

Planar and non-planar topologically consistent vector map simplification

Padraig Corcoran*, Peter Mooney and Adam Winstanley

Department of Computer Science, National University of Ireland Maynooth, Maynooth, Ireland

(Received 10 March 2010; final version received 25 November 2010)

This article contains a mathematical analysis of strategies for determining topological consistency of vector map simplifications. Such techniques exploit assumptions that can be made regarding the similarity of corresponding objects in successive simplifications. We propose that all topological relationships may be classified as planar or non-planar. A formal analysis of techniques for determining topological consistency of a simplification in terms of such relationships is presented. For each technique we analyse any corresponding constraints that are imposed. This provides a unified understanding of the benefits and limitations of individual techniques and the relationships that exist between techniques. Subsequently, a new strategy for determining the topological consistency of a simplification is proposed. This technique integrates the benefits all methods studied to provide a solution which is subject to less constraints. The effectiveness of this approach is demonstrated through fusion with an existing simplification technique resulting in simplifications that have equal topology and similar shaped features to the original map.

Keywords: generalization; simplification; topology

1. Introduction

Geospatial data contain information about the geographic locations of features and their boundaries on the surface of the earth. When represented in map form an important attribute of such data is that its content and geometric detail may be adapted using a process known as generalization (Jones and Ware 2005). This is a context-dependent process where the scale of generalization required can vary greatly. Traditionally, generalization was performed manually by cartographers but the growth of Geographical Information Systems (GIS) and an increase in the volume of spatial data have introduced a need for automation. The Internet is fast becoming the most popular medium for obtaining and analysing up-to-date map data. Such tools are commonly referred to as Web GIS and examples include Google Maps and OpenStreetMap (OSM). This growth is being driven partly by a corresponding growth in location-based services that often involve the downloading of map data usually to mobile devices. The data stored on GIS web servers generally represent the finest level of detail available. For Web GIS and location-based service applications, it is often desirable to reduce the detail of this representation using generalization. There are two main reasons for this. First, the spatial extent and detail of maps is ever increasing

*Corresponding author. Email: padraigc@cs.nuim.ie

and the bandwidth required to transmit them can be significant. Any client that attempts to obtain such data over a network will have finite bandwidth and this may prevent the map being transmitted in its original form (Yang and Weibel 2009). Second, many mobile devices used to display map data have limited screen resolution, processing power and storage capabilities. They are more applicable to maps that have been previously generalized and contain a less amount of detail (Li 2009).

(Jones 1997) identified eight categories of generalization algorithms. These are elimination, simplification, typification, exaggeration, enhancement, collapse, amalgamation and displacement. Simplification methods, which are the focus of this work, attempt to generalize polygon and line features by reducing the number of points or vertices used to represent them. The intention of any generalisation process is to produce the best result possible subject to a set of objectives (Jones and Ware 2005). Weibel (1996) identified four classes of objectives that such a process may aim to satisfy. These are shape (Gestalt), semantic, metric and topological objectives. To satisfy shape objectives, successive levels of generalization should represent an intuitive shape evolution of features (Latecki and Lakmper 1999). Semantic objectives integrate information about a feature's semantics when deriving generalization. For example, a line feature may be generalized differently if it represents a road as opposed to a river. Metric objectives aim to achieve the best possible result in terms of some error criterion. For example, this could be the result that minimizes the overall deviation from the original map. Finally, topological objectives are primarily concerned with the need to ensure that the simplified representations of features retain original relationships of containment and connectivity (Jones and Ware 2005). Two maps with equal topology are said to be topologically equivalent (Kuijpers *et al.* 1995, Cicerone *et al.* 2002).

Determining whether two maps are topologically equivalent has been the focus of much research in the fields of GIS and computational topology (Egenhofer and Franzosa 1994, Clementini and Di Felice 1998, Li and Liu 2010). In these domains all techniques attempt to determine the following: Given two topological spaces X and Y and two pairs of objects A_x, B_x and A_y, B_y in X and Y respectively, determine whether the topological relation between A_x and B_x is equivalent to the topological relation between A_y and B_y (Egenhofer and Franzosa 1994). In this context, no assumptions apart from identity can be made regarding the relationships between objects in X and Y . This is illustrated in Figure 1a where no assumptions regarding the relationships between A_x and A_y or between B_x and B_y can be made. In the context of determined topological equivalence between successive map simplifications such assumptions can be made. This is illustrated in Figure 1b where Y is the simplification of X obtained by removing the vertex v_i of A_x . Because A_y is the simplification of A_x , it is a sub-chain of A_x . Also because B_y is not a simplified version of B_x both objects are equal. Exploiting the existence of such relationships between objects in X and Y has led to the development of a distinct field of research. This exploitation results in techniques that are easier to implement and have significantly less computational complexity.

We propose that all topological relationships may be classified as planar or non-planar relationships. Consider the simple map in Figure 2a that contains a polygon, line and point feature. No lines or edges in this map cross without forming a vertex; therefore, we say that the topological relationships between all features are planar. Next consider the simple map in Figure 2b that contains a single polygon and line feature. Because the line crosses the polygon without forming a vertex, we say the topological relationship between these features is non-planar.

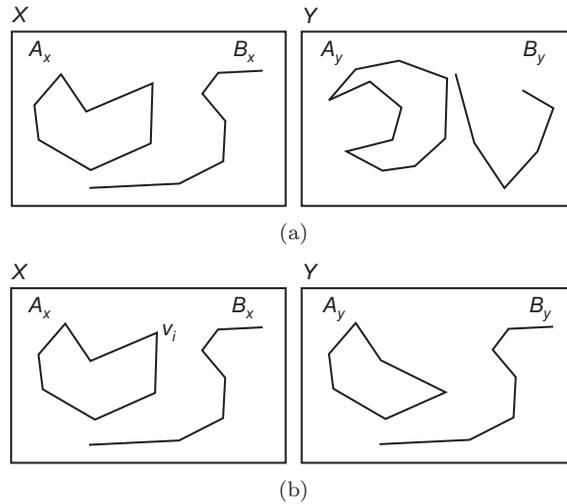


Figure 1. Two pairs of topological spaces are shown in (a) and (b). In (a) no assumptions can be made regarding the relationships between corresponding objects in X and Y . In (b) B_x and B_y are equal although A_y is a sub-chain of A_x .

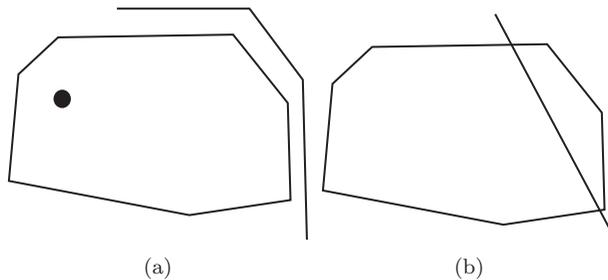


Figure 2. Planar and non-planar topology are shown in (a) and (b), respectively.

In this article an in-depth analysis of methods for determining the topological consistency of a map simplification is presented. We propose that any method for determining the topological consistency can be summarized in terms of the following constraints:

- (1) Constraints on the types of topology for which the technique can determine consistency without returning a false positive; that is, classifying a simplification as topologically correct when in fact it is not.
- (2) Constraints on the types of topology for which the technique can determine consistency without returning a false negative; that is, classifying a simplification as topologically incorrect when it is in fact correct.
- (3) Constraints on the types of simplification to which the technique can be applied.

An in-depth analysis of methods for determining the planar topological consistency of a simplification in terms of these constraints is presented. In particular, we focus on three techniques representing the current state of the art. Despite being equally important

as planar relationships, techniques for determining the non-planar topological consistency of a simplification have received much less research attention. From extensive literature searches the authors have identified only two existing techniques for determining such topological consistency and both are analysed in detail. Using the knowledge gained through this analysis, a new topological consistent simplification strategy is proposed. This technique integrates the benefits all methods studied to provide a simplification which is subject to less relative constraints. The effectiveness of this technique is demonstrated in the context of a simplification technique that satisfies shape objectives.

The layout of this article is as follows: Sections 2 and 3 analyse the existing techniques for determining planar and non-planar topological consistencies of a simplification, respectively. Section 4 describes the algorithm used to perform the integration of a simplification method that satisfies shape objectives with the proposed methodology for determining the topological consistency. Finally, in Sections 5 and 6 we present the results and draw conclusions, respectively.

2. Planar topological consistency

In this section we analyse three seminal methods for determining the planar topological consistency in terms of which relationships they can determine consistency and the constraints they impose. These are the works of De Berg *et al.* (1998), Saalfeld (1999) and da Silva and Wu (2006). Through this analysis we develop an improved strategy for determining the planar topological consistency. At this point it is necessary to introduce some background concepts. The *Jordan Curve Theorem* states that any simple closed curve or polygon C divides the points of the plane not on C into two distinct domains of inside and outside of which C is the common boundary. Let $I(p, C)$ be a function that returns an integer number representing the number of times a ray from p in any fixed direction intersects the chain C . The modules of this function can be taken to give the ‘parity’ of the point in question relative to the contour in question. The *point-in-polygon* criterion allows us to determine whether a point is inside or outside a simple polygon C as follows: The point p is outside C if $I(p, C)$ is an even number: that is, $I(p, C) \% 2 = 0$. The point p is inside C if $I(p, C)$ is an odd number: that is, $I(p, C) \% 2 = 1$ (O’Rourke 1998). Another important concept is that of a monotone chain. A chain C is monotone with respect to a line L if every line orthogonal to L intersects C in at most one point (O’Rourke 1998). A chain that is monotone with respect to the x -axis is x -monotone. The degree of a chain vertex is the number of edges incident to it. A vertex of degree 1 is a leaf; a vertex of degree 2 is an interior vertex; and a vertex of degree greater than or equal to 3 is a junction (De Berg *et al.* 1998). Generally the number of leaves and junctions is small compared to the number of interior vertices in a given map. All simplification techniques presented in this article are only applicable to interior vertices. All leaf and junction vertices are preserved in the corresponding simplifications. This strategy has been previously used by Estkowski and Mitchell (2001) and De Berg *et al.* (1998). A technique that exploits vertex degree to determine a suitable simplification was presented by Yang (2005).

2.1. The De Berg strategy

De Berg *et al.* (1998) proposed a simplification technique based on finding the shortest consistent path in a graph where line or polygon vertices represent graph nodes whereas the line segments joining these vertices represent graph edges. Let C be an x -monotone chain with vertices v_1, \dots, v_n and C' a corresponding simplification. Let

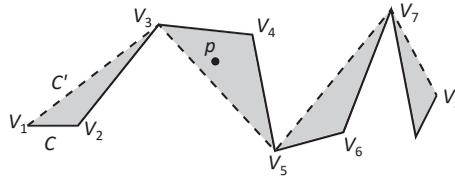


Figure 3. The chains C and C' are represented by solid and dashed lines, respectively. The point p lies inside a bounded face formed by C and C' .

P be a set of points p_1, \dots, p_m . Figure 3 shows two such chains and a single point p . If $(I(p, C) + I(p, C')) \% 2 = 1$, we say p lies inside a bounded face formed by C and C' . In the context of Figure 3, any point that lies in a grey region will be determined as lying in a bounded face. If $(I(p, C) + I(p, C')) \% 2 = 0$, we say p lies outside a bounded face formed by C and C' . The concept of applying the function I to an open contour will be introduced in Theorem 2.1. This theorem is the original theorem of De Berg *et al.* (1998) but we supply a new proof. The motivation for this new proof will be presented in Section 2.2.

Theorem 2.1: C' is a consistent simplification of C with respect to P if and only if no point of P lies in a bounded face formed by C and C' .

Proof: Proving the if and only if logical connection of this theorem requires two steps. First, we must prove that if a simplification is consistent it is classified as so and no false negatives are possible. Second, we must prove that if a simplification is inconsistent it is classified as so and no false positives are possible. Let p be a point in the set P . There exists a chain B that completes both C and C' to simple polygons denoted BC and BC' , respectively, such that the p lies inside BC . This is illustrated by the point p_1 in Figure 4. If the point p lies inside BC' , this represents a consistent simplification. On the contrary, if the point p lies outside BC' , this represents an inconsistent simplification. This specifies the two cases that may occur. Using two lemmas we now prove the following: First, for the inconsistent case the point p lies in a bounded face formed by C and C' . Second, for the consistent case the point p does not lie in such a bounded face. \square

Lemma 2.1: If a point p lies inside BC and outside BC' , it lies in a bounded face formed by C and C' .

Proof: By the point-in-polygon criterion, p lies inside BC so $I(p, BC) \% 2 = 1$. On the contrary, p lies outside BC' so $I(p, BC') \% 2 = 0$. Therefore, $(I(p, BC) + I(p, BC')) \% 2 = 1$. The two chains B in the polygons BC and BC' overlap. Therefore, the ray from p will intersect both or neither of these chains resulting in an increment of zero or two crossings, respectively; in turn not changing the corresponding parity. Therefore,

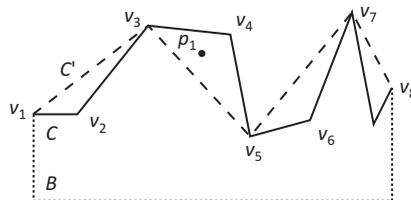


Figure 4. The chains C , C' and B are represented by the solid, dashed and dotted lines, respectively.

$(I(p, BC) + I(p, BC')) \% 2 = (I(p, C) + I(p, C')) \% 2$; that is, we need only count the number of intersections with C and C' and all others may be ignored. This implies $(I(p, C) + I(p, C')) \% 2 = 1$ and consequently that p lies in a bounded face formed by C and C' . \square

Lemma 2.2: *If a point feature p lies inside BC and inside BC' , it does not lie in a bounded face formed by C and C' .*

Proof: By the point-in-polygon criterion, p lies inside BC giving $I(p, BC) \% 2 = 1$. On the contrary, p lies inside BC' giving $I(p, BC') \% 2 = 1$. Therefore, $(I(p, BC) + I(p, BC')) \% 2 = 0$. From the same reason as in Lemma 2.1, $(I(p, BC) + I(p, BC')) \% 2 = (I(p, C) + I(p, C')) \% 2$. This implies $(I(p, C) + I(p, C')) \% 2 = 0$ and consequently that p does not lie in a bounded face formed by C and C' . \square

Lemmas 2.1 and 2.2 are true for all points p proving Theorem 2.1. The if and only if logical connection of Theorem 2.1 ensures that it exhibits no constraint in terms of returning false-positive or false-negative classifications. Therefore, this method is optimal in terms of these two forms of constraints. \square

De Berg *et al.* (1998) proposed that Theorem 2.1 may be used to determine whether a simplification is consistent with respect to maintaining the non-self-intersection of polygons. We paraphrase directly to allow understanding in the current context. ‘*There is a simple remedy, add all vertices as extra points to the set P. Now we can show that any line segment in C' that is consistent with C with respect to the extra points cannot destroy the condition that output chain be simple.*’ What De Berg *et al.* (1998) states here is that maintaining the consistency of polygon line segments not intersecting each other can be achieved by representing the line segments in question by their corresponding endpoints and maintaining the consistency of these using Theorem 2.1. However, De Berg *et al.* (1998) did not provide any proof of this fact. Line segments and points are two completely different types of vector data structure. To this end, we now prove the De Berg *et al.* (1998) proposal true if and only if the chain C is monotone. Let $\overline{l_i l_j}$ be a line segment that does not intersect a monotone chain C .

Theorem 2.2: *C' is a consistent simplification of C with respect to $\overline{l_i l_j}$ if and only if neither endpoint of $\overline{l_i l_j}$ lies in a bounded face formed by C and C' .*

Proof: There exists a chain B that completes both C and C' to simple polygons, denoted BC and BC' , respectively, such that BC contains l_i and l_j and $\overline{l_i l_j}$ does not intersect BC . Such a chain will always exist because, by definition, $\overline{l_i l_j}$ does not intersect C . Consider the x -monotone chain C and its corresponding simplification C' in Figure 5. C' is a sub-chain of C and it will therefore also be x -monotone. A bounded face will exist between each segment in C' and its corresponding chain in C . Due to both chains being x -monotone, each of these regions will be bounded by a single segment of C' . This is illustrated in Figure 5. The remainder of this proof is divided into two possible cases that may occur.

Case 1: Consider the case where the line segment $\overline{l_i l_j}$ intersects C' . This is the case when $\overline{l_i l_j}$ corresponds to $\overline{l_1 l_2}$ in Figure 5, which intersects C' a single time along the segment $\overline{v_3 v_5}$. This single intersection causes the endpoint l_1 to lie in a bounded region between C and C' . In order for l_1 not to lie in this region, $\overline{l_1 l_2}$ must intersect a segment on boundary of the region a second time. This region is bounded by a single line segment of C' that $\overline{l_1 l_2}$

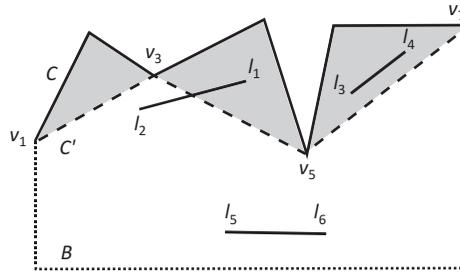


Figure 5. Three bounded regions between C and C' exist where each is bounded by a single segment of C' . Three possible line segments are also shown.

has previously intersected and cannot intersect a second time. The remainder of the region boundary consists of segments from C . By definition, $\overline{l_1 l_2}$ does not intersect C . Therefore, if $\overline{l_1 l_2}$ intersects $\overline{v_3 v_5}$ of C' one endpoint, in this case l_1 , must lie in the bounded face formed by C and C' . By Theorem 2.1, this endpoint will be classified as inconsistent. This proof is well defined for any line segment $\overline{l_i l_j}$ that intersects any line segment of C' .

Case 2: Consider the case where the line segment $\overline{l_i l_j}$ does not intersect C' . Both endpoints of $\overline{l_i l_j}$ will lie inside or outside BC' . If $\overline{l_i l_j}$ corresponds to $\overline{l_3 l_4}$ in Figure 5, both endpoints lie outside BC' . This corresponds to an inconsistent simplification because both endpoints are inside BC . In this case both endpoints will lie in a bounded face formed by C and C' and be classified as inconsistent by Theorem 2.1. If $\overline{l_i l_j}$ corresponds to $\overline{l_5 l_6}$ in Figure 5 both endpoints lie inside BC' . This corresponds to a consistent simplification. In this case no endpoint will lie in a bounded face formed by C and C' and this simplification will be classified as consistent by Theorem 2.1. \square

Similar to Theorem 2.1, Theorem 2.2 exhibits no constraint in terms of returning false-positive or false-negative classifications. This method is therefore optimal in terms of these two forms of constraints. Theorems 2.1 and 2.4 prove that the strategy proposed by De Berg *et al.* (1998) is optimal when determining the planar topological consistency for monotone chains. De Berg *et al.* (1998) showed that this strategy may be generalized to arbitrary chains by applying it to sub-chains which do not cycle or contain backward tangents. Intuitively, no backward tangent means that if v_i is the first vertex of a chain, there should not be an edge $\overline{v_j v_{j+1}}$ that is closer to v_i than the preceding edge $\overline{v_{j-1} v_j}$ (De Berg *et al.* 1998). The major drawback of this approach for determining the topological consistency is that it may only be applied to a subset of possible simplifications.

2.2. The Saalfeld strategy

In this section we analyse a second method for determining the planar topological consistency. Saalfeld (1999) proposed to apply the strategy of De Berg *et al.* (1998) for determining the consistency of monotone chains to arbitrary chains. In Section 2.1, using a new proof, we showed this approach to be well defined for monotone chains. In Saalfeld (1999), this was never proven for arbitrary chains. In this section we will prove that when determining the consistency of point features, this approach actually is well defined for arbitrary chains. Section 2.3 will discuss the work of Saalfeld (1999) in the context of determining the consistency of non-intersecting lines and non-self-intersecting polygons.

Let C be a simple arbitrary chain, as opposed to a monotone chain as in Theorem 2.1, with vertices v_1, \dots, v_n and C' its corresponding simple simplification. Let P be a set of points p_1, \dots, p_m .

Theorem 2.3: C' is a consistent simplification of C with respect to P if and only if no point of P lies in a bounded face formed by C and C' .

Proof: In Theorem 2.1 we proved the current theorem to be true if C is an x -monotone chain. This proof was not based on the assumption that C is in fact x -monotone and therefore is well defined for arbitrary chains also.

In the context of the simplification C' of the arbitrary chain C in Figure 6, any point feature that lies within the grey regions will be classified topological inconsistent by Theorem 2.3. For example, the points p_1 and p_2 in this figure will be determined inconsistent. Theorem 2.3 has no constraint in terms of returning false-positive or false-negative classifications. This method is optimal in terms of these two forms of constraints. It also possesses no constraint in terms of the simplifications to which it may be applied. It therefore represents an optimal technique, in terms of all three types of constraints, for determining the topological consistency of point features.

Theorem 2.3 shows the original strategy proposed by De Berg *et al.* (1998) for determining the consistency of point features to be well defined for monotone as well as arbitrary chains. Saalfeld (1999) proposed that the consistency of arbitrary chains not self-intersecting could be achieved through representing line segments by their corresponding endpoints and maintaining the consistency of these endpoints using Theorem 2.3. Theorem 2.2 proved this approach to be well defined in the case of monotone chains. Crucially Saalfeld (1999) did not prove this theorem for arbitrary chains. The following section addresses this issue.

2.3. The da Silva strategy

da Silva and Wu (2006) demonstrated the strategy proposed by De Berg *et al.* (1998) for determining the topological consistency of non-intersecting chains to be ill-defined for arbitrary chains. da Silva and Wu (2006) subsequently proposed a method that is well defined for arbitrary chains. In this section we illustrate the reasons why the original strategy proposed of De Berg *et al.* (1998) is ill-defined for arbitrary chains and analyse the inconsistencies it may introduce. We prove that the method of da Silva and Wu (2006) introduces unnecessary constraints if used to determine the consistency of point features.

We proved in Section 2.1 that for monotone chains the consistency of non-intersecting lines may be determined by evaluating the consistency of corresponding line endpoints. The following example demonstrates this strategy to be ill-defined for arbitrary chains.

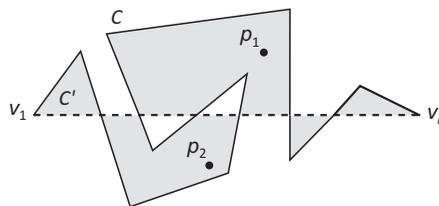


Figure 6. The bounded regions between C and C' are represented by grey.

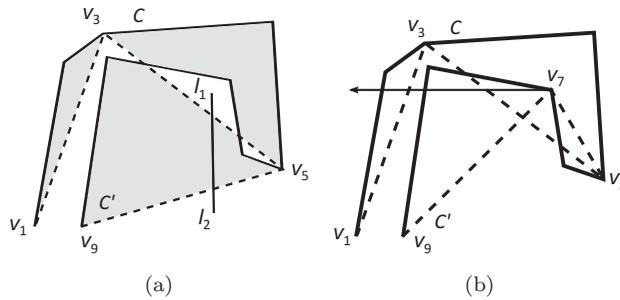


Figure 7. Two possible topologically inconsistent simplifications.

Consider the non-monotone chain C and its corresponding simplification C' in Figure 7a. The bounded regions between these two chains are coloured grey. Next consider the line segment $\overline{l_1l_2}$ that intersects C' but not C . This represents an inconsistent simplification of C relative to this line segment despite the fact that neither l_1 nor l_2 lies in a bounded face formed by C and C' . The simplification is consistent with respect to these endpoints. The reason for this is that when proving this strategy for determining the consistency to be well defined for monotone chains in Theorem 2.2 we exploited the fact that the chains were monotone as follows.

In Figure 7a if $\overline{l_1l_2}$ intersects C' it enters a bounded region. If the chains were monotone, $\overline{l_1l_2}$ could only intersect the boundary of this region once. In doing so a single segment endpoint of $\overline{l_1l_2}$ would lie in this bounded region and be determined inconsistent. This property does not hold for non-monotone chains. For example, a bounded region Figure 7a is bounded by two segments in C' allowing $\overline{l_1l_2}$ to intersect the boundary of this region twice. This can also result in self-intersecting polygons being incorrectly determined consistent. Consider Figure 7b where the simple chain C is simplified to a self-intersecting chain C' . A ray from every vertex in a westerly direction, as demonstrated for a single vertex in this figure, results in an even number of intersections with BC and BC' . In this case every vertex will be classified as being consistent and the simplification classified consistent.

The solution to this problem lies in requiring that each bounded region is bounded by only a single line segment in C' through another means other than the monotone property. da Silva and Wu (2006) proposed to achieve this by determining whether the endpoints of $\overline{l_1l_2}$ are topologically consistent with respect to each individual line segment in C' and its corresponding chain in C , in isolation. If it is determined that $\overline{l_1l_2}$ does not intersect each individual line segment of C' using this procedure, it can be inferred that $\overline{l_1l_2}$ does not intersect C' . Lemma 2.3 and Theorem 2.4 offer an improved version of the proof originally proposed by da Silva and Wu (2006), which is presented in the context of all previous analysis in this article. Let C be a simple arbitrary chain and C' its corresponding simple simplification, which is a single line segment. Let $\overline{l_1l_2}$ be a line segment with endpoints l_1 and l_2 that does not intersect C .

Lemma 2.3: C' is a consistent simplification of C with respect to $\overline{l_1l_2}$ if and only if neither endpoint of $\overline{l_1l_2}$ lies in a bounded face formed by C and C' .

Proof: There exists a chain B that completes both C and C' to simple polygons, denoted by BC and BC' , respectively, such that BC contains l_1 and l_2 and $\overline{l_1l_2}$ does not intersect BC . Such a chain will always exist because, by definition, $\overline{l_1l_2}$ does not intersect C . The remainder of this proof is divided into two possible cases.

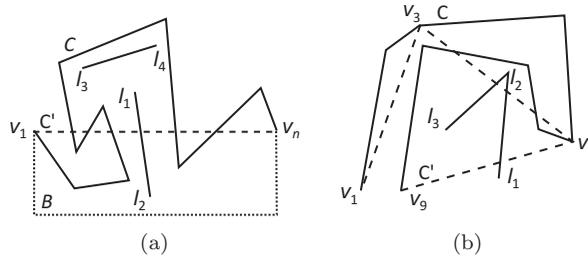


Figure 8. Two possible line segments are shown in (a). A chain containing a series of line segments is shown in (b).

Case 1: Consider the case where the line segment $\overline{l_i l_j}$ intersects the line segment C' . This is the case if $\overline{l_i l_j}$ corresponds to $\overline{l_1 l_2}$ in Figure 8a. Line segments that do not overlap can only intersect a single time. By the point-in-polygon criterion, a line that intersects a polygon a single time will have one endpoint inside and the other endpoint outside the polygon, in question. Therefore, one endpoint of $\overline{l_i l_j}$ will lie outside BC' . This point will lie in a bounded face formed by C and C' and will be classified inconsistent by Theorem 2.3.

Case 2: Consider the case where the line segment $\overline{l_i l_j}$ does not intersect C' . In this case both endpoints of $\overline{l_i l_j}$ will lie inside or outside BC' . If both endpoints lie outside BC' , which is illustrated by $\overline{l_3 l_4}$ in Figure 8a, this simplification is inconsistent because both endpoints lie inside BC . In this case both endpoints will lie in a bounded face formed by C and C' and will be classified inconsistent by Theorem 2.3. If both endpoints lie inside BC' , no endpoint will lie in a bounded face formed by C and C' ; this simplification is consistent and will be classified as such by Theorem 2.3. \square

Lemma 2.3 exhibits no constraint in terms of returning false-positive or false-negative classifications. This method is therefore optimal in terms of these two forms of constraints. Generalization of Lemma 2.3 to simplifications that are not a single line segment but a chain of segments is achieved as follows: Let C be a simple chain and C' its corresponding simple simplification. Let $\overline{l_i l_j}$ be a line segment with endpoints l_i and l_j that does not intersect C .

Theorem 2.4: *The chain C' is a consistent simplification of C with respect to $\overline{l_i l_j}$ if and only if each line segment in C' is determined consistent with respect to $\overline{l_i l_j}$ by Lemma 2.3.*

Proof: If each individual line segment in C' is determined consistent with respect to $\overline{l_i l_j}$ by Lemma 2.3 then $\overline{l_i l_j}$ does not intersect C' . If $\overline{l_i l_j}$ intersects one or more line segments in C' , this will be determined by the corresponding evaluation of Lemma 2.3. \square

In the context of Figure 7a, Theorem 2.4 would involve applying Lemma 2.3 individually to the sub-chains $\overline{v_1 v_3}$, $\overline{v_3 v_5}$ and $\overline{v_5 v_9}$ in C and C' . In this case the simplification C' would be correctly determined as inconsistent with respect to $\overline{l_1 l_2}$. Theorem 2.4 can be generalized to maintain the consistency of a chain, as opposed to a single line segment, not intersecting a simplification which may also be a chain. For example, in Figure 8b the consistency of the simplified chain C' with respect to the chain containing the line segments $\overline{l_1 l_2}$ and $\overline{l_2 l_3}$ must be determined. Theorem 2.4 can also be generalized to determine the consistency of a non-self-intersecting chain. The following two theorems prove these facts.

Theorem 2.5: *The chain C' is a consistent simplification of C with respect to the chain C_S not intersecting C if and only if each line segment in C_S is consistent by Theorem 2.4.*

Proof: If every line segment in C_S is consistent by Theorem 2.4 the chain C_S will be determined consistent. If a single line segment in C_S is inconsistent by Theorem 2.4, the chain C_S will be determined inconsistent. \square

Theorem 2.6: *The chain C' is a consistent simplification of C with respect to C not self-intersecting if and only if each line segment in C' is determined consistent with respect to every other line segment in C' by Theorem 2.4.*

Proof: Applying Theorem 2.4 individually to each line segment in C' and its corresponding sub-chain in C ensures that if an intersection occurs a corresponding endpoint will lie in a bounded face of C and C' . \square

The strategy of da Silva and Wu (2006) can be used to determine the consistency of point features as well as line features but doing so introduces unnecessary constraints. This strategy will not introduce any false positives but can introduce false-negative consistency evaluations. Consider the chain C , its corresponding simplification C' and the point feature p in Figure 9. By Theorem 2.3 C' is a consistent simplification of C with respect to p ; that is, p does not lie in a bounded face formed by C and C' . This evaluation is correct. Using the strategy of da Silva and Wu (2006) would incorrectly classify this simplification with respect to p as inconsistent; this is a false negative. Consequently, to impose the minimal constraints, the strategy of da Silva and Wu (2006) should not be used when determining the consistency of point features.

Table 1 presents a summary of the results of all analyses performed in this section. For each technique we indicate whether it determines a given simplification is topologically inconsistent with respect to the three types of planar topological relationships. An asterisk indicates that the technique in question can determine the topological consistency without constraint and is optimal. An x_i symbol indicates that the technique in question can determine the topological consistency but is subject to some form of constraint and therefore is not optimal. The final column defines what each constraint x_i represents. From this table the following conclusions can be drawn. The consistency of point features should be determined by the strategy of Saalfeld (1999), whereas the consistency of line features should be determined by the strategy of da Silva and Wu (2006). This strategy imposes the least possible constraints.

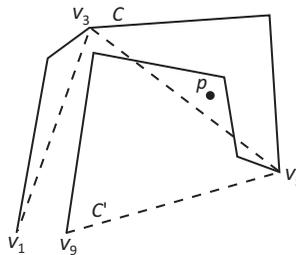


Figure 9. C' is a consistent simplification of C with respect to the point p .

Table 1. The forms of planar topology for which each technique can determine consistency and any corresponding constraints.

	Point	Line	Non-self-intersection	Constraints
De Berg	x_1		x_1	x_1 -monotone chains
Saalfeld	*			
da Silva	x_2	*	*	x_2 -Point features

2.4. Computational complexity

In this section we determine the computational complexity to determine whether a given simplification is topologically consistent with respect to each type of planar topological relationship. In all cases the original chain C and its simplification contain n and m vertices, respectively, where $m \leq n$.

Theorem 2.7: *Determining whether C' is a consistent simplification of C with respect to the point p using the strategy of Saalfeld (1999) can be computed in $O(n)$ time complexity.*

Proof: Computing the parity of p with respect to C and C' requires $O(n)$ and $O(m)$ time complexity. The overall time complexity is therefore $O(n + m)$. Because $m \leq n$ this reduces to $O(n)$ time complexity.

Theorem 2.8: *Determining whether C' is a consistent simplification of C with respect to the planar topological relationship with the chain L that contains l vertices using the strategy of da Silva and Wu (2006) can be computed in $O(nl)$ time complexity.*

Proof: Determining whether the simplification C' is a consistent simplification of C with respect to a single line segment in L requires $O(n)$ time complexity. This must be repeated for each line segment in L giving a total time complexity of $O(nl)$. \square

2.5. Implementation note

In many simplification algorithms individual vertices are iteratively removed from the map in question resulting in a progressively simplified result (Latecki and Lakmper 1999). This fact allows planar topological equivalence between successive simplifications to be easily computed. Removing a single vertex from a contour C results in a single bounded face containing three vertices between it and the corresponding simplification; this is illustrated in Figure 10. To determine the topological equivalence with respect to point features using the strategy of Saalfeld (1999), we simply determine whether any point feature lies inside this triangle. To determine the topological equivalence with respect to lines features using the strategy of da Silva and Wu (2006) we simply determine whether any line vertex lies inside this triangle. Similarly, to determine the topological equivalence with respect to a contour not self-intersecting, we determine whether any line vertex of the contour in question lies inside the triangle.

3. Non-planar topology

Determining the non-planar topological consistency between two maps has been significantly studied in the domain of GIS (Egenhofer and Franzosa 1994, Clementini and Di Felice 1998). Despite this fact approaches for determining the non-planar topological consistency between successive map simplifications have received less focus. Although there

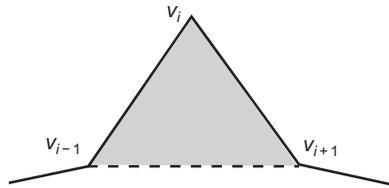


Figure 10. Removing the vertex v_i from C , represented by the solid line, results in a single bounded face containing the vertices v_{i-1} , v_i and v_{i+1} .

is potential to simply apply existing techniques in the context of map simplification, the authors are unaware of a single case where this strategy has been used. We believe this can be attributed to a number of factors. Existing works on determining the non-planar topological consistency can in some cases be very theoretical and lack concrete implementations details (Egenhofer and Franzosa 1994). In cases where such details are provided many techniques are very complex and re-implementation can present a significant challenge (Kuijpers *et al.* 1995, Clementini and Di Felice 1998). Also the computational complexity of such techniques can be high. Determining how to exploit the known relationships between objects in successive simplifications to reduce the complexity of determining the non-planar topological consistency is an open research question. In Section 2.3 we demonstrated that the planar topological consistency of non-intersecting lines can be determined by representing each line segment by its corresponding endpoints and determining their consistency. Unfortunately this strategy cannot be used to determine the non-planar topological consistency of intersecting lines. In Section 3.1 we demonstrate why this is the case. In Section 3.2 we analyse the techniques that have been previously used to maintain non-planar topological consistency between map simplifications.

3.1. Representing intersections explicitly

In Lemma 2.3 we exploited the fact that when simplifying a chain C by a sub-chain C' , it is always possible to form a set of bounded regions formed by C and C' such that the boundary of each region contains only a single line segment of C' . This is because C' is a sub-chain of C where each line segment in C' simplifies one or more line segments of C . If a line intersection is introduced with C' , a segment endpoint would lie in the corresponding bounded region. This endpoint would then be determined inconsistent. To apply this concept to detect the removal of a line intersection with C in the simplification C' , we would need to be able to form a set of bounded regions between C and C' such that the boundary of each region contains only a single line segment of C as opposed to C' ; this is illustrated in Figure 11a. Then if a line intersection between a line segment and C was not represented in the corresponding simplification C' , a segment endpoint would lie in the corresponding bounded region and in turn be determined inconsistent. Because C is not a sub-chain of C' , this is not possible. Instead each region bounded by C and C' will be bounded by least two line segments of C . This is illustrated in Figure 11b where the chain C containing two line segments is simplified by the chain C' containing a single line segment. C' is an inconsistent simplification of C with respect to the intersection with $\overline{l_1 l_2}$. This is despite the fact that no endpoint of $\overline{l_1 l_2}$ lies in a bounded face.

This demonstrates that the consistency of line intersections cannot be maintained by representing the intersecting lines by their corresponding endpoints and maintaining the

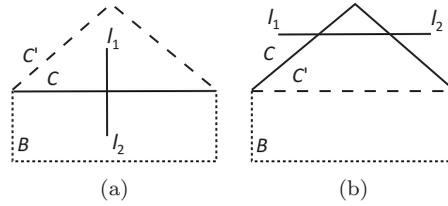


Figure 11. An endpoint of $\overline{l_1 l_2}$ lies in a bounded face of C and C' in (a) but not in (b).

consistency of these. Instead the line segments must be represented explicitly and their corresponding intersections maintained explicitly.

3.2. Determining non-planar consistency

In this section we review two approaches to maintaining non-planar topological consistency between simplifications. In essence both approaches transform the problem of determining the non-planar consistency to one of determining the planar consistency by imposing constraints on the simplification process.

The first approach, to best of our knowledge, was originally proposed by Agrawala and Stolte (2001). Using this strategy non-planar relationships are converted to planar relationships using a process known as planar enforcement (Wise 2002). All segments that intersect another segment are replaced by two corresponding segments, which share a common endpoint equal to the intersection point. These intersection points are then marked as unremovable by the simplification process and therefore all intersection points are maintained. To prevent the introduction of new intersection points, the strategy of da Silva and Wu (2006) (Section 2.3) can be used. Because all intersection points are maintained and no new intersection points are introduced, this strategy maintains non-planar topological consistency. A second strategy for maintaining non-planar topological consistency was proposed by Kulik *et al.* (2005) and Weihua (2008). In this strategy line intersections are maintained by marking all line segments that contain intersections as unremovable by the simplification process.

Both the above strategies for determining the non-planar topological consistency have a time complexity equal to that of determining all segment intersections in the map. This can be determined in $O(n \log n + I \log n)$ where n is the number of vertices and I the number of intersections (Berg *et al.* 2008). It is important to note that both strategies be severely constrained with respect to the possible simplifications for which consistency can be determined. For example, consider the simplification in Figure 12. Despite the fact that this simplification is topologically consistent, it would be determined inconsistent by both strategies for the following reasons: The strategy of Kulik *et al.* (2005) and Weihua (2008)

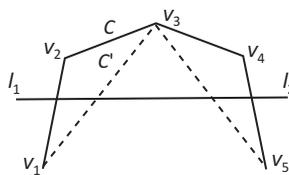


Figure 12. C' is a consistent simplification of C relative to $\overline{l_1 l_2}$.

would determine the simplification inconsistent because the line segments in C that $\overline{l_1l_2}$ intersects are not present in C' ; that is, the line segments $\overline{v_1v_2}$ and $\overline{v_4v_5}$. The strategy of Agrawala and Stolte (2001) would determine the simplification inconsistent because the intersection points between C and $\overline{l_1l_2}$ are not maintained. The development of techniques for determining the non-planar topological consistency of a simplification without such constraints is an open research question.

4. Map simplification

In Section 2 we demonstrated how the planar topological consistency of a map simplification can be determined without constraint. In Section 3 we demonstrated how the non-planar topological consistency of a map simplification can be determined while being subject to constraints. In this section we describe how these strategies can be used by an existing simplification method to generate simplifications that satisfy topological and shape objectives (see Section 1).

Many generalization techniques exist that attempt to satisfy a single objective. For example, Douglas and Peucker (1973) and Saalfeld (1999) proposed techniques that attempt to satisfy metric and topological objectives, respectively. Despite their importance few generalization techniques exist that attempt to satisfy multiply objectives. Galanda (2003) presented a number of metrics for determining the consistency of a generalization with respect to each type of objective. The author noted that future work should focus on the integration of these metrics such that a generalization which satisfies multiple objectives may be realized. Kulik *et al.* (2005) also noted that none of the current generalization algorithms satisfy multiple objectives. He proposed a simplification technique that attempts to satisfy shape, semantic and topological objectives. In Sections 2 and 3, a methodology that can determine the planar and non-planar topological consistency of a given simplification was presented. We propose to fuse this with an existing simplification technique that satisfies shape objectives. Latecki and Lakmper (1999) proposed a polygon simplification technique based on contour evolution. The basic concept of this technique is to replace two consecutive line segments with a single line segment formed by joining their endpoints in each evolution step to obtain a shape hierarchy. To produce an intuitive shape evolution, a suitable order of substitution must be used. Latecki *et al.* proposed to perform substitution in an order where line segments that contribute the least to the overall shape are substituted first and the process terminates when a convex contour is formed.

The proposed simplification process operates as follows: To ensure all non-planar topological relationships are maintained during the simplification process, one of the two pre-processing steps reviewed in Section 3.2 must be applied to the map. That is, either planar enforcement is applied or intersecting line segments are marked unremovable. Next the least significant line or polygon vertex such that its removal will result in a topologically consistent simplification is removed. This step is repeated until no more vertices can be removed from any feature. A vertex cannot be removed from a feature if the feature is convex or the removal results in a topologically inconsistent simplification. The use of convexity to halt simplification was originally proposed by Latecki and Lakmper (1999) who showed that such a simplification represents a high-level shape description of the contour in question.

5. Results

In this section we demonstrate the effectiveness of the proposed simplification methodology at satisfying both shape and topological objectives. All implementation was performed

in the C++ programming language. The point, line and polygon data structures from the Computational Geometry Algorithms Library (CGAL) (Giezeman and Wesselink 2008) were used to represent all map features. This section is divided into a series of parts where each evaluates the individual elements of the overall simplification algorithm. In Section 5.1 we evaluate the original simplification algorithm of Latecki and Lakmper (1999) without any topological constraints. In Section 5.2 we evaluate this simplification algorithm with the addition of planar topological constraints. In Section 5.3 we evaluate this simplification algorithm with the addition of both planar and non-planar topological constraints.

5.1. *Shape*

All data used in our analysis were taken from OSM which, since its beginnings in 2004, has grown to become an important source of geospatial data. An OSM data set containing a set of polygon and line features is plotted in Figure 13a. The polygons represent a forest park, a car park and a shopping centre. The line features represent roads. This map exhibits the following forms of planar topological relationships: the containment or non-containment of one polygon inside another, the non-self-intersection of polygons and the non-intersection of polygons and lines. A number of non-planar topological relationships in the form of lines intersecting polygons are also evident in the original map. For example, consider the relationships between the red line that crosses the blue polygon and the grey line that crosses the blue polygon. The map in Figure 13a is progressively simplified in Figure 13b–f using the strategy of Latecki and Lakmper (1999). No check for topological consistency is performed during the simplification process. The only constraint imposed is that vertices of degree not equal to 2 cannot be removed. The original map in Figure 13a contains a total of 641 polygon and line vertices. The simplifications in Figure 13b–f contain 341, 241, 141, 91 and 62 vertices, respectively. Figure 13f represents a 90% reduction in data size relative to Figure 13a. Our implementation achieved this result in a time of 4 running on an Intel 2.8 GHz dual core processor. It is evident from this series of simplifications that this process generates a suitable shape evolution of polygon and line shapes and satisfies shape objectives. On the contrary, this process fails to satisfy topological objectives. In Figure 13f it can be seen that a number of intersections between polygons have been introduced by the simplification process; this represents a planar topological inconsistency. Also the non-planar topological relationships mentioned earlier between the blue polygon and the red and grey lines are inconsistent in the final simplification.

5.2. *Planar consistent simplification*

The map in Figure 14a is progressively simplified in Figure 14b–f using the strategy of Latecki and Lakmper (1999). At each simplification step the strategies of Saalfeld (1999) and da Silva and Wu (2006) are used to ensure whether each simplification is consistent with respect to planar topological relationships. This reduces to checking whether any corresponding line vertices fall within a single bounded region (see Section 2.5). The original map in Figure 14a contains 641 polygon and line vertices total. The simplifications in Figure 14b–f contain 341, 241, 141, 91 and 62 vertices, respectively. Figure 14f represents a 90% reduction in data size relative to Figure 14a. Our implementation achieved this result in a time of 4 s running on an Intel 2.8 GHz dual core processor. It is evident from this series of simplifications that this process generates a suitable shape evolution of polygon and line shapes and satisfies shape objectives. It is also evident that each simplification is consistent with respect to planar topological relationships. No polygons or lines that did

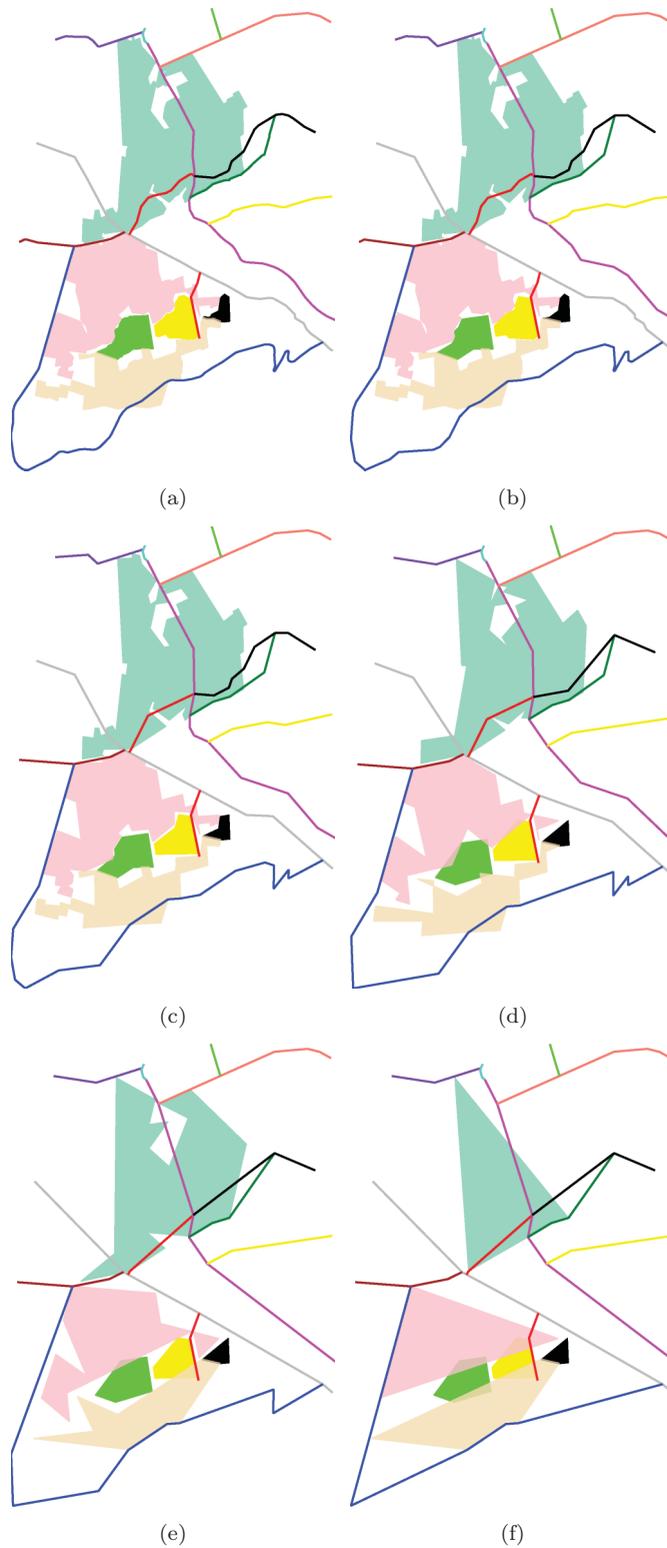


Figure 13. The set of features in (a) are simplified by an increasing amount in (b)–(f). All simplifications satisfy shape objectives.

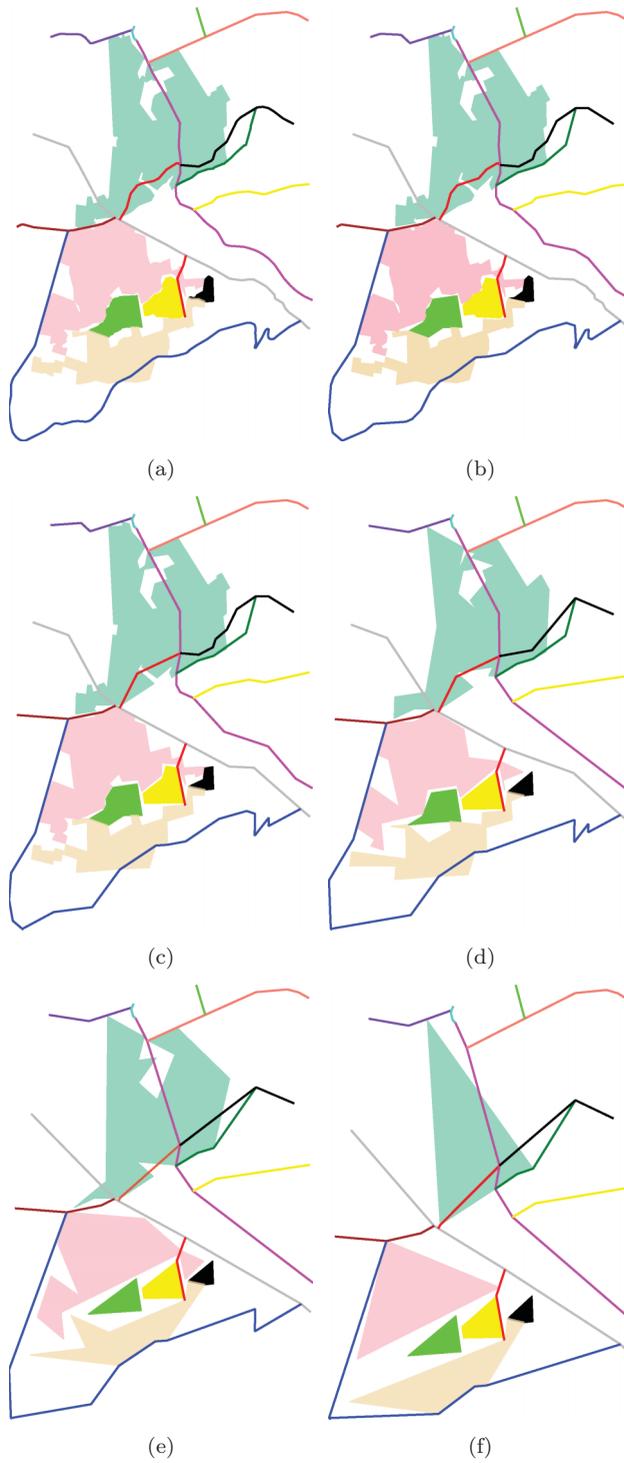


Figure 14. The set of features in (a) are simplified by an increasing amount in (b)–(f). All simplifications satisfy shape and planar topological objectives.

not intersect in the original map intersect in any of the simplified results. On the contrary, the simplification process does not produce simplifications that are consistent with respect to non-planar topological relationship. For example, the topological relationships between the blue polygon and the red and grey roads are inconsistent in Figure 14f.

5.3. Non-planar consistent simplification

The map in Figure 15a is progressively simplified in Figure 15b–f using the strategy of Latecki and Lakemper (1999). At each simplification step the strategies of Saalfeld (1999) and da Silva and Wu (2006) are used to ensure whether each simplification is consistent with respect to planar topological relationships. The strategy of Kulik *et al.* (2005) and Weihua (2008) is used to maintain all non-planar topological relationships where line segments containing an intersection are marked unremovable. The simplifications in Figure 15b–f contain 321, 271, 221, 171 and 141 vertices, respectively. Figure 15f represents a 78% reduction in data size relative to Figure 14a. Our implementation achieved this result in a time of under 5 s running on an Intel 2.8 GHz dual core processor. This reduction in data size is less than that achieved in the previous sections. This is because maintaining non-planar topology inherently requires a greater number of vertices. It is also partly because the current strategy for maintaining non-planar topological consistency is subject to the constraints discussed in Section 3.2, which further reduces the number of suitable simplifications. Given the rapidly growing size of spatial data sets, it is important to evaluate the performance of the proposed technique on a large data set. Such a data set extracted from OSM, which contains a total of 13,755 polygon and lines vertices, is displayed in Figure 16a. This was simplified to the map in Figure 16b, which contains 4755 vertices, and is both planar and non-planar topologically consistent. Our C++ implementation achieved this result in a time of 261 s running on an Intel 2.8 GHz dual core processor.

6. Conclusions

This article has presented a formal mathematical analysis of strategies for determining the topological consistency of a given vector map simplification. This analysis was divided into two parts. First, we considered techniques for determining the planar topological consistency and in particular the three seminal methods of De Berg *et al.* (1998), Saalfeld (1999) and da Silva and Wu (2006). A unified analysis of these methods was presented in terms of the forms of topology each individual technique can determine the consistency and any corresponding constraints imposed. It was proven that in terms of determining the consistency of point features, the method of Saalfeld (1999) imposes zero constraints and is optimal. It was also proven that in terms of determining the consistency of non-intersecting lines, the method of da Silva and Wu (2006) imposes zero constraints and is optimal. Next we considered techniques for determining the non-planar topological consistency. It was shown that the existing methods of Kulik *et al.* (2005), and Weihua (2008) and Agrawala and Stolte (2001) impose significant constraints on the simplifications to which they may be applied. Overcoming this issue represents a challenge for future research. A new simplification algorithm that satisfies shape and topological objectives was subsequently proposed. A number of data sets from OSM were simplified using this methodology. In all cases a large reduction in the number of vertices used to represent features was achieved although still maintaining shape and topological consistency.

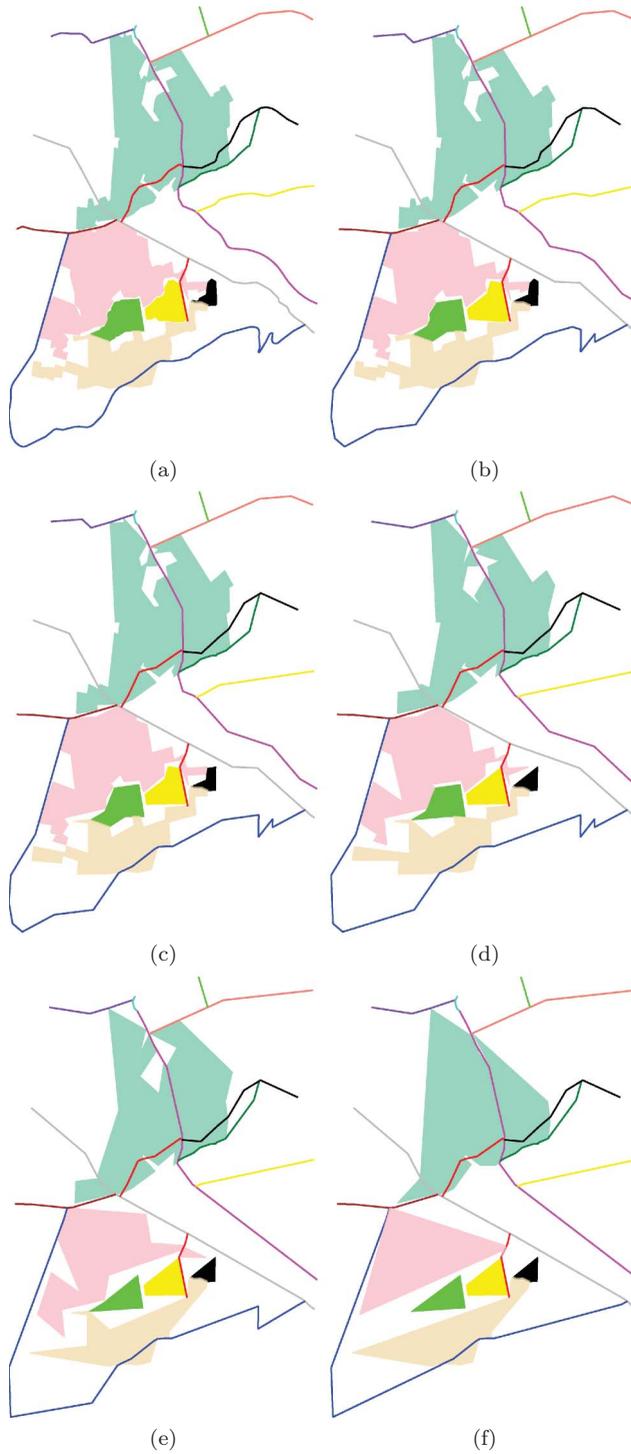


Figure 15. The set of features in (a) are simplified by an increasing amount in (b)–(f). All simplifications satisfy shape, planar and non-planar topological objectives.



(a)



(b)

Figure 16. The large-scale map in (a) is simplified in (b).

References

- Agrawala, M. and Stolte, C., 2001. Rendering effective route maps: improving usability through generalization. In: *Proceedings of SIGGRAPH*, Los Angeles, 12-17 August. New York: ACM, 241–249.
- Berg, M.D., et al., 2008. *Computational geometry: algorithms and applications*. 3rd ed. Berlin, Heidelberg: Springer.
- Cicerone, S., Frigioni, D., and Felice, P.D., 2002. A general strategy for decomposing topological invariants of spatial databases and an application. *Data and Knowledge Engineering*, 42 (1), 57–87.
- Clementini, E. and Di Felice, P., 1998. Topological invariants for lines. *IEEE Transactions on Knowledge and Data Engineering*, 10 (1), 38–54.
- da Silva, A.C.G. and Wu, S.T., 2006. A robust strategy for handling linear features in topologically consistent polyline simplification. In: *GeoInfo*, 19–34.
- De Berg, M., van Kreveld, M., and Schirra, S., 1998. Topologically correct subdivision simplification using the bandwidth criterion. *Cartography and Geographic Information Systems*, 25 (4), 243–257.
- Douglas, D. and Peucker, T., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The Canadian Cartographer*, 10 (2), 112–122.
- Egenhofer, M.J. and Franzosa, R.D., 1994. On the equivalence of topological relations. *International Journal of Geographical Information Systems*, 8 (6), 133–152.
- Estkowsky, R. and Mitchell, J.S.B., 2001. Simplifying a polygonal subdivision while keeping it simple. In: *17th annual symposium on computational geometry*, Medford, MA, 3-5 June, 40–49.
- Galanda, M., 2003. Modelling constraints for polygon generalization. In: *Fifth workshop on progress in automated map generalization*, Paris, France.
- Giezeman, G.J. and Wesselink, W., 2008. 2D Polygons. In: C.E. Editorial Board. *CGAL User and Reference Manual*. 3.4 ed.
- Jones, C.B., 1997. *Geographical information systems and computer cartography*. London: Prentice Hall.
- Jones, C.B. and Ware, J.M., 2005. Map generalization in the Web age. *International Journal of Geographical Information Science*, 19 (8–9), 859–870.
- Kuijpers, B., Paredaens, J., and Van den Bussche, J., 1995. Lossless representation of topological spatial data. In: *SSD '95: Proceedings of the 4th International Symposium on Advances in Spatial Databases*. London: Springer-Verlag, 1–13.
- Kulik, L., Duckham, M., and Egenhofer, M., 2005. Ontology-driven map generalization. *Journal of Visual Languages and Computing*, 16 (3), 245–267.
- Latecki, L.J. and Lakmper, R., 1999. Convexity rule for shape decomposition based on discrete contour evolution. *Computer Vision and Image Understanding*, 73 (3), 441–454.
- Li, Q., 2009. Variable-scale representation of road networks on small mobile devices. *Computers and Geosciences*, 35 (11), 2185–2190, Progressive Transmission of Spatial Datasets in the Web Environment.
- Li, S. and Liu, W., 2010. Topological relations between convex regions. In: *AAAI Conference on Artificial Intelligence* Atlanta, Georgia, 11-15 July.
- O'Rourke, J., 1998. *Computational geometry in C*. New York, USA: Cambridge University Press.
- Saalfeld, A., 1999. Topologically consistent line simplification with the Douglas–Peucker algorithm. *Cartography and Geographic Information Science*, 26 (1), 7–18.
- Weibel, R., 1996. A typology of constraints to line simplification. In: *Advances in GIS research II (Proceedings 7th International Symposium on Spatial Data Handling)*. London: Taylor & Francis, 533–546.
- Weihua, D., 2008. Generating on-demand web mapping through progressive generalization. In: *Proceedings of the 2008 International workshop on education technology and training & 2008 international workshop on geoscience and remote sensing. ETT and GRS 2008*, Shanghai, China, 21-22 December Vol. 2, December, 163–166.
- Wise, S., 2002. *GIS basics*. Boca Raton: CRC Press.
- Yang, B., 2005. A multi-resolution model of vector map data for rapid transmission over the Internet. *Computers & Geoscience*, 31 (5), 569–578.
- Yang, B. and Weibel, R., 2009. Editorial: some thoughts on progressive transmission of spatial datasets in the web environment. *Computers and Geosciences*, 35 (11), 2175–2176, Progressive Transmission of Spatial Datasets in the Web Environment.

Copyright of International Journal of Geographical Information Science is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.