Fuzzy Topological Predicates, Their Properties, and Their Integration into Query Languages

Markus Schneider
FernUniversität Hagen, Praktische Informatik IV
58084 Hagen, Germany
markus.schneider@fernuni-hagen.de

ABSTRACT

For a long time topological relationships between spatial objects have been a main focus of research on spatial data handling and reasoning. They have especially been integrated into query languages of spatial database systems and geographical information systems. One of their fundamental features is that they operate on spatial objects with precisely defined, sharp boundaries. But in many geometric and geographic applications there is a need to model spatial phenomena and their topological relationships rather through vague or fuzzy concepts due to indeterminate boundaries. This paper presents a model of fuzzy regions and focuses on the definition of topological predicates between them. Moreover, some properties of these predicates are shown, and we demonstrate how the predicates can be integrated into a query language.

Keywords

Fuzzy region, fuzzy spatial query language

1. INTRODUCTION

Representing, storing, quering, and manipulating spatial information is important for many non-standard database applications. Specialized systems like geographical information systems (GIS), spatial database systems, and image database systems to some extent provide the needed technology to support these applications. For these systems the development of formal models for spatial objects and for topological relationships between these objects is a topic of great importance and interest, since these models exert a great influence on the efficiency of spatial systems and on the expressiveness of spatial query languages.

In the past, a number of data models and query languages for spatial objects with precisely defined boundaries, so-called *crisp spatial objects*, have been proposed with the aim of formulating and processing spatial queries in databases (e.g., [8, 9]). *Spatial data types* (see [9] for a survey) like *point*, *line*, or *region* are the central concept of these approaches. They provide fundamental abstractions for modeling the structure of geometric entities, their relationships, properties, and operations. *Topological predicates* [6] between crisp objects have been studied intensively in disciplines like spatial analysis, spatial reasoning, and artificial intelligence.

Increasingly, researchers are beginning to realize that the current mapping of spatial phenomena of the real world to exclusively crisp spatial objects is an insufficient abstraction process for many spatial applications and that the feature of *spatial vagueness* is inherent to many geographic data [3]. Moreover, there is a general consensus that applications based on this kind of indeterminate spatial data are not covered by current information systems. In this paper we focus on a special kind of spatial vagueness called *spatial fuzziness*. Fuzziness captures the property of many spatial objects in reality which do not have sharp boundaries or whose boundaries cannot be precisely determined. Examples are natural, social, or cultural phenomena like land features with continuously changing properties (such as population density, soil quality, vegetation, pollution, temperature, air pressure), oceans, deserts, English speaking areas, or mountains and valleys. The transition between a valley and a mountain usually cannot be exactly ascertained so that the two spatial objects "valley" and "mountain" cannot be precisely separated and defined in a crisp way. We will designate this kind of entities as *fuzzy spatial objects*. In the GIS community, a number of models based on fuzzy sets (e.g., [1, 5, 13, 14]) have been proposed, but these are not suitable for an integration into a database system, because they do not provide data types for fuzzy spatial data. The author himself has started to work on this topic and to design a system of *fuzzy spatial data types* including operations and predicates [10, 11] that can be embedded into a DBMS.

The goal of this paper is to give a definition of topological predicates on fuzzy regions, which is currently an open problem, and to discuss some properties of these predicates. Besides, we show the integration of these predicates into a query language. Section 2 presents a formal model of very generally defined crisp regions and sketches the design of a well known collection of topological predicates on crisp regions. In Section 3 we employ the concept of a crisp region for a definition of fuzzy regions. Fuzzy regions are described as collections of so-called *crisp α-level regions*. In practice, this enables us to transfer the whole formal framework and later all the well known implementation methods available for crisp regions to fuzzy regions. In Section 4 we present an approach for designing topological predicates between fuzzy regions that is based on fuzzy set theory. Section 5 discusses some properties of these predicates, and Section 6 deals with their integration into a query language. Finally, Section 7 draws some conclusions.

2. CRISP REGIONS AND TOPOLOGICAL PREDICATES

Our definition of crisp regions is based on point set theory and point set topology [7]. Regions are embedded into the two-

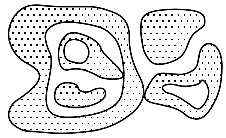

Figure 1: Example of a (complex) region object.

dimensional Euclidean space $\mathbb{R}^2$ and are thus point sets. Unfortunately, the use of pure point set theory for their definition causes problems. If regions are modeled as arbitrary point sets, they can suffer from undesired geometric anomalies. These degeneracies relate to isolated or dangling line and point features as well as missing lines and points in the form of cuts and punctures. A process called *regularization* avoids these anomalies.

We assume that the reader is familiar with some needed, well-known concepts of point set topology like topological space, open set, closed set, interior, closure, exterior, and boundary. The concept of regularity defines a point set S as *regular closed* if $S = \overline{S^\circ}$. We define a *regularization function* reg_c which associates a set S with its corresponding regular closed set as $reg_c(S) := \overline{S^\circ}$. The effect of the *interior* operation is to eliminate dangling points, dangling lines, and boundary parts. The effect of the *closure* operation is to eliminate cuts and punctures by appropriately supplementing points and to add the boundary. We are now already able to give a general definition of a type for *complex* crisp regions:

$$region = \{R \subseteq \mathbb{R}^2 \mid R \text{ is bounded and regular closed}\}$$

In fact, this very "structureless" and implicit definition models (complex) crisp regions possibly consisting of several components and possibly having holes (Figure 1). As a special case, a *simple* region is a region that does not have holes and does not consist of multiple components.

An important approach to designing topological predicates on *simple* crisp regions rests on the well-known 9-*intersection model* [6] from which a complete collection of mutually exclusive topological relationships can be derived. 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 combination of spatial objects just considered. For two simple regions eight meaningful configurations have been identified which lead to the eight predicates of the set $T_{cr} = \{$*disjoint*, *meet*, *overlap*, *equal*, *inside*, *contains*, *covers*, *coveredBy*$\}$. Each predicate is associated with a unique combination of nine intersections so that all predicates are mutually exclusive and complete with regard to the topologically invariant criteria of emptiness and non-emptiness. We explain the meaning of these predicates only informally here. Two crisp regions A and B are *disjoint* if their point sets are disjoint. They *meet* if their boundaries share points and their interiors are disjoint. They are *equal* if both their boundaries and their interiors coincide. A is *inside* B (B *contains* A) if A is a proper subset of B and if their boundaries do not touch. A *is covered by* B (B *covers* A) if A is a proper subset of B and if their boundaries touch. Otherwise, A and B *overlap*. Generalizations to topological predicates for *complex* crisp regions leading to the same (clustered) collection T_{cr} of predicates have been given in [12, 2]. We will base our definition of topological predicates for fuzzy regions on these topological predicates for complex crisp regions.

3. MODELING FUZZY REGIONS

A "structureless" definition of fuzzy regions in the sense that only "flat" point sets are considered and no structural information is revealed has been given in [10]. For our purposes we deploy a "semantically richer" characterization and approximation of fuzzy regions and define them in terms of special, nested α-cuts. A fuzzy region $\tilde{F}$ is described as a *collection of crisp* α-*level regions*[1] [10], i.e., $\tilde{F} = \{F_{\alpha_1}, \ldots, F_{\alpha_{n+1}}\}$ where $\alpha_i \in \Lambda_{\tilde{F}}$, which is the level set of $\tilde{F}$, and where

$$F_{\alpha_i} = reg_c(\{(x,y) \in \mathbb{R}^2 \mid \mu_{\tilde{F}}(x,y) \geq \alpha_i\})$$

We call F_{α_i} an α-*level region*. Clearly, F_{α_i} is a complex crisp region whose boundary is defined by all points with membership value α_i. In particular, F_{α_i} can have holes. The kernel of $\tilde{F}$ is then equal to $F_{1.0}$. An essential property of the α-level regions of a fuzzy region is that they are nested, i.e., if we select membership values $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_n > \alpha_{n+1} = 0$ for some $n \in \mathbb{N}$, then

$$F_{\alpha_1} \subseteq F_{\alpha_2} \subseteq \cdots \subseteq F_{\alpha_n} \subseteq F_{\alpha_{n+1}}$$

We here describe the finite, discrete case. If $\Lambda_{\tilde{F}}$ is infinite, then there are obviously infinitely many α-level regions which can only be finitely represented within this view if we make a finite selection of α-values. In the discrete case, if $|\Lambda_{\tilde{F}}| = n+1$ and if we take all these occurring membership values of a fuzzy region, we can even replace "$\subseteq$" by "$\subset$" in the inclusion relationships above. This follows from the fact that for any $p \in F_{\alpha_i} - F_{\alpha_{i-1}}$ with $i \in \{2, \ldots, n+1\}$, $\mu_{\tilde{F}}(p) = \alpha_i$. For the continuous case, we get $\mu_{\tilde{F}}(p) \in [\alpha_i, \alpha_{i-1})$. As a result, we obtain:

> A *fuzzy region* is a (possibly infinite) set of α-level regions, i.e., $\tilde{F} = \{F_{\alpha_i} \mid 1 \leq i \leq |\Lambda_{\tilde{F}}|\}$ with $\alpha_i > \alpha_{i+1} \Rightarrow F_{\alpha_i} \subseteq F_{\alpha_{i+1}}$ for $1 \leq i \leq |\Lambda_{\tilde{F}}| - 1$.

4. TOPOLOGICAL PREDICATES ON FUZZY REGIONS

In this section we introduce a concept of topological predicates for fuzzy regions. To clarify the nature of a *fuzzy* (*topological*) *predicate*, we can draw an analogy between the transition of a crisp set to a fuzzy set and the transition of a crisp predicate to a fuzzy predicate. In a similar way as we can generalize the characteristic function $\chi_A : X \to \{0,1\}$ to the membership function $\mu_{\tilde{A}} : X \to [0,1]$, we can generalize a (binary) predicate $p_c : X \times Y \to \{0,1\}$ to a (binary) fuzzy predicate $p_f : \tilde{X} \times \tilde{Y} \to [0,1]$. Hence, the value of a fuzzy predicate expresses the degree to which the predicate holds for its operand objects. In case of topological predicates, in this paper the sets X and Y are both equal to the type *region*, and the set $\{0,1\}$ is equal to the type *bool*. The sets $\tilde{X}$ and $\tilde{Y}$ are both equal to the type *fregion* for fuzzy regions, and for the set $[0,1]$ we need a type *fbool* for *fuzzy booleans*.

For the definition of fuzzy topological predicates, we take the view of a fuzzy region as a collection of α-level regions (Section 3), which are complex crisp regions (Section 2), and assume the set T_{cr} of topological predicates on these regions. This preparatory work now enables us to reduce topological predicates on fuzzy regions to topological predicates on collections of crisp regions.

The approach presented in this section is generic in the sense that any meaningful collection of topological predicates on crisp regions could be the basis for our definition of a collection of topological predicates on fuzzy regions. If the former collection additionally fulfils the properties of completeness and mutual exclusion

[1]Other structured characterizations given in [10] describe fuzzy regions as multi-component objects, as three-part crisp regions, and as α-partitions.

(which is the case for T_{cr}), the latter collection automatically inherits these properties.

The open question now is how to compute the topological relationships of two collections of α-level regions, each collection describing a fuzzy region. We use the concept of basic probability assignment [4] for this purpose. A *basic probability assignment* $m(F_{\alpha_i})$ can be associated with each α-level region F_{α_i} and can be interpreted as the probability that F_{α_i} is the "true" representative of F. It is defined as

$$m(F_{\alpha_i}) = \alpha_i - \alpha_{i+1}$$

for $1 \le i \le n$ for some $n \in \mathbb{N}$ with $\alpha_1 = 1$ and $\alpha_{n+1} = 0$. That is, m is built from the differences of successive α_i's. It is easy to see that the telescoping sum $\sum_{i=1}^{n} m(F_{\alpha_i}) = \alpha_1 - \alpha_{n+1} = 1 - 0 = 1$.

Let $\pi_f(F,G)$ be the value that represents a (binary) property π_f between two fuzzy regions F and G. For reasons of simplicity, we assume that $\Lambda_{\tilde{F}} = \Lambda_{\tilde{G}} =: \Lambda$. Otherwise, it is not difficult to "synchronize" $\Lambda_{\tilde{F}}$ and $\Lambda_{\tilde{G}}$ by forming their union $\Lambda := \Lambda_{\tilde{F}} \cup \Lambda_{\tilde{G}}$ and by reordering and renumbering all α-levels. Based on the work in [4] property π_f of F and G can be determined as the summation of weighted predicates by

$$\pi_f(F,G) = \sum_{i=1}^{n} \sum_{j=1}^{n} m(F_{\alpha_i}) \cdot m(G_{\alpha_j}) \cdot \pi_{cr}(F_{\alpha_i}, G_{\alpha_j})$$

where $\pi_{cr}(F_{\alpha_i}, G_{\alpha_j})$ yields the value of the corresponding property π_{cr} for two crisp α-level regions F_{α_i} and G_{α_j}. This formula is equivalent to

$$\pi_f(F,G) = \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_{i+1}) \cdot (\alpha_j - \alpha_{j+1}) \cdot \pi_{cr}(F_{\alpha_i}, G_{\alpha_j})$$

If π_f is a topological predicate of $T_f = \{disjoint_f, meet_f, overlap_f, equal_f, inside_f, contains_f, covers_f, coveredBy_f\}$ between two fuzzy regions, we can compute the degree of the corresponding relationship with the aid of the pertaining crisp topological predicate $\pi_{cr} \in T_{cr}$. The value of $\pi_{cr}(F_{\alpha_i}, G_{\alpha_j})$ is either 1 (*true*) or 0 (*false*). Once this value has been determined for all combinations of α-level regions from F and G, the aggregated value of the topological predicate $\pi_f(F,G)$ can be computed as shown above. The more fine-grained the level set Λ for the fuzzy regions F and G is, the more precisely the fuzziness of topological predicates can be determined.

It remains to show that $0 \le \pi_f(F,G) \le 1$ holds, i.e., π_f is really a fuzzy predicate. Since $\alpha_i - \alpha_{i+1} > 0$ for all $1 \le i \le n$ and since $\pi_{cr}(F_{\alpha_i}, G_{\alpha_j}) \ge 0$ for all $1 \le i, j \le n$, $\pi_f(F,G) \ge 0$ holds. We can show the other inequality by determining an upper bound for $\pi_f(F,G)$:

$$\begin{aligned}
\pi_f(F,G) &= \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_{i+1}) \cdot (\alpha_j - \alpha_{j+1}) \cdot \pi_{cr}(F_{\alpha_i}, G_{\alpha_j}) \\
&\le \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_{i+1}) \cdot (\alpha_j - \alpha_{j+1}) \\
&\quad (\text{since } \pi_{cr}(F_{\alpha_i}, G_{\alpha_j}) \le 1) \\
&= (\alpha_1 - \alpha_2)(\alpha_1 - \alpha_2) + \cdots \\
&\quad + (\alpha_1 - \alpha_2)(\alpha_n - \alpha_{n+1}) + \cdots \\
&\quad + (\alpha_n - \alpha_{n+1})(\alpha_1 - \alpha_2) + \cdots \\
&\quad + (\alpha_n - \alpha_{n+1})(\alpha_n - \alpha_{n+1}) \\
&= (\alpha_1 - \alpha_2)\big((\alpha_1 - \alpha_2) + \cdots + (\alpha_n - \alpha_{n+1})\big) + \cdots \\
&\quad + (\alpha_n - \alpha_{n+1})\big((\alpha_1 - \alpha_2) + \cdots + (\alpha_n - \alpha_{n+1})\big) \\
&= (\alpha_1 - \alpha_2) + \cdots + (\alpha_n - \alpha_{n+1}) \\
&\quad (\text{since } \sum_{i=1}^{n} (\alpha_i - \alpha_{i+1}) = 1) \\
&= 1
\end{aligned}$$

Hence, $\pi_f(F,G) \le 1$ holds.

An alternative definition of fuzzy topological predicates, which pursues a similar strategy like the one discussed so far, is based on the topological predicates π_{srh} on *simple regions with holes* (but without multiple components). If F_{α_i} is an α-level region, let us denote its connected components by $F_{\alpha_{i1}}, \ldots, F_{\alpha_{if_i}}$. Similarly, we denote the connected components of an α-level region G_{α_j} by $G_{\alpha_{j1}}, \ldots, G_{\alpha_{jg_j}}$. We can then define a topological predicate π'_f as

$$\pi'_f(F,G) = \sum_{i=1}^{n} \sum_{k=1}^{f_i} \sum_{j=1}^{n} \sum_{l=1}^{g_j} \frac{m(F_{\alpha_i}) \cdot m(F_{\alpha_j}) \cdot \pi_{srh}(F_{\alpha_{ik}}, G_{\alpha_{jl}})}{f_i \cdot g_j}$$

It is obvious that $\pi'_f(F,G) \ge 0$ holds since all factors have a value greater than or equal to 0. We can also show that $\pi'_f(F,G) \le 1$ by the following transformations:

$$\begin{aligned}
\pi'_f(F,G) &\le \sum_{i=1}^{n} \sum_{k=1}^{f_i} \sum_{j=1}^{n} \sum_{l=1}^{g_j} \frac{(\alpha_i - \alpha_{i+1}) \cdot (\alpha_j - \alpha_{j+1})}{f_i \cdot g_j} \\
&\quad (\text{since } \pi_{srh}(F_{\alpha_{ik}}, G_{\alpha_{jl}}) \le 1) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(\alpha_i - \alpha_{i+1}) \cdot (\alpha_j - \alpha_{j+1})}{f_i \cdot g_j} \cdot f_i \cdot g_j \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_{i+1}) \cdot (\alpha_j - \alpha_{j+1}) \\
&= 1
\end{aligned}$$

Hence, $\pi'_f(F,G) \le 1$ holds. As a rule, the predicates π_f and π'_f do not yield the same results. Assume that F_{α_i} and G_{α_j} fulfil a predicate $\pi_{cr} \in T_{cr}$. This fact contributes once to the summation process for π_f. But it does not take into account that possibly several faces $F_{\alpha_{ik}}$ (at least one) of F_{α_i} satisfy the corresponding predicate $\pi_{srh} \in T_{srh}$ with several faces $G_{\alpha_{jl}}$ (at least one) of G_{α_j}. This fact contributes several times (at most $f_i \cdot g_j$) to the summation process for π'_f. Hence, the evaluation process for π'_f is more fine-grained than for π_f.

Both generic predicate definitions reveal their quantitative character. If the predicate $\pi_{cr}(F_{\alpha_i}, G_{\alpha_j})$ and the predicate $\pi_{srh}(F_{\alpha_{ik}}, G_{\alpha_{jl}})$, respectively, is never fulfilled, the predicate $\pi_f(F,G)$ and $\pi'_f(F,G)$, respectively, yields *false*. The more α-level regions of F and G (simple regions with holes of F_{α_i} and G_{α_j}) fulfil the predicate $\pi_{cr}(F_{\alpha_i}, G_{\alpha_j})$ ($\pi_{srh}(F_{\alpha_{ik}}, G_{\alpha_{jl}})$), the more the validity of the predicate π_f (π'_f) increases. The maximum is reached if all topological predicates are satisfied.

5. PROPERTIES

An interesting issue relates to the effect of the number of α-level regions on the computation results for $\pi_f(F,G)$. What can we expect if we supplement Λ with an additional membership value? Can we make a general statement saying that the value for $\pi_f(F,G)$ will then always increase or decrease or stagnate? To answer this question, let $\pi^n_f(F,G)$ denote the predicate $\pi_f(F,G)$ if for its computation Λ contains n labels except for $\alpha_{n+1} = 0$. We now extend Λ by an additional label $\alpha_k = \alpha_{n+2} \in [0,1] \setminus \Lambda$ without rearranging the indices of the membership values. Without loss of generality we assume an $l \in \{1, \ldots, n\}$ such that $1 = \alpha_1 > \cdots > \alpha_l > \alpha_k > \alpha_{l+1} > \cdots > \alpha_{n+1} = 0$. This enables us to compute the difference $\pi^{n+1}_f(F,G) - \pi^n_f(F,G)$ and to investigate whether this difference is always greater than, less than, or equal to 0. The computation is simplified by the observation that all addends $(\alpha_i - \alpha_{i+1}) \cdot (\alpha_j - \alpha_{j+1}) \cdot \pi_{cr}(F_{\alpha_i}, G_{\alpha_j})$ of $\pi^{n+1}_f(F,G)$ and $\pi^n_f(F,G)$ with $i, j \in \{1, \ldots, n\} \setminus \{l\}$ neutralize each other. That is, for $\pi^{n+1}_f(F,G)$

$disjoint_{cr}(F,G)$	$\leq$	$disjoint_{cr}(F',G)$	$meet_{cr}(F,G)$	$\lesseqgtr$	$meet_{cr}(F',G)$
1	$=$	1	**1**	$\geq$	0 | 1
0	$\leq$	0 | 1	**0**	$\leq$	0 | 1
0 | 1	$\leq$	**1**	0 | 1	$\leq$	**1**
0	$=$	**0**	0 | 1	$\geq$	**0**
$overlap_{cr}(F,G)$	$\lesseqgtr$	$overlap_{cr}(F',G)$	$equal_{cr}(F,G)$	$\lesseqgtr$	$equal_{cr}(F',G)$
1	$\geq$	0 | 1	**1**	$\geq$	0 | 1
0	$\leq$	0 | 1	**0**	$\leq$	0 | 1
0 | 1	$\leq$	**1**	0 | 1	$\leq$	**1**
0 | 1	$\geq$	**0**	0 | 1	$\geq$	**0**
$contains_{cr}(F,G)$	$\geq$	$contains_{cr}(F',G)$	$covers_{cr}(F,G)$	$\lesseqgtr$	$covers_{cr}(F',G)$
1	$\geq$	0 | 1	**1**	$\geq$	0 | 1
0	$=$	0	**0**	$\leq$	0 | 1
1	$=$	**1**	0 | 1	$\leq$	**1**
0 | 1	$\geq$	**0**	0 | 1	$\geq$	**0**

Table 1: Evaluation of the predicate comparisons of the first class.

we have only to consider those addends having factors with an index equal to k. For $\pi_f^n(F,G)$ we have only to consider addends having factors with $i = l$ or $j = l$. With $\Delta = \pi_f^{n+1}(F,G) - \pi_f^n(F,G)$ we obtain:

$$\begin{aligned}
\Delta \;=\; & (\alpha_l - \alpha_k) \sum_{i=1}^{n} (\alpha_i - \alpha_{i+1}) \cdot \\
& \quad (\pi_{cr}(F_{\alpha_i}, G_{\alpha_l}) + \pi_{cr}(F_{\alpha_l}, G_{\alpha_i})) \\
& - (\alpha_l - \alpha_k)(\alpha_l - \alpha_{l+1}) \cdot 2 \cdot \pi_{cr}(F_{\alpha_l}, G_{\alpha_l}) \\
& + (\alpha_l - \alpha_k)(\alpha_l - \alpha_k) \cdot 2 \cdot \pi_{cr}(F_{\alpha_l}, G_{\alpha_l}) \\
& + (\alpha_k - \alpha_{l+1}) \sum_{i=1}^{n} (\alpha_i - \alpha_{i+1}) \cdot \\
& \quad (\pi_{cr}(F_{\alpha_i}, G_{\alpha_k}) + \pi_{cr}(F_{\alpha_k}, G_{\alpha_i})) \\
& - (\alpha_k - \alpha_{l+1})(\alpha_l - \alpha_{l+1}) \cdot \\
& \quad (\pi_{cr}(F_{\alpha_k}, G_{\alpha_l}) + \pi_{cr}(F_{\alpha_l}, G_{\alpha_k})) \\
& + (\alpha_k - \alpha_{l+1})(\alpha_l - \alpha_k) \cdot 2 \cdot \\
& \quad (\pi_{cr}(F_{\alpha_k}, G_{\alpha_l}) + \pi_{cr}(F_{\alpha_l}, G_{\alpha_k})) \\
& + (\alpha_k - \alpha_{l+1})(\alpha_k - \alpha_{l+1}) \cdot 2 \cdot \pi_{cr}(F_{\alpha_k}, G_{\alpha_k}) \\
& - (\alpha_l - \alpha_{l+1}) \sum_{i=1}^{n} (\alpha_i - \alpha_{i+1}) \cdot \\
& \quad (\pi_{cr}(F_{\alpha_i}, G_{\alpha_l}) + \pi_{cr}(F_{\alpha_l}, G_{\alpha_i}))
\end{aligned}$$

The first line computes the sum of all addends having the factor $\alpha_l - \alpha_k$. The fourth line does the same for all addends having the factor $\alpha_k - \alpha_{l+1}$. Unfortunately, these sums include addends having the factor $\alpha_l - \alpha_{l+1}$ so that these addends have to be subtracted (second and fifth line). The third, sixth, and seventh line insert the correct addends. The eighth line subtracts all those addends of $\pi_f^n(F,G)$ having the factor $\alpha_l - \alpha_{l+1}$. The whole expression can be restructured as follows:

$$\begin{aligned}
\Delta \;=\; & -(\alpha_k - \alpha_{l+1}) \sum_{i=1}^{n} (\alpha_i - \alpha_{i+1}) \cdot \\
& \quad (\pi_{cr}(F_{\alpha_i}, G_{\alpha_l}) + \pi_{cr}(F_{\alpha_l}, G_{\alpha_i})) \\
& + (\alpha_k - \alpha_{l+1}) \sum_{i=1}^{n} (\alpha_i - \alpha_{i+1}) \cdot \\
& \quad (\pi_{cr}(F_{\alpha_i}, G_{\alpha_k}) + \pi_{cr}(F_{\alpha_k}, G_{\alpha_i})) \\
& - (\alpha_k - \alpha_{l+1})(\alpha_k - \alpha_{l+1}) \cdot \\
& \quad (\pi_{cr}(F_{\alpha_k}, G_{\alpha_l}) + \pi_{cr}(F_{\alpha_l}, G_{\alpha_k})) \\
& - (\alpha_k - \alpha_{l+1})(\alpha_l - \alpha_k) \cdot 2 \cdot \pi_{cr}(F_{\alpha_l}, G_{\alpha_l}) \\
& + (\alpha_k - \alpha_{l+1})(\alpha_l - \alpha_k) \cdot \\
& \quad (\pi_{cr}(F_{\alpha_k}, G_{\alpha_l}) + \pi_{cr}(F_{\alpha_l}, G_{\alpha_k})) \\
& + (\alpha_k - \alpha_{l+1})(\alpha_k - \alpha_{l+1}) \cdot 2 \cdot \pi_{cr}(F_{\alpha_k}, G_{\alpha_k}) \\
=\; & (\alpha_k - \alpha_{l+1}) \cdot \\
& \Big[\sum_{i=1}^{n} (\alpha_i - \alpha_{i+1}) \cdot (\pi_{cr}(F_{\alpha_k}, G_{\alpha_i}) - \pi_{cr}(F_{\alpha_l}, G_{\alpha_i}) + \\
& \quad \pi_{cr}(F_{\alpha_i}, G_{\alpha_k}) - \pi_{cr}(F_{\alpha_i}, G_{\alpha_l})) \\
& + (\alpha_l - \alpha_k) \cdot (\pi_{cr}(F_{\alpha_l}, G_{\alpha_k}) - \pi_{cr}(F_{\alpha_l}, G_{\alpha_l}) + \\
& \quad \pi_{cr}(F_{\alpha_k}, G_{\alpha_l}) - \pi_{cr}(F_{\alpha_l}, G_{\alpha_l})) \\
& + (\alpha_k - \alpha_{l+1}) \cdot (\pi_{cr}(F_{\alpha_k}, G_{\alpha_k}) - \pi_{cr}(F_{\alpha_k}, G_{\alpha_l}) + \\
& \quad \pi_{cr}(F_{\alpha_k}, G_{\alpha_k}) - \pi_{cr}(F_{\alpha_l}, G_{\alpha_k})) \Big]
\end{aligned}$$

For a comparison of Δ with 0 we can observe that the values of the factors $\alpha_k - \alpha_{l+1}$, $\alpha_i - \alpha_{i+1}$ $(1 \leq i \leq n)$, $\alpha_l - \alpha_k$, and $\alpha_k - \alpha_{l+1}$ are all greater than 0. Hence, the result only depends on the predicate values. Analyzing the six differences of predicates, we can group them into two classes. The first class contains the differences $\pi_{cr}(F_{\alpha_k}, G_{\alpha_i}) - \pi_{cr}(F_{\alpha_l}, G_{\alpha_i})$, $\pi_{cr}(F_{\alpha_k}, G_{\alpha_l}) - \pi_{cr}(F_{\alpha_l}, G_{\alpha_l})$, and $\pi_{cr}(F_{\alpha_k}, G_{\alpha_k}) - \pi_{cr}(F_{\alpha_l}, G_{\alpha_k})$. Their common structure is $\pi_{cr}(F,G) - \pi_{cr}(F',G)$ with $F' \subseteq F$. The second class contains the remaining differences $\pi_{cr}(F_{\alpha_i}, G_{\alpha_k}) - \pi_{cr}(F_{\alpha_i}, G_{\alpha_l})$, $\pi_{cr}(F_{\alpha_l}, G_{\alpha_k}) - \pi_{cr}(F_{\alpha_l}, G_{\alpha_l})$, and $\pi_{cr}(F_{\alpha_k}, G_{\alpha_k}) - \pi_{cr}(F_{\alpha_k}, G_{\alpha_l})$. Their common structure is $\pi_{cr}(F,G) - \pi_{cr}(F,G')$ with $G' \subseteq G$.

Table 1 shows the result of comparing the predicates involved in the differences of the first class. For each predicate combination $\pi_{cr}(F,G)$ and $\pi_{cr}(F',G)$ we first set $\pi_{cr}(F,G)$ and then $\pi_{cr}(F',G)$ both to 1 (true) and 0 (false) (written in bold font). Afterwards we determine the result of the respective other predicates and thus obtain four pairs of values. For instance, if $disjoint_{cr}(F,G)$ is equal to **0**, $disjoint_{cr}(F',G)$ is either equal to 0 or to 1 (indicated by the expression "0 | 1"). Next, we assign the correct comparison operator $=$, $\leq$, or $\geq$ reflecting the relationship of a pair of values to each of the four cases. In the end, we form the combination of the four comparison operators and obtain the relationship between $\pi_{cr}(F,G)$ and $\pi_{cr}(F',G)$. The symbol $\lesseqgtr$ indicates that the equality or inequality of the two predicates cannot be generally decided. The only solution here is to compute it for each single case. We have omitted the comparisons for *inside* and *coveredBy*, since they are inverse to *contains* and *covers*, respectively. That is, $inside_{cr}(F,G) \leq inside_{cr}(F',G)$ and $coveredBy_{cr}(F,G) \lesseqgtr coveredBy_{cr}(F',G)$.

For an investigation of the second class we do not have to consider the predicates $disjoint_{cr}$, $meet_{cr}$, $overlap_{cr}$, and $equal_{cr}$; they are symmetric in their arguments, that is, $\pi_{cr}(F,G) = \pi_{cr}(G,F)$. We also need not consider the predicates $covers_{cr}$ and $coveredBy_{cr}$, since already their predicate combinations in the first class cannot be decided generally. Thus, only the predicates $contains_{cr}$ and the inverse $inside_{cr}$ remain. For $contains_{cr}$ we obtain Table 2.

$contains_{cr}(F,G)$	$\leq$	$contains_{cr}(F,G')$
1	=	1
0	$\leq$	0 | 1
0 | 1	$\leq$	**1**
0	=	**0**

Table 2: Evaluation of the *contains* comparison of the second class.

In summary, we obtain a rather negative result. Since the differences for $meet_{cr}$, $overlap_{cr}$, $equal_{cr}$, $covers_{cr}$, and $coveredBy_{cr}$ cannot be decided in general, no general statement can be made about the difference between $\pi_f^{n+1}(F,G)$ and $\pi_f^n(F,G)$ for these predicates. The computation of this difference also fails for $contains_{cr}$ and $inside_{cr}$. The problem is that the behavior of both predicates in the first and second class is opposite to each other. That is, in the first class $contains_{cr}(F,G) - contains_{cr}(F',G) \geq 0$ and in the second class $contains_{cr}(F,G) - contains_{cr}(F,G') \leq 0$ hold. For the computation of Δ these two differences are added, and we cannot generally decide whether the result is less than, greater than, or equal to 0. Hence, all these predicates do not satisfy some kind of "monotonicity criterion". A positive exception is only the predicate $disjoint_{cr}$ for which we obtain $disjoint_f^{n+1}(F,G) \leq disjoint_f^n(F,G)$.

6. QUERYING WITH FUZZY TOPOLOGICAL PREDICATES

In this section we demonstrate how fuzzy topological predicates can be integrated into an SQL-like spatial query language. The fact that the membership degree yielded by a fuzzy topological predicate is a computationally determined quantification between 0 and 1, i.e., a fuzzy boolean, impedes a direct integration. First, it is not very comfortable and user-friendly to use such a numeric value in a query. Second, spatial selections and spatial joins expect crisp predicates as filter conditions and are not able to cope with fuzzy predicates.

As a solution, we propose to embed adequate qualitative linguistic descriptions of nuances of topological relationships as appropriate interpretations of the membership values into a spatial query language. For instance, depending on the membership value yielded by the predicate $inside_f$, we could distinguish between *not* inside, *a little bit* inside, *somewhat* inside, *slightly* inside, *quite* inside, *mostly* inside, *nearly completely* inside, and *completely* inside. These fuzzy linguistic terms can then be incorporated into spatial queries together with the fuzzy predicates they modify. We call these terms *fuzzy quantifiers*, because their semantics lies between the universal quantifier *for all* and the existential quantifier *there exists*. It is conceivable that a fuzzy quantifier is either predefined and anchored in the query language, or user-defined.

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

$$\gamma\pi_f(F,G) = true \quad :\Leftrightarrow \quad (\mu_\gamma \circ \pi_f)(F,G) = 1$$

That is, only for those values of $\pi_f(F,G)$ for which μ_γ yields 1, the predicate $\gamma\pi_f$ is true. A membership function that fulfils this quite strict condition is, for instance, the crisp partition of $[0,1]$ into $|\Gamma|$ disjoint or adjacent intervals completely covering $[0,1]$ and the assignment of each interval to a fuzzy quantifier. If an interval $[a,b]$ is assigned to a fuzzy quantifier γ, the intended meaning is that $\mu_\gamma(\pi_f(F,G)) = 1$, if $a \leq \pi_f(F,G) \leq b$, and 0 otherwise. For example, we 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 by the fuzzy sets in Figure 2. While we can always find a fitting fuzzy quantifier for the partition due to the complete coverage of the interval $[0,1]$, this is not necessarily the case here. Each fuzzy quantifier is associated with a *fuzzy number* having a trapezoidal-shaped membership function. The transition between two consecutive fuzzy quantifiers is smooth and here modeled by linear functions. Within a fuzzy transition area μ_γ yields a value less than 1 which makes the predicate $\gamma\pi_f$ false. Examples in Figure 2 can be found at 0.2, 0.5, or 0.8. Each fuzzy number associated with a fuzzy 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 2 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 $\gamma\pi_f$ is only true if μ_γ yields 1. We can relax this strict condition by defining:

$$\gamma\pi_f(F,G) = true \quad :\Leftrightarrow \quad (\mu_\gamma \circ \pi_f)(F,G) > 0$$

In a crisp spatial database system this gives us the chance also to take the transition zones into account and to let them make the predicate $\gamma\pi_f$ true. When evaluating a fuzzy spatial selection or join in a fuzzy spatial database system, we can even set up a weighted ranking of database objects satisfying the predicate $\gamma\pi_f$ at all and being ordered by descending membership degree $1 \geq \mu_\gamma > 0$.

A special, optional fuzzy quantifier, denoted by *at all*, represents the existential quantifier and checks whether a predicate π_f 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 $x \in [0,1]$.

The following few example queries demonstrate how fuzzy spatial data types and quantified fuzzy 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 fuzzy regions as attribute values.

What we need first is a mechanism to declare user-defined fuzzy quantifiers and to activate predefined or user-defined fuzzy quantifiers. This mechanism should allow to specify trapezoidal-shaped and triangular-shaped membership functions as well as crisp partitions. In general, this means to define a *classification*, which could be expressed in the following way:

```
create classification fq
  (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))
```

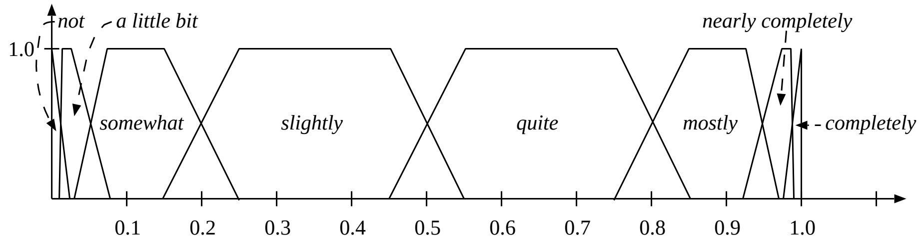

Figure 2: Membership functions for fuzzy quantifiers.

Such a classification could then be activated by

set classification fq

Assuming that we have a relation *pollution*, which stores among other things the blurred geometry of polluted zones as fuzzy regions, and a relation *areas*, which keeps information about the use of land areas and which stores their vague spatial extent as fuzzy 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 *fuzzy spatial joins*.

Another query could ask 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 vague extent is represented as a fuzzy 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 fuzzy regions and aggregates this set by repeated application of fuzzy geometric union):

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

7. CONCLUSIONS

We have presented a definition of topological predicates on fuzzy regions. We have shown that all these predicates with one exception do not fulfil some kind of "monotonicity criterion" which documents the independence of topological and metric properties. Moreover, we have sketched the integration of these predicates into fuzzy spatial query languages. For that purpose, fuzzy quantifiers are used that can be incorporated into spatial queries.

8. REFERENCES

[1] D. Altman. Fuzzy Set Theoretic Approaches for Handling Imprecision in Spatial Analysis. *Int. Journal of Geographical Information Systems*, 8(3):271–289, 1994.

[2] T. Behr and M. Schneider. Topological Relationships of Complex Points and Complex Regions. *Int. Conf. on Conceptual Modeling*, 2001.

[3] P. A. Burrough and A. U. Frank, editors. *Geographic Objects with Indeterminate Boundaries*. GISDATA Series, vol. 2. Taylor & Francis, 1996.

[4] D. Dubois and M.-C. Jaulent. A General Approach to Parameter Evaluation in Fuzzy Digital Pictures. *Pattern Recognition Letters*, pp. 251–259, 1987.

[5] S. Dutta. Qualitative Spatial Reasoning: A Semi-Quantitative Approach Using Fuzzy Logic. *1st Int. Symp. on the Design and Implementation of Large Spatial Databases*, LNCS 409, pp. 345–364. Springer-Verlag, 1989.

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

[7] S. Gaal. *Point Set Topology*. Academic Press, 1964.

[8] R. H. Güting. An Introduction to Spatial Database Systems. *VLDB Journal*, 3(4):357–399, 1994.

[9] M. Schneider. *Spatial Data Types for Database Systems - Finite Resolution Geometry for Geographic Information Systems*, volume LNCS 1288. Springer-Verlag, Berlin Heidelberg, 1997.

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

[11] M. Schneider. Finite Resolution Crisp and Fuzzy Spatial Objects. *Int. Symp. on Spatial Data Handling*, pp. 5a.3–17, 2000.

[12] M. Schneider. A Design of Topological Predicates for Complex Crisp and Fuzzy Regions. *Int. Conf. on Conceptual Modeling*, 2001.

[13] E. Stefanakis, M. Vazirgiannis, and T. Sellis. Incorporating Fuzzy Logic Methodologies into GIS Operations. *Int. Conf. on Geographical Information Systems in Urban, Regional and Environmental Planning*, pp. 61–68, 1996.

[14] F. Wang, G. B. Hall, and Subaryono. Fuzzy Information Representation and Processing in Conventional GIS Software: Database Design and Application. *Int. Journal of Geographical Information Systems*, 4(3):261–283, 1990.