresponding mapping attempts in smaller municipalities. Buckenhof municipality east of Erlangen, Middle Franconia, Bavaria is an example where the `is_sidepath:of` tag has been added for all sidewalks.

Check for `sidewalk` tag provision in selected cities

The following query has been used to analyze the `sidewalk` tag provision in selected cities that have been mentioned in previous works or in ongoing discussions of the OSM community as examples with extensively mapped separate sidewalk geometries. The query outputs CSV data instead of geometries.

This query was part of a step-wise test procedure to find suitable test areas. It was found out that even when separate sidewalk areas exist, the `sidewalk` tags are fully missing or the value has not yet been changed to `separate` after mapping the sidewalks. As the model presented in this work needs an attributive approach here, this was an exclusion criteria.

```
// Define fields for CSV output
[out:csv(
  city,
  "# of roads",
  "# of sidewalk geometries",
  "# of roads with sidewalk info",
  "percentage of roads with sidewalk info (%)",
  "sidewalk to road ratio"
)];

// Select area
(
  {{geocodeArea:Chicago}};
  {{geocodeArea:Seattle}};
  {{geocodeArea:New York City}};
  {{geocodeArea:Boston}};
  {{geocodeArea:Paris}};
  {{geocodeArea:Melbourne}};
  {{geocodeArea:Warsaw}};
  {{geocodeArea:Krakow}};
  {{geocodeArea:Berlin}};
);

// Query relevant tags for all selected cities
foreach->.c(
```

```
  way(area.c)[highway]->.highways;
  way(area.c)[footway=sidewalk]->.sidewalks;
  way(area.c)[highway][~"sidewalk"~"yes|left|right|both"]->.highwaysidewalks;

  // Count tagged features and percentages per selected city
  make count city = c.set(t["name"]),
    "# of roads" = highways.count(ways),
    "# of sidewalk geometries" = sidewalks.count(ways),
    "# of roads with sidewalk info" = highwaysidewalks.count(ways),
    "percentage of roads with sidewalk info (%)" =
      highwaysidewalks.count(ways) / highways.count(ways) * 100,
    "sidewalk to road ratio" =
      sidewalks.count(ways) / highways.count(ways);

  // Output CSV row
  out;
);
```

Tables 1 and 2 show the results of this query as of March 4, 2024. The row order is taken from the query result.

Table 1: Sidewalk tag distribution in selected cities (part 1)

city	# of roads	# of sidewalk geometries	# of roads with sidewalk info
Seattle	139597	48435	3088
Chicago	189579	48197	369
City of New York	319494	93903	4245
Boston	47595	9042	1250
Melbourne	463389	61945	4331
Paris	118977	42994	3914
Berlin	337824	38428	12672
Kraków	106625	7104	1095
Warszawa	231590	18138	394

Table 2: Sidewalk tag distribution in selected cities (part 2)

city	percentage of roads with sidewalk info (%)	sidewalk to road ratio
Seattle	2.21	0.35
Chicago	0.19	0.25
City of New York	1.33	0.29
Boston	2.63	0.19
Melbourne	0.93	0.13

Table 2: Sidewalk tag distribution in selected cities (part 2)

city	percentage of roads with sidewalk info (%)	sidewalk to road ratio
Paris	3.29	0.36
Berlin	3.75	0.11
Kraków	1.03	0.07
Warszawa	0.17	0.08

Check for relevant tags within selected cities of the DACH region

The following query is one of several tests that have been conducted in order to find an example test area within the DACH region (Germany, Austria, Switzerland) that is equally suitable as the Berlin use case. Buckenhof municipality seemed a promising case, but was rejected because of a too small area and relevant network.

```
// Define fields for CSV output
[out:csv(
  city,
  "way[highway]",
  "way[highway=footway]",
  "way[highway=path]",
  "way[footway=sidewalk]",
  "way[foot=use_sidepath]",
  "way[is_sidepath:of]",
  "way[is_sidepath:of=residential]"
)];

// Select area
(
  {{geocodeArea:Berlin}};
  {{geocodeArea:Hamburg}};
  {{geocodeArea:München}};
  {{geocodeArea:Köln}};
  {{geocodeArea:Wien}};
  {{geocodeArea:Zürich}};
  {{geocodeArea:Buckenhof}};
);

// Query relevant tags for all selected cities
foreach->.c(
  way(area.c)[highway]->.highways;
  way(area.c)[highway=footway]->.highwayFootway;
  way(area.c)[highway=path]->.highwayPath;
```

```
  way(area.c)[footway=sidewalk]->.footwaySidewalk;
  way(area.c)[foot=use_sidepath]->.footUseSidepath;
  way(area.c)["is_sidepath:of"]->.isSidepathOf;
  way(area.c)["is_sidepath:of"=residential]->.isSidepathOfResidential;

  // Count tagged features per selected city
  make count city = c.set(t["name"]),
    "way[highway]" = highways.count(ways),
    "way[highway=footway]" = highwayFootway.count(ways),
    "way[highway=path]" = highwayPath.count(ways),
    "way[footway=sidewalk]" = footwaySidewalk.count(ways),
    "way[foot=use_sidepath]" = footUseSidepath.count(ways),
    "way[is_sidepath:of]" = isSidepathOf.count(ways),
    "way[is_sidepath:of=residential]" = isSidepathOfResidential.count(ways);

  // Output CSV row
  out;
);
```

Tables 3 and 4 and show the results of this query as of March 4, 2024. The row order is taken from the query result.

Table 3: Distribution of relevant tags within selected cities of the DACH region (part 1)

city	way[highway]	way[highway=footway]	way[highway=path]	way[footway=sidewalk]
Zürich	54028	27649	3690	5504
München	119460	48817	14617	3591
Buckenhof	1243	348	193	157
Köln	87874	14627	18555	193
Berlin	337824	162196	12431	38428
Hamburg	177213	65639	14004	2118
Wien	122182	56287	7082	6352

Table 4: Distribution of relevant tags within selected cities of the DACH region (part 2)

city	way[foot=use_sidepath]	way[is_sidepath:of]	way[is_sidepath:of=residential]
Zürich	4208	0	0
München	5014	82	73
Buckenhof	108	136	123
Köln	296	3	0
Berlin	31481	8839	3541
Hamburg	1449	2	0

Table 4: Distribution of relevant tags within selected cities of the DACH region (part 2)

city	way[foot=use_sidepath]	way[is_sidepath:of]	way[is_sidepath:of=residential]
Wien	1171	0	0

Example geometric and attributive expressions

This section lists the most important geometric and attributive expressions used within the QGIS Graphic Modeler. Additional linebreaks have been added here to optimize layout and readability.

Extract foot network

This extracts the walkable features from the original layer containing all `highway` polyline features including all OSM tags. Roadways that pedestrians are allowed to use when sidewalks are missing are included if no separately mapped corresponding sidewalk exists.

```
"highway" IN
(
  'footway',
  'path',
  'pedestrian',
  'steps',
  'living_street',
  'track'
)
OR
(
  "highway" IN
  (
    'unclassified',
    'residential',
    'tertiary',
    'secondary',
    'secondary_link',
    'primary',
    'primary_link'
  )
  AND "foot" != 'use_sidepath'
)
OR
(
  "highway" IN
  (
```

```
    'unclassified',
    'residential',
    'tertiary',
    'secondary',
    'secondary_link',
    'primary',
    'primary_link'
  )
  AND
  (
    "sidewalk" IN
    (
      'no',
      'none',
      'left',
      'right',
      'both'
    )
    OR
    "sidewalk:both" IN
    (
      'no',
      'none'
    )
    OR
    (
      "sidewalk" IS NULL AND
      "sidewalk:both" IS NULL AND
      "sidewalk:right" IS NULL AND
      "sidewalk:left" IS NULL
    )
  )
)
```

Extract pedestrian crossings

```
(
  "crossing" <> '' AND
  "cycleway" = ''
)
OR
(
  "footway" IN
```

```
  (
    'crossing',
    'traffic_island'
  )
)
```

Extract sidewalks

```
"footway" = 'sidewalk'
```

Extract roadways

```
"highway" IN
(
  'unclassified',
  'residential',
  'tertiary',
  'secondary',
  'secondary_link',
  'primary',
  'primary_link'
)
```

Extract crossing lines left 90

```
make_line(
    $geometry,
    project(
        $geometry,
        @initial_crossing_length,
        radians("angle" - 90)
    )
)
```

To generate the right-hand part of the 90 degree crossings, `radians("angle" + 90)` has been used.

Extract crossing lines left 45 1

```
CASE
  WHEN @combined = TRUE THEN
      make_line(
        $geometry,
```

```
        project(
            $geometry,
            sqrt(@initial_crossing_length ^ 2) * 1.2,
            radians("angle" - 135)
        )
    )
  ELSE NULL
END
```

To generate the right-hand part, `radians("angle" + 45)` has been used. To generate the other direction of the 45-degree crossings, `radians("angle" - 45)` and `radians("angle" + 135)` have been used correspondingly.

Points left 45 1

This adds points at the intersections of the corresponding sidewalk edge and the newly created crossing lines starting at a roadway 10 m interval node. `eval()` has to be used due to a bug in the expression builder.

```
with_variable(
    'current_id',
    attribute('_temp_id'),
    make_point(
        x(eval(
            'overlay_nearest(
                layer:=@Intersect_sidewalks_left_45_1_OUTPUT,
                expression:=$geometry,
                filter:="_temp_id"='|| @current_id ||',
                limit:=1
            )[0]')),
        y(eval(
            'overlay_nearest(
                layer:=@Intersect_sidewalks_left_45_1_OUTPUT,
                expression:=$geometry,
                filter:="_temp_id"='|| @current_id ||',
                limit:=1
            )[0]'))
    )
)
```

This has been used for the right-hand part as well as the left-hand part and the right-hand part of both the other direction of the 45-degree crossings and the 90-degree crossings correspondingly.

Final crossings 45 1

This generates lines from the respective 10 m interval point at the roadway to the previously created corresponding point at the left-hand sidewalk.

```
with_variable(
    'current_id',
    "_temp_id",
    make_line(
        $geometry,
        geometry(
            get_feature(
                @Points_right_45_1_OUTPUT,
                '_temp_id',
                @current_id
            )
        )
    )
)
```

This has been used for the right-hand part as well as the left-hand part and the right-hand part of both the other direction of the 45-degree crossings and the 90-degree crossings correspondingly.

Create circular points

```
collect_geometries(
    with_variable(
        'radius',
        @test_distance,
        array_foreach(
            generate_series (0, @number_of_test_points - 1),
            with_variable (
                'azim',
                @element * pi() * 2 / @number_of_test_points,
                make_point(
                    x($geometry) + @radius * sin(@azim),
                    y($geometry) + @radius * cos(@azim)
                )
            )
        )
    )
)
```