Continuous Road Network Generalization

Radan Šuba[1], Martijn Meijers[1], Lina Huang[1,2] and Peter van Oosterom[1]

Abstract This paper presents our research efforts to obtain a sensible vario-scale representation of a road network. A vario-scale representation makes it possible to get maps at an arbitrary map scale, being more flexible with how much detail is shown during the use of geographic data. For this, an offline process creates a suitable vario-scale data structure (called tGAP), that then can be used for generating the maps at arbitrary map scale. Our intent is to design a generalisation process from (very) large scale data (e.g. 1:2,500,where roads are present as areas) to small scale data (where roads are present as lines, e.g. 1:1M). In this work we focus on the scale range where roads are present as lines (mid scale), but our intention is to start with large scale data where roads are represented as areas. At the heart of our approach we work with the area objects that are defined by the bounding road segments (the 'cycles' of the road network), but for determining the feature class of a merged edge, we fall back to an approach that is inspired by strokes. This together gives a solution for continuous road network generalization that aligns with our earlier attempts for creating vario-scale data structures.

[1]Section of GIS technology, OTB - Research for the Built Environment,
Delft University of Technology, The Netherlands
e-mail: R.Suba@tudelft.nl,
e-mail: B.M.Meijers@tudelft.nl,
e-mail: L.Huang-1@tudelft.nl,
e-mail: P.J.M.vanOosterom@tudelft.nl
[2]School of Resource and Environmental Science,
Wuhan University, Wuhan, 430079 China,

1 Introduction

Automated map generalization has been an important research area for many years now. It is a process of transformation from detailed maps to less detailed ones, while the essential topological, geometric and semantic characteristics of the geographic information presented are preserved (Chen et al, 2009). Focus of this paper is automated generalization of the road network, but instead of generating of a sequence road maps with different levels of detail (LOD) as has been common so far, we introduce incremental generalization through the spectrum of LODs. In this perspective the scale range for every map object needs to be introduced. As far as we know only a vario-scale data structure called tGAP (van Oosterom et al, 2014) can accommodate this functionality.

Our ultimate goal is to support input data, where roads are modelled as areas (large scale) and progressively are collapsed into lines (small scale), i.e. support a vario-scale representation for a large scale range. However, this paper covers only a specific part of this problem. We focus in this paper on roads already modelled as lines (mid-scale, 1:50K, to small scale). In more detail, we wish to achieve the road generalization where the network preserves its connectivity, all changes are as gradual as possible, small details are eliminated (not important dead-end, cul-de-sacs, are removed and no new dead-ends are created), road classification is taken into account and the overall map impression ("density of map") is preserved.

A number of earlier attempts have explored ways of generalizing road networks. At any given scale two main representations of the roads are found in these works: linear and area representations. Linear representations consider the network as a set of linear connected elements ('streets'), while the area representations consider the road network as a collection of areal 'building blocks', generated by minimal 'cycles' of streets around these blocks. Sometimes the term 'area partition' (Edwardes and Mackaness, 2000) or 'mesh density' (Chen et al, 2009; Li and Zhou, 2012) is used. We will use the term 'cycles' when we mean these blocks. With different representation the definition of the network itself has changed. In the first case, the network is defined from linear objects, which together form the network. In the second case, the focus is on the space between the linear objects and how these areal objects fill the space. Both representations have their own advantages and disadvantages (Edwardes and Mackaness, 2000).

Moreover, there are two main perspectives of the linear network representation; 1. strokes and 2. segments approach. The strokes approach groups road segments into longer lines based on some criteria (e.g. angle between segments) (Zhou and Li, 2012; Thomson and Richardson, 1999). Later the generalization process takes the stroke as smallest element, whereas the second approach considers road segments (from junction to junction) as the smallest atomic element for removal. The main advantage of using strokes is evident: It preserves information about connectivity between segments. This indicates that the stroke-based approach can be a useful generalization tool. However, Turner (2007) points out that segment analysis (the creation of a segment map as known in the space syntax community) can give comparable or better output compared to a stroke-based method and even produces more meaningful results (e.g. better in correlation with observed vehicular flow in reality). In addition, with using the segments approach the smallest element that can be removed from the map is exactly one road segment, this being useful from our continuous generalization perspective, where the geometric difference between two steps in the generalization process should be as minimal as possible, e.g. the omission of only one road segment makes a much smaller difference in the road network than the omission of a whole group of segments formed by one stroke.

Based on the above in our implementation the following approaches are used to get a good continuous generalization result: The road network is represented by cycles (area objects between the roads) and the whole process is driven by merging these cycles as this fits very well in our vario-scale approach (van Oosterom et al, 2014). Subsequently, the stroke-based approach is eventually used for determination of the feature class of the remaining segments that bound this cycle. The resulting segments and the information over their scale range is stored in a vario-scale data structure called tGAP (van Oosterom et al, 2014), which has been slightly modified for this purpose.

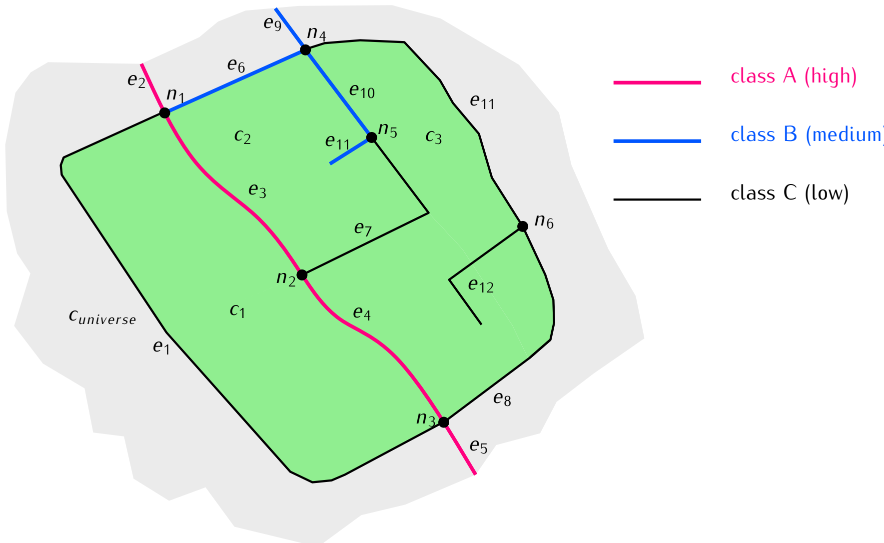
The remainder of this paper is structured as follows: Section 2 explains our method. Section 3 describes the used data structure in more details. Section 4 demonstrates and discusses obtained results. Finally, section 5 concludes the paper and present future work.

2 Method

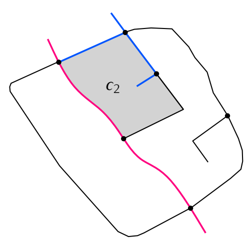
The initial step we perform is to convert all input geometry (lines) of the road network into a topological data structure consisting of nodes, edges, and faces. Nodes correspond to junctions (or ends of dead-ends). Edges correspond to road segments. Faces correspond to cycles of bounded areas; e.g. terrain, building, water, etc. Figure 1a shows an example of the resulting structure.

Then, we start the generalization process by viewing the road network as consisting of a set of enclosed areas. Figure 1 demonstrates the generalization actions for one step for a small dataset. A step can be seen as a merge of two cycles (which are faces in the topological data structure) together with some subsequent actions such as removing dead-ends (which are edges in the data structure).

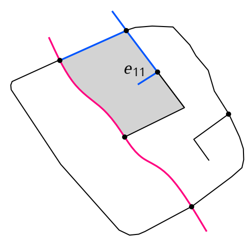
There are two aspects that require further explanation: Firstly, the selection of the least important road segment which defines the second cycle for the merge. Secondly, determining the classification of merged road segments.



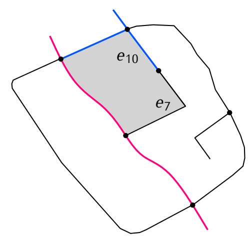
(a) An example of initial data.



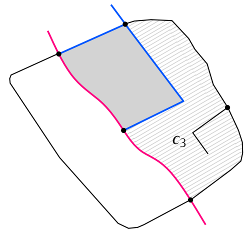
(b) Initial step: The least important cycle is selected (c_2).



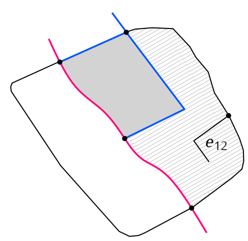
(c) Dead-end e_{11} is removed.



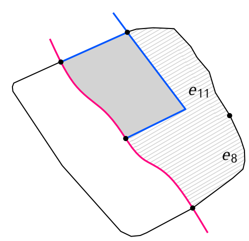
(d) Segments e_7 and e_{10} are merged. Reclassification of new segment takes place here.



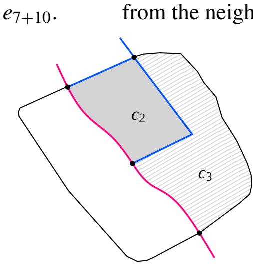
(e) The neighbouring cycle c_3 is selected based on the least important segment e_{7+10}.



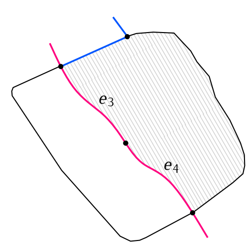
(f) Dead-end e_{12} is removed from the neighbouring cycle.



(g) Segments e_{11} and e_8 are merged.



(h) Cycle c_2 and c_3 are merged and segment between is removed.



(i) Segments e_3 and e_4 are merged.

Fig. 1: Example of one step in the process (i.e. two cycles are merged). Cycles are labelled with the prefix c and road segments with the prefix e.

2.1 The least important road segment

The least important cycle to be removed is defined by a set of bounding road segments. Then dead-ends inside this cycle are removed and the least important road segment must be chosen. This is done based on three criteria:

1. Classification of the segment (the segment having the 'lowest' classification);
2. How the segment is connected to other segments (*connectivity*);
3. Length of the road segment.

For every segment we look at its classification, see Figure 1a, where *class A* (high importance) > *class B* (medium) > *class C* (low importance). If this gives a winner, we remove this segment. In case, this results in a tie (segments having the same classification), we compute for every segment a *connectivity* value, and remove the segment that has the lowest *connectivity* value. As last resort, in case of segments also having the same *connectivity* value, the shortest of the segments is selected as least important segment.

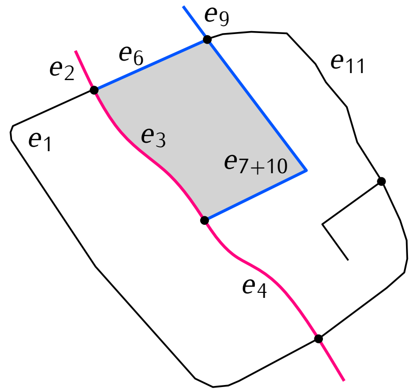
The *connectivity* value for one road segment is defined by how many routes between other segments go via this segment. More passing routes within segments means higher *connectivity*. This measure is known from space-syntax theory (Turner, 2007), which is used in analysis for urban planning and implemented in software, such as DepthMapX. We use a modified version of the connectivity value, called *radius connectivity*. Here a geometric criterion ε is used, where for the chosen segment only the other segments within a radius ε of the segment are considered. This eliminates the fact that segments on the edge of dataset have by definition a lower *connectivity* value than segments in the middle of the dataset.

In our case, we compute the *connectivity* locally. But instead of using a geometric radius ε, we use a topological measure, only considering $i-neighbours$. It is similar to a breadth-first search, e.g. $1-neighbours$ checks only the direct neighbours of the chosen segment, $2-neighbours$ checks the neighbours of the direct neighbours, $3-neighbours$ checks also the neighbours of the $2-neighbours$, etc.

Figure 2 illustrates the selection of the least important road segment in detail (it corresponds to the action in Figure 1e). Cycle c_2 is selected as the least important cycle. The possible candidates for least important road segments are e_3, e_6 and e_{7+10} Segment e_3 is eliminated first, because it has a higher classification than the others. Then the *connectivity* is calculated for the remaining segments. In this example, we only use the $1-neighbours$. Finally, segment e_{7+10} is selected to be removed, because it connects one segment less compared to e_6.

2.2 Merge of the road segments

When the dead-ends have been removed or cycles have been merged, a merge of two (or more) adjacent road segments (edges) takes place. For the merged road segments



Segment	$1 - neighbours[class]$
$e_3[A]$	$e_2[A], e_4[A], e_1[C]$
$e_6[B]$	$e_2[A], e_9[B], e_1[C], e_{11}[C]$
$e_{7+10}[B]$	$e_4[A], e_9[B], e_{11}[C]$

Fig. 2: Finding the least important segment for cycle c_2. The segment e_{7+10} is selected, connects other three 'A','B' and 'C' segments (e_6 connects 'A','B' and two 'C' segments, e_3 connects two 'A' and 'C' segments). Note that the segments of the cycles itself are not considered to be adjacent segments.

we determine the classification. We determine the new classification in the following order:

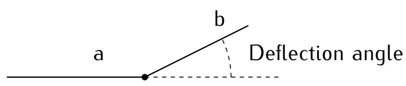
1. Consider the classification of the road segments;
2. Based on adjacent road segments with the best deflection angle;
3. Length of road segments.

First, the classification of two road segments which should be merged is checked. In case, the merged road segments all have the same classification, this is an easy task: The new merged road segment will carry the same class. However, Figure 1d and, in more detail, Figure 3 present that it can happen that two road segments having different classifications will merge and that the class for this new segment must be determined ('B' or 'C'). In our initial implementation, the class of the longest segment had been chosen. This resulted in a non-realistic process, where longer and longer segments dominate. Therefore, the direct neighbouring segments are involved in deciding the resulting classification(thus not only the segments to be merged).We find the best continuing neighbouring road segment and using the classification of this neighbour, we can determine the final classification of the merged segment.

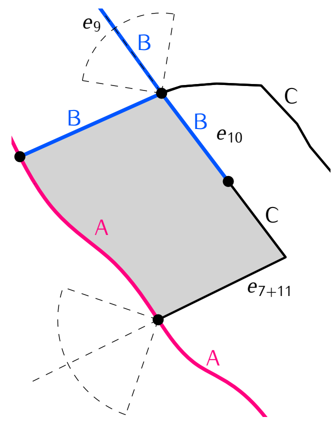
To find the best continuing neighbouring road segment, we use a simple geometric principle, that is called the deflection angle. Figure 3a illustrates that this is the deviation from 180° of the angle between two road segments. This is inspired by the strokes approach (note that many strategies for stroke building exist, cf. Zhou and Li (2012)). By using the deflection angle, we form locally a 'stroke' where a neighbouring segment appears to follow the same direction as the merged segment if the deflection angle is small. However, if the deflection angle is large (e.g. more 90°) the road segments will not be going in the same direction. We chose a 60°

threshold, if an adjacent, neighbouring segment is having a deflection angle larger than this, it will not be considered for determining the class.

In case the determination of the class based on deflection angle leads to a tie, the classification of the newly merged road segment will be based on the original classification of the *longest* segment in the group of road segments to be merged.



(a) The principle of deflection angle. A smaller angle means a better connectivity.



(b) An example of merging two road segment. The deflection angle is indicated with the dashed lines.

Fig. 3: Determining the classification of a merged segment, in case the original segments are classified differently

Figure 3 demonstrates the principles of determining the new classification based on the deflection angle. The road segments e_{10} and e_{7+11} should be merged. Both have different classification, e_{10} is 'B' and e_{7+11} is 'C'. The final classification should be between those two. Segment e_9 with 'B' is selected as the best continuing neighbouring road segment because it has the best deflection angle of all adjacent segments. Using the classification of e_9 we can determine the class of the newly merged segment. Note that the classification e_9 only indicates what the new classification should be and it thus not used directly.

One of the three possibilities can happen; class of $e_9 \geq$ maximum of classes e_{10}, e_{7+11}, class of $e_9 \leq$ minimum of classes e_{10}, e_{7+11}, or class of e_9 is between classes e_{10}, e_{7+11}. In the first case, the final class will be the maximum of the classes e_{10}, e_{7+11}. In the second case, the final class will be the minimum of classes e_{10}, e_{7+11} and in the third case, the class will be based on the *longest* segment from e_{10}, e_{7+11}. In example 3b, the new merged segment will carry 'B', because $e_9[B] \geq e_{10}[B]$.

Normally, the strokes approach is used at once for all road segments present in the dataset, but for our purpose only local information is needed, e.g. to find the road segment with best continuity to determine the new classification. Additionally, our road network is changing dynamically during the process and it would be useless to keep global information about strokes up-to-date. Therefore we only view a merged set of road segments as a stroke to determine its new classification, when there is no 'clearer' alternative to do this.

3 Data structure

The generalization process removes and modifies the road segments. The data structures that we use (tGAP), registers for every segment the scale range (importance low - importance high), which defines the visibility of the object. The reader is referred to (van Oosterom et al, 2014) for more details on the data structure (van Oosterom et al, 2014).

The geometry of the faces, in this case the 'cycles', are formed by the set of edges, that forms the boundary of the cycles. Geometry is stored only for edges (as polylines in *geometry* attribute), where each edge has a scale range given (by its importance low and importance high attribute value). The cycles are stored same way as in previous tGAP structure. For road network generalization the tGAP structure was extended for only one attribute *edge_class* in the *tgap_edge* table. This attribute defines classification of road segments. No other modification of the structure were needed.

```
CREATE TABLE tgap_edge(
  edge_id integer,
  start_node_id integer,
  end_node_id integer,
  left_face_id integer,
  right_face_id integer,
  imp_low bigint,
  imp_high bigint,
  edge_class integer,
  geometry geometry);
```

```
CREATE TABLE tgap_face(
  face_id integer,
  imp_low bigint,
  imp_high bigint,
  imp_own bigint,
  area numeric,
  bbox geometry);

CREATE TABLE tgap_node(
  node_id integer,
  imp_low bigint,
  imp_high bigint,
  geometry geometry);
```

4 Results

A road network dataset of the Netherlands was used as the test data (in Dutch: Nationaal Wegen Bestand). This is a dataset intended for use at a 1:25,000 map scale. A subset of the dataset was picked to experiment and demonstrate the principles. Figure 4 (p. 12) illustrates sequence of the generalization process.

We can observe, that even with a rather limited set of generalization operations (omitting dead-ends, merging cycles, removing road segments between merged cycles and re-classifying remaining road segments), we obtained relatively good results in terms of continuity of the road network.

The removal of multiple segments at once (e.g. all dead-ends) would give an artificial (strange) impression during zooming. Hence, we try to make changes as gradual as possible, see the dead-ends in Figure 4. However, when treating a cycle the removal of multiple segments can happen. Therefore, we have to perform the operation in a balanced manner based on various criteria: road classification, length, topology, etc. (e.g. T-type dead-ends, where only the last two segments are removed).

The least important cycle is selected just based on its size. This has been done to better demonstrate the effect of our road network generalization technique itself. On the contrary, we can see that this does not preserve the relative sizes of the cycles; small cycles of urban areas are removed as a first. On that account, we are considering to apply two additional techniques. First, computing densities in advance, and trying to keep these densities during the generalization process. Second, using a divide and conquer approach (based on our Field-tree implementation of tGAP creation as described in (Meijers, 2011a)

5 Discussion and future work

This paper has presented our approach to realize vario-scale road networks. For that target we have used the tGAP structure where we have applied our developed techniques. This approach can be improved when we involve some more operators, e.g. with a collapse operation it will be possible to eliminate roundabouts. The tGAP structure can support storage for the result of this types of split/collapse operations (Meijers and van Oosterom, 2013). However, it has not yet been implemented for the road network generalization procedure. Some other open points / future work are:

- In Figure 4 some of the segments are very detailed in the latter stages of the process. This can be solved by applying line simplification. Earlier we have experimented with simplifying those edges in the tGAP construction process that are merged (Meijers, 2011b).
- We like to formally validate the result. For such a validation, Li and Zhou (2012) compare the generalized results with a benchmark map that represents the same region at the same target map scale and every generalised segment has a corresponding segment in this benchmark. This cannot be used for our purposes, because our approach does not have such a clear correlation with other maps of a predefined and fixed map scale. Our initial idea for validation is based on the size and density of cycles. This can be done as a simple overlay of our generalized map and a reference map of a road network. When the size/density of the cycles on average corresponds to the benchmark map we can assume that our map scale correspond to the reference map and then it can be checked if the connectivity

of the roads on the two maps is similar. Other validation options can be based on visual checks of (expert) users.
- Missing classification or using data from more sources: The input dataset we used has roads with a good classification hierarchy. However, we also would like to have an approach when such a good classification is absent. The missing crucial information can be obtained by geometric analysis, e.g. road segments that are long and straight can be identified as being part of highways. Additionally, this extra information can be combined with an already existing classification to achieve even better results. We assume that some form of fuzzy logic can be used, e.g. Biljecki et al (2013).

6 Acknowledgment

This research is supported by the Dutch Technology Foundation STW (project number 11185), which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs.

References

Biljecki F, Ledoux H, van Oosterom P (2013) Transportation mode-based segmentation and classification of movement trajectories. International Journal of Geographical Information Science 27(2):385–407

Chen J, Hu Y, Li Z, Zhao R, Meng L (2009) Selective omission of road features based on mesh density for automatic map generalization. International Journal of Geographical Information Science 23(8):1013–1032

Edwardes A, Mackaness W (2000) Intelligent road network simplification in urban areas. GIS Research UK 2000 Conference (GISRUK 2000), University of York

Li Z, Zhou Q (2012) Integration of linear and areal hierarchies for continuous multi-scale representation of road networks. International Journal of Geographical Information Science 26(5):855–880

Meijers M (2011a) Cache-friendly progressive data streaming with variable-scale data structures. In: Proceedings of 14th ICA/ISPRS Workshop on Generalisation and Multiple Representation, pp 1–19

Meijers M (2011b) Simultaneous & topologically-safe line simplification for a variable-scale planar partition. In: Geertman S, Reinhardt W, Toppen F (eds) Advancing Geoinformation Science for a Changing World, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg, pp 337–358

Meijers M, van Oosterom P (2013) SplitArea: een algoritme om vlakken te splitsen voor de tGAP data structuren. In: Geo-Info, vol 10, Geo-Informatie Nederland, pp 14–18 van Oosterom P, Meijers M, Stoter J, Šuba R (2014) Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation, Lecture Notes in Geoinformation and Cartography, vol 2014, Springer, chap Data Structures for Continuous Generalisation: tGAP and SSC, pp 83–118

Thomson RC, Richardson DE (1999) The 'good continuation' principle of perceptual organization applied to the generalization of road networks. In: Proceedings of the 19th International Cartographic Conference, Ottwawa, pp 1215–1225

Turner A (2007) From axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis. Environment and Planning B: Planning and Design 34(3):539–555

Zhou Q, Li Z (2012) A comparative study of various strategies to concatenate road segments into strokes for map generalization. International Journal of Geographical Information Science 26(4):691–715

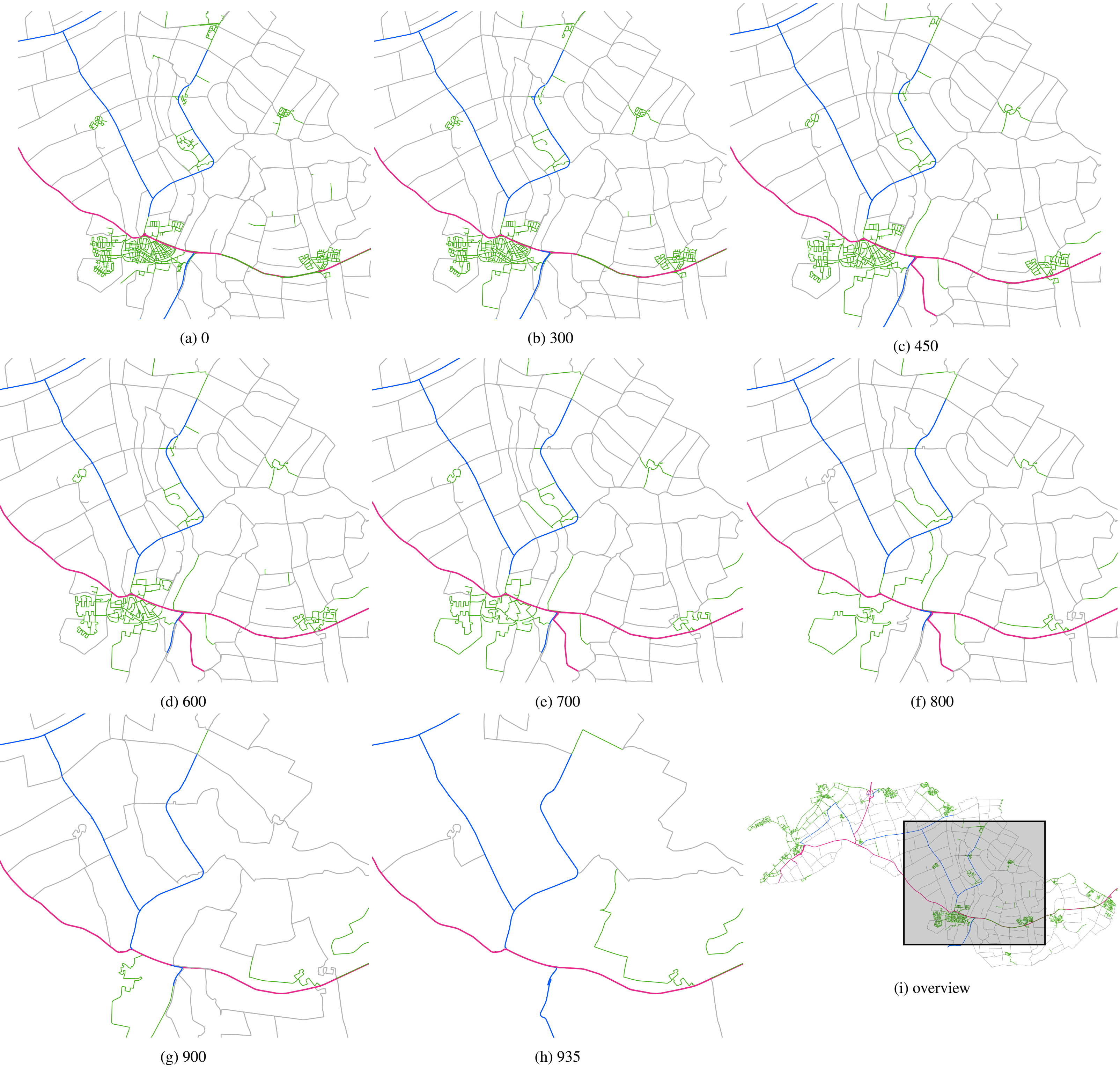


Fig. 4: An example of small testing area. 4i shows the zoomed in area present in all other figures. The figures cover the process from no 4a to maximal generalized 4h stage. The red, blue, green and grey colours present from more to less important roads.