Vague Topological Predicates for Crisp Regions through Metric Refinements

Markus Schneider*

University of Florida
Department of Computer and Information Science and Engineering
Gainesville, FL 32611, USA
`mschneid@cise.ufl.edu`

Abstract

Topological relationships between spatial objects have been a focus of research on spatial data handling and reasoning for a long time. Especially as predicates they support the design of suitable query languages for data retrieval and analysis in spatial databases and geographical information systems. Whereas research on this topic has always been dominated by qualitative methods and by an emphasis of a strict separation of topological and metric, that is, quantitative, properties, this paper investigates their possible coexistence and cooperation. Metric details can be exploited to refine topological relationships and to make important semantic distinctions that enhance the expressiveness of spatial query languages. The metric refinements introduced in this paper have the feature of being topologically invariant under affine transformations. Since the combination of a topological predicate with a metric refinement leads to a single unified quantitative measure, this measure has to be interpreted and mapped to a lexical item. This leads to vague topological predicates, and we demonstrate how these predicates can be integrated into a spatial query language.

Keywords. Vague topological relationship, metric refinement, quantitative refinement, 9-intersection model, lexical item, spatial data type, spatial query language

1 Introduction

In recent years, the exploration of topological relationships between objects in space has turned out to be a multi-disciplinary research issue involving disciplines like spatial databases, geographical information systems, CAD/CAM

---

systems, image databases, spatial analysis, computer vision, artificial intelligence, linguistics, cognitive science, psychology, and robotics. From a database perspective, their development has been motivated by the necessity of formally defined topological predicates as filter conditions for spatial selections and spatial joins in spatial query languages, both at the user definition level for reasons of conceptual clarity and at the query processing level for reasons of efficiency.

Topological relationships like *overlap*, *inside*, or *meet* describe purely qualitative properties that characterize the relative positions of spatial objects to each other and that are preserved (topologically invariant) under continuous transformations such as translation, rotation, and scaling. They deliberately exclude any consideration of metric, that is, quantitative, measures and are associated with notions like adjacency, coincidence, connectivity, inclusion, and continuity. Some well known, formal, and especially computational models for topological relationships have already been proposed for spatial objects, for example for regions. They permit answers to queries like "Are regions A and B *disjoint*?" or "Do regions A and B *overlap*?".

Unfortunately, these purely qualitative approaches (topology *per se*) are sometimes insufficient to express the full essence of spatial relations, since they do not capture all details to make important semantic distinctions. This is motivated in Figure 1 for the topological relationship *overlap*. Obviously, for all three configurations the predicate $overlap(A, B)$ yields *true*. But there is no way to express the fact that in the left configuration regions A and B *hardly* overlap, that the middle configuration represents a typical overlap, and that in the right configuration regions A and B *predominantly* overlap. In these statements and the corresponding resulting queries, the *degree of overlapping* between two spatial objects is of decisive importance. The crucial aspect is that this degree is a *relative metric*, and thus *quantitative*, feature which is topologically invariant under affine transformations. This leads to *metrically refined* topological relationships having a *vague* or *blurred* nature.

Transfering this observation to concrete applications, we can consider polluted areas, for example. Here it is frequently not only interesting to know the fact that areas are polluted but also to which degree they are polluted. If two land parcels are adjacent, then often not only this fact is interesting but also the degree of their adjacency.

Section 2 discusses some relevant related work about topological relationships. Our design is based on the 9-*intersection model*, an approach that uses point set topology to define a classification of binary topological relationships in a purely qualitative manner. The goals of this paper are then pursued in the following sections. In Section 3, we explore metrically refined topological relationships on spatial regions with precisely determined boundaries (so-called *crisp* regions) and show how qualitative descriptions (topological properties) can be combined with quantitative aspects (relative metric properties) into a single unified quantitative measure between 0 and 1. This leads us to *vague topological predicates*. In Section 4, we demonstrate how the obtained quan-
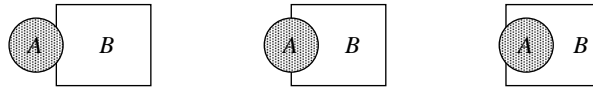
Fig. 1. Topological relationship *overlap*(A, B) with different degrees of overlapping.

titative measures can be mapped to lexical items corresponding to natural language terms like "*a little bit* inside" or "*mostly* overlap". This introduces a kind of vagueness or indeterminacy into user queries which is an inherent feature of human thinking, arguing, and reasoning. Section 5 deals with the integration of these indeterminate predicates into an SQL-like query language. Finally, Section 6 draws some conclusions.

2 Related Work

An important approach for characterizing topological relationships rests on the so-called 9-*intersection model* (Egenhofer et al. 1989). This model allows one to derive a complete collection of mutually exclusive topological relationships for each combination of spatial types. The model is based on the nine possible intersections of boundary (∂A), interior (A°), and exterior (A^-) of a spatial object A with the corresponding components of another object B. Each intersection is tested with regard to the topologically invariant criteria of emptiness and non-emptiness. $2^9 = 512$ different configurations are possible from which only a certain subset makes sense depending on the definition and combination of spatial objects just considered. For each combination of spatial types this means that each of its predicates p can be associated with a unique *boolean intersection matrix* BI_p (Table 1) so that all predicates are mutually exclusive and complete with regard to the topologically invariant criteria of emptiness and non-emptiness.

$$BI_p(A, B) = \begin{pmatrix} \partial A \cap \partial B \neq \varnothing & \partial A \cap B^\circ \neq \varnothing & \partial A \cap B^- \neq \varnothing \\ A^\circ \cap \partial B \neq \varnothing & A^\circ \cap B^\circ \neq \varnothing & A^\circ \cap B^- \neq \varnothing \\ A^- \cap \partial B \neq \varnothing & A^- \cap B^\circ \neq \varnothing & A^- \cap B^- \neq \varnothing \end{pmatrix}$$

Table 1. The boolean 9-intersection matrix. Each matrix entry is a 1 (*true*) or 0 (*false*).

Topological relationships have been first explored for simple regions (Clementini et al. 1993, Cui et al. 1993, Egenhofer et al. 1989). A *simple region* is a bounded, regular closed set homeomorphic (that is, topologi-

cally equivalent) to a two-dimensional closed disc[1] in $\mathbb{R}^2$. Regularity of a closed point set eliminates geometric anomalies possibly arising from dangling points, dangling lines, cuts, and punctures in such a point set (Behr & Schneider 2001). From an application point of view, this means that a simple region has a connected interior, a connected boundary, and a single connected exterior. Hence, it does not consist of several components, and it does not have holes. For two simple regions eight meaningful configurations have been identified which lead to the well known eight topological predicates of the set $T = \{disjoint, meet, overlap, equal, inside, contains, covers, coveredBy\}$. In a vector notation from left to right and from top to bottom, their (well known) intersection matrices are: $BI_{disjoint}(A, B) = (0, 0, 1, 0, 0, 1, 1, 1, 1)$, $BI_{meet}(A, B) = (1, 0, 1, 0, 0, 1, 1, 1, 1)$, $BI_{overlap}(A, B) = (1, 1, 1, 1, 1, 1, 1, 1, 1)$, $BI_{equal}(A, B) = (1, 0, 0, 0, 1, 0, 0, 0, 1)$, $BI_{inside}(A, B) = (0, 1, 0, 0, 1, 0, 1, 1, 1)$, $BI_{contains}(A, B) = (0, 0, 1, 1, 1, 1, 0, 0, 1)$, $BI_{covers}(A, B) = (1, 0, 1, 1, 1, 1, 0, 0, 1)$, $BI_{coveredBy}(A, B) = (1, 1, 0, 0, 1, 0, 1, 1, 1)$.

For reasons of simplicity and clear presentation, in this paper, we will confine ourselves to metric refinements of topological relationships for *simple* regions. An extension to general, *complex* regions (Schneider 1997), that is, regions possibly consisting of several area-disjoint components and possibly having area-disjoint holes, is straightforward. For this purpose, metric refinements have to be applied to the 33 generalized topological predicates between two complex regions (Behr & Schneider 2001).

Approaches dealing with metric refinements of spatial relationships on crisp spatial objects are rare. In (Hernandez et al. 1995) metric refinements of *distance* relationships are introduced to characterize indeterminate terms like *very close, close, far*, and *very far*. In (Peuquet & Xiang 1987, Goyal & Egenhofer 2004) *directional* relationships like *north, north-west*, or *southeast* are metrically refined. Two papers deal at least partially with metric refinements of topological relationships. In (Vazirgiannis 2000) refinements which are similar to our directed topological relationships are proposed. Metric details, which are similar to our metric refinements, are used in (Egenhofer & Shariff 1998) to refine natural-language topological relationships between a simple line and a simple region and between two simple lines. There are a number of differences to our approach. First, we deal with two regions. Second, they do not interpret the entries of a 9-intersection matrix for a given topological predicate as optimum values, as we do in Section 3.2. Third, our set of refinements is systematically developed and complete but not *ad hoc* (see Section 3.1). Fourth, their refinements are not combined with the 9-intersection matrix into a so-called similarity matrix, as in our case (see Section 3.2). Fifth, they do not employ our concept of applicability degree (see Section 3.2).

---

[1] $D(x, \epsilon)$ denotes a two-dimensional *closed disc* with center $x \in \mathbb{R}^2$ and radius $\epsilon \in \mathbb{R}^+$ iff $D(x, \epsilon) = \{y \in \mathbb{R}^2 \mid d(x, y) \leq \epsilon\}$ where d is a metric on $\mathbb{R}^2$.

Two completely different approaches to modeling indeterminate topological predicates rest on the concept of so-called *fuzzy topological predicates* (Schneider 2001*a*, Schneider 2001*b*), which are defined on *complex fuzzy* regions (Schneider 1999). That is, in contrast to the assumption in this paper, these predicates operate on (complex) regions whose extent cannot be precisely determined or is not precisely known.

3 Metric Refinements of Topological Relationships

Topological relationships are designed as binary predicates yielding a boolean and thus strict decision whether such a relationship holds for two spatial objects or not. Metric details based on the geometric properties of the two spatial objects can be used to relax this strictness. As we will see in Sections 4 and 5, they enable us to describe vague nuances of topological relationships, as they often and typically occur in natural language expressions, and hence they allow us to refine topological relationships in an indeterminate manner. Queries like "Which are the land parcels that are *hardly* adjacent to parcel X?" or "Which landscape areas are *mostly* contaminated (overlapped) with toxic substances?" can then be posed and answered.

To describe metrical details, we use relative area and length measures provided by the operand objects (in our case two simple regions). These measures are normalized values with respect to the areas of interiors and lengths of boundaries of two simple regions. Consequently, they are scale-independent and topologically invariant.

3.1 Refinement Ratio Factors

We now introduce six *refinement ratio factors* which are illustrated in Figure 2. For the definition of all factors we assume two simple regions A and B. The *common area* ratio

$$CA(A, B) = \frac{area(A^\circ \cap B^\circ)}{area(A^\circ \cup B^\circ)}$$

specifies the degree to which regions A and B share their areas. Obviously, $CA(A, B) = 0$, if $A^\circ \cap B^\circ = \varnothing$ (that is, A and B are *disjoint* or they *meet*), and $CA(A, B) = 1$, if $A^\circ = B^\circ$ (that is, A and B are *equal*). Like all the other factors to be presented, the common area factor is independent of scaling, translation, and rotation, and hence constant. This factor is also symmetric, that is, $CA(A, B) = CA(B, A)$.

The *outer area* ratio

$$OA(A, B) = \frac{area(A^\circ \cap B^-)}{area(A^\circ)}$$

computes the ratio of that portion of A with A that is not shared with B. Here, $OA(A, B) = 0$, if $A^\circ = B^\circ$, and $OA(A, B) = 1$, if $A^\circ \cap B^- = A^\circ$, that

is, A and B are disjoint or they meet. Obviously, the outer area ratio is not symmetric, that is, $OA(A,B) \neq OA(B,A)$.

The *exterior area* ratio

$$EA(A,B) = \frac{area(A^- \cap B^-)}{area(A^- \cup B^-)}$$

calculates the ratio between the area of the common exterior of A and B on the one hand and the area of the application reference system, where all our regions are located, minus the area of the intersection of A and B on the other hand. The reference system is usually called the *Universe of Discourse* (*UoD*). We assume that our UoD is bounded and thus not equal to but a proper subset of the Euclidean plane. This is not a restriction, since all spaces that can be dealt with in a computer are bounded. If A and B *meet* and their union is the UoD, $EA(A,B) = 0$. If $A = B$, $EA(A,B) = 1$. The exterior area ratio is symmetric, that is, $EA(A,B) = EA(B,A)$.

The *inner boundary splitting* ratio

$$IBS(A,B) = \frac{length(\partial A \cap B^\circ)}{length(\partial A)}$$

determines the degree to which A's boundary is split by B. If A and B are *disjoint* or *meet*, then $IBS(A,B) = 0$. If A is *inside* or *coveredBy* B, then $IBS(A,B) = 1$. The inner boundary splitting ratio is not symmetric, that is, $IBS(A,B) \neq IBS(B,A)$.

The *outer boundary splitting* ratio

$$OBS(A,B) = \frac{length(\partial A \cap B^-)}{length(\partial A)}$$

yields the degree to which A's boundary lies outside of B. If A is *inside* or *coveredBy* B, then $OBS(A,B) = 0$. If A and B are *disjoint* or *meet*, then $OBS(A,B) = 1$. The outer boundary splitting ratio is not symmetric, that is, $OBS(A,B) \neq OBS(B,A)$.

The *common boundary splitting* ratio

$$CBS(A,B) = \frac{length(\partial A \cap \partial B)}{length(\partial A \cup \partial B)}$$

calculates the degree to which regions A and B share their boundaries. Obviously, $CBS(A,B) = 0$, if $\partial A \cap \partial B = \varnothing$, and $CBS(A,B) = 1$, if $\partial A \cap \partial B = \partial A$,

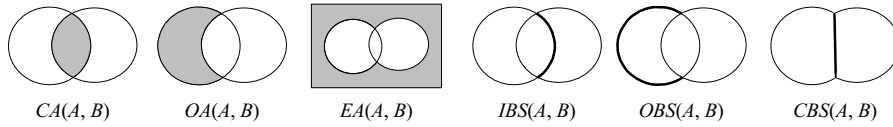$CA(A,B)$ $OA(A,B)$ $EA(A,B)$ $IBS(A,B)$ $OBS(A,B)$ $CBS(A,B)$
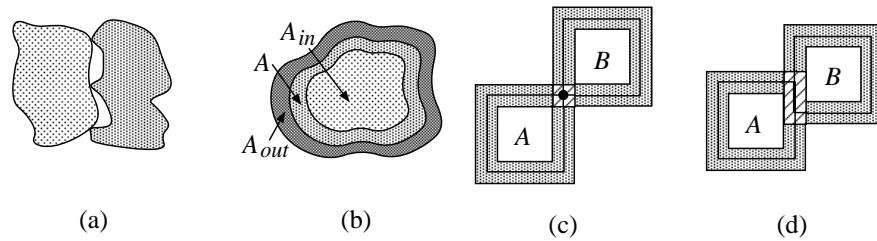
Fig. 2. Refinement ratio factors.

Fig. 3. Problem configuration for the common boundary splitting ratio (a), enlarged region A_{out} and reduced region A_{in} for region A, and 0-dimensional (c) and 1-dimensional (d) boundary intersections with their corresponding boundary areas.

which means that A and B are *equal*. The common boundary splitting ratio is also symmetric, that is, $CBS(A, B) = CBS(B, A)$.

The common boundary splitting ratio is especially important for computing the degree of meeting of two regions. A problem arises with this factor, if the common boundary parts do not have a linear structure but consist of a finite set of points. Figure 3a shows such a meeting situation. The calculation of $CBS(A, B)$ leads to 0, because due to regularization $\partial A \cap \partial B = \varnothing$ and the length is thus 0. Hence, common single points are not taken into account by this factor, which should be done for correctly evaluating (the degree of) a meeting situation. To solve this problem, for each simple region A we introduce two additional simple regions A_{out} and A_{in} which are slightly enlarged and reduced, respectively, by scale factors $1 + \epsilon$ and $1 - \epsilon$, respectively, with $\epsilon > 0$ (Figure 3b). We then consider $\Delta A = A_{out} - A_{in}$ as the extended boundary of A and redefine the common boundary splitting ratio as

$$CBS(A, B) = \frac{area(\Delta A \cap \Delta B)}{area(\Delta A \cup \Delta B)}$$

In Figures 3c and d, the dark shaded regions show the extended boundaries of A and B. The diagonally hatched regions correspond to the boundary intersection of A and B.

The refinement ratio factors have, of course, not been defined arbitrarily. They have been specified in a way so that *each* intersection occurring in a matrix entry in BI_p is contained as an argument of the area or length function of the numerator of a refinement ratio factor. As an example, consider the intersection $\partial A \cap B°$ included in the inequality of the first row and second column of $BI_p(A, B)$. This intersection reappears as argument of the length function of the numerator of $IBS(A, B)$. The intersection $A° \cap \partial B$ (second row, first column of $BI_p(A, B)$) is captured by $IBS(B, A)$. The purpose of the denominator of a refinement ratio factor then is to make the factor a relative and topologically invariant measure.

3.2 Evaluation of the Applicability of a Topological Relationship

In this subsection we show how the concept of metric refinement can be used to assess the applicability of a topological relationship for a given spatial configuration with a single, numerical value. This then leads us to the concept of a *vague topological relationship*.

The Similarity Matrix

The boolean intersection matrix $BI_p(A, B)$ contains nine strict, binary intersection tests leading either to *true* (1) or *false* (0). We now replace each intersection test (inequality) by the corresponding refinement ratio factor. This leads us to the *real-valued similarity matrix $RS_p(A, B)$* (Table 2) which we now employ in order to represent and estimate a topological relationship between two simple regions. Each matrix entry of $RS_p(A, B)$ represents a value between 0 and 1 and is interpreted as the *degree* to which the corresponding intersection in $BI_p(A, B)$ holds. That is, the statement about the *existence* of an intersection is replaced by a statement about the *degree* of an intersection.

$$RS_p(A, B) = \begin{pmatrix} CBS(A,B) & IBS(A,B) & OBS(A,B) \\ IBS(B,A) & CA(A,B) & OA(A,B) \\ OBS(B,A) & OA(B,A) & EA(A,B) \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{area(\Delta A \cap \Delta B)}{area(\Delta A \cup \Delta B)} & \dfrac{length(\partial A \cap B^\circ)}{length(\partial A)} & \dfrac{length(\partial A \cap B^-)}{length(\partial A)} \\[1.2em] \dfrac{length(A^\circ \cap \partial B)}{length(\partial B)} & \dfrac{area(A^\circ \cap B^\circ)}{area(A^\circ \cup B^\circ)} & \dfrac{area(A^\circ \cap B^-)}{area(A^\circ)} \\[1.2em] \dfrac{length(A^- \cap \partial B)}{length(\partial B)} & \dfrac{area(A^- \cap B^\circ)}{area(B^\circ)} & \dfrac{area(A^- \cap B^-)}{area(A^- \cup B^-)} \end{pmatrix}$$

Table 2. The real-valued similarity matrix. Each matrix entry is computed as a value between 0 and 1.

Seen from this perspective, each matrix entry 0 or 1 of BI_p can be interpreted in a new, different way, namely as the "optimum", "best possible", or sometimes "asymptotic" degree to which the corresponding intersection occurring as part of an intersection test in BI_p holds. On the other hand, this is not necessarily obvious. Hence, for each predicate p, Table 3 contains an analysis of the suitability of a matrix entry in BI_p for our interpretation.

The left column contains a list of the topological predicates. The first row uses shortcuts to represent the nine intersections. For example, ∂° means $\partial A \cap B^\circ$ ($\neq \varnothing$), and $^\circ\partial$ means $A^\circ \cap \partial B$ ($\neq \varnothing$).

An entry "+" in the table indicates that the respective 0 or 1 in BI_p is the optimum, perfect, and adopted value to fulfil predicate p. For example,

	$\partial\partial$	$\partial°$	∂^-	$°\partial$	$°°$	$°-$	$^-\partial$	$^-°$	$^-{}^-$
equal	+	+	+	+	+	+	+	+	(+)
meet	(+)	+	+	+	+	+	+	+	(+)
disjoint	+	+	+	+	+	+	+	+	(+)
inside	+	+	+	+	(+)	+	+	(+)	(+)
contains	+	+	+	+	(+)	(+)	+	+	(+)
covers	(+)	+	+	+	(+)	(+)	+	+	(+)
coveredBy	(+)	+	+	+	(+)	+	+	(+)	(+)
overlap	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)

Table 3. Suitability of a matrix entry in BI_p for interpreting it as the degree to which the respective intersection holds.

if for *disjoint* the intersection of the boundary of A and the interior of B is empty (matrix entry 0), this is the optimum that can be reached for the inner boundary splitting ratio $IBS(A, B)$. If for *covers* the intersection of the boundary of A and the exterior of B is non-empty (matrix entry 1), this is the optimum that can be reached for the outer boundary splitting ratio $OBS(A, B)$. This situation implies that the boundary of B touches the boundary of A only in single points and not in curves.

An entry "(+)" expresses that the respective 0 or 1 in BI_p is an asymptotic value for predicate p. That is, this value can be approached in an arbitrarily precise way but in the end it cannot and may not be reached. For example, for *meet* the common boundary splitting ratio $CBS(A, B)$ can be arbitrarily near to 1. But it cannot become equal to one, because then the relationship *equal* would hold. For our later computation, this is no problem. We simply assume the respective asymptotic 0's and 1's as optimum values.

Computing the Degree of Applicability

Whereas *exactly one* topological relationship applies with the boolean 9-intersection matrix, we will show now that the similarity matrix enables us to assess *all* topological relationships but with different *degrees of applicability*. For that purpose, we test for the similarity of RS_p with BI_p. Since we can interpret the matrix entries of BI_p for a predicate p as the ideal values, the evaluation of the similarity between RS_p and BI_p can be achieved by measuring for each matrix entry $RS_p^{(x,y)}(A, B)$ with $x, y \in \{\partial, °, ^-\}$ the deviation from the corresponding matrix entry $BI_p^{(x,y)}(A, B)$. The idea is then to condense all nine deviation values to a single value by taking the average of the sum of all deviations. We call the resulting value the *applicability degree* of a topological relationship p with respect to two simple regions A and B. The applicability degree is computed by a function μ taking two simple regions and the name of a topological predicate p as operands and yielding a real value between 0 and 1 as a result:

$$\mu(A,B,p) = \frac{\displaystyle\sum_{x\in\{\partial,\circ,^-\}}\sum_{y\in\{\partial,\circ,^-\}}\begin{array}{l}\textbf{if } BI_p^{(x,y)}(A,B) \textbf{ then } RS_p^{(x,y)}(A,B)\\ \textbf{else } 1 - RS_p^{(x,y)}(A,B)\end{array}}{9}$$

What we have gained is the relaxation of the strictness of a topological predicate $p : region \times region \to \{0,1\}$ (*region* shall be the type for simple regions) to an *applicability degree function* $\mu : region \times region \times T \to [0,1]$ (remember that T is the set of all topological predicates). The applicability degree $\mu(A,B,p)$ gives us the extent to which predicate p holds for two simple regions A and B. We abbreviate $\mu(A,B,p)$ by the *vague topological predicate value* $p_v(A,B)$ with $p_v : region \times region \to [0,1]$. The term p_v indicates the association to predicate p. Whereas the topological predicate p maps to the set $\{0,1\}$ and thus results in a strict and abrupt decision, the vague topological predicate p_v maps to the closed interval $[0,1]$ and hence permits a smooth evaluation. Codomain $[0,1]$ can be regarded as the data type for *vague booleans.*

4 Mapping Quantitative Measures to Qualitative, Lexical Items

The fact that the applicability degree yielded by a vague topological predicate is a computationally determined quantification between 0 and 1, that is, a vague boolean, impedes a direct integration into a query language. First, it is not very comfortable and user-friendly to use such a numeric value in a query. Second, spatial selections and spatial joins are not able to cope with vague predicates and expect strict and exact predicates as filter conditions that yield *true* or *false*.

As a solution which maintains this requirement, we propose to embed adequate qualitative linguistic descriptions of nuances of topological relationships as appropriate interpretations of the applicability degrees into a spatial query language. Notice that the linguistic descriptions given in the following are arbitrary and exchangeable, since it is beyond the scope of this paper to discuss linguistic reasons how to associate a meaning to a given applicability degree. In particular, we think that the users themselves should be responsible for specifying a list of appropriate linguistic terms and for associating an applicability degree with each of them. This gives them greatest flexibility for querying. For example, depending on the applicability degree yielded by the predicate $inside_v$, the user could distinguish between *not* inside, *a little bit* inside, *somewhat* inside, *slightly* inside, *quite* inside, *mostly* inside, *nearly completely* inside, and *completely* inside. These user-defined, vague linguistic terms can then be incorporated into spatial queries together with the topological predicates they modify. We call these terms *vague quantifiers*, because their semantics lies between the universal quantifier *for all* and the existential quantifier *there exists*.
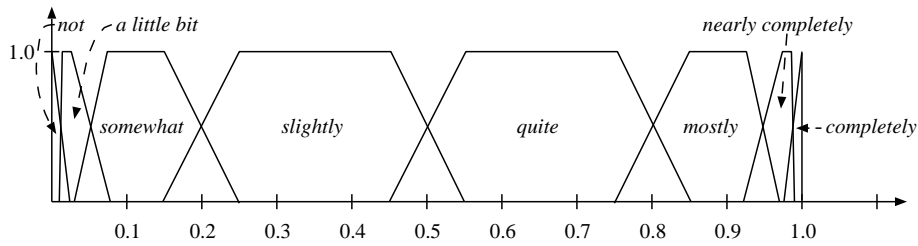
Fig. 4. Membership functions for vague quantifiers.

We know that a vague topological predicate p_v is defined as $p_v : region \times region \to [0, 1]$. The idea is now to represent each vague quantifier $\gamma \in \Gamma = \{not,\ a\ little\ bit,\ somewhat,\ slightly,\ quite,\ mostly,\ nearly\ completely,\ completely, \dots \}$ by an appropriate *membership function* $\mu_\gamma : [0, 1] \to [0, 1]$. Let $A, B \in region$, and let $\gamma\,p_v$ be a quantified vague predicate (like *somewhat inside* with $\gamma = somewhat$ and $p_v = inside_v$). Then we can define:

$$\gamma\,p_v(A, B) = true \quad :\Leftrightarrow \quad (\mu_\gamma \circ p_v)(A, B) = 1$$

That is, only for those values of $p_v(A, B)$ for which μ_γ yields 1, the predicate $\gamma\,p_v$ is true. A membership function that fulfils this quite strict condition is, for instance, the partition of $[0, 1]$ into $n \leq |\Gamma|$ disjoint or adjacent intervals completely covering $[0, 1]$ and the assignment of each interval to a vague quantifier. If an interval $[a, b]$ is assigned to a vague quantifier γ, the intended meaning is that $\mu_\gamma(p_v(A, B)) = 1$, if $a \leq p_v(A, B) \leq b$, and 0 otherwise. For example, the user could select the intervals $[0.0, 0.02]$ for *not*, $[0.02, 0.05]$ for *a little bit*, $[0.05, 0.2]$ for *somewhat*, $[0.2, 0.5]$ for *slightly*, $[0.5, 0.8]$ for *quite*, $[0.8, 0.95]$ for *mostly*, $[0.95, 0.98]$ for *nearly completely*, and $[0.98, 1.00]$ for *completely*.

Alternative membership functions are shown in Figure 4. While we can always find a fitting vague quantifier for the partition due to the complete coverage of the interval $[0, 1]$, this is not necessarily the case here. Each vague quantifier is associated with a *vague number* having a trapezoidal-shaped or triangular-shaped membership function. The transition between two consecutive vague quantifiers is smooth and here modeled by linear functions. Within a vague transition area, μ_γ yields a value less than 1 which makes the predicate $\gamma\,p_v$ false. Examples in Figure 4 can be found at 0.2, 0.5, or 0.8. Each vague number associated with a vague quantifier can be represented as a quadruple (a, b, c, d) where the membership function starts at $(a, 0)$, linearly increases up to $(b, 1)$, remains constant up to $(c, 1)$, and linearly decreases up to $(d, 0)$. Figure 4 assigns $(0.0, 0.0, 0.0, 0.02)$ to *not*, $(0.01, 0.02, 0.03, 0.08)$ to *a little bit*, $(0.03, 0.08, 0.15, 0.25)$ to *somewhat*, $(0.15, 0.25, 0.45, 0.55)$ to *slightly*, $(0.45, 0.55, 0.75, 0.85)$ to *quite*, $(0.75, 0.85, 0.92, 0.96)$ to *mostly*, $(0.92, 0.96, 0.97, 0.99)$ to *nearly completely*, and $(0.97, 1.0, 1.0, 1.0)$ to *completely*.

So far, the predicate $\gamma\, p_f$ is only true if μ_γ yields 1. We can relax this strict condition by defining:

$$\gamma\, p_f(A, B) = true \quad :\Leftrightarrow \quad (\mu_\gamma \circ p_f)(A, B) > 0$$

In a spatial database system this gives us the chance also to take the transition zones into account and to let them make the predicate $\gamma\, p_v$ true. When evaluating a spatial selection or join in a spatial database system on the basis of a vague topological predicate, we can even set up a weighted ranking of database objects satisfying the predicate $\gamma\, p_v$ at all and being ordered by descending membership value $1 \geq \mu_\gamma(x) > 0$ for some value $x \in [0, 1]$.

A special, optional vague quantifier, denoted by *at all*, represents the existential quantifier and checks whether a predicate p_v can be fulfilled to any extent. An example query is: "Do regions A and B (*at all*) overlap?" With this quantifier we can determine whether $\mu_\gamma(x) > 0$ for some value $x \in [0, 1]$.

5 Querying

In this section we briefly demonstrate with a few example queries how spatial data types and quantified vague topological predicates can be integrated into an SQL-like spatial query language. It is not our objective to give a full description of a specific language. We assume a relational data model where tables may contain regions as attribute values in the same way as integers or strings.

What we need first are mechanisms to declare and to activate *user-defined* vague quantifiers. These mechanisms should allow the user to specify trapezoidal-shaped and triangular-shaped membership functions as well as partitions. In general, this means to define a (possibly overlapping) *classification*, which for our example in Section 4 could be expressed by the user in the following way:

create classification fq
$$
\begin{array}{ll}
(not & (0.00, 0.00, 0.00, 0.02), \\
a\ little\ bit & (0.01, 0.02, 0.03, 0.08), \\
somewhat & (0.03, 0.08, 0.15, 0.25), \\
slightly & (0.15, 0.25, 0.45, 0.55), \\
quite & (0.45, 0.55, 0.75, 0.85), \\
mostly & (0.75, 0.85, 0.92, 0.96), \\
nearly\ completely & (0.92, 0.96, 0.97, 0.99), \\
completely & (0.97, 1.0, 1.0, 1.0))
\end{array}
$$

Such a classification could then be activated by

set classification   fq

We assume that we have a relation *pollution*, which stores among other things the geometry of polluted zones as regions, and a relation *areas*, which keeps

information about the use of land areas and which stores their spatial extent as regions. A query could be to find out all inhabited areas where people are rather endangered by pollution. This can be formulated in an SQL-like style as (we here use infix notation for the predicates):

> **select** areas.name
> **from**   pollution, areas
> **where** area.use = inhabited **and**
>   pollution.region *quite overlaps*
>   areas.region

This query and the following two ones represent *vague spatial joins.*

Another query asks for those inhabited areas lying almost entirely in polluted areas:

> **select** areas.name
> **from**   pollution, areas
> **where** areas.use = inhabited **and**
>   areas.region *nearly completely inside*
>   pollution.region

Assume that we are given living spaces of different animal species in a relation *animals* and that their indeterminate extent is represented as a vague region. Then we can search for pairs of species which share a common living space to some degree:

> **select** A.name, B.name
> **from**   animals A, animals B
> **where** A.region *at all overlaps* B.region

As a last example, we can ask for animals that usually live on land and seldom enter the water or for species that never leave their land area (the built-in aggregation function **sum** is applied to a set of vague regions and aggregates this set by repeated application of vague geometric union):

> **select** name
> **from**   animals
> **where** (**select sum**(region) **from** areas)
>   *nearly completely covers* **or**
>   *completely covers* region

6 Conclusions

In this paper we have presented a simple but expressive and effective concept showing how metric details can be leveraged to make important semantic distinctions of topological relationships on simple regions. The resulting vague topological predicates are often more adequate for expressing a spatial situation than their coarse, strict counterparts, because they are multi-faceted and

much nearer to human thinking and questioning. Consequently, they allow a much more natural formulation of spatial queries than we can find in current spatial query languages.

We are currently working on a prototype implementation for demonstrating the concepts presented in this paper and validating their relevance to practice.

In the future we plan to extend the concept of metric refinement to complex regions. We will also investigate metric refinements between two complex line objects and between a complex line object and a complex region object.

References

Behr, T. & Schneider, M. (2001), Topological Relationships of Complex Points and Complex Regions, *in* 'Int. Conf. on Conceptual Modeling', pp. 56–69.

Clementini, E., Di Felice, P. & Oosterom, P. (1993), A Small Set of Formal Topological Relationships Suitable for End-User Interaction, *in* '3rd Int. Symp. on Advances in Spatial Databases', LNCS 692, pp. 277–295.

Cui, Z., Cohn, A. G. & Randell, D. A. (1993), Qualitative and Topological Relationships, *in* '3rd Int. Symp. on Advances in Spatial Databases', LNCS 692, pp. 296–315.

Egenhofer, M. J., Frank, A. & Jackson, J. P. (1989), A Topological Data Model for Spatial Databases, *in* '1st Int. Symp. on the Design and Implementation of Large Spatial Databases', LNCS 409, Springer-Verlag, pp. 271–286.

Egenhofer, M. J. & Shariff, A. R. (1998), 'Metric Details for Natural-Language Spatial Relations', *ACM Transactions on Information Systems* **16**(4), 295–321.

Goyal, R. & Egenhofer, M. (2004), 'Cardinal Directions between Extended Spatial Objects', *IEEE Trans. on Knowledge and Data Engineering* . In press.

Hernandez, D., Clementini, E. C. & Di Felice, P. (1995), Qualitative Distances, *in* '2nd Int. Conf. on Spatial Information Theory', LNCS 988, Springer-Verlag, pp. 45–57.

Peuquet, D. J. & Xiang, Z. C. (1987), 'An Algorithm to Determine the Directional Relationship between Arbitrarily-Shaped Polygons in the Plane', *Pattern Recognition* **20**(1), 65–74.

Schneider, M. (1997), *Spatial Data Types for Database Systems - Finite Resolution Geometry for Geographic Information Systems*, Vol. LNCS 1288, Springer-Verlag, Berlin Heidelberg.

Schneider, M. (1999), Uncertainty Management for Spatial Data in Databases: Fuzzy Spatial Data Types, *in* '6th Int. Symp. on Advances in Spatial Databases', LNCS 1651, Springer-Verlag, pp. 330–351.

Schneider, M. (2001*a*), A Design of Topological Predicates for Complex Crisp and Fuzzy Regions, *in* 'Int. Conf. on Conceptual Modeling', pp. 103–116.

Schneider, M. (2001*b*), Fuzzy Topological Predicates, Their Properties, and Their Integration into Query Languages, *in* 'ACM Symp. on Geographic Information Systems', pp. 9–14.

Vazirgiannis, M. (2000), Uncertainty Handling in Spatial Relationships, *in* 'ACM Symp. for Applied Computing'.