RESEARCH ARTICLE

Road network selection for small-scale maps using an improved centrality-based algorithm

Roy Weiss and Robert Weibel

Department of Geography, University of Zurich, Zurich, Switzerland

Abstract: The road network is one of the key feature classes in topographic maps and databases. In the task of deriving road networks for products at smaller scales, road network selection forms a prerequisite for all other generalization operators, and is thus a fundamental operation in the overall process of topographic map and database production. The objective of this work was to develop an algorithm for automated road network selection from a large-scale (1:10,000) to a small-scale database (1:200,000). The project was pursued in collaboration with swisstopo, the national mapping agency of Switzerland, with generic mapping requirements in mind. Preliminary experiments suggested that a selection algorithm based on betweenness centrality performed best for this purpose, yet also exposed problems. The main contribution of this paper thus consists of four extensions that address deficiencies of the basic centrality-based algorithm and lead to a significant improvement of the results. The first two extensions improve the formation of strokes concatenating the road segments, which is crucial since strokes provide the foundation upon which the network centrality measure is computed. Thus, the first extension ensures that roundabouts are detected and collapsed, thus avoiding interruptions of strokes by roundabouts, while the second introduces additional semantics in the process of stroke formation, allowing longer and more plausible strokes to built. The third extension detects areas of high road density (i.e., urban areas) using density-based clustering and then locally increases the threshold of the centrality measure used to select road segments, such that more thinning takes place in those areas. Finally, since the basic algorithm tends to create dead-ends—which however are not tolerated in small-scale maps—the fourth extension reconnects these dead-ends to the main network, searching for the best path in the main heading of the dead-end.

Keywords: map generalization, road network selection, strokes, network centrality, betweenness centrality

1 Introduction

One of the ultimate goals of many national mapping agencies (NMAs) is to derive smaller-scale data sets from a single, detailed database [1, 11]. One of the key operators of map generalization is the selection operator, which is usually carried out as the first step of the generalization process, in order to select an initial subset of map objects that will subsequently be subjected to further generalization operators, such as shape simplification and smoothing, aggregation, exaggeration, and displacement [27]. Due to the overriding importance of the road network among the feature classes of topographic maps, in particular small scale maps, road network selection is a particularly important generalization operator, and has thus attracted considerable research interest over the past two decades. Road network selection, however, is a non-trivial process, as it has to cope with maintaining a coherent network of roads, each having a particular shape, angle, orientation, and length [7]. On the global level, the road network exhibits density variation across the map, ranging from high densities in urban areas to low densities in rural areas. This density variation should also be maintained. Not surprisingly, different approaches to road network selection exist, but so far no single algorithm has clearly outperformed others. Moreover and crucially, algorithms have usually only been tested on few, small datasets. To our knowledge, none has been tested and optimized on large datasets with diverse characteristics, and against real production requirements. Thus, most mapping agencies still carry out the selection process interactively (i.e., manually), or use simple selection algorithms that necessitate subsequent interactive cleaning.

The objective of this work was to develop an algorithm for automated road network selection from a large-scale (e.g., 1:10,000) to a small-scale database (1:200,000 or smaller). The project was pursued in collaboration with the Federal Office of Topography of Switzerland (swisstopo). This collaboration gave us access to large datasets, real-life production requirements, and thorough cartographic expertise in the evaluation phase. However, due to the universal role of the transportation networkconnecting places and providing accessibility of places it was possible to define a set of rather generic constraints for the envisaged algorithm that have the potential of being adaptable to the requirements of other map producers (Section 4.1).

Swisstopo thus served as case study in this work, but the aim was to develop a more generic solution. Like many other NMAs, swisstopo has developed a high-resolution topographic database as a basis for the derivation of a complete suite of map and database products at a series of scales. This basic database is called TLM3D. TLM3D is the large-scale topographic landscape model of Switzerland. It includes natural and artificial topographic features and is the most extensive and accurate 3-D vector official data set of the country, with a nominal scale between 1:5,000 and 1:25,000 (depending on the area and feature class). From TLM3D, smaller scale data sets are derived, such as VECTOR200, a cartographic product at a scale of 1:200,000. Owing to the vast scale difference and the complexity of map generalization involved, the derivation process is currently still carried out manually in an interactive system. Figure 1 gives an impression of the detail contained in TLM3D, as opposed to VECTOR200. TLM3D basically contains every single footpath, while VECTOR200 is restricted to the major network connections. The aim of this work was then to develop an algorithm that would allow road network selection between products exemplified by TLM3D and VECTOR200, respectively.
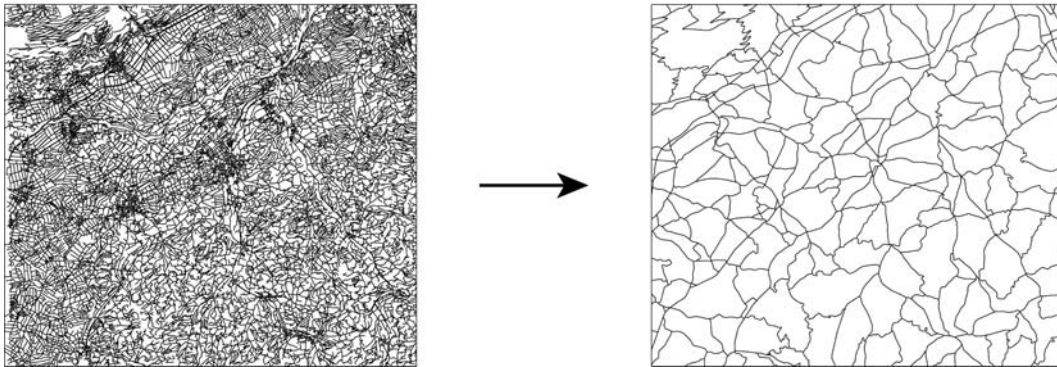
Figure 1: Test area of 26 km × 23 km around Langenthal, Switzerland. Source dataset (TLM3D) and the manually produced target dataset (VECTOR200). (Data © swisstopo)

Experiments with several candidate algorithms showed that the algorithm by Jiang and Claramunt [15], which is based on a topological measure of network centrality, had considerable potential (Section 4.2). However, several difficulties remained and not all of the requirements could be fulfilled. Therefore, an in-depth analysis was carried out as to how additional constraints and extensions could improve the results.

The main contribution of this paper thus consists of four extensions that address deficiencies of the basic version of the centrality-based algorithm by [15] and lead to a significant improvement of the results. In contrast to most previous studies, this research is solidly rooted in requirements defined by production cartography, and based on experimental evaluation using several large and diverse datasets from a real production environment. While this paper provides a condensed version of our proposed algorithm, the full detail as well as additional experiments and map examples can be found in [33].

The remainder of this paper is organized as follows: in Section 2, an overview of related work is introduced. The basic centrality-based algorithm underlying our work is then explained in Section 3. In Section 4, the requirements posed to an automatic selection algorithm, as well as the resulting problems of the basic approach by [15], as well as other pertinent algorithms [8, 19], are shown. In Section 5, four solutions to these problems extending the basic algorithm of [15] are described. Section 6 reports on the achieved results and Section 7 discusses them thoroughly. Finally, Section 8 ends the paper with concluding remarks.

2   Related work

According to [20], the existing road network selection algorithms can be categorized into three major groups: (1) semantics-based selection, (2) graph-based selection, and (3) stroke-based selection.

Semantics-based selection is the simplest of the three approaches. Streets are selected in a ranked order according to their relative importance of attributes [20], such as road class, length, or other qualifying attributes. While commonly applied in practice, particularly for the selection of discrete feature classes such as settlements, such methods do not produce

adequate results for most scales if applied to networks, as topological relationships are largely ignored. This weakness explains why the approach is rarely used in the literature.

Graph-based methods manipulate road networks as connected graphs (for a short introduction to graph theory, see Section 3). Graph-based selection was first introduced by Mackaness and Beard [21], who also point out that map generalization requires understanding and modeling at the geometric, topological, and attribute levels. They have used a so-called minimum spanning tree (MST) to select the most important streets while maintaining connectivity between cities, which is a major advantage over a simple semantics-based approach.

[29] used a graph-based approach that employs shortest path algorithms between nodes in the network, which produce a set of rankings reflecting the importance of the segments. In addition to MSTs, which are again used to maintain connectivity, the rankings are thus used in map density reduction and generalization.

Jiang and Claramunt [15, 16] introduced a graph-theoretic approach that uses a line graph (also referred to as dual or connectivity graph in transportation science [24]) and centrality measures to determine the importance of roads.

Stroke-based selection algorithms were introduced by Thomson and Brooks [30] and are based on the perceptual grouping principle of good continuation [31]. Segments are grouped together such that each pair forms the smallest angle possible [31], which results in long, linear segments that extend through junctions [28]. The generated strokes are then sorted according to some attributes, such as length or road class and the actual selection is then performed according to this hierarchy [28].

Next to algorithms belonging to one of the aforementioned groups, there also exist other approaches. Mesh-based algorithms, which make use of the areas formed between roads, were introduced more recently [8, 14]. Li and Zhou [19] improved upon the approach by using mesh-based algorithms in conjunction with a stroke-based algorithm to handle areal and linear features separately. An entirely different approach is the agent-based approach for road selection by [23].

Based on thorough testing in collaboration with a related project [3, 4], we decided to use a stroke-based centrality approach as a basis for our work, as it has shown the most promising results for *small-scale* target maps. Interestingly, if the target maps are in the medium scale range, such as in the project by Benz [3,4], where a target scale of 1:50,000 was used, it was found that the best among the existing algorithms was the integrated stroke-mesh algorithm by Li and Zhou [19]. Intuitively this makes sense, since at medium scales, rather fine-grained structures such as individual meshes or strokes of the road network still need to be maintained, which favors the stroke-mesh algorithm by Li and Zhou [19]. Conversely, at small scales, the road network is increasingly focused on the main hubs, and thus a topological centrality measure used by Jiang and Claramunt [15] provides a good estimate of these hubs.

3 The stroke-based centrality approach

The approach that serves as the main basis for our methodology was first proposed by Jiang and Claramunt [15, 16] and is based on graph-theoretic principles discussed by Mackaness and Beard [21], as mentioned in Section 2. In graph-theoretic approaches, the road network is seen as a graph, where the road segments represent the edges, while junctions (or end-
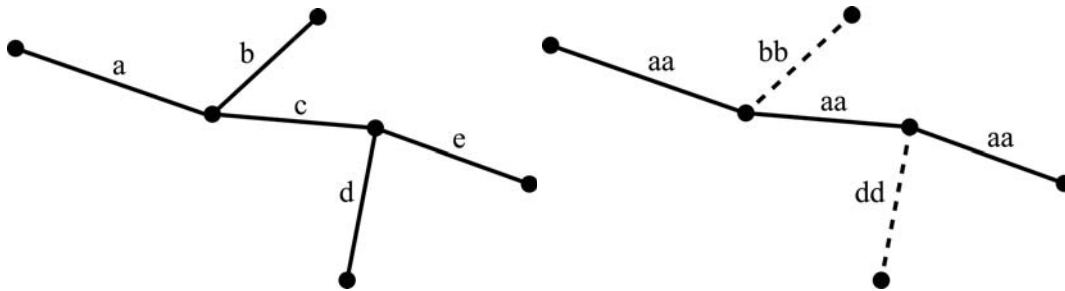
Figure 2: Following the principle of good continuation, the segments of the network to the left are concatenated as shown on the right (based on [35]).

points of segments) represent nodes. Although it would be possible to build an algorithm on the segment level (i.e., on the primal graph of the road network [24]), this does not represent the network well compared to how humans see a road network. A cartographer does not assess the importance of individual segments, but of interconnected roads as a whole—the algorithm should do the same. It should not keep or eliminate individual segments, but entire roads. However, road names—which were used in the original algorithm by Jiang and Claramunt—are not always available in topographic databases and it is not always advisable to generate roads based on road names alone [35]. That is the reason why the approach presented in this paper builds upon strokes, which were introduced by Thomson and Richardson [31]. Hence, our algorithm is a combination of the graph-based and stroke-based approaches.

3.1 Strokes

Strokes group together individual road segments by means of the conceptual grouping principle of good continuation [30]. If translated to an algorithm, this essentially means that at each intersection, the angle between all segments is calculated and pairs of segments are grouped together based on their angle. In Figure 2, an example of such a concatenation is shown. For deciding which segments to concatenate, the algorithm examines the angle between each possible pair of segments and chooses the pair-combinations having the smallest angle in between. While the original proposal consisted of solely evaluating the angle of two segments, others have included additional attributes, such as the class of the road [35].

3.2 Line graph

After generating the strokes, a line graph is created. In such a representation of a network graph, edges become nodes and nodes become edges, as commonly understood in transportation networks [24]. This means that each node represents a stroke, potentially consisting of multiple segments. The edges represent the intersections between the strokes. Such an example is shown in Figure 3, where a road network G is shown, consisting of various segments and colored strokes. The line graph representation H, based on the strokes of the
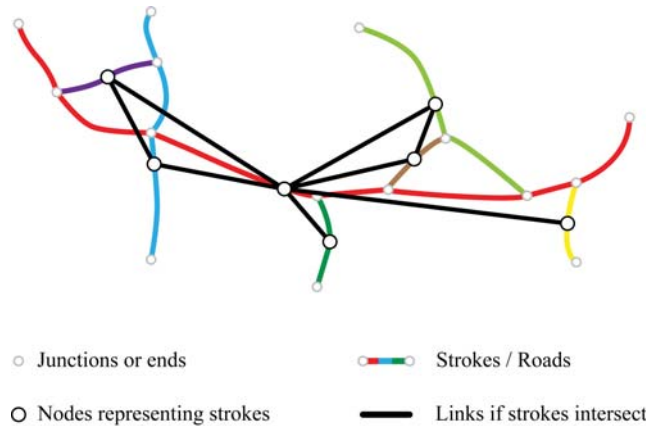
Figure 3: A road network with colored strokes and its line graph (based on [18]).

primal graph G, is shown in black. It is important to note that each stroke is mapped to exactly one node, as this forms the basis for the centrality calculation.

3.3 Centrality and centrality measures

Based on this line graph, different centrality measures can be computed which describe and rate a node's importance in the network. Because nodes represent the generated strokes, they directly determine their importance. While there exist various centrality measures, which were primarily developed for the analysis of social networks [5, 12], the most useful one for the selection of road networks is the so-called betweenness centrality.

Betweenness centrality expresses to what extent a node is located in between the shortest paths that connect a pair of nodes [15]. It measures how important a node is for the shortest paths formed between the other nodes. Thus, it indicates whether the node has a bridging role in a graph. According to Brandes [6], the betweenness $C_B(v)$ of a vertex $v \in V$, where V is the set of nodes and E the set of edges in a graph G, is defined as:

$$C_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

where $\sigma(s,t)$ is the number of shortest (s,t)-paths (also called geodesics) and $\sigma(s,t|v)$ the number of shortest (s,t)-paths passing through node v. If $s = t$, let $\sigma(s,t) = 1$, and if $v \in \{s,t\}$, let $\sigma(s,t|v) = 0$ [6]. By convention, let $0/0 = 0$. The measure can be interpreted as the degree to which a node has control over pair-wise connections between other nodes, assuming that the importance of connections is equally divided among all shortest paths of each pair [6].

Other measures, such as closeness centrality (the average distance of a node to all other nodes) and degree centrality (the direct links of a node) can also be used to further describe the rank of a road in a network. However, both closeness and degree centrality are not especially reliable in determining the hierarchical importance of a stroke in a network [34]. Degree centrality can be used to locally determine the importance of a stroke, but it is not suited for a more global evaluation of the network. While closeness centrality is a global

measure, it is not well suited to determining whether a road is structurally important. Betweenness centrality, on the other hand, can be used to identify roads which take a bridging role between different (topological) shortest paths, and thus in the entire network [18]. The actual selection is then achieved by simply setting a threshold on the centrality values, and selecting those strokes that exceed the threshold.

4 Cartographic requirements and initial assessment

4.1 Cartographic requirements

The cartographic requirements of the final product (i.e., the pruned road network for a target scale of 1:200,000) act as constraints under which the algorithm must operate [2, 13]. Constraints do not define how a solution has to be reached; they simply define aspects that may not be violated: the fewer constraints a particular solution violates, the better the algorithm that produced it [13].

Extensive discussions with swisstopo revealed several hard and soft constraints. Hard constraints can be verified relatively easily using computational methods, while soft constraints allow a wide spectrum of different solutions and need to be evaluated using qualitative methods. For the target scale of 1:200,000, a total of eight constraints were formulated (Table 1). These constraints have been defined in the context of swisstopo map production. However, it is important to note that they are nonetheless typical of the kinds of constraints and requirements that are commonly used in production cartography [11, 26, 27]. The function of transportation networks is universal: to connect places and provide accessibility. Thus, the constraints have been formulated with a focus on connectivity, and even seemingly "Swiss-style" constraints such as S1 and S2 can be adjusted to the particular mapping situation at hand. We therefore expect that our set of constraints can be easily adapted to other environments of topographic map and database production.

The inclusion of all highways and expressways is a requirement set by swisstopo and should be met under all circumstances, as the highways represent the most important roads in a small-scale map. Additionally, one should obviously be able to reach the entries and exits of the highways, which is why an additional constraint was added that ensures the accessibility of the highways. Furthermore, no roads should be disconnected, which means that no isolated roads are allowed. As a last hard constraint, the inclusion of dead-ends is prohibited as they deteriorate the navigability of the road network. A topological algorithm can check all of these hard constraints easily.

As soft constraints, mainly rather general statements were used. Nevertheless, this does not mean that they are in any way less important. In fact, the first soft constraint, which ensures that the overall thinning of the road network is appropriate for the target scale, could be considered as the most important constraint. The second and third soft constraints, which consider the general structure of the network, are intended to ensure that also the appropriate roads are selected in the result. Finally, as a last constraint, the inclusion of important link roads needs to be ensured. The soft constraints cannot be checked easily by an algorithm. One has to resort to visual inspection instead.

H1	Hard	All highways and expressways and everything related to them (such as entries and exits) must be included in the generalized map.
H2	Hard	The entries and exits to and from the highways must be directly connected to the road network.
H3	Hard	The road network must be completely connected. That is, from every point in the network, one has to be able to reach every other point in the network.
H4	Hard	No dead-ends may exist in the final result.
S1	Soft	The overall thinning of the network needs to be appropriate for the target scale of 1:200,000, comparable to swisstopo's VECTOR200 database product.
S2	Soft	The general structure of the road network needs to be maintained.
S3	Soft	Larger towns and cities should still be identifiable in the road network of the final map.
S4	Soft	Important link roads need to be maintained in the pruned road network.

Table 1: Constraints used as requirements to obtain the road network at the target scale of 1:200,000.

4.2   Initial assessment

Figure 4 shows the source data of the transportation network in one of the four test areas used in this project. Figure 5 depicts the corresponding result that was obtained by computing the betweenness centrality for each stroke in the network and selecting strokes above an appropriate threshold value, as in the basic centrality approach of Jiang and Claramunt [15, 16]. However, as explained in Section 3, strokes were used for centrality computation instead of named roads, since road names are often not available in spatial databases.

At first glance, this basic algorithm performed rather well in this test area, which contains rural areas in the northwest, a city (Lucerne) in the eastern part, and a rather mountainous area (manifested by a network of trails and paths) towards the south. It can be seen that most of the highways as well as major roads have been retained. Thus, the algorithm succeeded in finding and selecting some of the most important roads. However, the result is clearly of insufficient quality; it could not be used in a map production process. The basic algorithm is unable to meet the requirements defined in Table 1. One of the main problems is that many dead-ends are created as a result of the stroke-based selection that relies solely on thresholding the centrality measure. It is evident that such dead-ends are unwanted in a high-quality map product. Yet, they cannot simply be deleted as this would result in the loss of a large and important portion of the road network. Another problem is the heterogeneous density variation of the road network. While a slightly higher road density in urban areas is not a bad thing, the resulting density of the city is clearly too high compared to the more rural areas. Many unimportant and rather small roads have been retained in the urban area.

Figure 6 shows the result obtained with the integrated stroke-mesh algorithm introduced in [19]. Obviously, problems can be seen here as well. While the problem of dead-ends appears to be even more serious at a first glance, this is not really the case, as they are, in contrast to Figure 5, "real" dead-ends (i.e., they also appear in the source dataset). A simple truncation of these would in fact be identical to a purely mesh-based approach
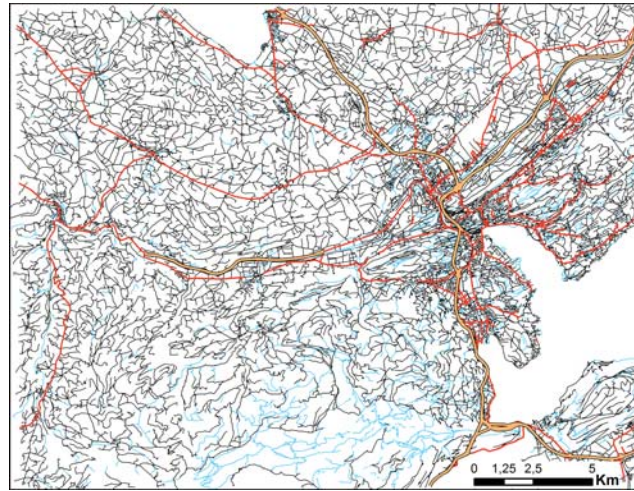
Figure 4: Transportation network of the TLM3D source dataset in the Lucerne test area. Highways shown by broad, orange lines; major roads (> 4 m) by red lines; minor roads (2–4 m) and paths by black lines; and trails (< 2 m) by light blue lines. (Data © swisstopo)

as in [8, 14]. Thus, an integrated approach at this scale does not make a lot of sense, as its main feature—the handling of line-segments in addition to meshes—is unwanted at smaller scales, despite the fact that it is an advantage in large to medium scale maps [4]. Another weakness of the integrated approach is that the density variation of the network is not adequately represented. While the purely centrality-based approach of [15, 16] has difficulties in reducing the density of urban regions, the integrated approach [19] reduces it too severely: the city of Lucerne effectively disappears from the network. Finally, it can be seen that several major roads end abruptly. Thus, the algorithm fails to identify important roads in many cases.

In sum, the cartographic performance of both algorithms is mixed: on the one hand, many relevant roads were extracted from the source dataset; but on the other hand, many problems are still present, to which no simple solutions exist within the scope of the existing algorithms. Nevertheless, the analysis of the preliminary experiments enabled us to decide to use the stroke-based centrality algorithm [15, 16] as the basis of our further developments: its results as well as the underlying concepts better reflect the requirements of small-scale mapping, and thus showed more potential for the scale range targeted by this project. As mentioned in Section 2, we found in a related project that, conversely, the integrated stroke-mesh algorithm [19] has more potential for medium-scale maps. In the next section, several enhancements to the basic algorithm of [15, 16] are introduced to improve upon the above problems.
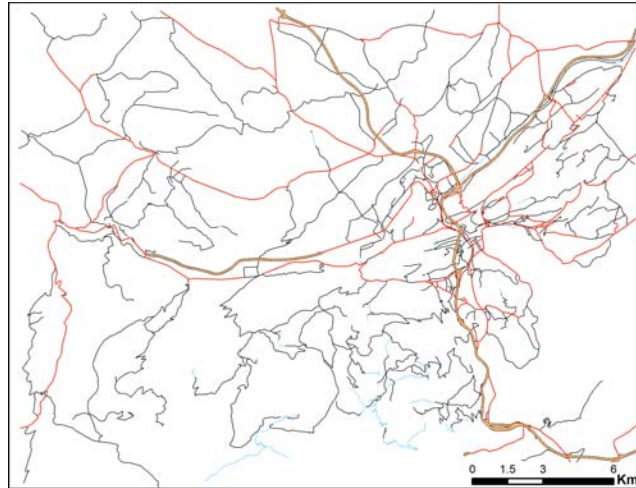
Figure 5: Result of the basic centrality algorithm [15, 16] in the Lucerne test area, based on the original data shown in Figure 4. (Data © swisstopo)
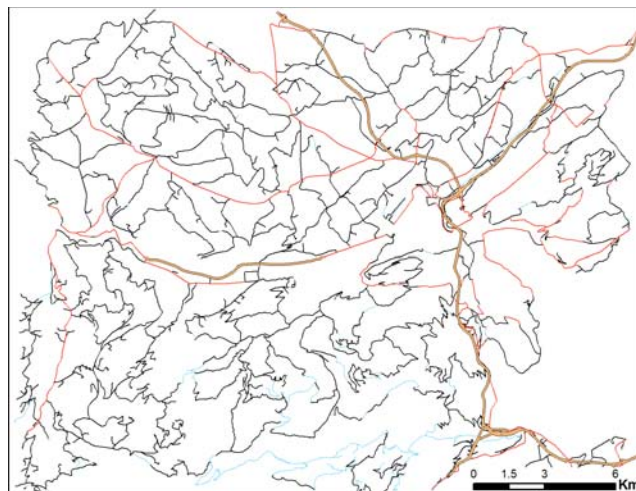

Figure 6: Result of the integrated stroke-mesh algorithm by Li and Zhou [19] in the Lucerne test area, based on the original data shown in Figure 4. (Data © swisstopo)

5 Enhancements to the basic approach

In order to enhance the stroke-based centrality approach, solutions to a number of different problems have been developed. The methods as well as the effect they have on the selection result will be expanded on in this section.

5.1 Detection and collapse of roundabouts

5.1.1 Problem definition

Roundabouts pose a problem in the stroke generation process. Figure 7 demonstrates their detrimental effect: each roundabout forms a stroke by itself, if the purely geometric approach of stroke generation is used. As a result, incoming strokes are disrupted and their continuity impeded.

[32] and [34] also recognized this problem and incorporated two different methods in their approaches to deal with roundabouts. The one presented in [32] uses a measure of polygon compactness to identify roundabouts. [34] use cluster analysis to detect complex junctions, including roundabouts. A similar method was already used by Mackaness and Mackechnie [22] and several other algorithms have been published (e.g., [25]). However, the computing and cartographic performance of these methods was often not determined systematically by their authors. Additionally, and most importantly, the existing methods require additional algorithms (e.g., the creation of meshes), which would lead to additional complexity of both the algorithm and the running time. Hence, a completely different approach was pursued in our work, which exploits and re-uses the already generated strokes to detect roundabouts in the data set, thus enabling a highly efficient solution.

5.1.2 Solution

As has been shown in Figure 7, roundabouts tend to generate single, isolated strokes. Hence, this circumstance can be exploited to analyze each stroke individually. As a first step, each stroke is analyzed and tested whether it forms a loop. This can be done easily by using graph-based methods. Because this also results in false positives, as there also exist strokes that form loops but have nothing to do with roundabouts, the detected loops need to be filtered afterwards. Three parameters, which are shown in Figure 8, are used to describe roundabouts and filter the obtained list of loops:

(1) **TotalLength**: The total length of the segments forming the roundabout may not exceed 200 m.
(2) **MaxLength**: The maximum length of a single segment inside a roundabout may not exceed 100 m.
(3) **Connected Nodes**: There must be at least two nodes that are connected to three edges in the primal graph of the road network.

Experiments in four different test areas have shown that the first two parameters are sufficient to extract the actual loops. The third parameter is needed to further reduce the result set, as it ensures that the loop does not simply consist of a dead-end loop that loops back to the same node.

After the roundabouts have been filtered, they are collapsed in the second stage of the algorithm. This is a straight-forward operation, which simply computes the centroid of the

Figure 7: The roundabout problem: they disrupt important strokes and impede their continuity. It is also important to note that the roundabouts form exactly one stroke in most cases, which is a property exploited in the detection algorithm. (Data © swisstopo)
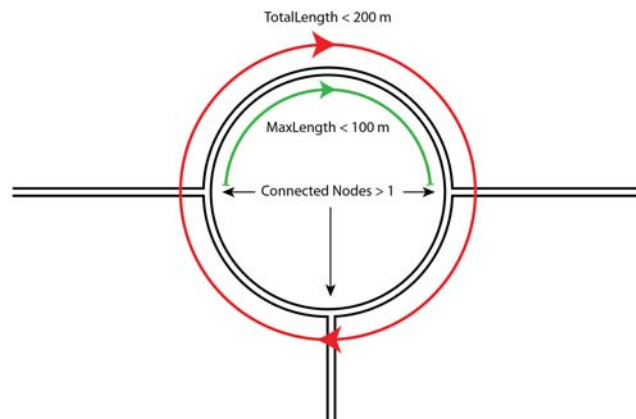

Figure 8: The three parameters used to define a loop as a roundabout.

Figure 9: The resulting strokes exhibit a much better continuity after the roundabouts have been identified and collapsed. (Data © swisstopo)

roundabout and extends the relevant neighboring strokes to this newly generated node. In a last step, the strokes are reevaluated at the relevant positions. Figure 9 shows the same extract as Figure 7 after the reevaluation. It shows that the strokes maintain a much better continuity and are no longer split up. This is an important improvement, as it reduces a multitude of splits in the selected result and generates more accurate centrality values. The described algorithm was evaluated on four different map samples with a combined segment count of over 130,000. It was able to detect 221 roundabouts in total, with no false positives. The algorithm missed no roundabout in a strict sense, but was not able to also identify plazas in cities which do not form circles, as the incoming streets form a flat angle with segments of the plaza itself (three such cases have been identified in our test areas). Because the running time of the algorithm is fast even for larger datasets of several hundred thousand road segments (<1s, including the stroke generation and adaptation), it is also viable for other applications where a quick detection of roundabouts is necessary.

5.2   Enhanced semantics for stroke generation

5.2.1   Problem definition

Traditionally, strokes are generated using only the purely geometric principle of good continuation. [35] came to the conclusion that the usage of thematic attributes in addition to the principle of good continuation improves the result. However, the TLM dataset used in this paper stores the road classes with a high degree of thematic granularity: instead of the commonly used ordinal road classes ("Highway," "Major Road," "Minor Road" etc.), classes are formed denoting the width of a road. Hence, what in other systems may be called a "Main Road" often changes its class multiple times as it widens or narrows. This
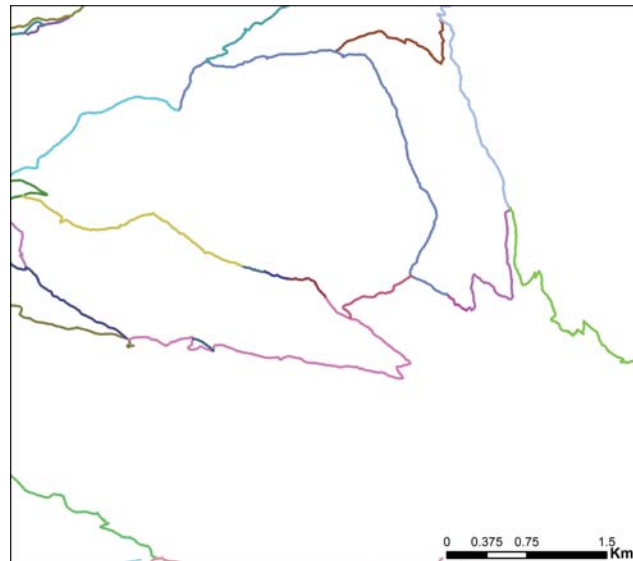
Figure 10: Detail of the mountainous Davos test area that contains many smaller paths and trails. This snippet highlights the problem of the traditional stroke generation approach, where the road class is largely ignored and a small angle threshold is used. The paths and trails are broken up into many different strokes. (Data © swisstopo)

is also the case at many intersections, where the road temporarily widens to accommodate more lanes.

Another problem lies in the selection of the threshold angle. While many authors suggested a threshold angle between $40°$ and $60°$ to maintain the principle of good continuation [18, 31, 34, 35], an inspection of the data revealed that this might not be enough. Figure 10 depicts such a situation in a mountainous area with winding roads, where the strokes are split up multiple times: the road classes change at these positions and the angle between the segments is above $60°$, which results in this picture.

As such, it is necessary to adapt the algorithm that builds the strokes to take into account road classes, while allowing different classes to join together, as proposed by Thomson [28]. At the same time, the angle threshold is increased—depending on the road classes involved—to allow strokes to continue even when sharp turns occur.

5.2.2 Solution

As a first step, the detailed road classes defined in the TLM3D dataset were assigned to one of the groups depicted in Figure 11. For example, roads having a width between 6 and 10 meters were assigned to the "Major Road" group. After each segment has been put into a group, the concatenations between them are constrained such that only segments belonging to the same group are allowed to be connected to the same stroke. This way, it is possible to eliminate unrealistic concatenations of trails and major roads or even highways. In addition, a distinct angle threshold is used for each group. While an angle threshold of $60°$ produced good results for highways, major roads, and minor roads, the threshold

Figure 11: This diagram shows the five primary road class groups, which were created based on the width-based road classes used in the TLM3D dataset. Each group uses a separate angle threshold in the stroke building process. The algorithm can be adapted in order to either constrain or relax the conditions under which strokes are generated.

was clearly too low for the narrower paths and trails. As such, the angle threshold was increased to $75°$ and $90°$, respectively.

The proposed hierarchical method describes a way to include thematic attributes (such as the road class) during the stroke building process in a meaningful way. Using such a hierarchical method, it is possible to constrain the different concatenation possibilities (e.g., we want highways to produce separate strokes), but it also allows different road classes to be interconnected dynamically using separate thresholds (e.g., major and minor roads are allowed to be concatenated, but with a stricter threshold). In addition, the algorithm was implemented in such a way that an easy replacement of the strategies is possible. Hence, it is entirely possible to use different stroke generators for urban and mountainous areas, to adapt to the different road network characteristics. While the use of differentiated thresholds for the various road classes may not represent a fundamentally new concept, it nevertheless improves the strokes tremendously. As the strokes themselves build the foundation of the centrality algorithm, the accuracy of the stroke formation will crucially influence the quality of the road selection result.

The results of this hierarchical approach are visible in Figure 12, which depicts the same area as Figure 10. It is evident that the continuity of the strokes was maintained much better than in the purely geometric approach.

5.3 Adaptive threshold based on road density

5.3.1 Problem definition

It has been shown in Section 4.2 that the rural areas were thinned out excessively in the pruned result when the basic approach is used. While a density difference in urban and rural areas is desired to some extent in order to retain the structure of the original network, urban areas remain still too dense after the initial pruning, considering the small target scale (Figure 13). The reason for this lies in the way the betweenness centrality is calculated. Because in cities there exist more (but short) strokes the betweenness centrality will reach higher values in those areas. A solution to this problem is not straightforward, as there only exists one global betweenness threshold, according to the basic algorithm. If the threshold is increased more in order to reduce the number of selected streets, the rural areas, which are already sufficiently pruned, will be thinned out even more.
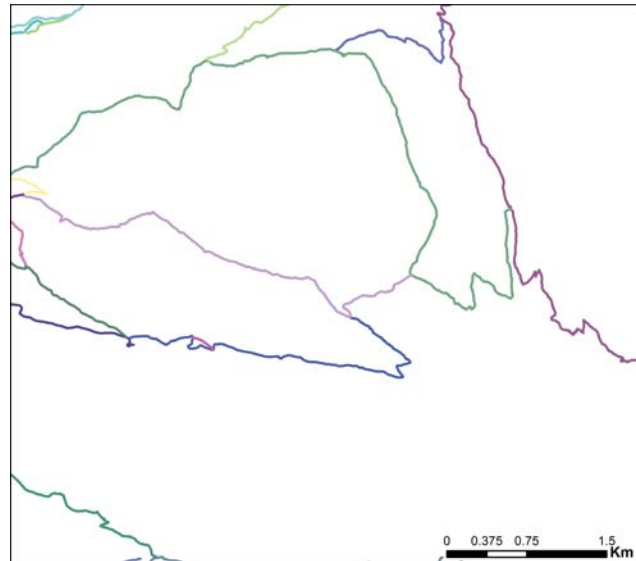
Figure 12: Same area as shown in Figure 10. In contrast to the purely geometric stroke approach, the enhanced approach successfully concatenates the paths and trails in this area and significantly improves the continuity of the strokes. (Data © swisstopo)
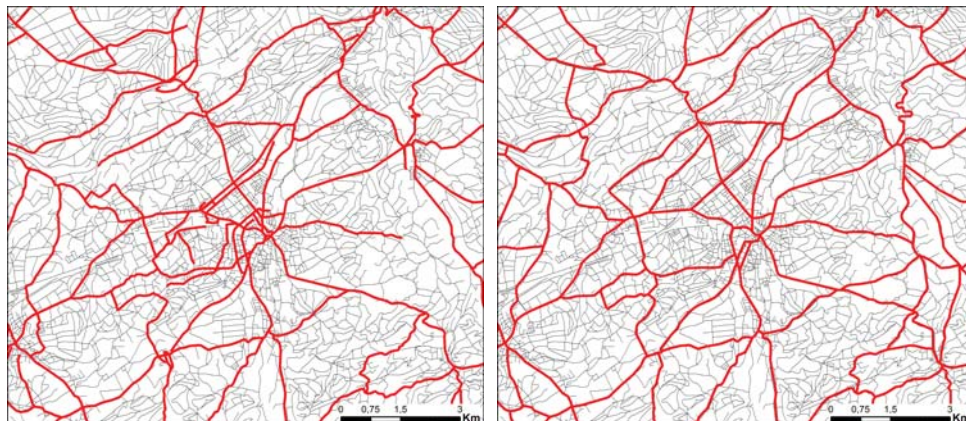

Figure 13: (a) The area around the city of Langenthal (population: 15,000) after an initial selection was performed using the improvements presented in Sections 5.1 and 5.2. (b) For comparison the same area in VECTOR200, the digital 1:200,000 map of swisstopo. Note that the density in the city remains too high compared to the surrounding rural area. (Data © swisstopo)
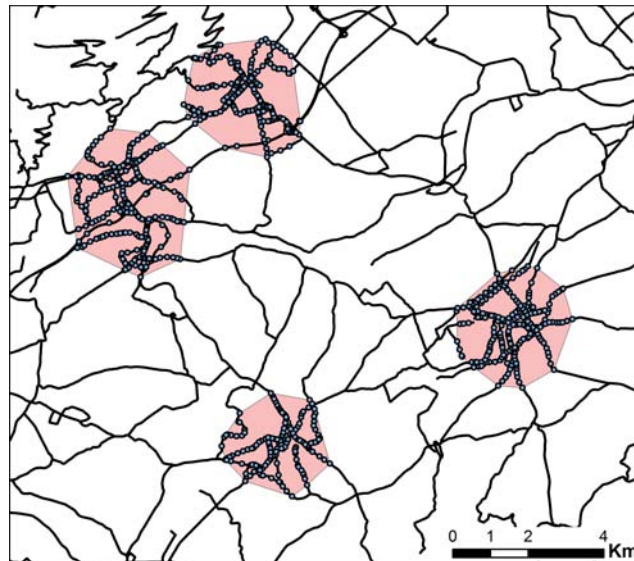
Figure 14: Result of the DBSCAN algorithm in an extract from the Langenthal test area, showing four different clusters. The centroid points of the individual street segments, which were detected as lying in dense areas by the DBSCAN algorithm, are shown with the respective convex hulls. (Data © swisstopo)

5.3.2 Solution

To tackle this problem, one has to find a way of controlling the density of rural and urban areas separately, by using different thresholds for the betweenness centrality. The initial approach tested was to use an additional layer of settlement areas to indicate where a different centrality threshold should be used. However, this proved to produce poor results (not *every* settlement area is affected). Thus, an approach using cluster analysis was developed. It identifies dense areas of the pre-selected road network (i.e., after the steps described in 5.1 and 5.2 by using the DBSCAN algorithm [10]. It thus identifies strokes which are potentially affected, by clustering the centroid points of the individual street segments, and adapts the betweenness threshold for these strokes. Figure 14 shows an extract from one of the test areas and the clusters identified as being too dense by their convex hulls. The centrality threshold of these areas is increased by a factor of 2 to 8, depending on the test area, such that the most unimportant roads are excluded from the result.

Hence, this method offers a way to regulate dense areas (mostly towns and cities) and rural areas independently of each other.

A disadvantage of this method is that it still relies on a stroke-based selection. This makes it difficult to handle specific areas in isolation: short strokes, which are completely contained in the areas can be handled easily, but longer strokes, which simply pass through the areas cannot be simply removed, as this could potentially affect long strokes and hence also areas which are far away from the dense cluster. As such, the adapted threshold was constrained to strokes which are mostly inside of dense areas (in this case, at least 60%).

5.4   Reconnection of dead-ends and disconnected parts

5.4.1   Problem definition

The generation of new dead-ends and disconnected network parts is a problem inherent to all stroke-based selection algorithms. While solutions to the removal of disconnected parts exist (for example, the usage of MSTs [34]), they do not solve the problem of dead-ends. Even for large-scale products, a strict requirement for swisstopo is that no new dead-ends are created except for special circumstances (e.g., to provide access to points of interest [3, 4]). This holds also for smaller scales, with the exception that cartographers are allowed to insert dead-ends to hint at certain access roads or special relationships between roads. The main reason for this treatment of dead-ends is that they deteriorate the ability of a map to offer guidance to a user in finding appropriate routes. Furthermore, in a small-scale map in particular, dead-ends are often not of high importance (e.g., access roads in settlements) and offer no important additional information.

5.4.2   Solution

After pre-selection of the network according to the methods described in Sections 5.1 to 5.3, the proposed algorithm additionally tries to reconnect dead-ends to the main network in three steps, using multiple rules:

(1) The first step consists of simply finding all dead-ends in the test area. This is done by finding nodes of degree 1 (i.e., nodes that only have one adjacent edge).
(2) In a second step, the algorithm generates all possible reconnecting paths from the dead-end to the main network, using the unpruned network.
(3) After the reconnecting paths have been generated, the algorithm tries to find an optimal solution. For this purpose, the list of the reconnecting paths found in 2) is sorted according to the following criteria:

   (a) The class of the reconnecting path should be at least as high as the dead-end path. If such paths are found, they are considered first.
   (b) If only paths of a lower road class are found, then the length of the reconnecting paths is compared (i.e., shorter paths are preferred).
   (c) As a final criterion, the direction of the path is considered (see Figure 15). The reconnecting path should lead in the same general direction as the dead-end path. This criterion acts as a hard constraint for the algorithm, as it would otherwise select undesirable solutions. Consider a short dead-end path of a minor road. If the algorithm would only look at the road class and the length of the reconnecting path, a possible solution could be a simple U-turn in the opposite direction. This is obviously not a desirable solution in most cases. Thus, the proposed algorithm only considers reconnecting paths that do not violate the direction requirement (Figure 15).

If the dead-end path cannot be reconnected (because the algorithm was unable to find a suitable reconnection), the whole path will be deleted. Depending on the test area, the algorithm showed a reconnection rate of up to 57%. This number includes dead-ends in the proper sense (i.e., they already existed in the source database), for which no reconnection is possible. Thus, the actual reconnection rate would be even higher.
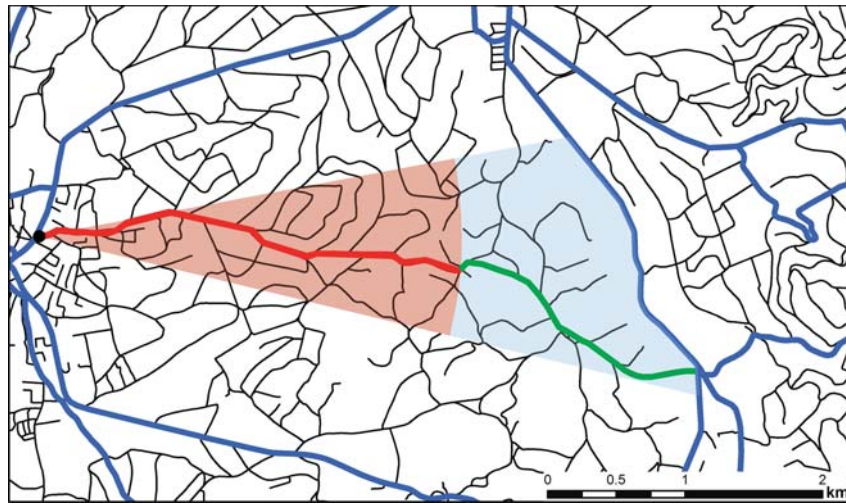
Figure 15: Conceptual illustration showing the principle behind the direction criterion used to find a suitable reconnecting path. The end-node of the reconnecting path (green), which reconnects the dead-end path (red) to the main network (blue), needs to fall inside the blue area. If it falls outside this area, the reconnecting path is discarded, as it does not fulfill the direction criterion. (Data © swisstopo)

6 Results

In Section 4.2, we have shown that the basic stroke-based centrality algorithm was unable to fulfill the requirements defined by swisstopo. This section reports on the results of the improved approach, showing how the presented enhancements were able to reduce the number of problems encountered with the basic algorithm.

The analysis of the results is primarily based on the hard and soft constraints presented in Section 4.1. Table 2 shows various key statistics of the four test areas that were used to evaluate the hard constraints.

	Davos		Lucerne		Langenthal		Zurich	
	Basic	**Impr.**	**Basic**	**Impr.**	**Basic**	**Impr.**	**Basic**	**Impr.**
TS[1]	9,757		26,232		31,318		65,712	
RS[2]	2,416	2,496	5,887	4,603	5.229	6,074	11,742	10,091
HS[3] **total**	0		355		87		853	
HS missing	0	0	86	0	87	0	491	0
HE[4] **total**	0		46		16		95	
HE disconnected	0	0	22	0	16	0	74	1
DP[5]	1	0	1	0	0	0	1	0
DE[6]	52	0	176	0	102	0	339	1
[1]Total Segments, [2]Retained Segments, [3]Highway Segments, [4]Highway Entries and Exits, [5]Disconnected Parts (i.e., a path or sub-graph which is entirely disconnected from the main network), [6]Dead-ends (real and generated)								

Table 2: Key figures of the basic and improved algorithms for four test areas.

While a quantitative analysis is often sufficient to evaluate hard constraints, it is not possible to test whether the soft constraints were fulfilled in the same way. Hence, expert cartographers conducted a qualitative evaluation of the results. The evaluation was conducted by two swisstopo cartographers with several years of experience in the field. For the analysis, they were given an entire day to be able to thoroughly analyze the results in digital form and compare them to existing, generalized data. So, the approach taken was to benefit from the deep knowledge and experience of the expert cartographers, rather than achieving high numbers of test persons. We wanted the experts not only to rate the results, but also let us know what was good or bad, and why and where. For each of the four test areas, three different variants with varying thresholds were analyzed. The experts were asked to fill out a questionnaire consisting of six questions, each of which aimed to evaluate a particular soft constraint (Table 3). They were asked to rate each question according to the rating scheme of [9], where the quality level is either rated as a) good, b) acceptable, c) bad, or d) unusable, and provide reasons and examples for their rating. An overview of the results for the best variants of each test area is shown in Table 4.

In the scope of this work, a *good* rating means that the swisstopo experts would consider the resulting pruned map of the road network as usable in their production process with none or only minor adjustments. A bad rating on the other hand means that the road network needs major adjustments in the particular area. For instance, based on Table 4, the rural areas in the Langenthal data set would need major adjustments if they were to be used in swisstopo products. Note that none of the results was rated as "unusable" regarding any of the soft constraints.

Related Constraint	Topic	Question (translated)
S1	General thinning	How do you judge the general thinning of the road network for the target scale of 1:200,000?
S1a	Thinning in dense areas	How do you judge the thinning in denser areas (e.g., urban areas)?
S1b	Thinning in less dense areas	How do you judge the thinning in less dense areas (depending on the region either rural or mountainous areas)?
S2	Structure preservation	How do you judge the general preservation of the structure in the final result?
S3	Recognizability of urban areas	How well are urban areas recognizable in the final result?
S4	Link roads	How do you judge the preservation of link roads?

Table 3: Questions posed to the swisstopo experts in the qualitative assessment.

Figure 5 shows the result of the basic centrality-based algorithm of [15, 16], while Figure 16 presents the result of the proposed improved algorithm for the same test area (i.e., the Lucerne test area depicted in Figure 4). Figures 17 and 18 illustrate the differences between the results of the basic versus the improved algorithm, respectively, for another test area (a map of the source data is included in the Supplementary Material). It should be noted that segments along the edges of the test areas (within a margin of 1 km to 5 km, depending on the test area) were excluded from the reconnection algorithm and the affected dead-ends were neither removed nor counted in the statistics in Table 2, as these are merely unavoidable edge effects and artifacts when dealing with cut out test areas.

Topic	Da[1]	Lu[2]	La[3]	Zu[4]
General thinning	🟩	🟦	🟦	🟦
Thinning in dense areas	🟥	🟩	🟦	🟥
Thinning in less dense areas	🟩	🟩	🟥	🟦
Structure preservation	🟩	🟩	🟩	🟩
Recognizability of urban areas	🟩	🟩	🟩	🟦
Link roads	🟦	🟩	🟦	🟩

[1]Davos, [2]Lucerne, [3]Langenthal, [4]Zurich

Table 4: An overview of the experts' opinions on the fulfillment of the soft constraints. *Good* is shown in green, *acceptable* in blue and bad in *red*.
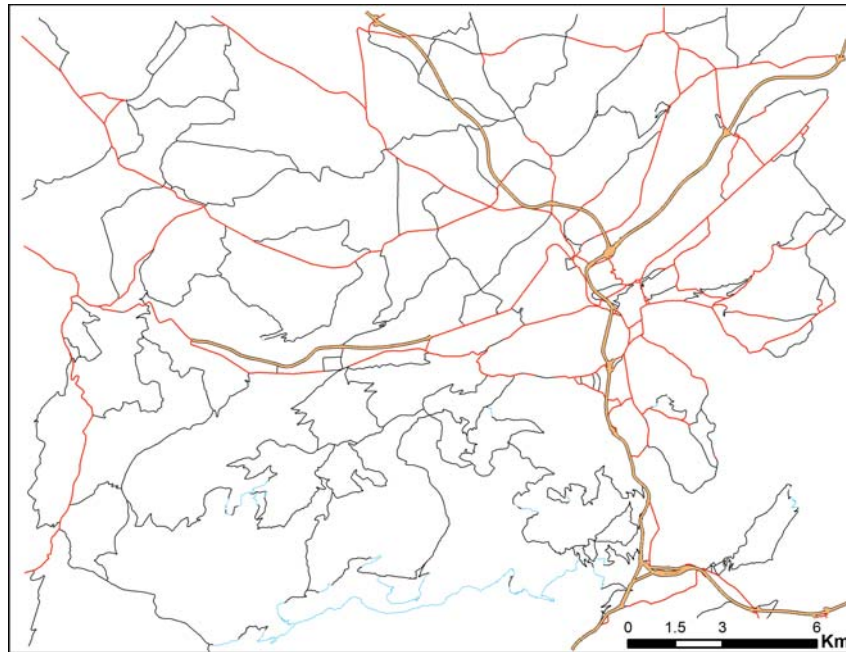

Figure 16: Result of the **improved** centrality algorithm in the Lucerne test area. Compare with Figure 4 (source data) and Figure 5 (result of the basic algorithm). (Data © swisstopo)

Supplementary Material. The Supplementary Material contains illustrations of the original TLM3D data sets for all four test areas, as well as the results for the basic and the improved version of the centrality-based road selection algorithm. The reader is invited to have a look at these supplements, as they provide a better impression of the results when applied to large test areas.
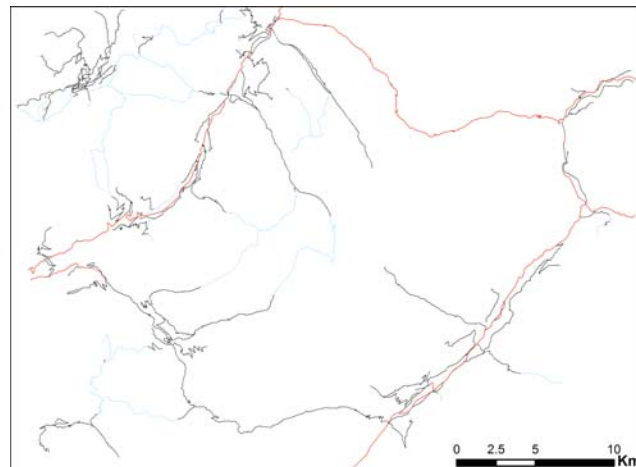
Figure 17: Result of the **basic** centrality algorithm in the Davos test area (a map of the source data is included in the Supplementary Material). (Data © swisstopo)
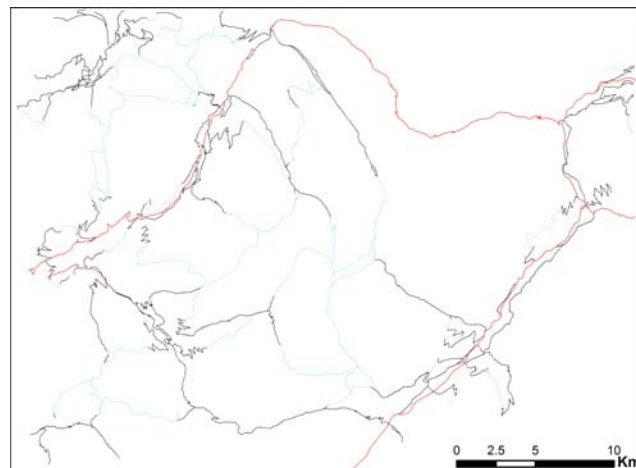

Figure 18: Result of the **improved** centrality algorithm in the Davos test area. (Data © swisstopo)

7 Discussion

The discussion of the results mainly focuses on the level of fulfillment of the constraints defined in Section 4.1. The discussion is divided into an analysis of the hard constraints, mainly based on the quantitative results of Table 2, and a qualitative analysis of the soft constraints, based on the evaluation by swisstopo experts.

7.1 Hard constraints

The first two hard constraints H1 and H2 deal with the highways and the related entries, exits and access roads. The most important features in road networks of small-scale maps are indeed the highways, as they offer the user important navigational cues. As can be seen in Table 2, the basic algorithm is unable to retain all highway segments for all of the test areas (with the exception of Davos, which simply does not contain any highways in the original dataset). Hence, highways do not necessarily exhibit a high betweenness centrality value. If they are counted as an important, non-deletable feature class, as is the case at swisstopo, it is necessary to treat them separately. While a fix to this problem can be implemented rather easily by simply forcing highway segments to be retained in the pruned map, this approach further aggravates the problem of unconnected highway access roads, which again make the highways less useful, as the user is unable to find routes to gain access to the main network. As is evident in Table 2, the test areas that contain highways all have many disconnected access points. In fact, the number of disconnected access points is vastly larger than the number of connected access points. This makes the highways that were retained even less useful. In contrast, the improved approach retains all highway segments and was able to connect all highway access points in a reasonable way, which in turn makes the whole highway network accessible and suitable for navigational purposes. Hence, while the two highway related constraints were violated by the basic algorithm quite heavily, they were fulfilled perfectly in the improved version, mainly because of the reconnection algorithm presented in Section 5.4. Only one highway access point in the test area of Zurich remained disconnected.

The hard constraints H3 and H4 deal with the general connectivity of the road network. While many authors have quickly recognized that disruptions in graph-based and especially stroke-based selection algorithms pose a major problem (e.g., [34]), the creation of dead-ends was largely ignored by the authors of previous work. However, when looking at Table 2 and Figures 5 and 17, it becomes evident that dead-ends pose much larger problems than disconnected network parts. They often disrupt long and important roads, hide the true connectivity of the road network, and make them useless for navigational purposes. Popular solutions to disconnected parts, such as MSTs, are unable to remedy this flaw in the pruned network. By improving the stroke generation process and reconnecting dead-ends and disconnected parts, it was possible to significantly improve the road networks in all test areas. The quantitative analysis shows that while the basic approach was unable to fulfill the connectivity constraints, the enhanced approach fulfilled them nearly perfectly, with the exception of one remaining dead-end (which equals to the unconnected highway access point mentioned in the discussion of the highway constraints).
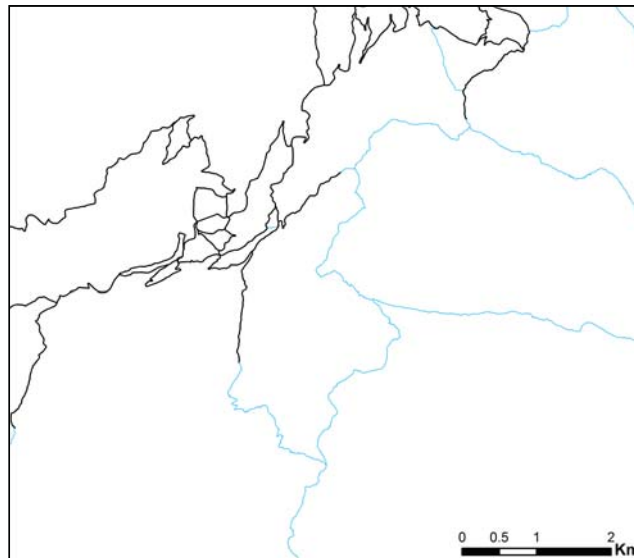
Figure 19: The road network in the town of Arosa in the Davos test area remains too dense. As this has a considerable impact on a large part of the test area (cf. Figure 18), the thinning in dense areas was not rated as good as in other test areas. (Data © swisstopo)

7.2   Soft constraints

The qualitative evaluation, which was conducted by swisstopo cartographers (Section 6), enables us to assess the fulfillment of the soft constraints laid out in Section 4.1. Table 4 showed an overview of the results of this evaluation.

Soft constraint S1 is about the general density of the resulting road network, which should be comparable to the current 1:200,000 map product of swisstopo, VECTOR200. To assess this constraint, three questions were used: the first one relates to the overall density of the result and the other two to the density in urban and rural areas, respectively. For most test areas, the general network thinning was rated as acceptable. For each test area, different reasons were stated by the experts for this rating, which also relate to different, non-optimal parts of the algorithm. In Zurich, for example, the density of several urban areas was rated as too low, which could mean that the parameterization of the density adaptation described in Section 5.3 needs further fine-tuning. While the overall density of the Davos test area was rated as good, one of the urban areas of this mountainous region was rated as too dense, which again points to a sub-optimal parameterization of the adaptive density algorithm (Figure 19). Because only a small part of the test area is affected, it had no influence on the general rating, however. In the Langenthal test dataset, on the other hand, the rural areas exhibited some unnecessary road network structures, which could potentially be alleviated by further enhancing the stroke generator and the betweenness calculation. In general, however, the experts were quite satisfied with the fulfillment of the overall density constraint S1.

Soft constraint S2, structure preservation, was rated as very good across all test areas. Important road structures were still recognizable as well as smaller paths and trails (Fig-
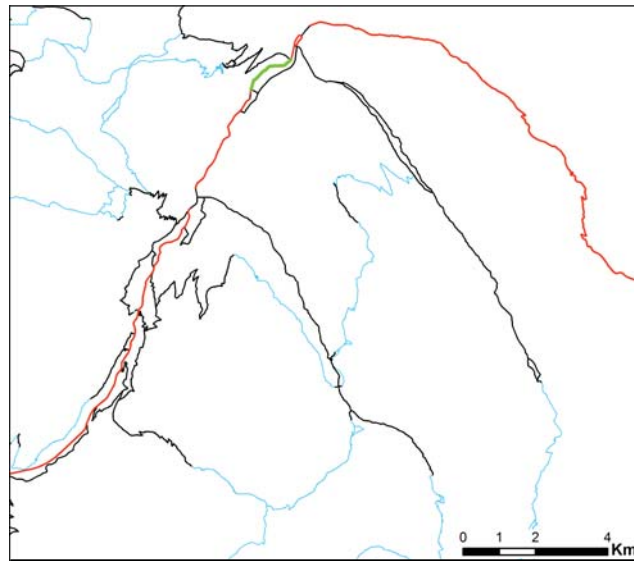
Figure 20: Extract of the Davos test area with missing part of an important link road marked in green. (Data © swisstopo)

ure 18). The non-optimal density of urban areas in Zurich did not impact the overall structure to the degree to have an influence on this rating.

Soft constraint S3 deals with the recognizability of urban areas. With the exception of some urban areas in Zurich, this constraint was fulfilled in all the test areas. However, the experts stated that this criterion is not especially important for small-scale topographic maps, as other feature classes (such as land use or buildings) are used to support the identification of urban areas.

The last soft constraint S4, the preservation of link roads, is of particularly high importance. Generally, the fulfillment of this criterion was rated as very good by the experts. They were surprised how well the algorithm worked, considering the fact that it did not use any additional information except the geometry and the road classes, as link roads are not always part of a major road class. Depending on the area, even small paths can be considered as link roads (e.g., in the mountains). Hence, the experts were especially impressed with the Davos and Lucerne test areas, which combine several major roads, but also small paths and trails, resulting in a rather complex and particularly diverse road network structure. The reason for the lower rating in the Davos dataset is due to the fact that one small segment of the link road connecting an important major road was missing (Figure 20). This also highlights the importance of this constraint and the thorough and critical assessment process by the experts. The same is also true for the Langenthal test area, which also received a lower rating as a result of a relatively short removed road that is considered as an important link road.

8   Conclusion

The automation of the selection of road networks has attracted considerable research interest over the past two decades, as many NMAs are hoping to be able to derive smaller-scale maps from a single, detailed database. Furthermore, the selection of the road network is of fundamental importance in map generalization, owing to the fact that the road network is among the dominant feature classes in topographic maps, and that the selection operator forms the basis of all subsequent generalization operations. Nevertheless, road network selection is a non-trivial process, as one has to deal with a coherent network of roads, each having a particular shape, angle, orientation and length [7]. Consequently, different approaches to this problem exist and no single algorithm has been able to establish itself at the top so far. Moreover, most algorithms have not been tested on large, diverse datasets and against real production requirements.

The work reported in this paper resulted from a collaboration with the Swiss national mapping agency, swisstopo, and has thus been grounded in production-driven requirements, which is different from previous research. Despite the collaboration with swisstopo, however, the aim was to define requirements and develop algorithms that were as generic as possible, and thus transferable to the requirements of other mapping agencies, with different products. The project started off by evaluating several methods from the literature on large datasets and against the requirements defined in Section 4.1. Among these existing algorithms the centrality-based algorithm by Jiang and Claramunt [15, 16] showed considerable promise, yet did not meet all of the requirements.

After careful analysis of the weaknesses of the basic algorithm by Jiang and Claramunt [15, 16], we developed and evaluated several improvements to the basic approach, four of which have been presented in this paper. It is important to note that some of the proposed enhancements, such as the enhanced semantics or the detection and collapse of roundabouts, can also be used in entirely different approaches, as they solve problems not strictly linked to centrality-based selection. Thanks to our extensions, the basic centrality approach was significantly improved to such a level that the swisstopo experts were largely satisfied with the results.

While we were able to improve the basic approach, not all problems were solved satisfactorily. In future work, the parameterization of the stroke generator as well as the adaptive threshold should be optimized. Additionally, research should focus on finding other ways to adapt the network density in urban and rural areas separately in order to achieve optimal results. One promising approach could be the combination of the stroke-based centrality approach with the mesh-based approach [8].

Acknowledgments

References

[1] BEARD, K. M. How to survive on a single detailed database. In *Proc. International Symposium on Computer-Assisted Cartography (Auto-Carto 8)* (1987), pp. 211–220.

[2] BEARD, K. M. Constraints in rule formation. In *Map Generalization: Making Rules for Knowledge Representation*, B. P. Buttenfield and R. B. McMaster, Eds. Longman Group, 1991, pp. 121–135.

[3] BENZ, S. A. Algorithmen zur automatischen Selektion von Strassennetzwerken: Lösungen für den Übergang vom Topografischen Landschaftsmodell (TLM) zu Massstab 1:50'000. Master's thesis, Department of Geography, University of Zurich, 2013.

[4] BENZ, S. A., AND WEIBEL, R. Road network selection for medium scales using an extended stroke-mesh combination algorithm. *Cartography and Geographic Information Science 41*, 4 (2014), 323–339. doi:10.1080/15230406.2014.928482.

[5] BRANDES, U. A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology 25*, 2 (2001), 163–177. doi:10.1080/0022250X.2001.9990249.

[6] BRANDES, U. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks 30*, 2 (2008), 136–145. doi:10.1016/j.socnet.2007.11.001.

[7] CHAUDHRY, O., AND MACKANESS, W. A. Rural and urban road network generalisation: Deriving 1:250,000 from OS MasterMap. In *Proc. 22th International Cartographic Conference* (La Coruña, Spain, 2005), pp. 1–10.

[8] CHEN, J., HU, Y., LI, Z., ZHAO, R., AND MENG, L. Selective omission of road features based on mesh density for automatic map generalization. *International Journal of Geographical Information Science 23*, 8 (2009), 1013–1032. doi:10.1080/13658810802070730.

[9] EHRLIHOLZER, R. Methoden für die Bewertung der Qualität von Generalisierungslösungen. Master's thesis, Department of Geography, University of Zurich, 1996.

[10] ESTER, M., KRIEGEL, H.-P., SANDER, J., AND XU, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. 2nd International Conference on Knowledge Discovery and Data mining* (1996), AAAI Press, pp. 226–231.

[11] FOERSTER, T., STOTER, J., AND KRAAK, M.-J. Challenges for automated generalisation at European mapping agencies: A qualitative and quantitative analysis. *The Cartographic Journal 47*, 1 (2010), 41–54.

[12] FREEMAN, L. C. A set of measures of centrality based on betweenness. *Sociometry 40*, 1 (1977), 35–41. doi:10.2307/3033543.

[13] HARRIE, L., AND WEIBEL, R. Modelling the overall process of generalisation. In *Generalisation of Geographic Information: Cartographic Modelling and Applications*, W. A. Mackaness, A. Ruas, and L. T. Sarjakoski, Eds. Elsevier Science, 2007, pp. 67–87.

[14] HU, Y., CHEN, J., LI, Z., AND ZHAO, R. Selection of streets based on mesh density for digital map generalization. In *Proc. 4th International Conference on Information Science* (Chengdu, China, 2007), pp. 903–908.

[15] JIANG, B., AND CLARAMUNT, C. A structural approach to the model generalization of an urban street network. *GeoInformatica 8*, 2 (2004), 157–171. doi:10.1023/B:GEIN.0000017746.44824.70.

[16] JIANG, B., AND CLARAMUNT, C. Topological analysis of urban street networks. *Environment and Planning B: Planning and Design 31*, 1 (2004), 151–162. doi:10.1068/b306.

[17] JIANG, B., AND HARRIE, L. Selection of streets from a network using self-organizing maps. *Transactions in GIS 8*, 3 (2004), 335–350. doi:10.1111/j.1467-9671.2004.00186.x.

[18] JIANG, B., ZHAO, S., AND YIN, J. Self-organized natural roads for predicting traffic flow: A sensitivity study. *Journal of Statistical Mechanics: Theory and Experiment*, 7 (2008), P07008. doi:10.1088/1742-5468/2008/07/P07008.

[19] LI, Z., AND ZHOU, Q. Integration of linear and areal hierarchies for continuous multi-scale representation of road networks. *International Journal of Geographical Information Science 26*, 5 (2012), 855–880.

[20] LIU, X., ZHAN, F. B., AND AI, T. Road selection based on Voronoi diagrams and "strokes" in map generalization. *International Journal of Applied Earth Observation and Geoinformation 12*, Supplement 2 (2010), S194–S202.

[21] MACKANESS, W. A., AND BEARD, K. M. Use of graph theory to support map generalization. *Cartography and Geographic Information Systems 20*, 4 (1993), 210–221.

[22] MACKANESS, W. A., AND MACKECHNIE, G. A. Automating the detection and simplification of junctions in road networks. *GeoInformatica 3*, 2 (1999), 185–200. doi:10.1023/A:1009807927991.

[23] MORISSET, B., AND RUAS, A. Simulation and agent modelling for road selection in generalisation. In *Proc. 18th International Cartographic Conference* (Stockholm, Sweden, 1997), pp. 1376–1380.

[24] PORTA, S., CRUCITTI, P., AND LATORA, V. The network analysis of urban streets: A dual approach. *Physica A: Statistical Mechanics and its Applications 369*, 2 (2006), 853–866. doi:10.1016/j.physa.2005.12.063.

[25] SAVINO, S., RUMOR, M., AND LISSANDRON, I. Road junction generalization in large scale geodatabases. In *Urban and Regional Data Management*, A. Krek, M. Rumor, S. Zlatanova, and E. M. Fendel, Eds. Taylor & Francis, 2009. doi:10.1201/9780203869352.ch19.

[26] STOTER, J., VAN SMAALEN, J., BAKKER, N., AND HARDY, P. Specifying map requirements for automated generalization of topographic data. *The Cartographic Journal 46*, 3 (2009), 214–227. doi:10.1179/174327709X446637.

[27] SWISS SOCIETY OF CARTOGRAPHY. *Topographic Maps—Map Graphics and Generalization*, vol. 17 of *Cartographic Publication Series*. Swiss Society of Cartography, 2005.

[28] THOMSON, R. C. The "stroke" concept in geographic network generalization and analysis. In *Progress in Spatial Data Handling*, A. Riedl, W. Kainz, and G. Elmes, Eds. Springer Berlin, Heidelberg, 2006, pp. 681–697. doi:10.1007/3-540-35589-8_43.

[29] THOMSON, R. C., AND BROOKS, R. A graph theory approach to road network generalisation. In *Proc. 17th International Cartographic Conference* (Barcelona, Spain, 1995), pp. 1871–1880.

[30] THOMSON, R. C., AND BROOKS, R. Efficient generalisation and abstraction of network data using perceptual grouping. In *Proc. 5th International Conference on GeoComputation* (2000), University of Greenwich, UK.

[31] THOMSON, R. C., AND RICHARDSON, D. E. The "good continuation" principle of perceptual organization applied to the generalization of road networks. In *Proc. 19th International Cartographic Conference* (Ottawa, Canada, 1999), pp. 1215–1223.

[32] TOUYA, G. A road network selection process based on data enrichment and structure detection. *Transactions in GIS 14*, 5 (2010), 595–614. doi:10.1111/j.1467-9671.2010.01215.x.

[33] WEISS, R. Road network selection for small-scale maps using an improved centrality approach. Master's thesis, Department of Geography, University of Zurich, 2013.

[34] YANG, B., LUAN, X., AND LI, Q. Generating hierarchical strokes from urban street networks based on spatial pattern recognition. *International Journal of Geographical Information Science 25*, 12 (2011), 2025–2050. doi:10.1080/13658816.2011.570270.

[35] ZHOU, Q., AND LI, Z. A comparative study of various strategies to concatenate road segments into strokes for map generalization. *International Journal of Geographical Information Science 26*, 4 (2012), 691–715. doi:10.1080/13658816.2011.609990.