Metric Details of Topological Line-Line Relations

KONSTANTINOS A. NEDAS

National Center for Geographic Information and Analysis
Department of Spatial Information Science and Engineering
Boardman Hall, University of Maine, Orono, ME 04469-5711, USA
kostas@spatial.maine.edu

MAX. J. EGENHOFER

National Center for Geographic Information and Analysis
Department of Spatial Information Science and Engineering,
Department of Computer Science
Boardman Hall, University of Maine, Orono, ME 04469-5711, USA
max@spatial.maine.edu

and DOMINIK WILMSEN

National Center for Geographic Information and Analysis
Department of Spatial Information Science and Engineering
Boardman Hall, University of Maine, Orono, ME 04469-5711, USA
dominik@spatial.maine.edu

Metric Details of Topological Line-Line Relations

Abstract

Many real and artificial entities in geographic space, such as transportation networks and trajectories of movement, are typically modeled as lines in geographic information systems. In a similar fashion, people also perceive such objects as lines and communicate about them accordingly as evidence from research on sketching habits suggests. To facilitate new modalities like sketching that rely on the similarity among qualitative representations, oftentimes multi-resolution models are needed to allow comparisons between sketches and database scenes through successively increasing levels of detail. Within such a setting, topology alone is sufficient only for a coarse estimate of the spatial similarity between two scenes, whereas metric refinements may help extract finer details about the relative positioning and geometry between the objects. The 9-intersection is a topological model that distinguishes 33 relations between two lines based on the content invariant (empty-nonempty intersections) among boundaries, interiors, and exteriors of the lines. This paper extends the 9-intersection model by capturing *metric details* for line-line relations through *splitting ratios* and *closeness measures*. *Splitting ratios*, which apply to the 9-intersection's non-empty values, are normalized values of lengths and areas of intersections. *Closeness measures*, which apply to the 9-intersection's empty values, are normalized distances between disjoint object parts. Both groups of measures are integrated into compact representations of topological relations, thereby addressing topological and metric properties of arbitrarily complex line-line relations.

1. Introduction

Modern geographic information systems (GISs) still rely heavily on quantitative descriptions of spatial objects and phenomena, both for storage and querying. There is significant evidence, however, that people think of space and communicate about spatial concepts using qualitative rather than quantitative terms (Lynch 1960; Hernández 1994; Regier 1995). An example is the approximate way in which people communicate directions to one another (e.g., the church is in the center square, which is a couple blocks down and to the left). The persistence on the classic quantitative paradigm renders GIS packages usable only for professionals or sophisticated users who often receive extensive training so that they become proficient in the formalizations of underlying spatial data models and their terminology. Non-expert users typically feel alienated, since they lack the necessary background and the technical jargon needed to comprehend and employ these tools, even for relatively simple tasks such as way-finding or spatial querying in order to find objects of interest around them.

Recent studies addressed the lack of commonsense formalizations of geographic knowledge in computers, by proposing formal and sound theories that allow reasoning about spatial relations, primarily in a qualitative manner (Egenhofer and Franzosa 1991; Randell *et al.* 1992). One such developed theory is the 9-intersection model (Egenhofer and Herring 1990), which focuses on binary topological relations between two regions, two lines (Egenhofer 1994), and a region and a line. The 9-intersection is an effort to incorporate *Naive Geography* concepts and reasoning into GISs (Egenhofer and Mark 1995). The internal representations of spatial relations and the mathematical operations that take place within this model are transparent to users, who are able to formulate queries by employing spatial predicates that correspond to natural-language terms such as *inside* or *overlap*, and also receive answers in a similar fashion.

The prominence of topology in the 9-interesection as the most critical aspect that people refer to when assessing spatial relationships in geographic space, has been confirmed by experiments in psychology and cartography (Lynch 1960; Stevens and Coupe 1978; Mark 1992; Mark and Egenhofer 1994). A critical factor that reinforces this view is that errors about spatial relations in human cognition are typically of metric,

rather than topological nature (Tversky 1981; Talmy 1983). Despite its importance, however, topology *per se* is often insufficient in addressing people's needs. Metric details—though considered to be of lesser importance—are still required to capture the essence of spatial relations. Such circumstances arise when topology-based results to queries—even though exact—are underdetermined (i.e., do not provide enough detail so as to help accomplish the task at hand). Another typical situation of the usefulness of metric enhancements is exemplified by people's tendency to occasionally complement qualitative with quantitative information in order to resolve ambiguities in the description of spatial scenes. To reflect better human behavior, GISs that rely on models such as the 9-intersection need to incorporate mechanisms that will allow metric, in addition to the topological inferences among spatial entities. We follow the premise that *topology matters, while metric refines* (Egenhofer and Mark 1995); hence, the metric enhancements should be viewed only as extensions and supplements to the theory and not as the core of a qualitative geographic information system.

This paper focuses on binary relations between linear objects. The intent is to develop a comprehensive model for capturing metric details about such relations. Examples of entities that people often conceptualize as lines include road networks, sewer systems, rivers and streams, irrigation networks, aerial navigation routes, and satellite orbits. The critical components for line-line relations are the interiors and boundaries of the lines. When the interior or boundary of one line interacts with either the interior or boundary of the other line, certain metric properties can be captured about this interaction. For instance, a line may cross the interior of another, thus separating its interior into two distinct segments, the length of which could be measured. Even when parts of the two lines do not interact (i.e., their intersection is empty), it is still possible to measure the distance among those parts.

Purely quantitative measures, however, are undesirable because they do not take into consideration the relation to the objects for which they were derived. To describe details about topological relations, we consider the metric concept of (1) *splitting*, which determines how a line's interior and exterior are partitioned by the other line's interior or boundary, and (2) *closeness*, which determines how close or far apart are disjoint line

parts (boundary-boundary, boundary-interior, and interior-interior). Both types of measures are normalized (i.e., scale-independent) values with respect to metric properties of line relations, such as the lengths of common parts or the area enclosed by two lines.

The remainder of this paper presents in detail the topological and metric models used to specify the geometry of spatial relations. Section 2 summarizes previous work on line-line relations, focusing mainly on concepts of the 9-intersection model which we extend with metric details. Section 3 introduces the rationale for splitting ratios and defines three types of line-splitting ratios: *line alongness*, *interior splitting*, and *exterior splitting*. Section 4 is concerned with the development of closeness measures and defines three types of closeness measures: *boundary closeness*, *interior closeness*, and *interior-boundary closeness*. The next two sections are concerned with the representation of the derived metrics in database structures. Section 5 describes the integration of splitting ratios into the same tabular representation that was used for the 9-intersection-based detailed topological relations (Clementini and di Felice 1998), yielding a *metrically enhanced classifying invariant*, whereas Section 6 discusses the encoding of closeness measures in a 4-intersection matrix, a more compact version of the 9-intersection. Section 7 provides conclusions and suggests topics for future research.

2. Topological Measures for Line-Line Relations

From a purely quantitative standpoint, line algorithms and representations have been omnipresent in the classic GIS literature and date back to the genesis of GISs. Lines and points are at the core of representational structures for objects in conventional GISs, whereas line algorithms (e.g., intersection of lines, line in polygon) are used to perform fundamental GIS operations relating to the querying and updating of those systems (Laurini and Thompson 1992).

The first formalisms for representing and reasoning qualitatively about spatial information, however, have mostly been concerned with qualitative relations among regions. Such theories were primarily developed within the AI and the GIS communities. Besides the 9-intersection (Egenhofer and Herring 1990), notable alternative models for spatial relations include symbolic projections (Chang and Jungert 1996) and the region-

connection calculus (Randell *et al.* 1992). *Symbolic projections* model spatial relations based on directions captured independently along the coordinate axes. Unlike the 9-intersection, however, they refer to the objects' minimum bounding rectangles, rather than to their actual shapes, which provides an approximation that depends on the objects' orientations. The *region-connection calculus*, based on the part-whole theory of *mereology* (Simons 1987) and Clarke's (1981) calculus of individuals, identifies for region-region configurations the same set of binary relations as the 9-intersection; however, it has not been developed for relations involving line-like objects.

Building on the basic notions of qualitative reasoning for relations between regions, several respective formalisms for lines have been proposed. Such an effort is the theory of *path relations*, a special type of spatial relations which are concerned with forming the geometry of trajectories and defining the exact semantics of relations such as "along", "past", or "around" (Krüger and Maaß 1997; Kray and Blocher 1999). Other methods focus on reasoning about oriented line segments. Schlieder (1995) developed a calculus that resulted in a set of 14 spatial relations among ordered linear segments, based on an extension of Allen's interval relations to two-dimensional space and the order information of triples of points, where points are the vertices of the segments. Moratz *et al*. (2000) identified an extended set of 24 atomic relations by considering additional possible configurations and documented a relation algebra for those relations. Cristani's (2003) work is similar, but also tackles the problem of line-segment relations in 3-dimensional space. Isli (2002) developed a ternary algebra for reasoning about orientations. The algebra consists of a rotational and a translational component of knowledge of 3-valued relations among directed lines. Most of these efforts have implicit the assumption that line segments are straight, in contrast to the 9-intersection, which applies to arbitrary shapes and thus has a wider applicability. Qualitative reasoning with lines has also been used outside the context of binary and ternary relations to describe individual object properties. An example includes qualitative description of the shape of spatial entities (Schlieder 1996; Gottfried 2003).

This research focuses on relations between lines derived by the 9-intersection model, and complements them with metric details. Our work in this sense is parallel to that of

Shariff (1996) and Egenhofer and Shariff (1998) who identified, for the 9-intersection, metric refinements for region-region and line-region relations, respectively. The measures they developed, however, are not directly applicable to line-line relations, due to the particularity of linear objects (i.e., each with two 0-dimensional boundary points and a 1-dimensional interior). For example, an overlap relation between two regions has metric refinements of common interior areas and common boundary lengths, whereas an overlap relation between two lines has common interior lengths but no boundary length measures. Furthermore, our set of closeness metrics is stricter in the sense that they are object-identity invariant (i.e., the choice of the reference object is immaterial). This is a basic assumption of previous approaches, however, where a different labeling of the objects would change all metric quantities derived for the description of a spatial scene (Blaser 2000).

The 9-intersection model (Egenhofer and Herring 1990), upon which we base this work, provides a comprehensive framework and a relation algebra for the description of topological relations between objects of type area, line, and point. The topological relation between two such geometric objects, A and B, is characterized by the binary value (empty, non-empty) of the set intersections of A's interior (A°), boundary (∂A), and exterior (A^-), with the interior, boundary, and exterior of B (Equation 1).

$$I(A,B) = \begin{pmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^- \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^\circ & A^- \cap \partial B & A^- \cap B^- \end{pmatrix} \quad \textbf{(1)}$$

The content of the set intersections is a topological invariant (i.e., a topological property that is preserved under groups of topological transformations such as rotation, scaling, and skewing). With nine set intersections and two possible values for each, the model distinguishes 512 possible topological relations, some of which cannot be realized depending on the dimensions of the objects and the dimension of the embedding space. Those that cannot be realized are eliminated through a set of consistency constraints (Egenhofer and Franzosa 1991; Egenhofer 1994). One that applies to line-line relations, for example, is that the intersection of the exteriors of two lines in $\mathbb{R}^2$ can never be

empty. Eliminating impossible relations through constraints results in a set of 33 relations that can be realized between linear objects in $\mathbb{R}^2$ (Figure 1). These relations are the focus of this work. The only assumption we make is that the relations pertain to simple lines, defined as lines with exactly two boundary nodes and without any self-intersections. We also use the term *boundary of a line* to refer strictly to both points that make up a line's boundary; to refer to a specific point the term *boundary point* is used instead.

The content invariant of empty and non-empty intersections, although attractive due to its simplicity, is only a coarse measure, incapable of differentiating several situations that people often do. For example, the two spatial configurations in Figure 2 are distinct, while they are represented by the same 9-intersection matrix; therefore, in order to capture such finer details one has to consider additional invariants. Early work for invariants of line-line relations suggested using the type of interior intersections (touching or crossing) as an invariant (Herring 1991). Egenhofer and Franzosa (1995) developed a set of invariants that help establish topological equivalence between a model representation and a spatial configuration for region-region relations. Based on this model, Clementini and di Felice (1998) derived a complete set of invariants for line-line relations. Examples of those invariants are the type of intersection, the dimension, and the order in which intersections are encountered when their cardinality is larger than one. Each of these invariants is explained in depth in section 5, where we describe how metric information derived from splitting measures and detailed topological information for a relation can be stored in a single, uniform representation.

An important invariant is the *number of components*. The definition of a component is based on the topological concepts of *separation* and *connectedness* (Egenhofer and Franzosa 1995). For a set *Y*, a component is the largest connected (non-empty) subset of *Y*. Whenever any of the nine set intersections is separated into disconnected subsets, these subsets are the *components* of this set intersection. Hence, any non-empty intersection may have several distinct components, each of which may be characterized by its own topological properties. The number of components of an intersection is denoted by $\#(A \cap B)$. For example, for the configuration in Figure 2a, $\#(L_1^\circ \cap L_2^\circ) = 1$, whereas for Figure 2b, $\#(L_1^\circ \cap L_2^\circ) = 2$. In addition, for Figure 2b, component c_o is a 0-dimensional

component whereas c_1 is a 1-dimensional component. An obvious dependency between the content and the component invariants is that any empty intersection has zero components, and every non-empty intersection has at least one component.

3. Splitting Measures

Splitting determines how a line's interior is divided by another line's interior or boundary. A special case of splitting pertains to the separation of the common exterior of the lines into one unbounded and one or more bounded components. To describe the degree of splitting, the metric concepts of the *length of a line* and the *area of a bounded exterior* are used. Among the entries of the 9-intersection for two simple lines, there are five intersections—between two boundaries, between boundary and interior, and between boundary and exterior—that cannot be evaluated with a length or area measure, because these intersections are 0-dimensional (Table 1). The intersection of the two interiors can be evaluated with a length measure only when it is 1-dimensional. If non empty, the two intersections of one line's interior with the other line's exterior are always 1-dimensional. The intersection of the exteriors of the lines is always 2-dimensional.

To normalize the length of the common interior we compare it with the length of L_1 (or the length of L_2). The length of the intersection between L_1's interior and L_2's exterior is normalized by the length of L_1. Similarly, the length of the intersection between L_2's interior and L_1's exterior is normalized by the length of L_2. The area of a bounded exterior is normalized by the area of a circle whose perimeter is equal to the sum of the lengths of the two lines. Such a circle encloses the largest bounded exterior area that two lines can form.

Two simple lines may form a topological configuration of arbitrary complexity with multiple components of the same or different intersection types; therefore, the metric refinements in the form of the splitting measures operate at the component level describing adequately the different metric properties of each component. For instance, for the configuration in Figure 3a one calculates the metric properties separately for each intersection between the interiors of the lines. A global measure that would rely on the

sum of all common interior segments would not help distinguish between the two topologically equivalent configurations depicted in Figures 3a and 3b.

3.1 Line Alongness

Line alongness captures how much of one line coincides with another line. In order to consider line alongness, the intersection of the interiors of two lines must be non-empty $(L_1^\circ \cap L_2^\circ = \neg\varnothing)$ and 1-dimensional. The interior of one line interacts with the interior of the other such that each line is separated into two sets of line parts: (1) line segments that are in the common interior (i.e., common interior components) and (2) line segments that are in the exterior of the other line. This separation makes a 1-dimensional object split another 1-dimensional object into two or more 1-dimensional parts (Figure 4).

As the measure for the separation we employ the notion of the *line alongness ratio* (*LA*) as the ratio between the length of the common interior and the length of a line. There are two possible ratios: one with respect to the length of L_1 and another with respect to the length of L_2 (Equation 2). The range of the line alongness ratio is $0 \le LA \le 1$. When the common interior segment degenerates to a point, *LA* reaches 0. If L_1 is entirely contained within the interior of L_2, then LA_1 becomes 1, and the same occurs for LA_2, when L_2 is entirely contained within the interior of L_1. If both LA_1 and LA_2 are 1, then the lines are equal. For arbitrarily complex configurations with multiple interior-interior intersections, a separate measure of line alongness is derived for each component c where $c \in \mathbb{N}^+$ (the set of positive integers).

$$LA_{i,j(c)} = \frac{length(L_i^\circ \cap L_j^\circ)}{length(L_i)} \quad \textbf{\textit{with}} \; i, j \in \{1,2\}, i \neq j \;,\; c \in \mathbb{N}^+ \tag{2}$$

3.2 Interior Splitting

If the interior or boundary of one line interacts with the interior of the other line, it separates the interior into left and right line segments according to some predetermined orientation. This involves a 1-dimensional object (i.e., common interior segment) or a 0-dimensional object (i.e., interior or boundary point) splitting a 1-dimensional object into

two 1-dimensional parts, both of which intersect with the exterior of the splitting line (Figure 5).

In order to consider interior splitting for a line L_1, the intersection of that line's closure with the closure of the other line L_2 must be non-empty (i.e., $L_1^\circ \cap L_2^\circ = \neg\varnothing$, or $L_1^\circ \cap \partial L_2 = \neg\varnothing$, or $\partial L_1 \cap \partial L_2 = \neg\varnothing$) and part of L_1's interior must intersect with L_2's exterior (i.e., $L_1^\circ \cap L_2^- = \neg\varnothing$). A normalized measure for interior splitting is the *interior splitting ratio* (*IS*) between the line segment of the split line that is located in the exterior of the splitting line, and the length of the split line (Equation 3). This measure is evaluated separately for each applicable component intersection *c*. For example, in a typical *cross*-like configuration (Figure 5a), there are four components.

$$IS_{i(c)} = \frac{length(component(L_i^\circ \cap L_i^-))}{length(L_i)} \quad \textbf{with} \; i, j \in \{1,2\}, i \neq j, \; c \in \mathbb{N}^+ \tag{3}$$

The interior splitting ratio is complementary to the line alongness ratio and its range is $0 < IS \leq 1$. It would be 0 if one line was entirely contained within another, or if the lines were equal, which means that either the intersection $L_1^\circ \cap L_2^-$, or the intersection $L_1^- \cap L_2^\circ$ or both, would be empty. It reaches 1 for one line when the interior-interior intersection becomes empty, for instance, when the configuration in Figure 5a degenerates to Figure 5b. It becomes 1 for both lines when the two lines meet only at a common boundary (Figure 5c).

3.3 Exterior Splitting

Exterior splitting occurs if parts of the two lines (interiors or boundaries or both) interact such that they form one or more closed regions (Figure 6). Hence, exterior splitting involves two 1-dimensional objects splitting a 2-dimensional object into two or more parts. Specifically, this type of splitting implies a partitioning of the common exterior of the two lines into two or more components: an unbounded exterior component and one or more *bounded* exterior components. The term *bounded* refers to the exterior-exterior intersections that are completely surrounded by the interiors of the two lines.

A normalized measure for this property is the *exterior splitting ratio* (*ES*) as the ratio between the area of the bounded exterior that is formed by two lines, and the maximum bounded exterior that could possibly be formed by the same lines (Equation 4). For arbitrarily complex configurations with multiple bounded exteriors, a separate measure of the exterior splitting ratio is derived for each component *c*.

$$ES_{(c)} = \frac{4\pi(area(BoundedComponent(L_i^- \cap L_j^-))}{(length(L_i) + length(L_j))^2} \textbf{ with } i, j \in \{1,2\}, i \neq j \textbf{,} \quad c \in \mathbb{N}^+ \tag{4}$$

The area of the maximum possible bounded exterior is equal to the area of a circle with a perimeter that is equal to the sum of the lengths of the two lines. The range of the exterior splitting ratio is $0 < ES \leq 1$. It would reach 0 if the bounded area was nonexistent. It becomes 1 if the two lines form only one bounded area, and there are two non-empty boundary-boundary ($\partial L_1 \cap \partial L_2$) intersections (Figure 6e).

4. Closeness Measures

Closeness involves considerations of distances among points and lines. Unlike splitting ratios, which require coincidence, closeness describes how far apart disjoint parts are. The object parts involved are the boundaries and the interiors of the lines. Shape-descriptive measures, such as the distance between two boundary points that belong to the same line, would specify properties of individual objects, not relations. The focus here is on the metric relations between disjoint object parts that belong to different lines. For all possible configurations between two linear objects there are three types of closeness measures of interest:

- the closeness of a line's boundary to another line's boundary that involves distances between points;
- the closeness of a line's interior to another line's boundary that involves distances between lines (1-dimensional objects) and points (0-dimensional objects); and
- the closeness between a line's interior and another line's interior that involves distances between two lines (1-dimensional objects).

Considering exteriors would require definitions of distances between 2-dimensional objects and it would not offer additional refinements of the spatial configuration between two lines in $\mathbb{R}^2$ (i.e., the distance between the exterior of one line and parts of the other line would always be 0). As in the case of splitting ratios, the distances expressing the closeness measures need to be normalized in order to become scale-independent.

Closeness measures are based on a generalization of the typical distance concept, which is defined between two 0-dimensional objects (i.e., points). Two lines, however, also require distance measures between 1-dimensional components, such as the lines' interiors, and, therefore, the distance definitions need to accommodate such cases as well. The extended distance definitions used here are based on the concepts of *largest non-containing circle*, *smallest containing circle*, *largest non-containing buffer*, and *smallest containing buffer*:

Definition 1: Given a point P and a geometric object O, the *largest non-containing circle* around P with respect to O—$LNCC(P,O)$—is a circle of maximum radius R, such that nothing of O is contained in the circle's interior (Figure 7a).

Definition 2: Given a point P and a geometric object O, the *smallest containing circle* around P with respect to O—$SCC(P,O)$—is a circle of minimum radius R, such that all of O is contained in the circle's interior (Figure 7b).

Definition 3: Given a line L and a geometric object O, the *largest non-containing buffer* around L with respect to O—$LNCB(L,O)$—is a buffer of maximum offset distance R, such that nothing of O is contained in the buffer's interior (Figure 7c).

Definition 4: Given a line L and a geometric object O, the *smallest containing buffer* around L with respect to O—$SCB(L,O)$—is a buffer of minimum offset distance R, such that all of O is contained in the buffer's interior (Figure 7d).

Definitions 1-2 pertain to boundary-boundary and boundary-interior distances, whereas definitions 3-4 pertain to interior-boundary and interior-interior distances where the terms

circle and *radius* are substituted with the more general concepts of *buffer* and *offset distance*, respectively. A buffer of offset distance R around a simple line L is the area formed by continuously moving a circle of radius R from the starting to the ending boundary of the line. Based on these definitions we discuss the measures of *boundary closeness*, *interior closeness*, and *interior-boundary closeness*. For simplicity, the illustrated examples for each measure are given for disjoint lines, although closeness measures apply to any of the 33 relations realizable between two simple lines.

4.1 Boundary Closeness

The *boundary closeness* is a measure of the remoteness of one line's boundary ∂L_1 from the boundary of another line ∂L_2 and should be expressed as some function of an appropriately selected set of distances that can be realized among four boundary points. There are eight such possible distances, realized by drawing *LNCC*s and *SCC*s from the boundary points of one line to the boundary points of the other line. This set can be reduced to four, because it comprises four symmetric pairs. Our goal is to isolate from these four distances an appropriate subset such that the derived metrics will always be conceptually equivalent and thus directly comparable for similarity. It is also desirable for the metrics to be independent of any labeling choices for the lines (object identity) or of their orientation (sequence of their boundary points), so that they can apply to more generalized cases when such information is unavailable. We describe the methodology for accomplishing this goal along with a justification of why a selection of different measures would fail to satisfy the stated requirements.

We apply the concept of *largest non-containing circle* to derive only two of the four possible distances: (1) the *minimum distance* and (2) the *maximum distance* between the boundaries of two lines. In the case of *boundary closeness* the definition of those distances is restricted by the condition that *the sets of boundary points between which minimum and maximum distances are derived must be mutually exclusive*.

The approach is as follows: (1) select arbitrarily one boundary point P of a line; (2) create the *LNCC* circle from P with respect to the boundary of the other line and measure its radius; (3) repeat step 2 for the remaining boundary point Q of the initially selected

line to create a second *LNCC* and measure its radius; (4) select the smallest of the two *LNCC*s and assign the value of its radius to the variable D_{min}; (5) mark the two boundary points that define the radius of the selected *LNCC* as reserved; and (6) calculate the distance from the remaining point of the initially selected line to the remaining point of the other line and assign the value of that distance to the variable D_{max}. This theoretical description of the methodology can be concisely expressed by means of the following pseudo code (Figure 8).

This algorithm ensures that the value assigned to the variable D_{min} is indeed the smallest realizable distance between two boundary points. The distance between the remaining set of boundary points is called the maximum boundary-boundary distance. Although the maximum distance is always larger than the minimum distance, it is not necessarily the largest realizable distance between two boundary points. The approach suggested results in a set of boundary-boundary metrics that are always conceptually equivalent for any two configurations consisting of two simple lines, and can directly be compared for similarity (Figure 9).

Such a conceptual equivalence of the measures in different line-line configurations would not hold if the condition of disjoint sets of boundary points, for which the minimum and maximum distances are derived, was dropped. In that case the minimum distance would still be derived as before, but the maximum distance would be produced by applying the concept of the smallest containing circle. This method, however, yields problematic measures, as shown in the configurations in Figure 10. While they are quite different metrically with respect to their boundary closeness, their minimum and maximum boundary-boundary distances would imply that they are metrically very similar. The discrepancy occurs because the calculations of minimum and maximum measures, based on the concepts of *LNCC*s and *SCC*s, assume that such distances are captured between contiguous geometric objects. The boundary of a line, however, is unique as it consists of two disconnected points; therefore, relying only on definitions 1-4 leads to a comparison of maximum distances that represent conceptually different quantities.

A similar problem would arise if one chooses to calculate all four possible distances between boundary points (when all boundary points are disjoint) by applying the concepts of *largest non-containing circle* and *smallest containing circle* from any boundary point of one line to the boundary points of the other line. It would create an undesirable reliance on the chosen labeling scheme for the points (Figure 11). This justifies the choice of selecting two global measures that can always be compared reliably and produce a similarity score indicative of the actual similarity of the spatial configurations, without any reliance on labeling choices.

The two remoteness measures (D_{min} and D_{max}) are actual distances and, therefore, scale-dependent. For instance, a scaling by a factor of two would make any two lines appear to be twice as much remote. The normalization by the length of an arbitrarily selected line should be avoided, however, because it may distort similarity inferences (Figure 12). We choose to normalize by the sum of the lengths of the two lines, thus distinguishing two boundary closeness measures: (1) the *minimum boundary closeness* ($BC_{\min}$) as the ratio between the minimum distance (D_{min}) and the sum of the lengths of the two lines (Equation 5a), and (2) the *maximum boundary closeness* ($BC_{\max}$) as the ratio between the maximum distance (D_{max}) and the sum of the lengths of the two lines (Equation 5b).

The boundary closeness measures apply to all 33 topological relations (Figure 1).The possible values for both boundary closeness measures are greater than or equal to 0 without an upper bound. The boundary closeness measure $BC_{\min}$ assumes a value of 0 for relations 24-33 and both $BC_{\min}$ and $BC_{\max}$ become 0 in relations 21-23, thus adding essentially no refinement to those relations.

$$BC_{\min} = \frac{D_{\min}}{length(L_i) + length(L_j)} \quad \textbf{with} \quad i, j \in \{1,2\}, i \neq j \qquad \textbf{(5a)}$$

$$BC_{\max} = \frac{D_{\max}}{length(L_i) + length(L_j)} \quad \textbf{with} \quad i, j \in \{1,2\}, i \neq j \qquad \textbf{(5b)}$$

4.2 Interior Closeness

The *interior closeness* is a measure of the remoteness of one line's interior L_1° from the interior of another line L_2°. There are four distances that can be realized between interiors, a minimum and a maximum distance from each line's interior to the other line's interior. Those distances are derived based on the definitions of *largest non-containing buffer* and *smallest containing buffer*. The minimum distances are always symmetric and, therefore, equal, but the maximum distances may differ (Figure 13). We choose the larger of the two maximum distances as the measure of the maximum interior-interior closeness.

The process of retrieving the minimum and maximum distances between two interiors (Figure 13) consists of the following steps: (1) select arbitrarily one line, say L_1; (2) create the *LNCB* and the *SCB* from the interior of L_1 with respect to the interior of L_2; (3) assign the offset distance of the $LNCB(L_1^\circ, L_2^\circ)$ to the variable D_{min} (*minimum distance*); (4) create the *SCB* from the interior of the other line L_2 with respect to the interior of the first line L_1 (5) compare the radii of the two SCB buffers—$SCB(L_1^\circ, L_2^\circ)$, $SCB(L_2^\circ, L_1^\circ)$—and select the *SCB* with the largest radius; (6) assign the value of that *SCB* to the variable D_{max} (*maximum distance*). Although those steps describe the theory of the method, in practical computer representations lines are implemented as series of connected points and, therefore, the actual code for the interior-interior distances (Figure 14) can be based on a function that samples line points with a specified step (i.e., function *pointSequence*).

This procedure results in a set of interior-interior metrics that are always conceptually equivalent for any two configurations consisting of two simple lines and can be directly compared for similarity. It also avoids any dependence on the labeling scheme chosen for the lines (the order of the lines as indicated by their labeling is irrelevant). To maintain that independence we choose to normalize the lengths by the sum of the lengths of the two lines, similarly to the boundary closeness measures. The two normalized closeness measures for interiors are: (1) the *minimum interior closeness* (IC_{min}) as the ratio between the minimum distance (D_{min}) and the sum of the lengths of the two lines (Equation 6a),

and (2) the *maximum interior closeness* (BC_{max}) as the ratio between the maximum distance (D_{max}) and the sum of the lengths of the two lines (Equation 6b).

$$IC_{\min} = \frac{D_{\min}}{length(L_i) + length(L_j)} \quad \textbf{with} \quad i, j \in \{1,2\}, i \neq j \tag{6a}$$

$$IC_{\max} = \frac{D_{\max}}{length(L_i) + length(L_j)} \quad \textbf{with} \quad i, j \in \{1,2\}, i \neq j \tag{6b}$$

The interior closeness measures apply to all 33 topological relations (Figure 1).The possible values for both interior closeness measures are greater than or equal to 0 without an upper bound. The interior closeness measure $IC_{\min}$ is practically useful only for distinguishing among disjoint relations, since it assumes a value of 0 for all other relations. Both $IC_{\min}$ and $IC_{\max}$ become 0 when the lines are equal.

4.3 Interior-Boundary Closeness

The *interior-boundary closeness* describes the remoteness of one line's interior L_2° to another line's boundary ∂L_1. Previous closeness measures were derived between geometric objects of the same type, therefore, they have a global character because they are not specific to a single line and they provide an overall indication of the remoteness between two boundaries or two interiors for the whole configuration consisting of two simple lines. On the other hand, boundary-interior closeness measures are derived for geometric objects of different type (i.e., interiors and boundaries); therefore, they are specific to a single line. For example, the closeness of the interior of L_1 to the boundary of L_2 is different from the closeness of the interior of L_2 to the boundary of L_1 (Figure 15).

The process of retrieving the minimum and maximum distances between one line's interior and another line's boundary consists of the following steps (Figure 16): (1) select arbitrarily one line, say L_1; (2) create the *LNCB* and the *SCB* from the interior of L_1 with respect to the boundary of L_2; and (3) assign the offset distance of the $LNCB(L_1^\circ, \partial L_2)$ to the variable D_{min} (*minimum distance*) and the offset distance of the $SCB(L_1^\circ, \partial L_2)$ to the variable D_{max} (maximum *distance*). Alternatively, we may choose to calculate these

measures as the distances from the boundary of L_2 to the interior of L_1 similarly to the methodology given for boundary closeness. The end result will be the same. The pseudo code for the interior-boundary distances (Figure 17) is similar to the code for the interior-interior case. The distances here, however, are derived between 2 boundary points of one line, and a sampled set of the interior points of the other line.

The minimum and maximum distances need to be normalized by a quantity that will yield scale-independent measures. We choose to normalize an interior-to-boundary distance by dividing it through the length of the line to which the interior belongs. The two normalized interior boundary closeness measures for each line are (1) the *minimum interior-boundary closeness* (IBC_{min}) as the ratio between the minimum distance ($D_{\min}(L_i^{\circ}, \partial L_j)$) and the length of L_i (Equation 7a), and (2) the *maximum interior-boundary closeness* (IBC_{max}) as the ratio between the maximum distance ($D_{\max}(L_i^{\circ}, \partial L_j)$) and the length of L_i (Equation 7b).

$$IBC_{\min}(L_i) = \frac{D_{\min}}{length(L_i)} \quad \textbf{with} \quad i \in \{1,2\} \tag{7a}$$

$$IBC_{\max}(L_i) = \frac{D_{\max}}{length(L_i)} \quad \textbf{with} \quad i \in \{1,2\} \tag{7b}$$

The interior-boundary closeness measures apply to all 33 topological relations (Figure 1). The possible values for both interior closeness measures are greater than or equal to 0 without an upper bound. The interior-boundary closeness measure $IBC_{\min}$ will become 0 if a line's interior intersects a boundary point of the other line. Both $IC_{\min}$ and $IC_{\max}$ become 0 for a line when its interior intersects both boundary points of the other line.

5. Representational Structures for Metric Details

To become practically useful, the metrics need to be encoded in complete and efficient representational structures. Completeness requires that all applicable measures for a spatial configuration between two lines be encoded, whereas efficiency requires that the form of representation be organized such that it can be easily understood. In the context of efficiency, it is also desirable to combine the topological and metric properties for a

scene into a single form of representation. An important distinction between the two classes of measures is that all of the eight closeness measures apply for any line-line configuration, whereas the number of splitting ratios is variable. It is 0 in the case of disjoint lines, while there is theoretically no upper limit in the case of arbitrarily complex relations with multiple component intersections. Furthermore, the two types of measures are conceptually different since one of them depends on the interaction of intersecting components, whereas the other concerns distances between disjoint object parts. Therefore, we choose two different schemes for storing splitting and closeness measures. Both of these schemes are integrated into previously developed representational structures for storing topological details, thus maintaining compatibility with previous work.

5.1 Representation of Splitting Ratios for Arbitrarily Complex Relations

For a complex configuration, with many intersections of the same or different type between two simple lines, all of the splitting ratios may apply one or multiple times, depending on the number of existing components. In this case one needs to develop a representational structure, such that it allows a smooth transition from the representation of simple to arbitrarily complex line-line relations. We base our representation technique on the concept of the *classifying invariant* (Clementini and di Felice 1998). The classifying invariant captures in a matrix the values of the topological properties needed to describe a scene involving two simple intersecting lines. In this section we extend this matrix to include the splitting ratios in addition to the topological invariants. We call the resulting matrix a *metrically-enhanced classifying invariant*.

The general structure of the classifying invariant for two simple lines, denoted as $Cl(L_1, L_2)$, is a matrix of four columns and *m* rows where $m \in \mathbb{N}^+$ (Table 2). Each row describes an interior-interior, interior-boundary, or boundary-boundary intersection between the two lines. These are the most essential intersections since they determine how the two lines interact. If all these intersections are empty then the lines are simply disjoint and no splitting measure or topological property applies. The four columns give the qualitative values of several topological properties, which are the *intersection*

sequence $S(L_2)$, the *collinearity sense* CS, the *intersection type* T, and the *link orientation* LO_{L2}. The generic entry k_i represents the label of the intersection component. This set of topological invariants has been proven sufficient and necessary in order to establish topological equivalence with any configuration for a pair of simple lines.

The intersection sequence describes the order in which the various components of the intersections occur. One first follows line L_1 from its first point and assigns numeric labels to the intersections until the last point is reached. The *intersection sequence* is then the sequence of numbers established by traversing line L_2 and recording the labels that were previously assigned to L_1. For example, the intersection sequence in Figure 18 is [0,1,3,2].

First establishing a clockwise orientation and then recording at the intersection node the sequence of incoming and outgoing arcs, starting from the boundary of one line, defines the intersection type. For instance, for Intersection 0 in Figure 18 the sequence is $< i_1, i_2, o_1, o_2 >$, assuming that we record the arcs starting from the incoming arc of L_1. The choice of the sequence's starting arc is irrelevant, since the sequence remains invariant under cyclic permutations. The number of arcs in the sequence can be less than four. For example, for Intersection 2 in Figure 18 the sequence is $< i_1, o_1, i_2 >$. Although the choice of the first arc to start the sequence is arbitrary, their order must be preserved as it implicitly stores information about whether the intersections are crossing or touching (Herring 1991).

For 1-dimensional intersections, the collinearity sense distinguishes whether the segments that make these components are traversed following the same or the reverse orientation in the two lines. If the former is true the value of the collinearity sense is 1, if the latter holds it is -1, whereas for 0-dimensional intersections it takes the value of 0. For instance, since the 1-dimensional Intersection 2 (Figure 18) is traversed in reserve orientation, its collinearity sense is –1; Intersections 0, 1, and 3, however, are 0-dimensional, therefore, their collinearity sense is 0.

The link orientation depends on the notion of a link, which is the part of line L_2 located between two consecutive—according to the intersection sequence—intersections (h,k). If the cycle obtained by traversing the link $L_2(h,k)$ and coming back to *h* traversing the line L_1 has a clockwise orientation, the link orientation value becomes *r* (i.e., right), otherwise *l* (i.e., left). For example, the link orientation for the pair of intersections (0,1) is *l* whereas for the pair (3,2) it is *r* (Figure 18). Since the link orientation invariant depends on two consecutive intersections, its value is undefined for the last row of the classifying invariant matrix. Figure 18 demonstrates how these concepts apply for a complex configuration of two simple lines and the construction of its classifying invariant (Table 3).

Besides the intersection sequence, which records the order at which intersections occur and assigns a number and a row to each intersection of interest, an examination of the three remaining invariants suggests a one-to-one correspondence with the three splitting ratios. The first correspondence is between the collinearity sense and the line alongness measure. Instead of using 1 and -1 to denote whether the segments along the common interior have the same or reverse orientation, respectively, we use a positive or negative value between 0 and 1, equal to the line alongness ratio. For 0-dimensional intersections, the value of the collinearity sense remains 0 and represents the extreme case of the line alongness measure, where the common segment degenerates to a single point.

The second correspondence is between the intersection type and the interior splitting ratio. The encoding sequence of the arcs can be extended with numeric information that relates each arc to an interior splitting ratio measure between 0 and 1. The interior splitting ratio for each arc is derived by dividing the length of the arc through the length of the line that contains it. The length of each arc is taken equal to the length of the line between the intersection that is being recorded and the immediate previous intersection or starting boundary of the line for incoming arcs (i.e., arcs entering the intersection), or the immediate next intersection or finishing boundary of the line for outgoing arcs (i.e., arcs exiting the intersection). The labels of arcs (i.e., i_1, i_2, o_1, o_2) must also be recorded, because they may occur in different orders, depending on the intersection type, and such information must be maintained in order to distinguish different topological relations.

The third correspondence between a topological invariant and a splitting measure is between the link orientation and the exterior splitting ratio. The link orientation describes the orientation of the circular section between two consecutive intersections. This circular section, however, always creates a bounded exterior. Therefore, one can combine the link orientation and the exterior splitting measure by recording only the value of the exterior splitting ratio for each bounded exterior component. The value is preceded by a *plus* sign if the link orientation is clockwise and by a *minus* sign if it is counter-clockwise. The topological configuration between two simple lines (Figure 18) is annotated with metric details (Figure 19). Table 4 displays the matrix for this scene's *metrically-enhanced classifying invariant*.

The *metrically-enhanced classifying invariant* matrix stores both topological and metric information about a spatial configuration between two simple lines. For the interior splitting ratios the topological information is stored explicitly along with the metric, whereas for the line alongness and the exterior splitting ratios it is encoded in the signs that precede the ratios. A closer examination of this matrix reveals that some of the line splitting measures may be repeated in two consecutive rows. The repetition happens because an outgoing arc from a previous intersection may change its label and become an incoming arc in the next; however, the length of the arc remains the same. The small redundancy in number storage is a ramification resulting from the original structure of the classifying invariant matrix, and not a shortcoming inherent to the extension with metric details. In addition, the redundancy may be used as a sentinel for consistency maintaining and error-checking for the metric description of a spatial scene.

Another characteristic of the classifying invariant that affects the metrically enhanced classifying invariant is that its values may differ depending on labeling choices. Since we are concerned with scenes with two lines, where each line has two boundary points, then, depending on the labeling scheme, there can be 2^3 or 8 constructed classifying invariants. Unlike the 9-intersection, where one matrix may correspond to two or more topologically different configurations, in the classifying invariant model two topologically equivalent representations may be represented by different matrices (Figure 20).

This one-to-many correspondence between a spatial configuration and a set of possible classifying invariants for that configuration does not prevent us from representing the topology of a scene or even from obtaining metric enhancements. It does have ramifications, however, when comparing scenes for topological equivalence and, consequently, for metric similarity. In such a case, deriving the classifying invariant for two spatial scenes might only be the first step in assessing topological equivalence. If the resulting matrices are different, a second step is required. This step consists of finding a relabeling of lines and their boundary points such that the classifying invariant of one scene becomes identical to the other one. A brute-force approach is possible; however, more efficient algorithms have been developed to address this issue (Clementini and di Felice 1998). Only if the scenes are found identical during the second step is it possible to proceed to the calculation of the splitting ratios and the comparison of the corresponding measures so as to yield a similarity ranking.

5.2 Representation of Closeness Measures

For a spatial configuration consisting of two simple lines we defined eight closeness metrics, forming four pairs of minimum and maximum closeness measures: one describing the closeness of the boundaries, one describing the closeness of the interiors, and two describing the closeness between the interiors and the boundaries (one pair for each line). Each pair of minimum and maximum closeness measures can be stored in its corresponding cell of a 4-intersection matrix (Egenhofer and Franzosa 1991). The 4-intersection is a subset of the 9-intersection, excluding considerations about the exteriors of the lines. We call the metrically-annotated 4-intersection matrix the *closeness matrix* for two simple lines. The topological configuration of two disjoint lines (Figure 21) is now annotated with metric details (Table 5).

Similarly to the classifying invariant, the closeness matrix for a spatial configuration is not unique, because the interior-boundary closeness measures may differ depending on the labels of the lines. For instance, the metrics represented by the interior-boundary closeness measures of Table 5 would be transposed if the labels of the lines in Figure 21 were switched. This dependency is undesirable, because it prohibits a reliable comparison

of the metrics of two scenes for similarity (i.e., a sketch-query vs. a database configuration). For every topological configuration besides the disjoint relation, the problem is solved by trying the different permutations of a classifying invariant for one scene and checking whether one of the permutations would be identical to the classifying invariant of the other scene. If that is indeed the case, then the splitting ratios as well as the closeness measures can consequently be reliably compared for similarity.

In the case of disjoint relation, however, the classifying invariant is an empty matrix and only the closeness measures apply. The solution for similarity comparisons in this scenario consists of the following three steps: (1) perform an arbitrary labeling on the lines in both the sketched and the database scene; (2) estimate the closeness matrices for both scenes and compare them for similarity; (3) invert the labels of the lines in the database scene and repeat the second step; and (4) from the two similarity scores select the highest as the actual closeness similarity score between the two scenes (Figure 22). The second and third steps imply the existence of some sort of similarity. As a very crude estimate of similarity we choose to derive the *difference table* (absolute values) of two closeness matrices. Obviously, the smaller the sum of the elements in the difference table, the more similar the two configurations will be.

The presented methodology ensures that the comparison of the closeness measures will yield a similarity score that is representative of the actual closeness between the lines and not some exaggerated and misleading value due to incompatible labeling for the two scenes. The labeling repercussions discussed pertain in essence to problems of object identity. Such problems were not explicitly addressed in previous work on metric enhancements, which focused on splitting ratios and closeness measures for topological configurations involving a region and a line (Egenhofer and Shariff 1998) or topological configurations involving two regions (Shariff 1996). In the case of a line vs. a region, the problem is not apparent or even relevant, because there already exists an *a-priori* labeling of the objects due to their different types (i.e., line vs. region) and, hence, the ability to uniquely determine them. However, the problem persists in the measures developed for region-region configurations as both objects are of the same type. Perhaps it is assumed that object identity is resolved through semantic annotations that demarcate the objects'

nature (i.e., one line is denoted as a river and another as a highway). For an unambiguous comparison, however, such annotations need to be present in both the sketch query and the database representations. In our work, we provided extensions that allow the employment of metric measures even in the absence of knowledge about the identity of the objects.

6. Conclusions and Future Work

This paper introduced a computational model that extends topological information about binary relations between simple lines based on the 9-intersection with metric information in terms of *splitting ratios* and *closeness measures*. Three splitting ratios refine 30 of the 33 topological relations distinguished by the 9-intersection: *line alongness*, which applies for 19 of the 33 relations and describes lengths of common paths; *interior splitting,* which applies for 30 relations and deals with the partitioning of lines through intersections; and *exterior splitting*, which can be applied to 23 relations, addressing areas enclosed by two lines with two or more common components. They all take values between 0 and 1 and grow linearly with the size of the intersection component that they measure. Closeness measures describe how far apart disjoint object parts are and their values are positive numbers with no upper bound. Three classes of closeness measures were derived: *boundary closeness*, *interior closeness*, and *interior-boundary closeness*. All closeness measures apply to all 33 relations since there is no constraint on their usage; however, values of 0 add in essence no refinement. Closeness measures are particularly useful in the case of disjoint relations where the splitting ratios are not applicable.

To encode splitting ratios we converted Clementini's and di Felice's (1998) matrix, which stores values of topological properties for detailed topological relations between lines, into the *metrically-enhanced classifying invariant*. Each of the values of topological properties that describe an intersection has a natural correspondence with one of the splitting ratios. This allows capturing both topological and metric information for a scene, in a single representation. Closeness measures are stored in the *closeness matrix*, which is a table derived from the 4-intersection. Each cell of this matrix contains an ordered pair of numbers corresponding to the minimum and maximum measures.

The metric refinements may complement both coarse and detailed relations in the description of a spatial scene adding progressively more information about the scene. Metric refinements need only be calculated when necessary. An example of such a case is spatial similarity retrieval, where metric details may be used in order to sort query results when a user requests a scene by sketching (Egenhofer 1997). If the database returns more than one scene where objects have the same topology as those in the drawn sketch, metric details will enable an ordering of the most similar scene to less similar scenes. In addition to similarity retrieval, metric information can be used for preprocessing sketch queries in order to correct unintended errors (i.e., overshoots, undershoots) by the users and restore the proper topology for a query. Such corrections may also be facilitated by making semantic inferences from the annotation of the objects in the sketch. For instance, a user may draw a parcel incorrectly, having two edges of the parcel only slightly intersecting, instead of exactly meeting at the boundary. Absence of a topological match for this query may trigger a topology restoration operation, where metric details help detect small inaccuracies (Ubeda and Egenhofer 1997). The query with the new topology may now be submitted again to the database for retrieval of a new topological match. The exact details of such operations and the theory to support them are topics for future work.

An important distinction between closeness measures and splitting ratios is that the former may be compared for scenes that have identical or similar topology, whereas the latter can only be compared for scenes with identical topology. This difference is a result of the closeness measures having a more global character since they deal with disjoint object parts, whereas the splitting ratios operate at the component level and are heavily dependent on the labels of the lines and their boundary points. It would be desirable to compare splitting ratios for scenes of different topologies. An additional topic of interest that is germane to this problem is the establishment of a set of invariants that create a one-to-one correspondence for a scene and its classifying invariant matrix. Finally, a comprehensive similarity model for spatial configurations is needed so that it can simultaneously consider topological, metrical, and directional information, determine when finer levels of detail need to be employed in the similarity assessment, and also take under consideration semantic correlations and weighting issues.

7. Acknowledgments

This work was funded partially by the National Science Foundation under grant numbers IIS-9970123 and EPS-9983432; the National Geospatial-Intelligence Agency under grant numbers NMA201-00-1-2009, NMA201-01-1-2003, and NMA401-02-1-2009; and the National Institute of Environmental Health Sciences, NIH, under grant number 1 R 01 ES09816-01. Konstantinos Nedas was partially supported by a University of Maine Graduate Research Assistantship. Dominik Wilmsen has been supported by a Fulbright Fellowship. We appreciate feedback to an earlier version of the splitting ratios (Nedas and Egenhofer 2004), presented at FLAIRS 2004.

8. References

BLASER, A. (2000) Sketching Spatial Queries. Ph.D. Thesis. Department of Spatial Information Science and Engineering, University of Maine, Orono, Maine, USA.

CHANG, S.-K. and JUNGERT, E. (1996) Symbolic Projection for Image Information Retrieval and Spatial Reasoning, Academic Press, New York, NY.

CLARKE, B. (1981) A Calculus of Individuals Based on "Connection". Notre Dame Journal of Formal Logic 22(3): 204-218.

CLEMENTINI, E. and DI FELICE, P. (1998) Topological Invariants for Lines. IEEE Transactions on Knowledge and Data Engineering 10(1): 38-54.

CRISTANI, M. (2003) Reasoning about Qualitative Relations between Straight Lines. Technical Report RR 06/2003, Department di Informatica, University of Verona, Verona, Italy.

EGENHOFER, M. and HERRING, J. (1990) Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases. http://www.spatial.maine.edu/~max/9intReport.pdf. Technical Report, Department of Surveying Engineering, University of Maine, Orono, Maine, USA.

EGENHOFER, M. and FRANZOSA, R. (1991) Point-Set Topological Spatial Relations. International Journal of Geographical Information Systems 5(2): 161-174.

EGENHOFER, M. (1994) Definitions of Line-Line Relations for Geographic Databases. Data Engineering 16(11): 479-481.

EGENHOFER, M. and FRANZOSA, R. (1995) On The Equivalence of Topological Relations. International Journal of Geographical Information Systems 9(2): 133-152.

EGENHOFER, M. and MARK, D. (1995) Naive Geography. in: COSIT '95, Semmering, Austria. W. Kuhn (ed.), Lecture Notes in Computer Science, Vol. 988, pp. 1-15, Springer-Verlag September, 1995.

EGENHOFER, M. (1997) Query Processing in Spatial-Query-by-Sketch. Journal of Visual Languages and Computing 8(4): 403-424.

EGENHOFER, M. and SHARIFF, R. (1998) Metric Details for Natural-Language Spatial Relations. ACM Transactions on Information Systems 16(4): 295-321.

GOTTFRIED, B. (2003) Tripartite Line Tracks - Bipartite Line Tracks. in: Proceedings of KI 2003: Advances in Artificial Intelligence, Lecture Notes in Artificial Intelligence 2821, pp. 535-549, B. Neumann (ed.) Springer-Verlag, Berlin.

HERNÁNDEZ, D. (1994) Qualitative Representation of Spatial Knowledge, Springer-Verlag, New York.

HERRING, J. (1991) The Mathematical Modeling of Spatial and Non-Spatial Information in Geographic Information Systems. in: Cognitive and Linguistic Aspects of Geographic Space, pp. 313-350, A. Frank (ed.) Kluwer Academic Publishers, Dordrecht.

ISLI, A. (2002) A Ternary Relation Algebra of Directed Lines. Technical Report FBI-HH-M-313/02, Fachbereich Informatik, Universitaet Hamburg.

KRAY, C. and BLOCHER, A. (1999) Modeling the Basic Meanings of Path Relations. in: Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI), pp. 384-393 Morgan Kaufmann.

KRÜGER, A. and MAAß, W. (1997) Towards a Computational Semantics of Path Relations. in: Workshop on Language and Space at the 14th National Conference on Artical Intelligence (AAAI 97), Providence, RI.

LAURINI, R. and THOMPSON, D. (1992) Fundamentals of Spatial Information Systems, Academic Press, San Diego, CA.

LYNCH, K. (1960) The Image of a City, MIT Press, Cambridge, MA.

MARK, D. (1992) Counter-Intuitive Geographic Facts: Clues for Spatial Reasoning at Geographic Scales. in: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Lecture Notes in Computer Science, Vol. 639, pp. 305-317, U. Formentini (ed.) Springer.

MARK, D. and EGENHOFER, M. (1994) Calibrating the Meanings of Spatial Predicates from Natural Language: Line-Region Relations. in: Sixth International Symposium on Spatial Data Handling, Edinburg, Scotland. R. Healey (ed.), pp. 538-553.

MORATZ, R., RENZ, J., and WOLTER, D. (2000) Qualitative Spatial Reasoning about Line Segments. in: ECAI 2000. Proceedings of the 14th European Conference on Artificial Intelligence, W. Horn (ed.) IOS Press, Amsterdam.

NEDAS, K. and EGENHOFER, M. (2004) Splitting Ratios: Metric Details of Topological Line-Line Relations. in: The 17th International FLAIRS Conference, Miami Beach, FL.

RANDELL, D., CUI, Z., and COHN, A. (1992) A Spatial Logic based on Regions and Connection. in: Proceedings 3rd International Conference on Knowledge Representation and Reasoning, San Mateo, CA, pp. 165-176, Morgan Kaufmann.

REGIER, T. (1995) A Model of the Human Capacity for Categorizing Spatial Relations. Cognitive Linguistics 6(1): 63-88.

SCHLIEDER, C. (1995) Reasoning about Ordering. in: Spatial Information Theory: A Theoretical Basis for GIS, Lecture Notes in Computer Science, Vol. 988, pp. 341-350, W. Kuhn (ed.) Springer-Verlag, Berlin.

SCHLIEDER, C. (1996) Qualitative Shape Representation. in: Geographic Objects with Indeterminate Boundaries, pp. 123-140, P. Burrough (ed.) Taylor & Francis, London.

SHARIFF, R. (1996) Natural-Language Spatial Relations: Metric Refinements of Topological Properties. Ph.D. Thesis. Department of Spatial Information Science and Engineering, University of Maine, Orono, ME, USA.

SIMONS, P. (1987) Parts: A Study In Ontology, Clarendon, Oxford.

STEVENS, A. and COUPE, P. (1978) Distortions in Judged Spatial Relations. Cognitive Psychology 10: 422-437.

TALMY, L. (1983) How Language Structures Space. in: Spatial Orientation: Theory, Research, and Application, pp. 225-282, L. Acredolo (ed.) Plenum Press, New York.

TVERSKY, B. (1981) Distortions in Memory for Maps. Cognitive Psychology 13: 407-433.

UBEDA, T. and EGENHOFER, M. (1997) Correcting Topological Errors. in: Advances in Spatial Databases-Fifth International Symposium on Large Spatial Databases, SSD '97, Lecture Notes in Computer Science, Vol. 1262, pp. 283-297, A. Voisard (ed.) Springer-Verlag, Berlin.

Table 1: Area and length measures applied to the 9-intersections of two lines.

	L_2°	∂L_2	L_2^-
L_1°	$length(L_1^\circ \cap L_2^\circ)$	—	$length(L_1^\circ \cap L_2^-)$
∂L_1	—	—	—
L_1^-	$length(L_1^- \cap L_2^\circ)$	—	$area(L_1^- \cap L_2^-)$

Table 2: Representation of the general form of the classifying invariant in tabular form.

$S(L_2)$	CS	T	LO_{L_2}
k_0	$CS(k_0)$	$T(k_0)$	$LO_{L_2}(k_0, k_1)$
k_1	$CS(k_1)$	$T(k_1)$	$LO_{L_2}(k_1, k_2)$
...	...	...	...
...	...	...	$LO_{L_2}(k_{m-2}, k_{m-1})$
k_{m-1}	$CS(k_{m-1})$	$T(k_{m-1})$	-

Table 3: Representation of a classifying invariant in tabular form.

$S(L_2)$	CS	T	LO_{L_2}
0	0	(i_1, i_2, o_1, o_2)	l
1	0	(i_1, o_2, o_1, i_2)	r
3	0	(i_1, i_2, o_1, o_2)	r
2	-1	(i_1, o_1, i_2)	-

Table 4: Metrically enhanced classifying invariant matrix for the configuration in Figure 19.

$S(L_2)$	CS	T	LO_{L_2}
0	0	(i_1, i_2, o_1, o_2), (0.18, 0.10, 0.16, 0.23)	-0.05
1	0	(i_1, o_2, o_1, i_2), (0.16, 0.37, 0.17, 0.23)	0.14
3	0	(i_1, i_2, o_1, o_2), (0.23, 0.37, 0.15, 0.21)	0.11
2	-0.09	(i_1, o_1, i_2), (0.17, 0.23, 0.21)	-

Table 5: The closeness matrix for the two lines of Figure 21.

	L_2°	∂L_2
L_1°	$IC_{\min}$, $IC_{\max}$	$IBC_{\min}(L_1)$, $IBC_{\max}(L_1)$
∂L_1	$IBC_{\min}(L_2)$, $IBC_{\max}(L_2)$	$BC_{\min}$, $BC_{\max}$

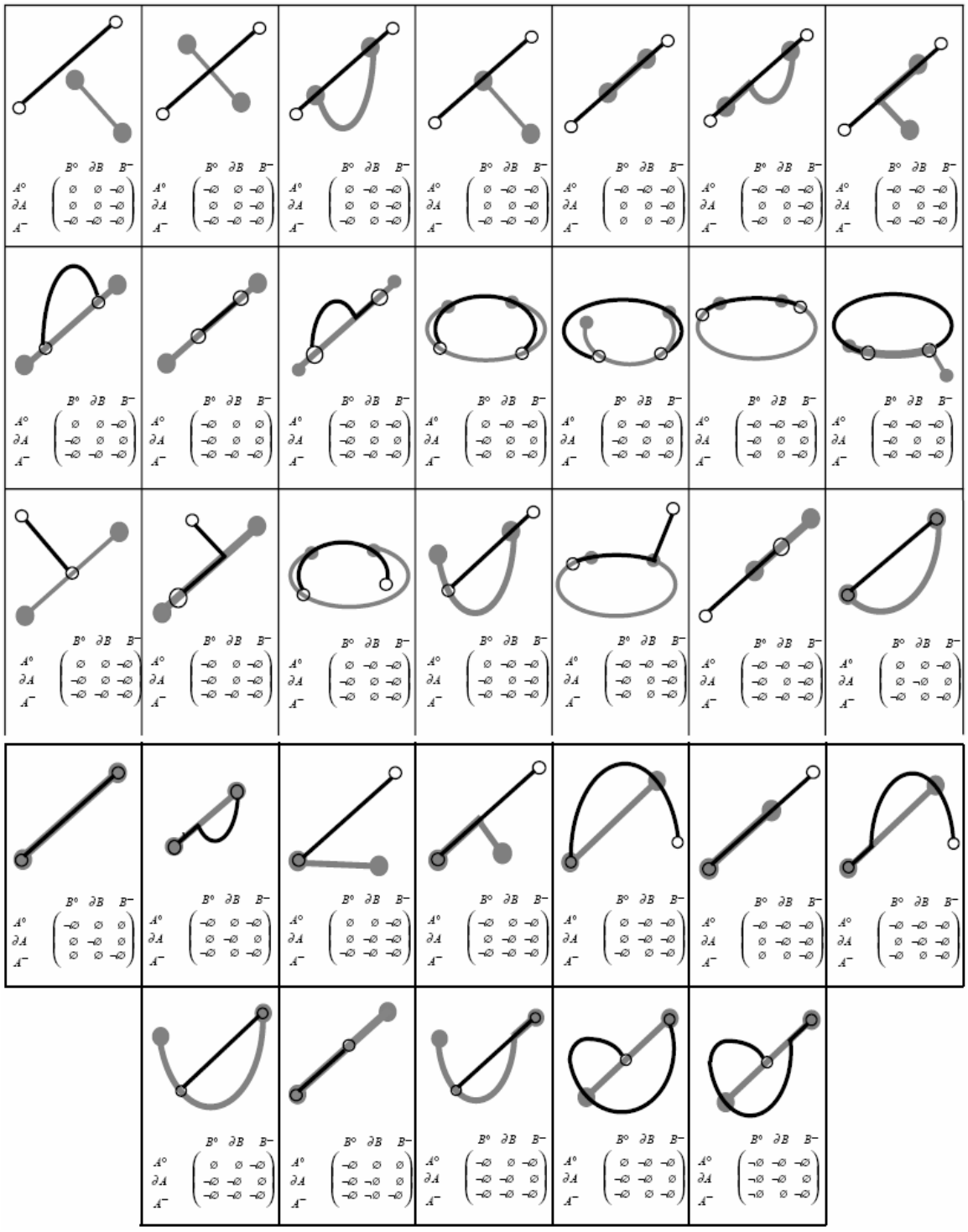
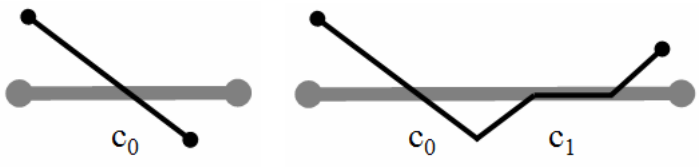


Figure 1: The set of 33 line-line relations realizable in $\mathbb{R}^2$ (Egenhofer 1994).



(a) (b)

Figure 2: Two configurations with different numbers of components.

(a) (b)

Figure 3: A global metric measure instead of one based on components would fail to add any refinement between these two topologically equivalent configurations.

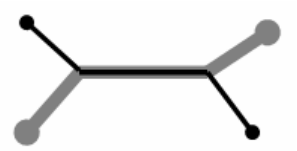
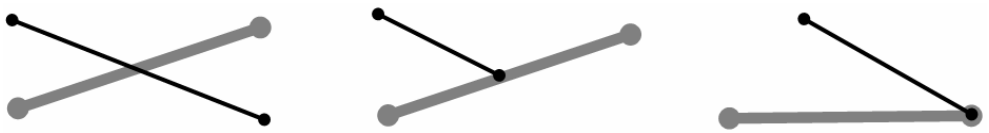


Figure 4: Line alongness: the common interior separates each line into parts of inner and outer segments (more complex configurations may have multiple components in the intersection of the line interiors).



(a) (b) (c)

Figure 5: Interior splitting: (a) one line's interior separates the other line's interior into two parts (the common interior could also be 1-dimensional); (b) one line's boundary separates the other line's interior into two parts; and (c) the extreme case where lines meet at a boundary and the split segments equal the lengths of the lines.

(a) (b) (c) (d) (e)

Figure 6: Exterior splitting: a bounded exterior formed by (a) two interior-interior intersections; (b) one boundary-interior and one interior-interior intersection; (c) one boundary-boundary and one interior-interior intersection; (d) one boundary-boundary and one boundary-interior intersection; and (e) two boundary-boundary intersections.

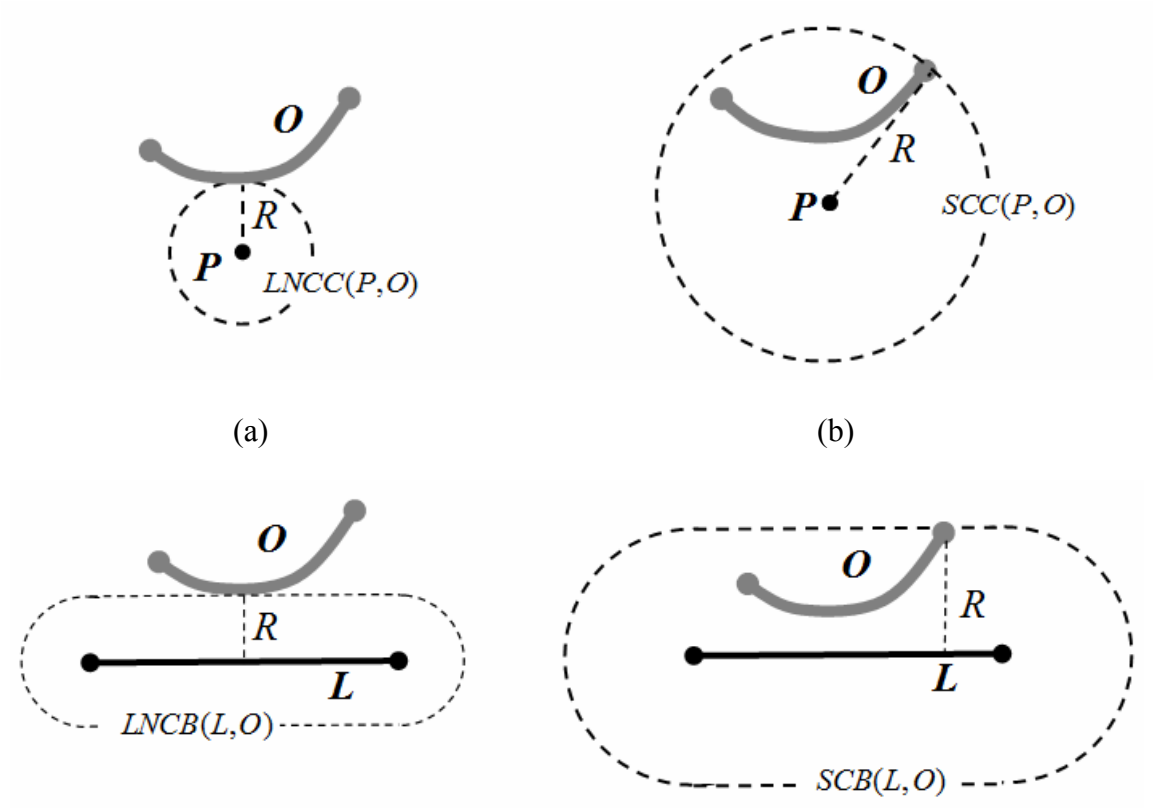


Figure 7: (a) The largest non-containing circle around *P* with respect to object *O*; (b) The smallest containing circle around *P* with respect to object *O*; (c) The largest non-containing buffer around line *L* with respect to object *O*; and (d) The smallest containing buffer around line *L* with respect to object *O*.

algorithm: boundaryBoundaryClosenessDistances
input: two lines: lineA and lineB
output: minimum boundary-boundary distance dMin, maximum boundary-boundary distance dMax
method:

```
    p1 := start(lineA)
    p2 := end(lineA)
    p3 := start(lineB)
    p4 := end(lineB)
    d1 := distance(p1,p3)
    d2 := distance(p1,p4)
    d3 := distance(p2,p3)
    d4 := distance(p2,p4)
    dMin := min(d1, d2, d3, d4)
    if dMin = d1 then dMax := d4
        elseif dMin = d2 then dMax := d3
            elseif dMin = d3 then dMax := d2
                else dMax := d1
    return dMin, dMax
end
```

Figure 8: Algorithm boundaryBoundaryClosenessDistances.

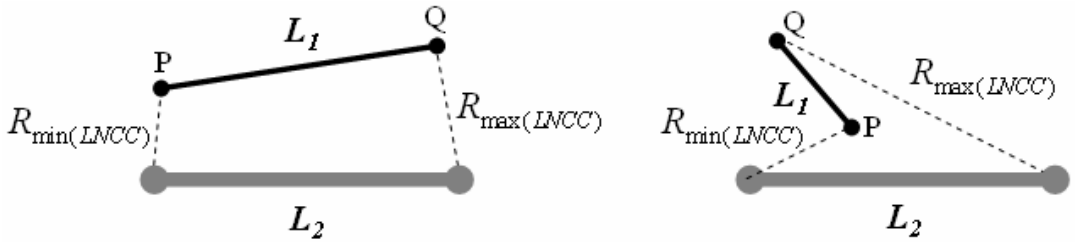


Figure 9: Boundary distances formed between disjoint sets of boundary points will yield metrics that are representative of the actual proximity of the boundaries of two lines.

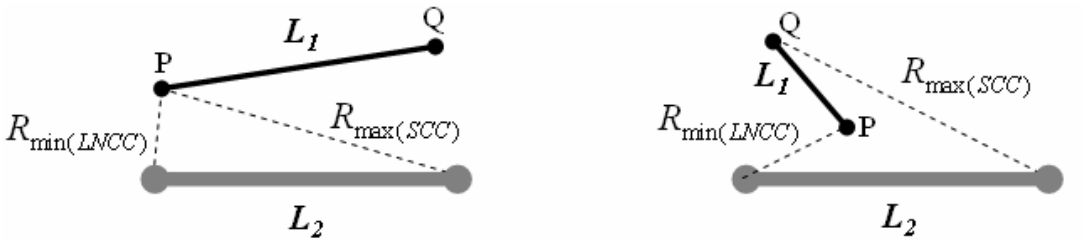


Figure 10: Two metrically-different configurations may appear very similar without the condition of disjoint sets of boundary points between which distances are derived.

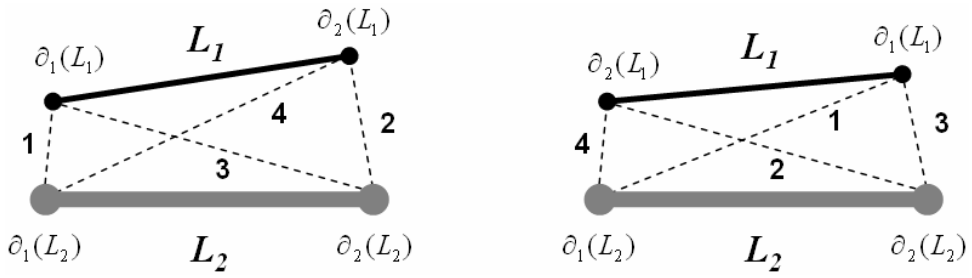


Figure 11: Considering all possible distances between boundary points creates label-dependent closeness measures. Even though the two configurations are very similar metrically, the different labeling schemes lead to comparisons (compared distances are indicated with the same integer numbers) that portray the two scenes as very different with respect to the metric property of boundary closeness.

(a) (b)

Figure 12: (a) a sketch query and (b) a database configuration. Normalizing by the length of an arbitrarily selected line (L_1 for Figure 12a and L_2 for Figure 12b) may distort similarity inference, thus making the two very similar configurations appear significantly different with respect to their boundary closeness.

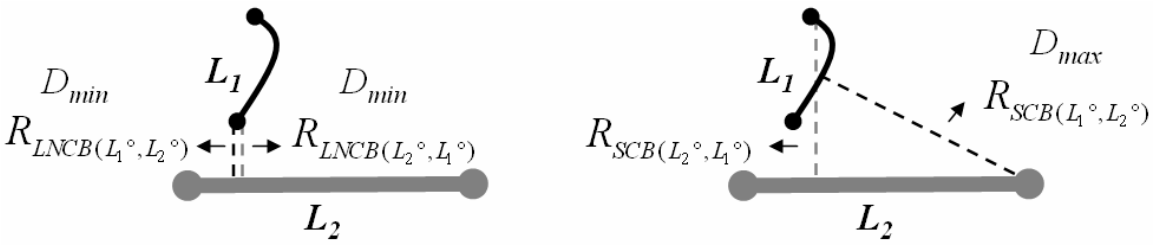


Figure 13: The process of deriving (a) the minimum and (b) the maximum interior distance. While the two minimum distances are symmetric, the same does not necessarily hold for the maximum distances.

```
algorithm:  interiorInteriorClosenessDistances
input:      two lines: lineA and lineB
output:     minimum interior-interior distance dMin, and maximum interior-interior
            distance dMax
method:
    pointsLineA := pointSequence(lineA)
    pointsLineB := pointSequence(lineB)
    for (i=1 to i<=pointsLineA.length)
        for (j=1 to j<=pointsLineB.length)
            d[i,j] := distance(pointsLineA[i], pointsLineB[j])
        end for
    end for
    dMin := min(d[i,j])
    dMax := max(d[i,j])
    return dMin, dMax
end
```

Figure 14: Algorithm interiorInteriorClosenessDistances.

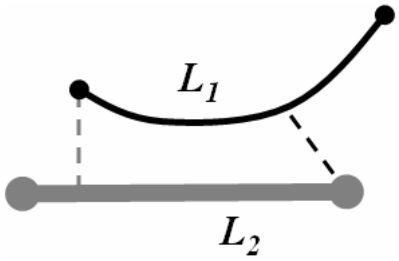


Figure 15: The shortest distance from the interior of L_1 to the boundary of L_2 is different from the shortest distance of the interior of L_2 to the boundary of L_1.

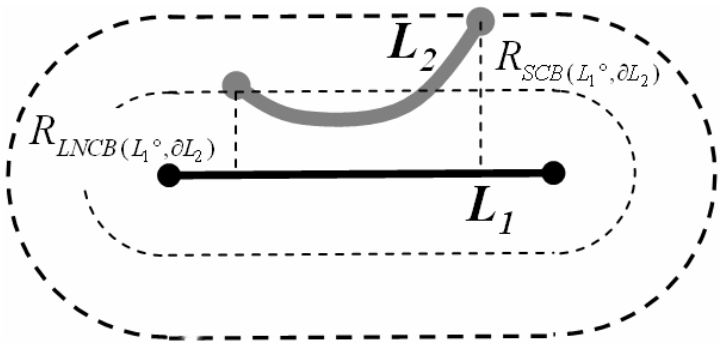


Figure 16: Retrieving the minimum and maximum distances from the interior of L_1 to the boundary of L_2.

```
algorithm:  interiorBoundaryClosenessDistances
input:      two lines: lineA and lineB
output:     minimum interior-boundary and maximum interior-boundary distances of
            lines A and B: dMinL1, dMaxL1, dMinL2, dMaxL2.
method:
    p1:= start(lineA)
    p2:= end(lineA)
    p3:= start(lineB)
    p4:= end(lineB)
    boundaryLineA := [p1, p2]
    boundaryLineB := [p3, p4]
    pointsLineA := pointSequence(lineA)
    pointsLineB := pointSequence(lineB)
    for (i=1 to i<=2)
        for (j=1 to j<=pointsLineA.length)
            dA[i,j] := distance(boundaryLineB[i], pointsLineA[j])
        end for
    end for
    dMinL1 := min(dA[i,j])
    dMaxL1 := max(dA[i,j])
    for (i=1 to i<=2)
        for (j=1 to j<=pointsLineB.length)
            dB[i,j] := distance(boundaryLineA[i], pointsLineB[j])
        end for
    end for
    dMinL2 := min(dB[i,j])
    dMaxL2 := max(dB[i,j]
    return dMinL1, dMaxL1, dMinL2, dMaxL2
end
```

Figure 17: Algorithm interiorBoundaryClosenessDistances.

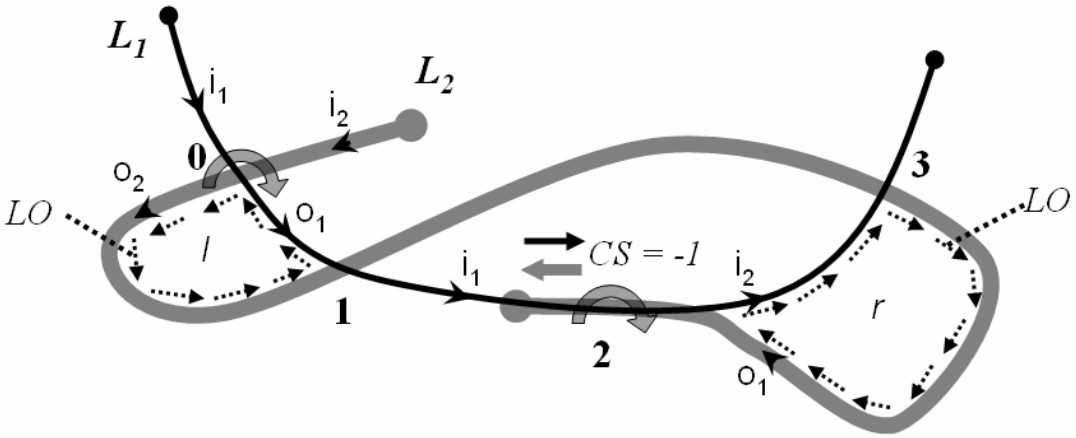


Figure 18: A complex configuration with four interior-interior intersection components formed by two simple lines.

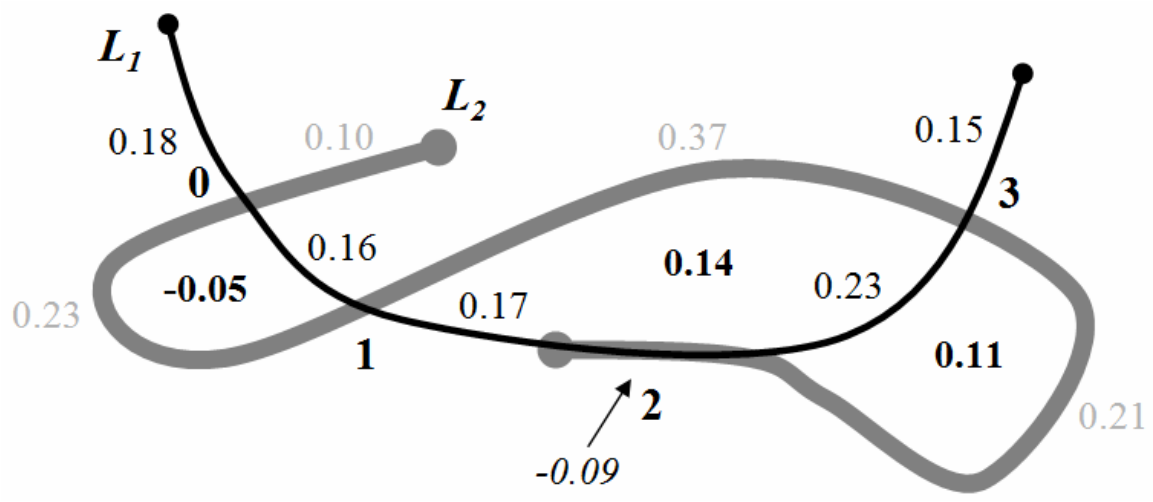
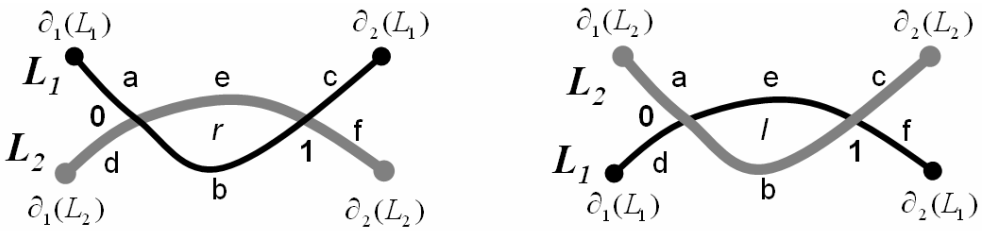


Figure 19: A complex configuration with four intersection components enhanced with metric details. Numbers in black represent the interior splitting ratios for line L_1; numbers in gray represent the interior splitting ratios for line L_2; numbers in **bold** black represent the exterior splitting ratio for each bounded exterior component; and the number in *italic* represents the line alongness ratio.



$S(L_2)$	CS	T	LO_{L_2}
0	0	a,e,b,d	r
1	0	b,e,c,f	0

$S(L_2)$	CS	T	LO_{L_2}
0	0	d,a,e,b	l
1	0	e,c,f,b	0

Figure 20: Two topologically equivalent configurations with different line labeling produce different classifying invariants (equivalent incoming and outgoing arcs are represented by the letters a,…, f).

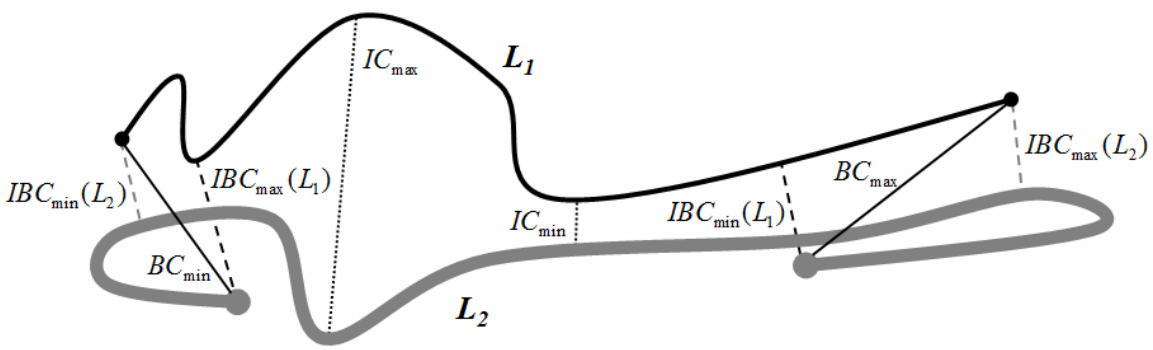


Figure 21: A configuration of two disjoint lines enhanced with metric details of closeness measures. Continuous segments represent distances derived for boundary closeness measures; dotted segments represent distances derived for interior closeness measures; black and gray dashed segments represent distances derived for interior-boundary closeness measures for L_1 and L_2, respectively.

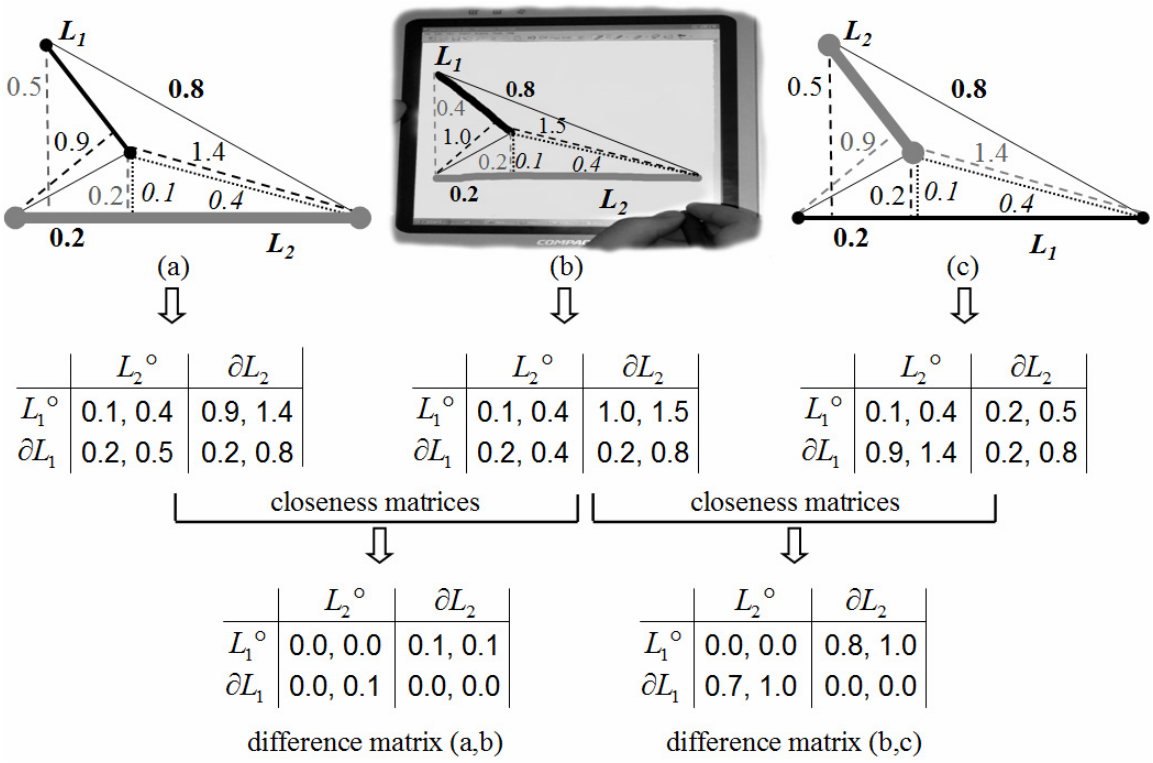


Figure 22: (a) a database configuration; (b) a sketch-query; and (c) the initial database configuration with inverse labeling of the lines. The difference matrix (a,b) should be chosen as a measure of the actual similarity between the sketch and the database configurations. (boundary-boundary distances are drawn with continuous black segments and annotated in **bold**; interior-interior distances are drawn with dotted segments and annotated in *italic*; and interior-boundary distances are drawn with black and gray dashed segments and annotated in black and gray for lines L_1 and L_2 respectively).