JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 22 (2021), pp. 1–31

doi:10.5311/JOSIS.2021.22.677

RESEARCH ARTICLE

Towards detecting, characterizing, and rating of road class errors in crowd-sourced road network databases

Johanna Guth[1], Sina Keller[1], Stefan Hinz[1], and Stephan Winter[2]

[1] Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Germany

[2] Department of Infrastructure Engineering, The University of Melbourne, Victoria, Australia

Received: July 23, 2020; returned: October 13, 2020; revised: November 10, 2020; accepted: December 14, 2020.

Abstract: OpenStreetMap (OSM), with its global coverage and Open Database License, has recently gained popularity. Its quality is adequate for many applications, but since it is crowd-sourced, errors remain an issue. Errors in associated tags of the road network, for example, are impacting routing applications. Particularly road classification errors often lead to false assumptions about capacity, maximum speed, or road quality, possibly resulting in detours for routing applications. This study aims at finding potential classification errors automatically, which can then be checked and corrected by a human expert. We develop a novel approach to detect road classification errors in OSM by searching for disconnected parts and gaps in different levels of a hierarchical road network. Different parameters are identified that indicate gaps in road networks. These parameters are then combined in a rating system to obtain an error probability to suggest possible misclassifications to a human user. The methodology is applied to an exemplar case for the state of New South Wales in Australia. The results demonstrate that (1) more classification errors are found at gaps than at disconnected parts, and (2) the gap search enables the user to find classification errors quickly using the developed rating system that indicates an error probability. In future work, the methodology can be extended to include available tags in OSM for the rating system. The source code of the implementation is available via GitHub.

Keywords: OSM, routing, detour, network analysis, gap detection, road hierarchy, GIS

1 Introduction

Road networks worldwide contain an inherent hierarchy of road classes that is linked to the movement needs of vehicles. High-capacity roads such as freeways and highways form the highest level in the road classification hierarchy and are designed to satisfy the highest traffic needs. They are followed by distributor or arterial roads with medium traffic, and then collectors and local access routes, which are lowest in the hierarchy and generally feature a low traffic volume [16].

Therefore, the class of a road is crucial in determining its purpose for the road network. Particularly routing applications often rely on the road class for information about the road network like maximum speed, capacity, or access limitations. Thus, errors in the road class can hinder routing applications and may lead to detours because of false assumptions about travel time or access limitations. These errors may also become obstructive for hierarchical route planning which uses the level of detail appropriate for the task to solve the task with the least amount of effort [38]. Finer levels of detail are not considered [38]. Hierarchical routing algorithms can result in large detours when there is a classification error in a high-level road. Furthermore, as high-level roads are generally more important and sparser than lower-ranked roads, class allocation errors for high-level roads have a larger impact on route planning. The frequency of classification errors in a road network is dependent on the quality of the underlying road data.

OpenStreetMap (OSM) is an example of a worldwide road network dataset prone to road classification errors because the data is collected by volunteers and by donations of agencies and corporations worldwide. Many studies have proven that the quality of the road network itself is adequate in most countries and can often compete with commercial or administrative datasets [3, 6, 10, 11, 15, 17, 25, 26]. However, additional attributes for its edges, such as the road class or the maximum speed, are needed for routing. In OSM, these additional attributes can be added as tags to every road element. The tag road class (*highway* in OSM) is mandatory for every road element such that there is a road class assigned to every edge representing a road. However, due to the crowd-sourced nature of OSM, errors often occur in the assignment of road classes [11, 41]. To improve the navigability of the OSM road dataset, these errors need to be detected and then corrected.

This study aims at finding potential classification errors automatically. Human experts can then check if the detected potential error is an error and, if necessary, correct it. The presented methodology is based on an extension of the definition by Liu [23]. He states that in a hierarchical road network, one can observe that major roads form a network themselves. This subnetwork of major roads is more sparse than the complete network, and while it may not form a connected network in a city, it may form a connected major road network in a state or country [23]. We define a subnetwork as a union of all roads with a level equal or higher to the subnetwork's level. As a result, multiple subnetworks for one road network are obtained with increasing level of detail, the more levels are included. We expand the assumption of Liu [23] and suggest that each subnetwork should (a) be connected and (b) have no gap of the sort that for any pair of origin and destination (OD) in a subnetwork, the shortest route in the subnetwork is significantly longer than in the complete network.

Under these assumptions, we formulate our hypothesis: *Both disconnected parts and gaps of subnetworks in the OSM road network are indicators for road classification errors if the disconnection or the gap can be resolved in the complete network.* In order to test the hypothesis, we formulate two main research questions:

1. Is an approach by searching for disconnected parts or gaps in subnetworks able to find potential road classification errors? Is this approach able to provide information about the likelihood that the result is an error?
2. Which parameters (such as lengths of detours on a subnetwork compared to the complete network) or combination of parameters indicate gaps in road networks best?

To answer the first research question, we develop a novel approach to detect road classification errors in OSM by searching independently for (a) disconnected parts and (b) gaps in subnetworks. It demonstrates—against expectations—that the error search at disconnected parts leads to fewer results than at gaps. The error search at gaps in subnetworks identifies different parameters that indicate gaps in road networks and combines them in a rating system to obtain an error probability. An efficient implementation of the error search is published on GitHub [12]. To answer the second research question, we provide a detailed analysis of parameters and combination of parameters that indicate gaps and their influence on the error probability. We test our error search with an exemplary case study on the OSM dataset of New South Wales (NSW) in Australia. A reference dataset of identified OSM classification errors compared to authoritative PSMA data is also published on GitHub [12].

We argue that, instead of finding all classification errors, the presented error search only finds the most important classification errors for routing applications and aims to improve the navigability. Identified errors have to be checked by a human expert because our assumption of connected subnetworks without gaps is an ideal that is not met in resource-strapped road infrastructures in all cases. Furthermore, this approach is not able to identify which errors cannot be detected by this method. The errors that can be found by the error search can cause routing applications to take large detours because they might imply that some road classes cannot be passed with all vehicles. Furthermore, classification errors at gaps and disconnected network components can lead to wrong travel time calculations because of the assumed low road class.

In this paper, we first provide a short overview of the related work on error detection in OSM data in Section 2. In Section 3, we elaborate on the theoretical foundation of the presented methodology and identify parameters that we suspect to indicate gaps. The error search contains two main parts, which are described in Section 4. First, the search for disconnected network components (Section 4.1) is presented, and then the different steps of the implementation for the error search at gaps (Section 4.2) are described. The collection of the reference data set is described in Section 4.3. The results are presented in Section 5 and discussed in detail in Section 6. Finally, a conclusion and an outlook are given in Section 7.

2 Related work

In this chapter, the related work regarding this study's aim is summarized. We first consider the quality of OSM data in general, both concerning features and attributes. Also, tools are presented to detect errors in the OSM road network (see Section 2.1). In Section 2.2, we focus on related studies on detecting road class errors, which can be classified into two categories: machine learning and rule-based approaches. Finally, the gaps of existing approaches are summarized, and the novelty of the approach presented in this paper is highlighted (see Section 2.3).

2.1 OSM quality and error detection in general

The quality of OSM data is an important research problem. Many studies exist that assess the geometric accuracy and feature completeness of OSM data. For this assessment, they often rely on reference data in the form of commercial or administrative datasets. Cipeluch et al. [6] assess the spatial coverage, currency, and positional accuracy of OSM for five cities in Ireland. They compare the OSM dataset with Google Maps and Bing Maps manually by GIS overlay. Similar studies exist for many parts of the world [9, 15, 26]. Since manual comparison of datasets is time-consuming, different automated methods to quantify geometric accuracy and feature completeness have been developed [3, 5, 10, 17, 22, 25].

In addition to geometric accuracy and feature completeness of road network data, other important quality indicators are semantic accuracy and attribute accuracy and completeness. These indicators concern the tags attributed to a road in OSM. Although guidelines exist in the OSM Wiki [29], how to tag roads, the compliance with these guidelines is generally average or poor [8]. Tags that are essential for routing are often missing. For example, the maximum speed is only available for 7.4 % of the road elements in OSM worldwide [14].

Some studies exist that assess semantic and attribute accuracy and completeness in OSM [9, 25]. They conclude that some tags are of relatively high accuracy like the road name but feature low completeness, especially for low-level classes. Ludwig et al. [25] also find that attribute accuracy and completeness are generally higher in inhabited areas. The semantic accuracy of the tag highway, which indicates the road class, is often high in the highest road classes but reduces already for the next road classes in the hierarchy. Girres et al. [11] analyze the quality of OSM data in France and find that the highest level road classes, *motorway* and *motorway*, are nearly 100 % correct. However, the next lower class, *secondary* is only 49 % correct. Similarly, over 40 % of the road network in Canada is misclassified in OSM [41]. The results of these studies demonstrate that semantic and attribute accuracy and completeness is often low. Especially errors in road classification frequently occur, which underlines the need to identify these errors and correct them.

To reduce the number of errors in OSM, tools have been developed by the OSM community that help the user while mapping OSM data to find the right tags. OSMRec [20] is an application that recommends categories such as road or building for spatial entities in OSM based on a Support Vector Machine classification. OSMantic is a similar tag recommender system that relies on relationships between tags based on semantic similarity [39]. Undocumented keys can be considered errors because they are missing a definition. To reduce the number of undocumented keys, Majic et al. [28] propose an unsupervised approach to identify equivalent documented keys to the used undocumented keys. They evaluate semantic similarity of keys based on the extensional definitions through their values, co-occurring keys, and geometries of the features they annotate. In a further study, they concentrate on the discovery of bridges by topological relations [27]. This approach enables detecting errors in the *bridge* tag, as a bridge in OSM must be defined with the tag *bridge=yes*. An algorithmic approach to error detection flags errors in names and speed limits is presented in [33] and bases on the comparison of two data sets. Londögård and Lindblad [24] employ deep learning to find spelling errors in tags and correct them. Sehra et al. [34] use a number of basic topological error detection methods available in the desktop GIS Openjump. They find many basic topological errors and conclude that the OSM data in the metropolitan area in Punjab (India) needs preprocessing before using it for navigation. These basic methods include checking minimum segment length, identifying duplicate lines, or finding nodes that almost touch a line. Keller [21] proposes the software ReMAPTCHA, a map-based anti-spam method that can correct *almost connections* in OSM. However, it is not able to detect these *almost connections*.

2.2 Detection of classification errors in OSM

Few studies have addressed the issue of road classification errors in the OSM road network. Within this field, two general approaches can be distinguished: An approach by machine learning and a rule-based approach.

In their master thesis, Stypa and Sandberg [35] use machine learning techniques to classify roads in OSM with intrinsic methods. The authors identify major challenges due to the incompleteness of OSM. To address these challenges, they use rule-based data imputation, for example, for the tags *oneway*, *maxspeed*, and *lanes*. Furthermore, they employ feature engineering and create synthetic attributes like node count, element length, and mean density. They achieve an overall accuracy of around 40 % with the original dataset and around 79 % using their data imputation and feature engineering methods. However, they test their model on a small dataset in Sweden and do not validate their model against reference data. Similarly, machine learning has been used to learn the road class in OSM networks in a series of studies [7, 18, 19]. The authors first develop a representation of the street network, which combines primal and dual graphs, called multi-granular street network representation [18]. Then, they propose an intrinsic machine learning model that learns the geometrical and topological characteristics of different semantic classes of streets [19]. They test the data set with the London OSM street network and conclude that the model's accuracy varies with the road class because some road classes are geometrically and topologically similar. In a similar study, the model of [7] achieves precision and recall values of 68 % and 65 %, respectively.

Rule-based approaches to error detection have mostly been proposed by the OSM community. Several tools exist to find various types of errors, such as Keep Right, Osmose, JOSM/Validator, OSM Inspector, Maproulette, and many others [29]. The error types they detect range from the validity of spatial objects like non-closed areas to topology related issues like dead-ended one-ways and attribute incompleteness like POIs without names. Osmose [30] includes two issue types for possible road classification errors: the issue *sudden highway type change* and *broken highway continuity*. The issue *sudden highway type change* is detected when a road connects directly to a road with a much lower level like a *primary* road connecting with a *residential* road. The issue *broken highway continuity* is raised when the classification of a highway is not consistent along a path, for example, if there is a *secondary* road that connects to a *residential* road and again to a *secondary* road. However, it is only detected if the misclassified part is shorter than 1000 m and if at least one end of the high-level road does not connect to another road besides the low-level road. These issues are presented together on a map and can be corrected by contributors.

2.3 Gaps in existing approaches

In summary, existing machine learning approaches such as [7, 19, 35] either need sufficient reference data or suffer from the incompleteness of OSM attributes. The presented studies classify roads with passable accuracy for many applications but create many false positives in the process. These false positives then have to be checked by humans manually, which is time-consuming. Furthermore, the presented machine learning approaches such as [19] are good for low-level road classes but worse for higher-level road classes, which are more important for routing. Rule-based approaches, like those in the Osmosis tool, can detect very specific classification issues. However, due to static rules, Osmosis detects only errors that are specifically described in the rule. Slight variations of the same type of error are not discovered. Also, it produces a large number of false positives. Additionally, the presented machine learning and rule-based methods cannot provide the error probability or select the most important errors for navigability.

Compared to existing approaches, the error search presented in this paper aims at finding only misclassifications that can cause large detours for routing algorithms. Using the detours, it can provide the user with an importance-based ranking of the errors. This reduces the number of false positives the user has to check to obtain a network with improved navigability. Furthermore, it can prioritize the search for classification errors in high-level road classes that are more important for routing. Because the presented approach does not rely on static rules but dynamic thresholds, it does not suffer from the limitations of a rule-based approach.

3 Theoretical background of the error search

A road network can be formally represented as a graph. G is a non-empty set of n nodes N connected by a set of m links L. The elements of $N \equiv n_1, n_2, ..., n_n$ and the elements of $L \equiv l_1, l_2, ..., l_m$. In a graph, each link is defined by two nodes i and j and denoted as l_{ij}. An alternating progression of adjacent nodes with no node visited more than once is called a path [4]. For this study, the graph G is an undirected graph. We use an undirected graph to avoid issues with the *oneway* tag in OSM.

The OSM road network graph is a multi-class graph with each link l_{ij} belonging to one of the classes in Table 1. Values of the tag highway beside the road classes and their link roads in Table 1 are not included in this study. For this study, OSM road classes are also categorized into different hierarchy levels (Table 1, right column). The levels range from L1 (top-level) to L7 (bottom level). Link roads are categorized into the respective level (e.g., *motorway_link* in L1 and *primary_link* in L2). We combine the classes *motorway* and *trunk* in one hierarchy level because in our study region in NSW in Australia, few *motorways* exist. The classification might have to be adapted for different study regions with a more dense *motorway* network. A road network graph may contain a union of multiple levels: A graph that contains, for example, *motorways* (L1), *trunks* (L1), and *primary* roads (L2) contains the union of L1 and L2.

Accordingly, we will refer to seven networks in total: six subnetworks and the complete road network. The most sparse subnetwork S1 consists only of L1 (*motorway*, *trunk*). The next subnetwork, S2, contains all L1 roads (*motorway*, *trunk*) and all L2 roads (*primary*). It is, therefore, less sparse than S1. The subnetworks S3 to S6 are formed correspondingly, as unions of all roads with a level smaller or equal to the subnetwork's level. The complete road network graph is formed as the union of all levels, $L1 \cup \ldots \cup L7$, equivalent to S7. These networks are illustrated in Figure 1 for a part of the study region in NSW, Australia. In each subnetwork (S1-S6) of the OSM road network, we search for (a) disconnected network components and (b) gaps to find potential road classification errors.

Road Class	Description [29]	Level
Motorway	Restricted access, major divided highway.	L1
Trunk	Most important roads in a country's system that are not motorways.	L1
Primary	Major highways, linking large towns.	L2
Secondary	Highways, not part of a major route, form a link in the national route network, often link towns.	L3
Tertiary	Connect smaller settlements and minor streets to more major roads.	L4
Unclassified	Minor public roads, lowest level of the network, often link villages and hamlets.	L5
Residential	Access roads to housing, without function of connecting settlements.	L6
Living Street	Residential street, pedestrians have legal priority over cars.	L6
Service	Access roads to or within an industrial estate, camp site, business park etc.	L6
Services	Roads in service areas, rest areas.	L6
Road	Road of unknown type, temporary.	L7
Track	Mostly agricultural or forestry use.	L7

Table 1: Road classes and hierarchy level in the OSM road network. The levels range from L1 (top level) to L7 (bottom level). The descriptions of the road classes are cited from the OSM Wiki [29].

Focusing first on the disconnected components, we search for disconnected network components in all subnetworks, respectively. Disconnected components are individual graphs that are not connected by any link with each other. Examples of disconnected network components are visualized in Figure 2. Often, a road network graph of a subnetwork in OSM consists of one large connected graph with many vertices and links that can be considered the main road network. Additionally, it may contain disconnected components that are not connected by any link to the main road network.

With the assumption that subnetworks are typically connected, disconnected network components can indicate four types of errors: *Connection error*, *Self error*, *Disconnected*, and *Border error*. If a disconnected network component is a *Connection error*, the connection(s) to the disconnected network component is wrongly classified (see Figure 2 for an example). *Self error* is assigned if the roads of the disconnected network component itself are in the wrong class (see Figure 2 for an example). Network components that are disconnected both on the subnetwork and on the complete network are called *Disconnected*. These disconnections might happen because of missing roads in OSM but also because of real-world disconnections like islands. Due to cuts at the region borders, disconnected network components can be generated, which are connected in the bordering region. These network components are called *Border error*.

Secondly, regarding the gaps, we search for gaps in the otherwise connected OSM road network graph. The gap search is the more challenging task because the identification of gaps is an unsolved problem. The challenge starts with a clear definition of a gap, which turns out to be context-dependent. We identify a gap between an origin O and a destination D in any connected subnetwork if the shortest path from O to D is substantially longer than on the complete network. The exact limit of how much longer it has to be cannot be de-

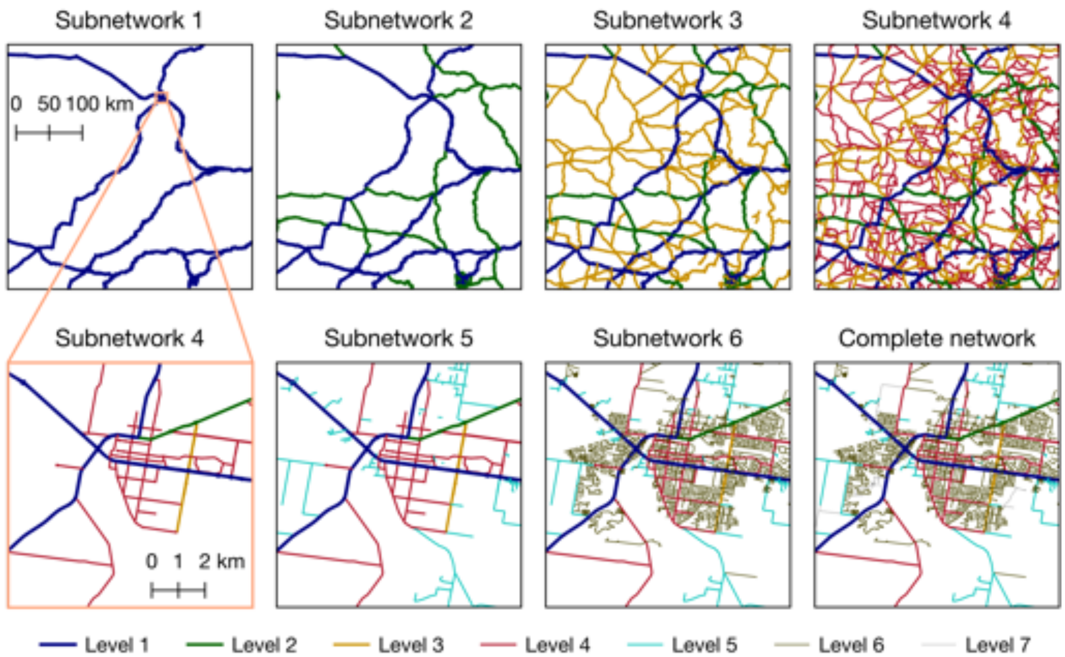


Figure 1: An example of subnetworks in the OSM road network in New South Wales (Australia).

termined universally because it varies by many factors such as the level of the subnetwork (S1-S6), the geography of the region, or the population density. Therefore, indicators have to be identified that point at possible gaps.

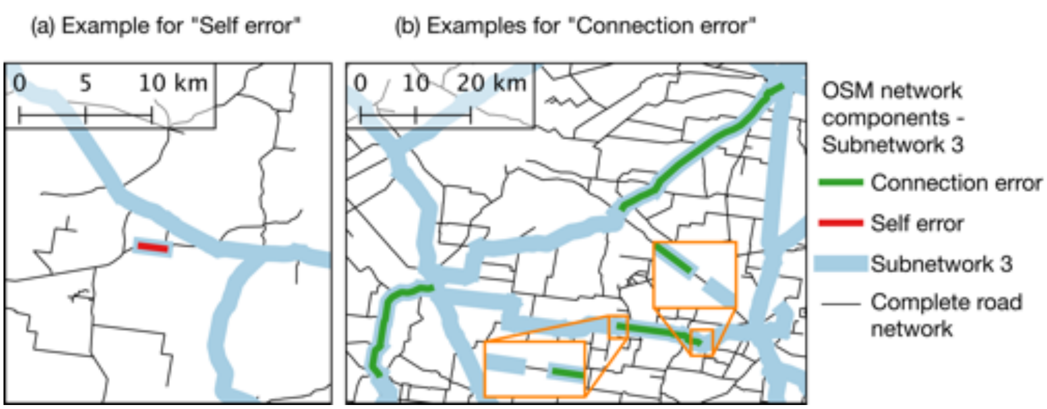


Figure 2: Examples of two types of errors in disconnected network components. In (a), the red network component itself is the wrong class. In (b), the green disconnected networks are not connected to subnetwork S3 because the connection is the wrong class.

A combination of an origin O and a destination D suspected to be a gap is herein called a gap candidate. Three distance measures are identified to find gap candidates: the shortest path distance on the subnetwork from O to D (Pd), the Euclidean distance from O to D (Ed), and the shortest path distance on the complete network from O to D (cPd). We analyze five different parameters (G1-G5) which might indicate a gap:

- G1 = Pd/Ed.
- G2 = Pd – Ed.
- G3 = number of destinations on the same spot (only gap candidates where G1 is highest per origin).
- G4 = Pd/cPd.
- G5 = Pd – cPd.

We are aware that some correlations may exist between the parameters G1 and G2, as well as G4 and G5. Figure 3 is an exemplary road network that helps to visualize these parameters. Calculations of G1, G2, G4, and G5 in the exemplary road network in Figure 3 are given in Table 2.

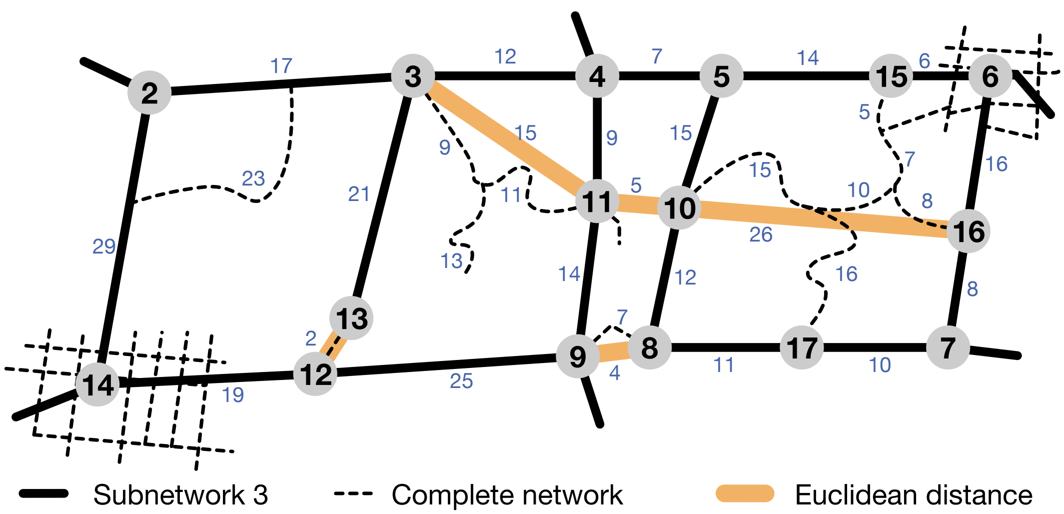


Figure 3: Exemplary road network with nodes (gray) and links (thick black lines) of the subnetwork S3. The complete network is represented by dotted thin black lines. The Euclidean distance is shown exemplary for four gap candidates. The relevant links are labeled with their cost factor (blue). The figure is not drawn to scale.

The parameter G3 is calculated by first selecting only the gap candidates where G1 is highest per origin. As a result, we obtain only gap candidates with distinct origins, but the destinations can still intersect. G3 is calculated as the number of gap candidates that have the same destination. This calculation is based on the observation that the more gap candidates with a high G1 per origin map to the same destination, the more likely this destination is located at a gap.

We distinguish the errors at gap candidates between *No error*, *Near error*, and *Error*. *Near error* is assigned if either O or D is not the start or end of the road connection on the complete network. If both O or D are not the start or end of the road connection on the

Gap candidate	Ed	Pd	cPd	G1	G2	G4	G5	Is gap?
3-11	15	21	20	1.4	6	1.1	1	No
8-9	4	57	7	14.3	53	8.1	50	Yes
10-11	5	31	31	6.2	26	1	0	No
10-16	26	41	33	1.6	15	1.2	8	No
12-13	2	81	2	40.5	79	40.5	79	Yes

Table 2: Exemplary G1, G2, G4 and G5 calculations for gap candidates in Figure 3.

complete network, *No error* is assigned. *Errors* are detected if both O and D are the start or end of the road connection on the complete network.

Multiple gap candidates often indicate a single class error as multiple OD pairs near a gap feature high ratings. Also, multiple connection possibilities for a gap might exist leading to multiple gap candidates labeled *Error* or *Near error*. We assign the same error id for every gap candidate that indicates the same class error. As a result, we obtain a count of *Unique errors* where all unique error ids are counted.

4 Implementation of the error search

As described in Section 3, the developed error search consists of two independent parts: (a) the search for disconnected network components and (b) the gap search. The search for disconnected network components is presented in Section 4.1. Part (b) of the error search is the gap search presented in Section 4.2. The results of both parts are compared against reference data and are described in Section 4.3. The implementation is realized in a PostgreSQL (version 11.5) database with PostGIS (version 2.5) and pgRouting (version 2.6) extensions. We apply the command-line tool *osm2pgrouting* (version 2.3) [40] to import the OSM road network into a pgRouting graph in the PostGIS database.

4.1 Search for disconnected network components

Disconnected network components are identified by using a depth-first search algorithm. This algorithm begins at a certain node and notes all connected vertices along each branch before backtracking such that each node in a connected network is visited. Then, it selects a node not yet visited and does the same with this network component until all nodes in the network have been visited. We run this algorithm on each subnetwork to identify disconnected network components for every subnetwork. As described in Section 3, these components likely indicate errors in the road classification. The identified disconnected network components can be checked and corrected by a human user if they indicate errors.

4.2 Gap search

This section describes the different steps of implementing the gap search, also illustrated in Figure 4. To prepare the networks for the gap search, meshes in all subnetworks are identified in Section 4.2.1. In the core module, the gap search is performed for each identified mesh (Section 4.2.2), and finally, a rating system is employed that rates gap candidates according to their likelihood of being an error (Section 4.2.3).

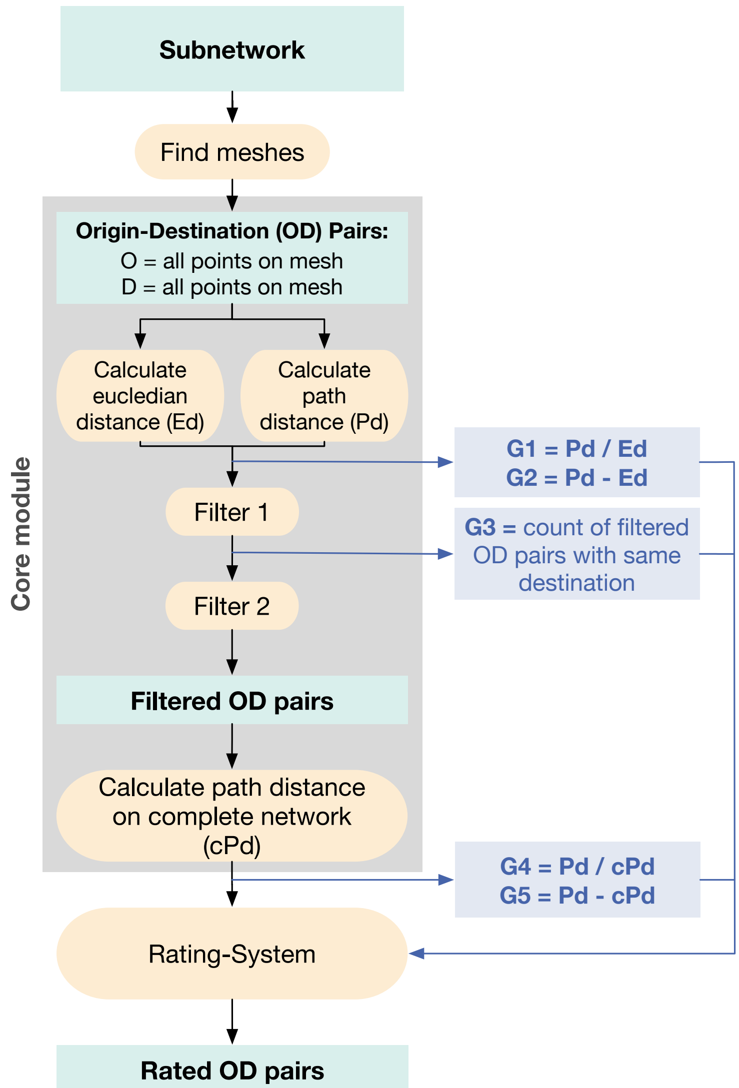
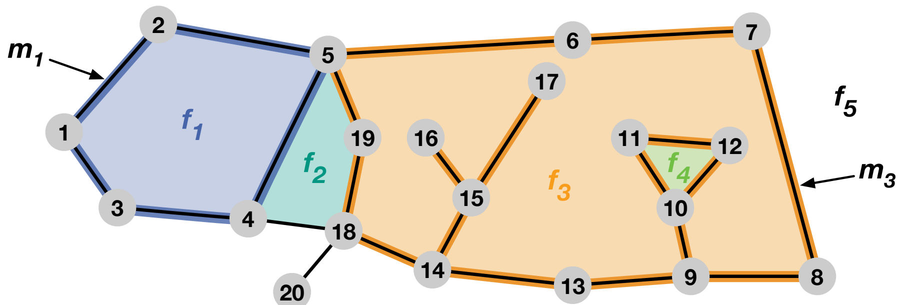


Figure 4: Overview of the implementation for the gap search.

4.2.1 Mesh identification

In order to find gaps in the road network, the shortest paths have to be calculated between Origin-Destination (OD) pairs. All pairs shortest paths are computationally expensive, especially in large networks such as the one for the state of NSW in Australia. However, to solve the problem of gap identification, only the shortest paths between specific OD pairs are required. To find these specific OD pairs, the theory of planar graphs has been considered.

The planar representation of a graph divides the pane into regions, called faces. One of these faces—the exterior one—is unbound and is called the infinite face. Faces are the maximal open, two-dimensional regions that are not further divided into sub-areas. Each face is bounded by a closed walk we herein call a mesh. Every link of the network belongs to one or at most two meshes, one mesh in each direction. Figure 5 is an exemplary planar representation of a graph.



Mesh	Corresponding links
m_1	1-2, 2-5, 5-4, 4-3, 3-1
m_3	5-6, 6-7, 7-8, 8-9, 9-10, 10-12, 12-11, 11-10, 10-9, 9-13, 13-14, 14-15, 15-17, 17-15, 15-16, 16-15, 15-14, 14-18, 18-19, 19-5

Figure 5: Planar graph with nodes (gray) and links between nodes (black). The graph has five faces, the inner faces f_1-f_4 and the infinite outer face f_5. The faces are bounded by meshes. Two meshes are illustrated exemplarily: the mesh m_1 is drawn with thick blue lines, the mesh m_3 is illustrated with thick orange lines. For illustration purposes, m_2, m_4 and m_5 are not colored.

In this example, five faces exist: the inner faces f_1-f_4 and the infinite outer face f_5. The face f_1 with its corresponding mesh m_1 (highlighted in blue in Figure 5) and the face f_3 with its corresponding m_3 (highlighted in orange in Figure 5) are illustrated exemplarily. The link $l_{4,5}$ belongs to two meshes: In direction 5-4 it belongs to m_1, and in direction 4-5, it belongs to m_2. Similarly, the link $l_{9,10}$ belongs to m_3 in both the direction 9-10 and in direction 10-9.

As stated above, to solve the problem of gap identification, only the shortest path between specific OD pairs, namely between OD pairs, where O and D are located on the same mesh, are required. To illustrate this, Figure 5 can be considered. A gap might exist between nodes n_{16} and n_{19} or any other pair of points corresponding to the mesh m_3 because a connection is possible. However, gaps between nodes located on different meshes, for example, nodes n_{16} and n_4, are already covered by calculating the shortest path between nodes n_{16} and n_{18} or n_{19}, both located in the same mesh as node n_{16}. Thus, instead of calculating all pairs of shortest paths, we reduce computing time radically by only calculating the shortest paths between OD pairs located on the same mesh.

The road network itself is not a planar graph because links like bridges or tunnels exist which cross other links without a node at their intersection. Thus, to create a planar representation of the road network, we create artificial nodes where two links intersect. Then, meshes in the road network can be identified, reducing computing time for the gap search. We can now limit the search for gaps to all pairs of shortest paths within one mesh instead of all possible OD pairs in the entire graph. Scenarios exist where gaps occur across meshes, for example, if a bridge over a road is missing. This special case can not be detected by this methodology. However, the radical reduction of computation time justifies the discharge of these rare scenarios.

With the planar representation of the road network graph, an algorithm to find meshes is then implemented based on maze solving algorithms. Algorithm 1 finds meshes by following a link from a starting node to the counterclockwise next link and continues to do so until the start node is reached again. For illustration, we show our implementation of Algorithm 1 in PL/Python in a PostGIS database with a road network graph created by the pgRouting extension (see [12]).

Algorithm 1 Finds all meshes in a planar road network graph.

```
node_id list = all node ids in road network graph
mesh id = 0
for each node id in node id list do
    link id list = all links with start = node id or end = node id
    for each link id in link id list that has mesh id = NaN in this direction do
        set mesh id of link with link id = mesh id
        next node = id of the node at the other end of the link link id
        link = link id
        while next node ≠ node id do
            next link = the id of the counterclockwise from link next link where the source
            or the target is next node
            link = next link
            set mesh id of link = mesh id
            next node = node id of the next node following the trail of link
        end while
        mesh id = mesh id + 1
    end for
end for
```

4.2.2 Core module of the gap search

After the identification of meshes, the gap search begins in the core module for each subnetwork individually. Although the meshes are identified for the network with artificial nodes, the gap search uses the original road network but calculates the shortest paths only for OD pairs located on the same mesh (see Section 4.2.1). First, the shortest paths on the respective subnetwork from all nodes of a mesh to all nodes of a mesh are calculated. Then, the Euclidean distance is calculated for these OD pairs. This is done for all meshes in a subnetwork. The parameters G1 and G2 are calculated from the resulting path distance and Euclidean distance. Figure 5 can be considered an example of the first step of the gap search in the core module. In this example, the methodology will calculate the shortest paths between all nodes in m_1, m_2, m_3, m_4, and m_5, respectively. Within m_3 all shortest path combinations and the Euclidean distance between all nodes five to nineteen are calculated. In this example, both resulting parameters G1 and G2 are highest for the OD pair 17-6.

Even though all-pairs shortest paths only have to be calculated within meshes on subnetworks, this still results in a high number of OD pairs. Shortest path calculations on the complete network are computationally more expensive than on subnetworks because of the higher level of detail. Thus, the number of OD pairs for the shortest path calculation on the complete network has to be reduced. Most of the resulting OD pairs are not gap candidates. The parameter that indicates gap candidates best at this stage is G1 such that OD pairs with a low G1 can probably be filtered out (for a detailed discussion, see Section 6).

We employ two filters on the data. In a first filter, for each start point S_n, only the OD pair with the highest value of G1 is kept. However, many OD pairs still exist, which are not gap candidates as many start points are not at gaps. The second filter reduces the data such that only those OD pairs above the 70 % quantile of G1 and above the 25 % quantile of G2 are kept. These values are chosen, so that much unnecessary information is filtered out, and at the same time, possible gap candidates are kept. The selection and the impact of the filters on the error search is also discussed in Section 6. Table 3 provides values for the number of OD pairs before and after filtering in the study region and exact values for the 70 % quantile of G1 and 25 % quantile of G2.

	S2	S3	S4	S5	Total
OD pairs (million)	5.7	7.6	7.9	6.4	27.7
OD pairs after filter 1	15275	39584	71762	78793	205414
OD pairs after filter 2	803	5906	15482	19046	41237
G1 threshold	6.39	4.09	4.21	5.07	4.56
G2 threshold [m]	757	852	645	565	645

Table 3: Effect of filtering on the OD pair dataset for New South Wales (Australia).

The parameter G3, the number of endpoints of gap candidates on one spot, is calculated after the first and before the second filter. If G3 is calculated before the first filter, it contains no information because the data set contains all possible OD pairs, so G3 is the same for every point in one mesh. If G3 is calculated after the second filter, the number of endpoints on one spot is much lower, and much of the information which makes G3 valuable for gap identification has been filtered out. Because of this, G3 is calculated after filtering only the OD pairs with the highest G1 per start point and before deleting all OD pairs under a certain threshold of G1 and G2.

Finally, the path distance on the complete network is calculated only for the filtered OD pairs. The parameters G4 and G5 result from the relation of the path distance on the complete road network to the path distance on the subnetwork. The workflow for the gap search is also illustrated in Figure 4 in the gray box.

4.2.3 Rating system

We employ a rating system to rate the parameters G1, G2, G3, G4, and G5. The rating system assigns points from 1 to 10 to the parameters mentioned above.

The parameter G3 contains discrete numbers with many low values and very few high values. To assign points, the distribution of the values has to be evaluated, and points are assigned according to how high the value is. The point rating is constructed with expert knowledge (for a more detailed evaluation, see Section 6) and is given in Table 4. It might have to be adapted for different study regions. A classification of the sorted data by deciles as for the parameters G1, G2, G4, and G5 (see below) is not applicable for G3 because of the discrete values.

$G3_R$	1	2	3	4	5	6	7	8	9	10
G3	1	2	3	4	5-6	7-8	9-10	11-12	13-15	>15

Table 4: Exemplary parameter rating for G3.

We calculate ten deciles for each parameter G1, G2, G4, and G5 (see Table 5), which divide the sorted data into ten equal parts so that each decile represents one-tenth of the data. Then, each value is assigned the point rating of the decile is located in. For example, if a value of G1 lies between the 0 % and 10 % quantile, the point rating 1 is assigned. Ten points are assigned if the value is above 90 %.

GX_R	1	2	3	4	5	6	7	8	9	10
Percent	< 10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100

Table 5: Deciles for parameter rating for GX_R where $X = 1, 2, 4, 5$.

The resulting point ratings are noted as $G1_R$ for all G1 ratings, and similarly for G2, G3, G4, and G5 as $G2_R$, $G3_R$, $G4_R$, and $G5_R$ (see Table 4 and Table 5). The combination of point ratings serves as an indicator of how likely a gap candidate is a classification error. In this study, both the importance of each point rating individually and the influence of different combinations of point ratings on the result are evaluated.

4.3 Reference data

We employ the authoritative PSMA Street Network data [31] as reference data, especially the road network data of the state of NSW. This dataset is also chosen because it is independent of the OSM road network data: Unlike other official road data, PSMA data has not been integrated into the OSM database.

The description of the road classes in the PSMA dataset can be found in [32]. The road classes in the PSMA dataset do not match the OSM road classes. For example, *secondary* roads in OSM are mostly (around 60 %) classified as Sub-arterial roads in PSMA. However, many cases also exist where they are categorized as Arterial roads or Collector roads. Similarly, Sub-arterial roads in PSMA include OSM *primary*, *secondary*, and *tertiary* roads. Therefore, a direct comparison of road classes to detect errors is not possible. We manually check both the resulting disconnected network components and gap candidates and decide if there is a classification error.

To facilitate the generation of reference data, we aim at applying some general rules. These rules are based on the assumption that even though the classification schemes do not match, the continuity of a road class in the PSMA dataset still contains some useful information. If both O and D of a gap candidate are located on roads with the same PSMA road class, the connection must be a road of the same or higher PSMA class to be an error. In this case, if the connection is a lower PSMA road class, the gap candidate is not marked as an error. Similarly, suppose O and D are located on roads with different PSMA road classes. In that case, the gap is an error if the connection is a road of the same or higher PSMA road class as the lower one of both PSMA road classes of O and D. This is illustrated with examples in Figure 6. However, in some cases, these rules do not apply because of the incompatible classification schemes. Then, we decided with expert knowledge and by comparison with additional data sources like Google Maps.

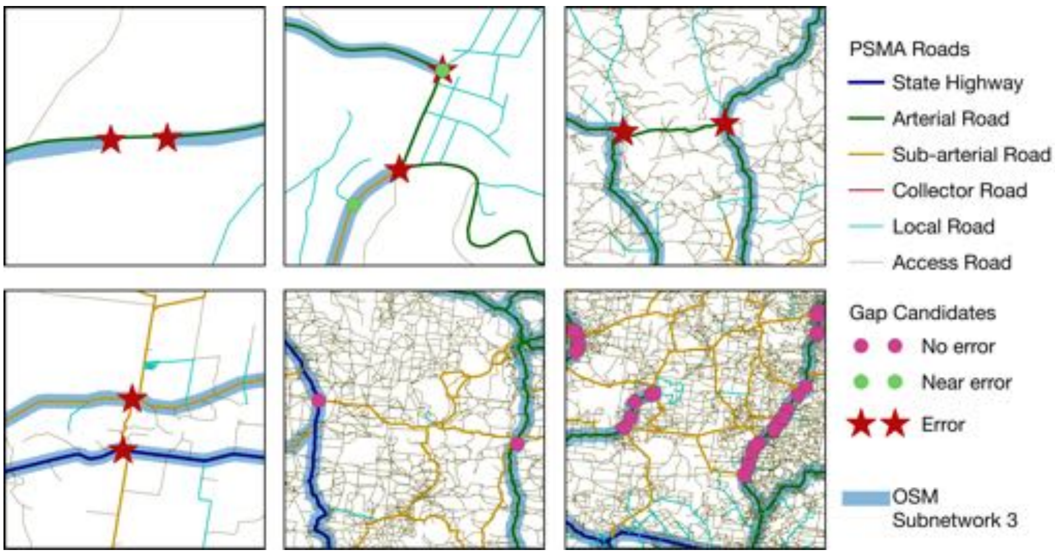


Figure 6: Examples of the collection of reference error data for gap candidates. Note that a gap candidate is always illustrated as two points, an origin and a destination.

Since it is not feasible to analyze every gap candidate manually, all gap candidates with high ratings are checked. Furthermore, many gap candidates with medium ratings and few gap candidates with low ratings are checked. The lower the ratings, the fewer *Errors* are found, which supports this methodology (see Section 6).

5 Results

In this section, we show the results of applying the presented error search on the OSM road network for the study region of NSW. The region is chosen because it is large enough so that all subnetworks form a network themselves and because of the availability of the PSMA data as ground truth. Generally, the population density is much higher on the coast in the east of the region than in the Outback in the west.

First, we present the results of the error search at disconnected components, then we focus on the results of the gap search. For the gap search, we calculate the presented point ratings and present the results; first each of the point ratings separately, then in combination with each other. The error types per subnetwork for all disconnected network components and gap candidates are presented in Table 6. We find 1991 disconnected network components on all levels, with 94.48 % of them in subnetwork S5. Most disconnected network components are *Self errors* (95.13 %), and few are *Connection errors* (3.37 %). Subnetworks

	Disconnected components				Gap candidates			
	Con. error	Self error	Discon.	Border error	Error	Near error	Unique error	No error
Subnetwork S1	0	3	0	1	0	0	0	151
Subnetwork S2	0	2	1	0	14	13	14	362
Subnetwork S3	13	12	0	4	66	110	51	1857
Subnetwork S4	22	48	0	4	304	444	214	3362
Subnetwork S5	32	1829	17	3	871	394	498	3514
Sum	**67**	**1894**	**18**	**12**	**1255**	**961**	**777**	**9131**

Table 6: Error types per subnetwork both at disconnected components and at gap candidates.

S1-S4 feature a total of 110 disconnected network components with 22.73 % *Connection errors* and 59.09 % *Self errors*.

In total, 11.06 % of all analyzed gap candidates are *Errors*, 8.47 % are *Near errors*. The lower the level, the more errors occur: 64.10 % of all *Unique errors* are in subnetwork S5,

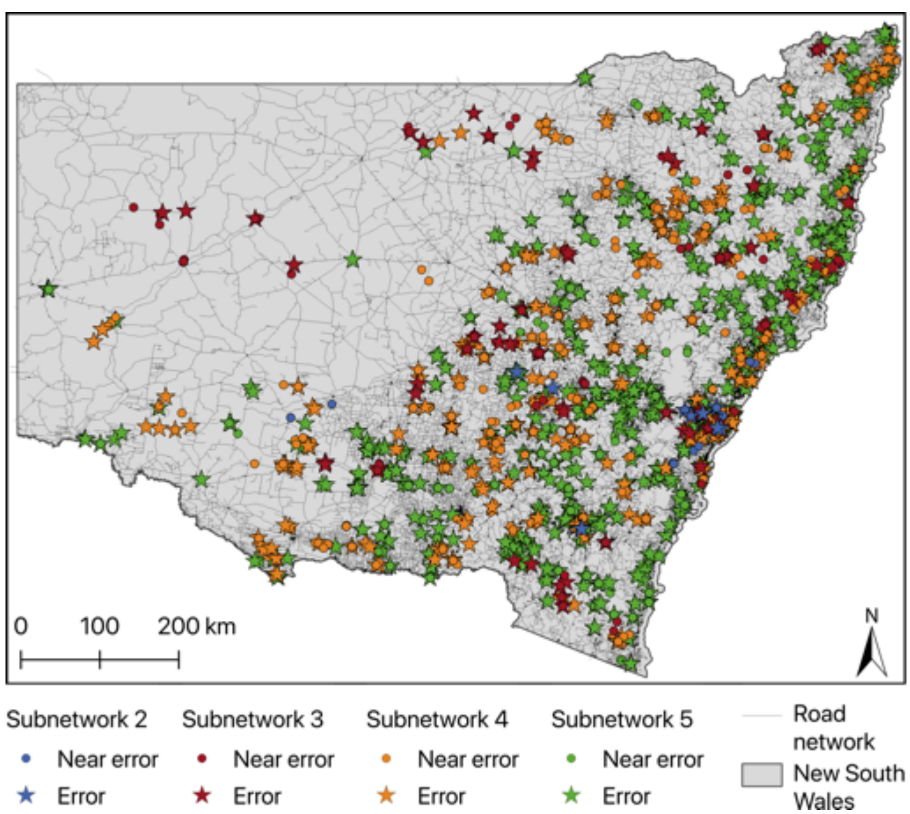


Figure 7: All *Errors* and *Near errors* in New South Wales (Australia) per subnetwork.

and only 1.80 % of all *Unique errors* are in subnetwork S2. No *Errors* or *Near errors* are found in subnetwork S1. In subnetworks S1-S4, there are many more unique errors (279) than disconnected network components (110). 69.23 % of all *Connection errors* in subnetwork S3 and 27.27 % of all *Connection errors* in subnetwork S4 are also identified by gap candidates. Figure 7 shows a map of all detected *Errors* and *Near errors* in gap candidates per subnetwork in NSW.

Figure 8 shows the distribution of the error types over the ratings 1-10 for all analyzed gap candidates. Most *Errors* and *Near errors* feature a high $G1_R$ and $G4_R$. On the other hand, $G1_R$ and $G4_R$ values indicate *No error* more often than *Error* or *Near error*. Regarding $G2_R$ and $G3_R$, *Errors* and *Near errors* are distributed approximately uniformly over the ratings 1-10, with slightly more *Errors* for the higher ratings of $G2_R$. $G3_R$ is the only point rating that has many *No errors* in the lower ratings, especially for the rating 1 (3603 *No errors*). $G5_R$ features more *Errors* and *Near errors* for medium and high ratings than for low ratings. Also, the number of *No errors* is high for high ratings for $G5_R$.

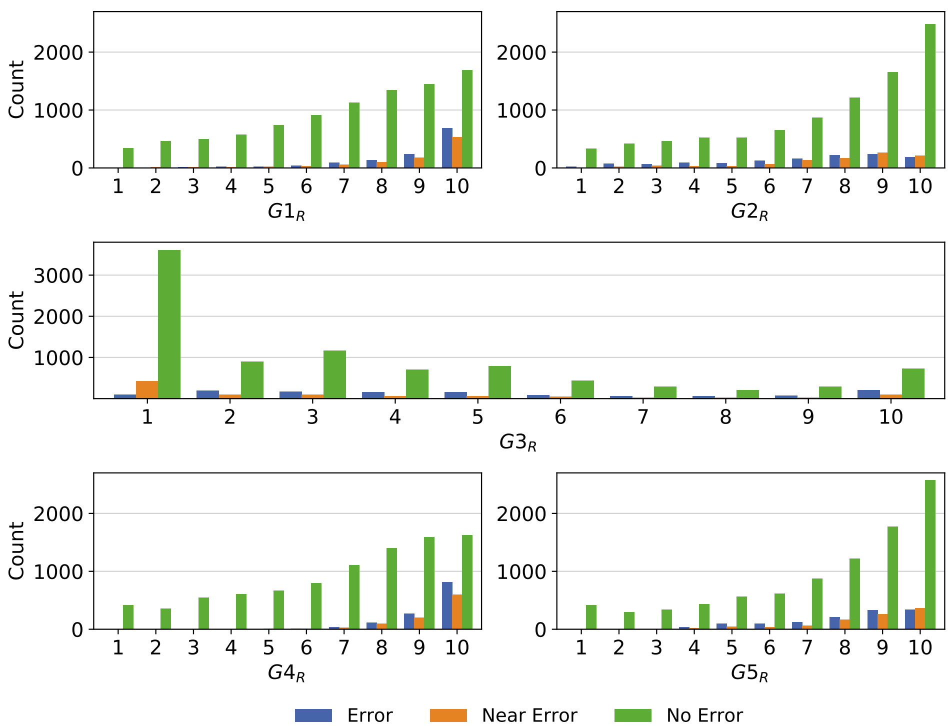


Figure 8: Error types per rating for all point ratings.

In Figure 9, $G1_R$ is combined with the other point ratings, respectively. The highest number of *Errors* in the high ratings can be observed by adding $G1_R + G4_R$. There, 88 % of all *Errors* and 41 % of *No errors* have a rating higher or equal to 16. This results in a rate of 88/41 = 2.14 of *Errors* versus *No errors* for all ratings higher or equal to 16. The y-axis on the right of Figure 9 shows the rate of all *Errors* and *No errors* equal or higher to the current rating. Note that the scale of the rate is different in each plot in Figure 9. The sum of $G1_R + G2_R$ and $G1_R + G5_R$ has fewer *Errors* and a lower rate in high ratings than

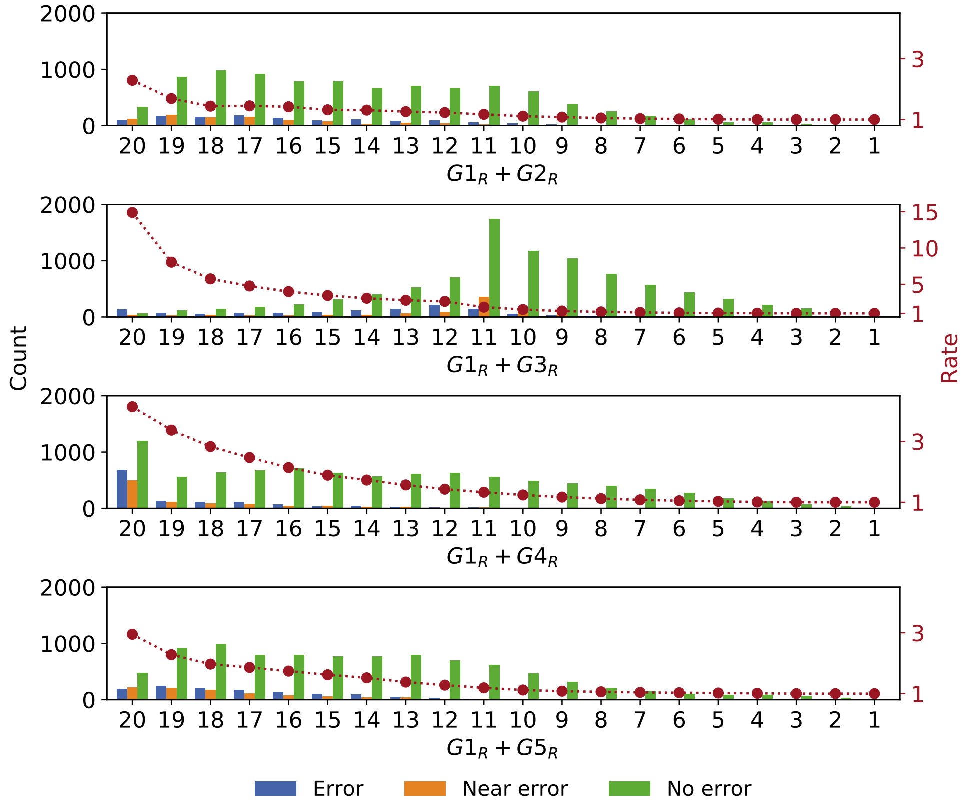


Figure 9: Combinations of $G1_R$ with all other point ratings, respectively. The y-axis on the right displays the rate of all *Errors* versus *No errors* equal or higher to the current rating. Note the different scale of the rate in the second plot.

$G1_R + G4_R$. However, $G1_R + G5_R$ has slightly higher ratings, and more *Errors* in high ratings than $G1_R + G2_R$. $G1_R + G3_R$ has a low absolute number of *Errors* in high ratings, but at the same time also a low absolute number of *No errors* in high ratings. This leads to a high rate of 2.65 for all ratings higher or equal to 12, where 78 % of all *Errors* and 29 % of all *No errors* are analyzed.

Figure 10 shows different combinations of point ratings. In Figure 10 (a), the point ratings $G1_R$, $G3_R$, and $G4_R$ are added up, resulting in a maximum of 30. Figure 10 (b) shows the combination of the point ratings $G1_R$, $G3_R$, $G4_R$, and $G5_R$, ranging from 0 to 40. The absolute numbers of gap candidates per rating are displayed in the upper plots of Figure 10. The middle plots illustrate the percentage of *Errors*, *Near errors*, *No errors*, and *NaN* per rating. *NaN* signifies gap candidates where it is unknown if they are *Errors*, *Near errors*, or *No errors*.

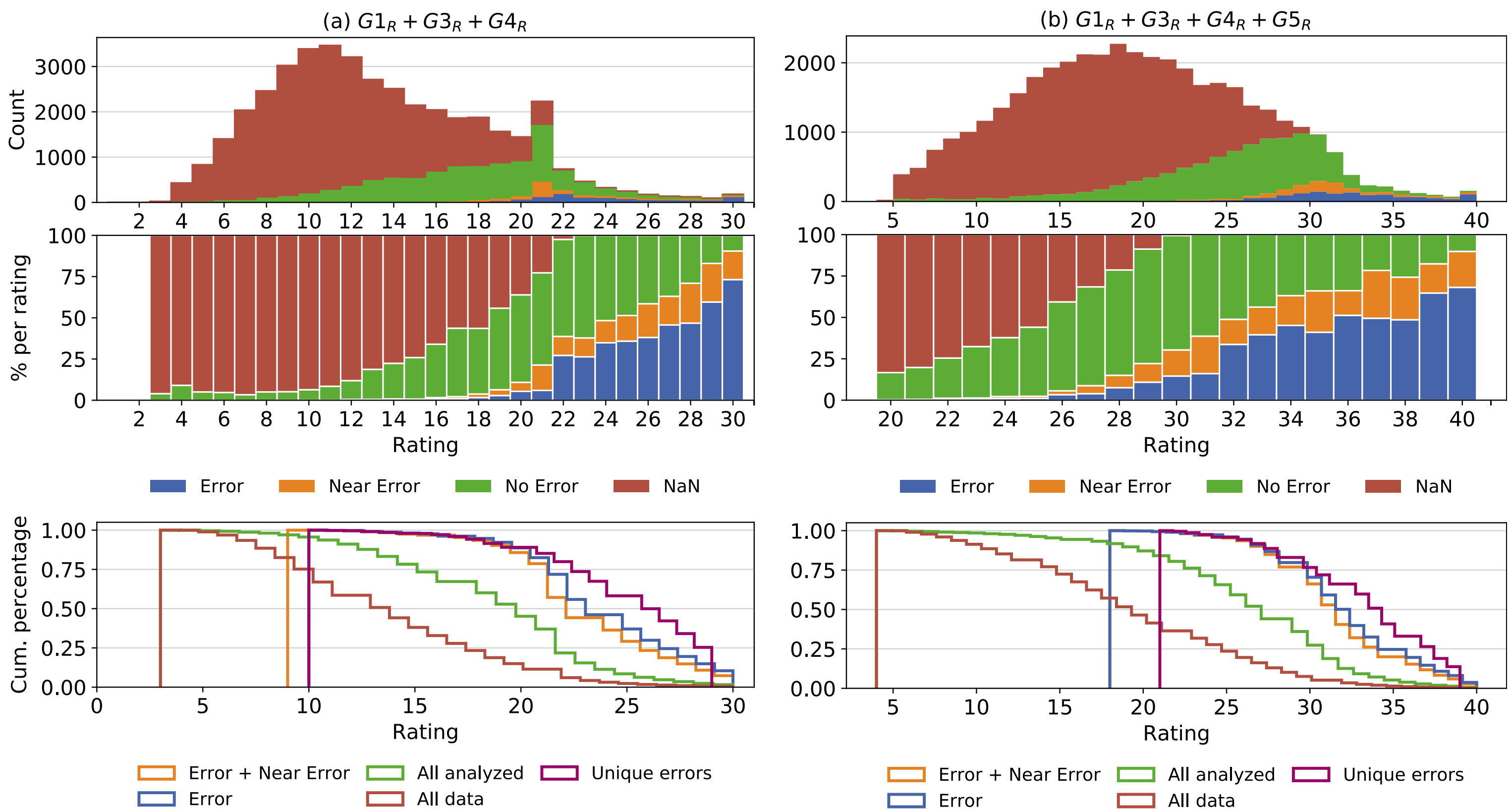


Figure 10: Combinations of the point ratings $G1_R$, $G3_R$ and $G4_R$ (left) and $G1_R$, $G3_R$, $G4_R$ and $G5_R$ (right). The absolute numbers of gap candidates per rating (stacked) are displayed in the upper plots. The plots in the middle illustrate the percentage of error types per rating. The lower plots show the cumulative percentage of error types per rating.

The lower plots of Figure 10 show the cumulative percentage of different values per rating, beginning with high ratings.

For the combination $G1_R$, $G3_R$ and $G4_R$, 91 % of all gap candidates with rating 30 are *Errors* or *Near errors*. This declines to 83 % for 29, 71 % for 28 and 63 % for 27. The trend continues until the percentage of *Errors* and *Near errors* is next to zero for ratings lower or equal than 18. Furthermore, 75 % of all data has a rating lower or equal to 18. Simultaneously, 94 % of all unique errors have a rating above or equal to 18. 50 % of all unique errors can be found by searching 1.73 % of all data. The combination $G1_R$, $G3_R$, $G4_R$, and $G5_R$ shows the same trend, but slightly more data has to be searched to obtain the same amount of unique errors. To obtain 50 % of all unique errors, 1.99 % of all data has to be searched in the combination $G1_R$, $G3_R$, $G4_R$, and $G5_R$. Figure 11 compares all presented combinations concerning the percentage of *Errors* and unique errors found against the percentage of all data searched.

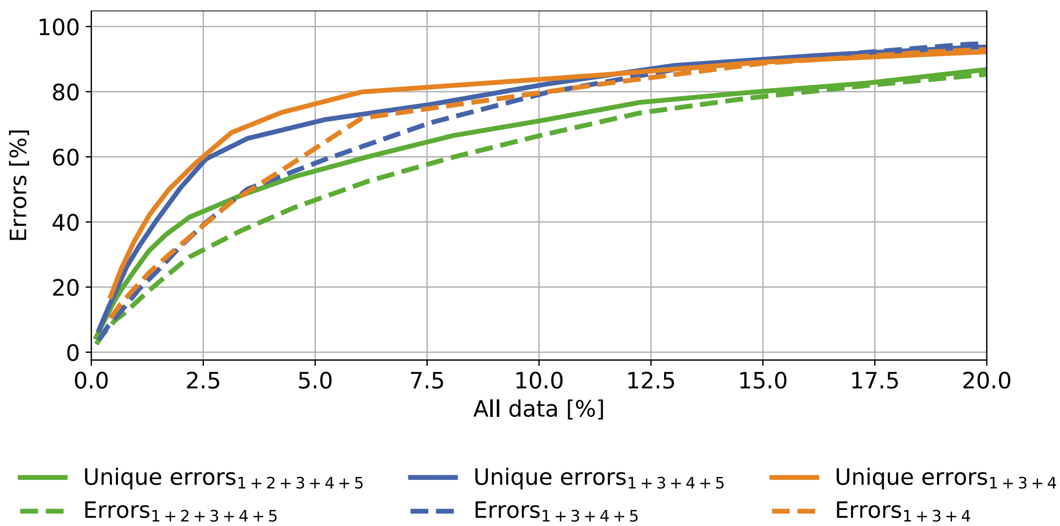


Figure 11: Percentage of *Errors* and unique errors found in respect to the percentage of all data searched for different combinations of ratings.

6 Analysis and discussion

In this section, we discuss and interpret the results shown in Section 5. We focus on the applied filters, the distribution of errors over the network levels and their spatial distribution in Section 6.1. In Section 6.2, we evaluate the parameters and the performance of different parameter combinations. Finally, we discuss the research hypothesis, the limitations and sources of errors, and the transferability to other regions in Section 6.3.

6.1 Analysis and discussion of resulting error types

Two filters are applied (see Section 4.2.2) to reduce the number of gap candidates prior to the path calculations on the complete network. The parameters G1, G2, and G3, can be calculated before the path calculations on the complete network. As shown in Figure 8, the point ratings $G1_R$, which results out of G1, is the most significant for the error search because *Errors* generally have a high $G1_R$. The lower the rating, the fewer *Errors* are found. *No errors* are found in the ratings 1, 2, or 3 of $G1_R$. It can be assumed that if *No errors* occur in the lowest ratings of $G1_R$, *No errors* will appear with even lower G1 values that are filtered out with the second filter. Furthermore, the first filter only keeps the gap candidate with the highest G1 per start node. This filter can be justified with the same argument that a high G1 has a higher probability for an *Error*. However, this filter also discharges *Errors* in some rare cases. By analyzing the highest G1 values of the gap candidates discharged by this filter, we find that for discharged *Errors* there are in almost all cases *Near errors* that are found by the methodology. The second filter also removes all gap candidates below the 25 % quantile of G2. In Figure 8, it seems that $G2_R$ and with that G2 are insignificant for error detection. However, in the definition of a gap, we state that the detour has to be significant for a gap candidate to be considered a gap. Table 3 shows that the 25 % quantile of G2 is between 565 m to 852 m. Any potential *Error* lower than this range is deemed not significant because the detour is too small.

Because the parameters G1 and G2, as well as G4 and G5, are calculated from the same values, a correlation might be expected. However, the Pearson correlation coefficient does not suggest a correlation in both cases: for G1 and G2 it is 0.039, and for G4 and G5 it is 0.048.

The error search finds both errors at disconnected components and at gaps. As we argue in Section 1, the important errors are the ones that can potentially cause large detours for routing applications. While *Self errors* are indeed classification errors, they are not important for routing as they do not cause detours. *Connection errors* are usually more important for routing because, like *Errors* at gaps, they can cause large detours (as visible in Figure 2). It can be seen in Table 6 that at disconnected network components, many more *Self errors* than *Connection errors* are found. In comparison to *Errors* at gap candidates, the *Connection errors* are few. This leads to the conclusion that the gap search finds more and also more important errors than the search for disconnected network components. Furthermore, some of the *Connection errors* in the subnetworks S3 and S4 are also identified with a gap search.

A hierarchical road network is constructed so that the importance of a road decreases from a high hierarchy level (e.g., *motorways*) to a low hierarchy level (e.g., *tracks*). Considering subnetworks, as a union of levels, the sum of road network kilometers in subnetworks is much lower in high-level subnetworks like S1 than in low-level subnetworks like S6. Furthermore, L5 of the road network are *unclassified* roads, which are technically defined as minor public roads and the lowest level of the network (see Table 1). However, mappers might often intuitively tag roads with unknown classification with *highway=unclassified*, such that there might be many classification errors in the L5 network. For these reasons, there are many more *Errors* (see Table 6) in low-level subnetworks than in high-level subnetworks. However, the *Errors* in high-level roads are more significant for a country's transportation network because high-level roads carry more traffic than low-level roads. Thus, more vehicles are affected by *Errors* in high-level roads than by *Errors* in low-level roads. We do not search for errors on S6 because we find that level L6 and L7 are often not distinguishable.

Regarding the spatial distribution of *Errors* at gap candidates, most *Errors* of S5 are in the east of the region where the population density is high. We often find *Errors* in S5 inside a city's road network where major roads in cities are classified as *residential*. *Errors* in S3 and S4 are also often located in rural regions. A significant amount of *Errors* in rural areas is also due to bridges that are classified in a different road class than the connecting roads.

6.2 Analysis and discussion of the rating system

Considering the point ratings $G1_R - G5_R$ separately in Figure 8 provides some information on how significant the parameters are for the error search at gap candidates. This suggests that the point ratings $G1_R$ and $G4_R$ might be the most significant point ratings for the error search. Looking only at the absolute number of *Errors* in high ratings, the point ratings can be ranked in the following order of significance: $G4_R > G1_R > G5_R > G2_R > G3_R$. This indicates that the calculated ratio of distance is much more significant for the error search than the mathematical difference of distance. However, the absolute number of *Errors* in high ratings is just one aspect. If the number of *No errors* is also high, many potential gap candidates have to be searched to find *Errors*. This is the reason why the point rating $G3_R$ is essential even if the number of errors in high ratings is low. Compared to the other point ratings, it features significantly more *No errors* in low ratings.

The significance of $G3_R$ becomes apparent when looking at the combinations of point rating $G1_R$ with all others in Figure 9. The rate of *Errors* and *No errors* is crucial because it is an indicator of how many *No errors* have to be searched in relation to the numbers of *Errors* that are found. Ideally, this rate is high such that most gap candidates that are searched are *Errors* and *Near errors*, and very few are *No errors*. Figure 9 demonstrates that the rate of *Errors* and *No errors* in high ratings is highest for the sum of the point ratings $G1_R$ and $G3_R$. The second-highest rate for high ratings can be seen in $G1_R + G4_R$. However, there are fewer *Errors* in high ratings in $G1_R + G3_R$ than in all other combinations. As can be expected by looking at the ratings of $G1_R$ and $G4_R$ individually, the sum $G1_R + G4_R$ features the highest number of absolute *Errors* in high ratings. We conclude that, out of the point ratings analyzed in this study, $G1_R, G3_R$, and $G4_R$ are the most significant indicators of *Errors* at gaps.

Therefore, we establish a rating system with the combination of $G1_R, G3_R$, and $G4_R$ and then compare it to the combination of $G1_R, G3_R, G4_R$, and$G5_R$ and to the combination of all point ratings. All three combinations show the desired result where many *Errors* occur in high ratings. As it turns out, the combination $G1_R, G3_R$, and $G4_R$ performs best as the rate of both *Errors* and unique errors versus analyzed data is highest. Adding $G5_R$ lowers this rate slightly, and further adding $G2_R$ lowers it significantly.

The underlying problem with the different input parameters is basically a multi-criteria decision problem. Our developed rating system implements a basic multi-criteria decision system with the parameters G1 – G5 as criteria. This basic decision system can still be improved. A weighting of the criteria could potentially enhance the rating system's performance, but finding the appropriate weights for a study region requires additional studies. Furthermore, outranking methods like ELECTRE or PROMETHEE could reduce the number of gap candidates before applying the rating system. However, because of the high number of potential gap candidates (41,237 for NSW, see Table 3), a pairwise comparison of gap candidates for the determination of the concordance and discordance matrix (ELECTRE) and for the determination of deviation (PROMETHEE) would probably result in huge matrices, and hence, remains the subject of future investigations.

The results of the combination $G1_R, G3_R$, and $G4_R$ suggest that the error probability is decreasing with the rating: all gap candidates of rating 30 have a 91 % probability of being an error, all gap candidates with a rating of 29 an error probability of 83 %, etc. This supports a methodology by gap detection where a human user can prioritize the error search by first checking the gap candidates with the highest ratings and then eventually continuing to the lower-rated gap candidates. When checking all gap candidates equal or higher than 22, 80 % of all errors can be detected by either an *Error* or a *Near error*. At the same time, 51 % of gap candidates with a rating equal or higher than 22 are *Errors* or *Near errors*. Moreover, the errors that significantly impact the accuracy of routing are the ones with the highest ratings as the possible detour is the largest for high ratings. Thus, a human user can quickly detect influential errors for routing applications by prioritizing high ratings in gap candidates.

6.3 Usage implications and limitations

The error search is based on the hypothesis that both disconnected parts and gaps of subnetworks in the OSM road network are indicators for road classification errors if the disconnection or the gap can be resolved in the complete network. Our results prove this hypothesis. Disconnected parts and gaps of subnetworks in the OSM road network prove reliable indicators for road classification errors. However, they are not guaranteed classification errors. In the real world, a *primary* road may turns into a lower quality road for a certain distance and then back into a *primary* road. This can have numerous reasons like different jurisdictions, traffic, or missing funding. For example, the US interstate highway system has some well-known true gaps [1]. These gaps occur mostly because the connecting roads fail to conform to interstate standards fully, and for some of these gaps plans to close them already exist. Therefore, a human expert has to check the results of the error search to confirm them.

When applying the error search, its limitations have to be considered as well. First and foremost, the method is not able, but also not designed to find all classification errors in a road network. It is only designed to find the errors which lead to detours when considering only a subnetwork. We can not clearly state how many *Errors* are missed to be detected because it is not feasible to analyze the entire network manually. We argue that the *Errors* missed are few and less influential on the accuracy of routing applications because of the distribution of *Errors* in Section 5. Roads that are wrongly classified as a higher class than they actually are can not be detected with this method. Also, using an undirected graph might cause the Error Search to miss gaps that would otherwise be detected. But, a directed graph makes use of the OSM tag *oneway*. The accuracy and completeness of this tag are often low [2, 8, 25], and its enhancement is not in the scope of this study. Some missing roads are found by the search for disconnected network components and gap search, but this is also not the scope of this study.

Generally, we observe two categories of false positive gap candidates, meaning gap candidates with high ratings, which are *No errors*. Both categories are visualized in Figure 12. On the one hand, the high number of *No errors* with high ratings is often due to gap candidates in a broader range along *Errors* that cause a large detour. While they are not classified as *Near error*, because they do not start or end at the gap, the point ratings are often high due to the gap in the vicinity. This phenomenon also leads to *Errors* near other *Errors* getting a high rating even though the detour caused by the first *Error* is very small. The combination of both *Errors* then leads to large detours, resulting in a misleading high rating for this *Error*. On the other hand, false-positive gap candidates sometimes occur because the hypothesis

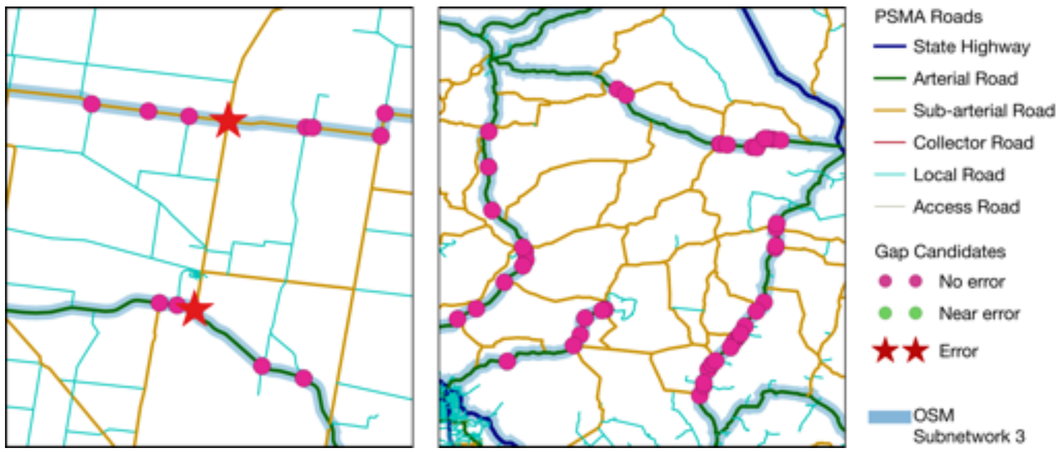


Figure 12: Examples of false positive gap candidates. On the left, *No errors* in the vicinity of an *Error* are visualized. The *Error* causes the high rating of the *No errors* which are not marked as *Errors* because they do not start or end at a gap candidate. On the right, *No errors* are visualized where there is no need for a high-level connecting road.

does not apply. As mentioned above, this can have numerous reasons. We observe that the most frequent reason is that there is no need for a high-level connecting road because it is not used frequently. Especially in the rural parts of NSW, the population is concentrated in towns, and large areas are uninhabited. Thus, these uninhabited parts of the country do not require good accessibility. Furthermore, sometimes a high-level connecting road is impossible because of difficult terrain, for example, in mountain ranges. The methodology cannot separate these cases, so a human user is required to confirm the result.

Generally, the presented error search can be applied for all road networks. However, it has to be considered that some values have to be adapted to fit the characteristics of a different region. The classification of the road network into hierarchy levels might have to be adapted to the country's circumstances. Especially the thresholds for Filter 2 have to be analyzed in detail and may be higher for regions with overall lower quality of the road network. Furthermore, as the rating system is based on relative thresholds, there is the underlying assumption that there are classification errors in every road network. For road networks with higher or lower quality of road classification, the resulting probability of error distribution will be different. To apply the error search, a region's road network has to be more or less complete such that there are few missing roads because this might hinder the error search.

The strengths and limitations of the Error Search are summarized in Table 7.

7 Conclusion and outlook

Errors in road classification that occur in crowd-sourced geographic data such as OSM can hinder routing applications because of false assumptions about travel time or access limitations. We develop a novel approach to detect these road classification errors by searching for disconnected parts and gaps in subnetworks. A detailed and efficient implementation

Strengths	Limitations
• Intrinsic methodology • Finds road class errors at disconnected components and gaps • Based on basic graph theory • Includes a probability-based ranking of identified gap candidates • Errors which might lead to large detours are found first because of the rating system • Applicable to road network databases worldwide • Expendable by the community because the implementation is freely available	• Does not find all road class errors • No information about missed errors available • Requires human user to check potential errors • Can not detect roads wrongly classified in higher class • Some errors might be missed because of the use of an undirected graph • A gap leading to a large detour can cause false positives in the surrounding area • Adaptation to different regions required

Table 7: Summary of strengths and limitations of the Error Search.

of the developed methodology is provided in this study. The methodology is successfully applied in an exemplary case study on the OSM road network dataset of NSW in Australia.

In the introduction, we formulated two main research questions:

1. Is an approach by searching for disconnected parts or gaps in subnetworks able to find potential road classification errors? Is this approach able to provide information about the likelihood that the result is an error?
2. Which parameters (thresholds such as lengths of detours on a subnetwork compared to the complete network) or combination of parameters indicate gaps in road networks best?

To answer our first research questions, we conclude that a search for disconnected parts finds fewer potential road classification errors than a search for gaps. A gap search can find a significant number of misclassifications together with an error probability that results from a multi-parameter rating system. As an answer to our second research question, our study has shown that three parameters are most relevant for the estimation of the error probability: G1 – the ratio of the shortest path distance on the subnetwork network divided by the Euclidean distance, G4 – the shortest path distance on the subnetwork network divided by the shortest path distance on the complete network, and G3 – the number of filtered destinations on the same spot. A combination of these parameters performs best as the rate of errors versus analyzed data is highest, meaning few data has to be checked by a human user to obtain many classification errors. In our case study, only 6 % of gap candidates have to be checked by a human user to find 80 % of identified road classification errors using the multi-parameter rating system.

A major advantage of this methodology is the worldwide transferability to all regions of the world, which have an almost complete road network in OSM. When a different region is analyzed, some values might have to be adapted to fit the characteristics of the new region. Furthermore, it can also be applied for road network data from other sources, as long as it is represented as a graph. The error search is intrinsic such that no additional data besides the road network is required to find misclassifications. The source code of the implementation is published on GitHub [12], such that the study can be easily repeated or applied to other datasets.

The findings of this study can be used in many different applications. On the one hand, it can generally improve OSM data quality by detecting and correcting the errors. On the other hand, it is also a valuable tool for routing algorithms to improve their underlying data and search for potential errors. In research on critical road infrastructure, often, only higher road network levels are analyzed because lower-level roads are less relevant and increase computing time [13, 36]. In these studies, a gap on a high-level subnetwork can cause false results. The presented search for misclassifications can be introduced to these studies as a data preprocessing step. Furthermore, the presented methodology can be applied to assess the quality of OSM by checking for navigability, an important quality aspect of road network data.

In future research, the methodology can be extended. As disconnected components might appear more often in other countries, a methodology to rate the error probability at disconnected network components could also ease the job of manually checking these disconnected network components for errors. Available tags of gap candidates and their connecting roads can be analyzed for continuity, such as the name, surface, or maximum speed of the road and could be included as additional parameters. Also, strokes [37] could be computed to observe their behavior at gap candidates. The information if a connection between gap candidates consists of a single or multiple strokes could, for example, be considered as a parameter for the rating system. These parameters could provide additional information on the error probability. Regarding the rating system as a multi-criteria decision system, it could still be extended, for example, by adding weights to the parameters or by implementing outranking methods to reduce the number of gap candidates. Furthermore, remote sensing can be applied to check if the shape of a road changes at gap candidates, indicating a class change. More case studies can be performed, including different study regions with different qualities of OSM data. These case studies would enable a detailed sensitivity analysis. Ideally, a reference dataset could be used where an automated matching of roads is possible. Then, it would also be possible to identify which types of classification errors can not be found by this methodology. It could also be interesting to apply the Error Search and test the human correction with real OSM contributors to check the applicability of the Error Search.

References

[1] ACADEMIC. List of gaps in Interstate Highways. https://enacademic.com/dic.nsf/enwiki/422412. Accessed 5 November 2020, 2020.

[2] ALMENDROS-JIMÉNEZ, J., AND BECERRA-TERÓN, A. Analyzing the Tagging Quality of the Spanish OpenStreetMap. *ISPRS International Journal of Geo-Information 7*, 8 (2018), 323. doi:10.3390/ijgi7080323.

[3] BARRINGTON-LEIGH, C., AND MILLARD-BALL, A. The world's user-generated road map is more than 80% complete. *PLOS ONE 12*, 8 (2017), 1–20. doi:10.1371/journal.pone.0180698.

[4] BONDY, J. A., AND MURTY, U. S. R. *Graph Theory with Applications*, vol. 290. Macmillan London, 1976.

[5] BROVELLI, M. A., MINGHINI, M., MOLINARI, M., AND MOONEY, P. Towards an automated comparison of OpenStreetMap with authoritative road datasets. *Transactions in GIS 21*, 2 (2017), 191–206. doi:10.1111/tgis.12182.

[6] CIPELUCH, B., JACOB, R., WINSTANLEY, A., AND MOONEY, P. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resuorces and Enviromental Sciences* (Leicester, UK, 2010), University of Leicester, p. 4.

[7] CORCORAN, P., JILANI, M., MOONEY, P., AND BERTOLOTTO, M. Inferring semantics from geometry: the case of street networks. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '15* (Bellevue, Washington, 2015), ACM Press, pp. 1–10. doi:10.1145/2820783.2820822.

[8] DAVIDOVIC, N., MOONEY, P., STOIMENOV, L., AND MINGHINI, M. Tagging in volunteered geographic information: An analysis of tagging practices for cities and urban regions in OpenStreetMap. *ISPRS International Journal of Geo-Information 5*, 12 (2016), 232. doi:10.3390/ijgi5120232.

[9] DEMETRIOU, D. Uncertainty of OpenStreetMap data for the road network in Cyprus. In *Fourth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2016)* (2016), vol. 9688, International Society for Optics and Photonics, p. 968806. doi:10.1117/12.2239612.

[10] DORN, H., TÖRNROS, T., AND ZIPF, A. Quality evaluation of VGI using authoritative data – A comparison with land use data in Southern Germany. *ISPRS International Journal of Geo-Information 4*, 3 (2015), 1657–1671. doi:10.3390/ijgi4031657.

[11] GIRRES, J.-F., AND TOUYA, G. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS 14*, 4 (2010), 435–459. doi:10.1111/j.1467-9671.2010.01203.x.

[12] GUTH, J. johannaguth/OSM-errorsearch-road-class: First release of the OSM road class error search — tested for the regions New South Wales (Australia) and Bio-Bio and Maule (Chile), 2020. doi:10.5281/zenodo.3957386.

[13] GUTH, J., WURSTHORN, S., BRAUN, A. C., AND KELLER, S. Development of a generic concept to analyze the accessibility of emergency facilities in critical road infrastructure for disaster scenarios: exemplary application for the 2017 wildfires in Chile and Portugal. *Natural Hazards 97*, 3 (2019), 979–999. doi:10.1007/s11069-019-03672-7.

[14] GUTH, J., WURSTHORN, S., AND KELLER, S. Multi-parameter estimation of average speed in road networks using fuzzy control. *ISPRS International Journal of Geo-Information 9*, 1 (2020), 55. doi:10.3390/ijgi9010055.

[15] HAKLAY, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design 37*, 4 (2010), 682–703. doi:10.1068/b35097.

[16] HANCOCK, M. W., AND WRIGHT, B. *A policy on geometric design of highways and streets*, 7 ed. American Association of State Highway and Transportation Officials, Washington, DC, USA, 2013.

[17] JACKSON, S., MULLEN, W., AGOURIS, P., CROOKS, A., CROITORU, A., AND STEFANIDIS, A. Assessing completeness and spatial error of features in volunteered geographic information. *ISPRS International Journal of Geo-Information 2*, 2 (2013), 507–530. doi:10.3390/ijgi2020507.

[18] JILANI, M., CORCORAN, P., AND BERTOLOTTO, M. Multi-granular street network representation towards quality assessment of OpenStreetMap data. In *Proceedings of the Sixth ACM SIGSPATIAL International Workshop on Computational Transportation Science - IWCTS '13* (Orlando, FL, USA, 2013), ACM Press, pp. 19–24. doi:10.1145/2533828.2533833.

[19] JILANI, M., CORCORAN, P., AND BERTOLOTTO, M. Automated highway tag assessment of OpenStreetMap road networks. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - SIGSPATIAL '14* (Dallas, Texas, 2014), ACM Press, pp. 449–452. doi:10.1145/2666310.2666476.

[20] KARAGIANNAKIS, N., GIANNOPOULOS, G., SKOUTAS, D., AND ATHANASIOU, S. Osmrec tool for automatic recommendation of categories on spatial entities in OpenStreetMap. In *Proceedings of the 9th ACM Conference on Recommender Systems* (New York, NY, USA, 2015), RecSys 15, Association for Computing Machinery, pp. 337–338. doi:10.1145/2792838.2796555.

[21] KELLER, S. ReMAPTCHA: A map-based anti-spam method that helps to correct OpenStreetMap. In *GI_Forum 2014 — Geospatial Innovation for Society* (Salzburg, 2015), Austrian Academy of Sciences Press, pp. 40–44. doi:10.1553/giscience2014s40.

[22] KOUKOLETSOS, T., HAKLAY, M., AND ELLUL, C. An automated method to assess data completeness and positional accuracy of OpenStreetMap. In *GeoComputation* (2011), vol. 3, Citeseer, pp. 236–241.

[23] LIU, B. Route finding by using knowledge about the road network. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 27*, 4 (1997), 436–448. doi:10.1109/3468.594911.

[24] LONDÖGÅRD, H., AND LINDBLAD, H. Improving the OpenStreetMap data set using deep learning. *Department of Computer Science, Faculty of Engineering, Lund University Master's thesis* (2018), 73.

[25] LUDWIG, I., VOSS, A., AND KRAUSE-TRAUDES, M. A comparison of the street networks of Navteq and OSM in Germany. In *Advancing Geoinformation Science for a Changing World*. Springer, Berlin, Heidelberg, 2011, pp. 65–84. doi:10.1007/978-3-642-19789-5_4.

[26] MAHABIR, R., STEFANIDIS, A., CROITORU, A., CROOKS, A., AND AGOURIS, P. Authoritative and volunteered geographical information in a developing country: A comparative case study of road datasets in Nairobi, Kenya. *ISPRS International Journal of Geo-Information 6*, 1 (2017), 24. doi:10.3390/ijgi6010024.

[27] MAJIC, I., NAGHIZADE, E., WINTER, S., AND TOMKO, M. Discovery of topological constraints on spatial object classes using an extended topological model. *Journal of Spatial Information Science 2019*, 18 (2019), 1–30. doi:10.5311/JOSIS.2019.18.459.

[28] MAJIC, I., WINTER, S., AND TOMKO, M. Finding equivalent keys in OpenStreetMap: Semantic similarity computation based on extensional definitions. In *Proceedings of the 1st Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery - GeoAI '17* (Los Angeles, California, 2017), ACM Press, pp. 24–32. doi:10.1145/3149808.3149813.

[29] OSM WIKI. Using OpenStreetMap – OpenStreetMap wiki. http://wiki.openstreetmap.org/wiki/Main_Page. Accessed on 24 June 2020., 2020.

[30] OSMOSE CONTRIBUTORS. Osmose — backend. https://github.com/osm-fr/osmose-backend. Accessed on 30 June 2020., 2020.

[31] PSMA AUSTRALIA. PSMA — street network (line) (August 2018) — AURIN data. https://data.aurin.org.au/dataset/psma-street-line-201808-na. Accessed on 16 March 2019.

[32] PSMA AUSTRALIA. Transport and topography — data product description. Tech. rep., Sydney (Australia), 2019.

[33] RÖMAN, L., AND FINNMAN, S. Algorithmic approach to error correction in map datasets using conflation techniques. *Department of Computer Science, Faculty of Engineering, Lund University Master's thesis* (2018).

[34] SEHRA, S., SINGH, J., AND RAI, H. Assessing OpenStreetMap data using intrinsic quality indicators: An extension to the QGIS processing toolbox. *Future Internet 9*, 2 (2017), 15. doi:10.3390/fi9020015.

[35] STYPA, M., AND SANDBERG, H. Improving the semantic accuracy and consistency of OpenStreetMap using machine learning techniques. *Department of Computer Science, Faculty of Engineering, Lund University Master's thesis* (2018).

[36] TAYLOR, M. A., AND SUSILAWATI. Remoteness and accessibility in the vulnerability analysis of regional road networks. *Transportation Research Part A: Policy and Practice 46*, 5 (2012), 761–771. doi:10.1016/j.tra.2012.02.008.

[37] THOMSON, R. C. The 'stroke' Concept in Geographic Network Generalization and Analysis. In *Progress in Spatial Data Handling: 12th International Symposium on Spatial Data Handling*, A. Riedl, W. Kainz, and G. A. Elmes, Eds. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 681–697. doi:10.1007/3-540-35589-8_43.

[38] TIMPF, S., AND FRANK, A. U. Using hierarchical spatial data structures for hierarchical spatial reasoning. In *International Conference on Spatial Information Theory* (1997), Springer, pp. 69–83. doi:10.1007/3-540-63623-4_43.

[39] VANDECASTEELE, A., AND DEVILLERS, R. Improving volunteered geographic information quality using a tag recommender system: the case of OpenStreetMap. In *Openstreetmap in GIScience*. Springer, 2015, pp. 59–80. doi:10.1007/978-3-319-14280-7_4.

[40] WENDT, D., PATRUSHEV, A., PAVIE, A., ANDERSON, J., KUMAR, J. K., DE SOUSA, L., AGARWAL, S., NAGASE, K., OBE, R., AND VERGARA, V. Osm2pgrouting — Import tool for OpenStreetMap data to pgRouting database. https://github.com/pgRouting/osm2pgrouting, 2018.

[41] ZHANG, H., AND MALCZEWSKI, J. Accuracy evaluation of the canadian OpenStreetMap road networks. *International Journal of Geospatial and Environmental Research 5*, 2 (2017).