Simplifying Sets of Events by Selecting Temporal Relations

Andrea Rodríguez[1], Nico Van de Weghe[2], and Philippe De Maeyer[2]

[1] Department of Computer Science, University of Concepción,
Center for Web Research, University of Chile, Chile
`andrea@udec.cl`
[2] Department of Geography, Ghent University, Belgium
{`nico.vandeweghe,philippe.demaeyer`}`@ugent.be`

Abstract. Reasoning about events or temporal aspects is fundamental for modeling geographic phenomena. This work concerns the analysis of events as configurations of temporal intervals. It presents two strategies to select relations that characterize configurations of temporal intervals: a strategy based on the algebraic property of composition and a strategy based on a neighboring concept in a vector representation. This type of analysis is useful for characterizing sets of events without the need of making an exhaustive specification of all temporal relations. This work complements a previous study about topological relations of regions in a 2D space and confirms the potential of using the algebraic properties of composition and the metric characteristics of intervals, even if only qualitative relations are considered.

1 Introduction

The ability to understand, represent, and manage temporal knowledge about the world is fundamental in humans and artificial agents [19, 22]. Reasons why time should be included in Geographic Information Systems (GISs) have already been discussed [13]. Such reasons are mostly associated with the fact that many applications involving the description of geographic phenomena require the treatment of dynamic aspects that are related to space and time. Examples of such applications are transportation and urban analysis, and the analysis of physical phenomena.

The temporal data model currently by far the most frequently used for understanding dynamic processes in the real world is based on the linear concept. In this concept, temporal intervals are projected onto a one-dimensional conceptual space by a one-dimensional segment. Relations between these intervals are called temporal relations. Allen [1] defined thirteen fine temporal relations, which later were extended by the sixteen coarse relations [8]. The Allen's basic thirteen relations define a relation algebra [23], which implies that the set of temporal relations is complete, each relation has a converse relation within the set, and there exists a composition operation that results in one or more relations within the set. The composition operation creates a reasoning mechanism that

allows one to derive the relation between two intervals through the intervals' combination with a common third interval.

This work concerns the temporal analysis of sets of events described by configurations of temporal intervals. In this paper we talk about events as something that occurs or happens at a particular time interval, not only at a specific point in time. Thus, we do not make a distinction between events, processes, states or actions. We consider events as everything that happens and is then gone. This work abstracts the semantics of events and focuses on temporal relations between events. In the spatial domain, the same idea is applied when one abstracts the meaning of spatial features and concentrates on spatial relations between these features. Such an abstraction has been useful for content-based retrieval where queries are expressed by sketches [5, 20]. In this work, however, sketches apply to the temporal domain. For example, if we have a database that stores the information about a diseases that occurred in a geographic area, one could be interested in searching for diseases that co-occur, diseases that precede or start to appear at the same time, and so on. In this search, the concern is not about the absolute positions of events on time, but rather on the relations between temporal intervals. This search could be extended to more than two diseases to become a search of a set of diseases, called in this paper a configuration of temporal intervals. The way the configurations of temporal intervals are expressed in a query language is out of the scope of this paper and left for future work.

The paper presents two strategies to select relations that characterize configurations of temporal intervals: a composition-based and a neighborhood-based strategy. The selection of temporal relations aims at minimizing the number of relations that are needed to characterize the temporal arrangement of events. Such temporal relations are the basis for content-based retrieval, where these relations represent query criteria or query constraints. For example, given three different events, we can have three binary temporal relations among them, without considering *equal* and *converse* relations. This study explores strategies that indicate whether or not one needs the three relations to characterize the temporal arrangement of events. The hypothesis of this work is that by minimizing the number of constraints in a query evaluation, one decreases the computational cost of content-based retrieval. In addition to applications in content-based retrieval with temporal criteria, this work can be interesting for compressing information about relations between events, or in discovering dynamic patterns by focusing on subsets of temporal relations. Associated with spatial information, where topological relations describe constraints about the spatial arrangement of objects [3, 5, 11, 20], selecting temporal relations may have an impact not only on modeling but also representing dynamic phenomena.

The composition-based strategy relies on the basic derivation property of composition operations, whereas the neighborhood-based strategy exploits the closeness of temporal intervals to select relations between intervals. In this sense, this work follows closely ideas from the work by Rodríguez *et al.* [20], which defines and compares strategies for query pre-processing of topological relations in a 2D space; however, to the best of our knowledge, there has not yet been

done some research in the temporal domain in this direction. Thus, a goal of this work was to explore how well strategies that were applied to topological relations between regions in 2D space could be applied to temporal intervals.

The organization of this paper is as follows: Section 2 describes temporal relations and their composition. Section 3 presents the composition- and neighborhood-based strategies for simplifying configurations of temporal intervals. Section 4 evaluates the applicability of these strategies for content-based retrieval of configurations of temporal intervals. Conclusions and future work are presented in Section 5.

2   Temporal Relations

The representation of time by means of intervals rather than points has a history in philosophical studies of time [9, 10]. In 1983, Allen [1] defined a calculus of time intervals as a representation of temporal knowledge that could be used in artificial intelligence. Nowadays, many researchers from different disciplines still use Allen's thirteen temporal relations. Allen's Interval Algebra provides a rich formalism for expressing qualitative relations between temporal intervals. Allen's interval algebra uses the notion of binary relations between convex intervals, which are intervals without gaps. An interval I is represented as a pair $[I-; I+]$ of real numbers with $I- < I+$, denoting the left and right end points of the interval, respectively. This means that Allen only deals with pure intervals without considering point intervals.

Let the beginnings and endings of two events have three possible relations: *smaller* ($<$), *equal* ($=$), and *larger* ($>$). Then events have thirteen possible qualitative relations. The relations that are shown in Table 1 capture the qualitative aspect of event pairs as *before*, *meets*, *overlaps*, *starts*, *finishes*, *during*, and *equal*, in terms of constraints on the end points of the constituent temporal intervals. These basic temporal relations have their corresponding converse relations, with *equal* being self-converse. The symbolic representation of temporal relations used in this paper modifies the proposal by Kulpa [12]. It considers symbols that are based on two horizontal lines, representing the two intervals, and one or two vertical lines, representing the temporal topology.

Like the composition table of topological relations between regions [4], there exists a composition table for the temporal relations between intervals (Table 2). In this table, a crisp result of a composition r_{ik} and (;) r_{kj} gives a unique possible relation r_{ij}. For example, the composition I_1 is *before* I_2 ; I_2 *contains* I_3 (⌐_ ; ⊤) has the crisp result that I_1 is *before* I_3 (⌐_). When the composition does not give any information (ni), all relations are possible.

Although there is a significant analogy between spatial and temporal relations, there are certain important differences. For example, when working with temporal intervals one has to differentiate between X *meets* Y and Y *meets* X, since time has a unidirectional nature.

4

Table 1. Topological temporal relations between two pure intervals

Relation	Symbol	Conditions	Converse
I_1 *before* I_2	⌐_	$I_1+ < I_2-$	*after* (_⌐)
I_1 *meets* I_2	⌐L	$I_1+ = I_2-$	*met_by* (_⌐)
I_1 *overlaps* I_2	⊓L	$I_2- > I_1- \wedge I_1+ < I_2+ \wedge I_1+ > I_2-$	*overlapped_by* (⊓⌐)
I_1 *starts* I_2	⌊L	$I_1- = I_2- \wedge I_1+ < I_2+$	*started_by* (⌊⌐)
I_1 *finishes* I_2	_⌐⌐	$I_1+ = I_2+ \wedge I_1- > I_2-$	*finished_by* (⌐⌐)
I_1 *during* I_2	⊓L	$I_1- > I_2- \wedge I_1+ < I_2+$	*contains* (⊓⌐)
I_1 *equal* I_2	⬜	$I_1- = I_2- \wedge I_1+ = I_2+$	*equal*

Table 2. The composition table for the thirteen relations

3 Simplifying Configurations of Temporal Intervals

A configuration of temporal intervals is considered to be a set of temporal intervals that holds particular temporal relations. In this context, a configuration can be seen as a graph G that is complete, directed and labeled, where nodes N are temporal intervals and arcs A are binary relations between these intervals (Equation 1).

$$G = (N, A)$$
$$N = \{I_1, I_2, \ldots, I_n\}$$
$$A = \{r_{ij} | \forall i, j \in [1..n], \exists I_i, I_j \in N\} \tag{1}$$

A graph G that describes a set of n intervals can be represented as a matrix of $n \times n$ elements, where these elements identify binary temporal relations r_{ij}. Elements r_{ii} along the diagonal are the identity relation (i.e., *equal*) (Figure 1).
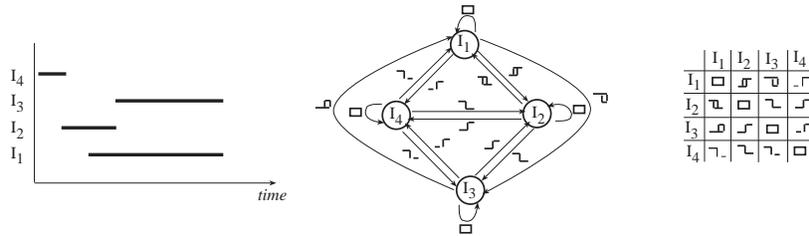
Fig. 1. Temporal configurations as a directed graph represented by an $n \times n$ matrix

The idea of simplifying configurations of temporal intervals aims at determining the subset of temporal relations that can characterize temporal configurations or events. Thus, the idea is to reduce the number of temporal relations without (significantly) reducing the amount of information conveyed by the whole set of relations. To do so, two different strategies are analyzed: (1) composition-based and (2) neighborhood-based strategies.

3.1 Composition-Based Strategy

The composition-based strategy (CBS) for analyzing configurations of temporal intervals is solely based on the composition of binary relations so that no explicit information about the duration and absolute moment on time of intervals are needed. CBS starts with a graph representing a configuration of temporal intervals and selects a minimal subgraph from which one can derive the complete and original graph without losing information about the relations between temporal intervals (i.e., relations defined in Table 1). The strategy for finding this subgraph follows the principles of logical consistency in a graph [14].

Logical consistency: Logical consistency is expressed by the composition of binary relations. Given a configuration expressed as a graph, consistency is formulated as a constraint satisfaction problem [16] over a network of binary relations [15]. To be a consistency network of binary relations, a graph must fulfill three constraints: node-consistency, arc-consistency, and path-consistency.

- Node-consistency: each node must have a self-loop arc, denoting the identity relation.
- Arc-consistency: for each directed arc there must be an arc in the reverse direction, denoting the converse binary relation.
- Path-consistency: although a variety of paths can lead from one node to another, in order to infer the path consistency of a relation it is sufficient to consider all composition paths of length two that connect the relation between nodes. Having a consistent graph, a relation must coincide with its induced relation determined by the intersection of all possible composition paths of length two (Equation 2, where ';' represents the composition operator).

$$\forall_{i,j} r_{ij} = \bigcap_{k=1}^{n} r_{ik}; r_{kj} \tag{2}$$

Following the principle of path consistency, a relation can be completely derived if, and only if, it is the only relation that results from the intersection of all possible composition paths of length two in the graph. Consider the example in Figure 2 with three temporal intervals. The graph in the figure corresponds to a node-, arc-, and path-consistent network. The path consistency is checked with the determination of the nine induced relations (Table 3).
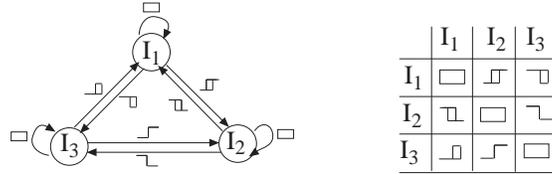
Fig. 2. Node-, arc-, path- consistency network of temporal relations

Composition-based simplification: Unlike consistency checking [6], where one is interested in completing a partial graph, a strategy for simplifying configurations of temporal intervals starts from a complete and consistent graph and eliminates relations that can be consistently derived. In such a process of simplifying temporal configurations, it is important to analyze whether or not by eliminating relations we obtain a unique minimal subgraph.

The work in [20] showed that there exists one minimal subgraph that results from the application of the composition-based strategy to simplify topological configurations of regions in a 2D space. In the temporal domain with thirteen interval relations, in contrast, there may be more than one minimal subgraph. A

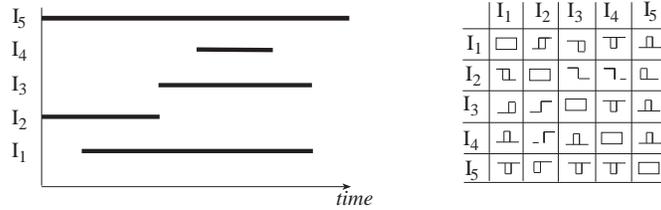**Table 3.** All possible derivations from path-consistency

Relation	Derivation
$r_{11} = \square$	$r_{12}; r_{21} \bigcap r_{11}; r_{11} \bigcap r_{13}; r_{31} = (\ \sqcup \lor \ \sqcap \lor \ \square \lor \ \sqsubset \lor \ \sqsupset \lor \ \sqcap \lor \ \sqcup \lor \ \lrcorner \lor \ \square \ )$
	$\bigcap \ \square \ \bigcap \ (\ \lrcorner \lor \ \square \lor \ \square \ )$
$r_{12} = \ \sqcap$	$r_{11}; r_{12} \bigcap r_{12}; r_{22} \bigcap r_{13}; r_{32} = \ \sqcap \ \bigcap \ \sqcap \ \bigcap \ (\ \sqcap \lor \ \sqcap \lor \ \sqsupset \ )$
$r_{13} = \ \square$	$r_{11}; r_{13} \bigcap r_{12}; r_{23} \bigcap r_{13}; r_{33} = \ \square \ \bigcap \ (\ \sqcap \lor \ \sqcap \lor \ \square \ ) \bigcap \ \square$
$r_{21} = \ \sqcup$	$r_{21}; r_{11} \bigcap r_{22}; r_{21} \bigcap r_{23}; r_{31} = \ \sqcup \ \bigcap \ \sqcup \ \bigcap \ (\ \sqcup \lor \ \sqcup \lor \ \sqsubset \ )$
$r_{22} = \square$	$r_{21}; r_{12} \bigcap r_{22}; r_{22} \bigcap r_{23}; r_{32} = (\ \sqcup \lor \ \sqcap \lor \ \square \lor \ \sqcap \lor \ \sqcap \lor \ \sqsubset \lor \ \sqsupset \lor \ \square \lor \ \square \ )$
	$\bigcap \ \square \ \bigcap \ (\ \lrcorner \lor \ \square \lor \ \square \ )$
$r_{23} = \ \llcorner$	$r_{21}; r_{13} \bigcap r_{22}; r_{23} \bigcap r_{23}; r_{33} = (\ \llcorner \lor \ \sqcup \lor \ \llcorner \ ) \bigcap \ \llcorner \ \bigcap \ \llcorner$
$r_{31} = \ \square$	$r_{31}; r_{11} \bigcap r_{32}; r_{21} \bigcap r_{33}; r_{31} = \ \square \ \bigcap \ (\ \sqcap \lor \ \sqcup \lor \ \square \ ) \bigcap \ \square$
$r_{32} = \ \lrcorner$	$r_{31}; r_{12} \bigcap r_{32}; r_{22} \bigcap r_{33}; r_{32} = (\ \llcorner \lor \ \sqcap \lor \ \lrcorner \ ) \bigcap \ \lrcorner \ \bigcap \ \lrcorner$
$r_{33} = \square$	$r_{31}; r_{13} \bigcap r_{32}; r_{23} \bigcap r_{33}; r_{33} = (\ \square \lor \ \square \lor \ \square \ ) \bigcap \ (\ \square \lor \ \square \lor \ \square \ ) \bigcap \ \square$

reason for having more than one minimal subgraph is the asymmetric property of temporal relations, which has an effect on the composition of relations. Unlike the spatial domain of 2D regions, where the relation *overlaps* is a relation whose participation in a composition does not result in a crisp result, in the temporal domain each relation participates in at least one composition with a crisp result. Both *overlaps* and *overlapped_by* relations in the temporal domain participate in three different compositions with crisp results. Thus, in the temporal domain, if either *overlaps* or *overlapped_by* results from the intersection of composition paths of length two, these relations may still be used in other composition paths that derive other relations. For instance, consider a configuration with four temporal intervals and the derivable relations presented in Table 4. The relation r_{ij} is derived from the intersection of two composition paths of length two. One of these compositions involves the relation r_{ik}. In addition, r_{ik} can be derived by using a composition path that involves r_{ij}. Thus, one could use r_{ij} or r_{ik} in the composition path and get a crisp result.

Table 4. Example of alternative derivable relations

Relation	Derivation
$r_{ij} = \ \sqcap$	$r_{il}; r_{lj} \bigcap r_{ik}; r_{kj} = \ \square \ ; \ \lrcorner \ \bigcap \ \sqcup \ ; \ \sqsupset$
$r_{ik} = \ \sqcup$	$r_{ij}; r_{jk} \bigcap r_{il}; r_{lk} = \ \sqcap \ ; \ \sqcup \ \bigcap \ \square \ ; \ \sqcup$

A more complete example illustrates how the graph of a configuration of temporal intervals, represented as a matrix, can be reduced to a minimum number of relations. Figure 3 shows the set of temporal intervals and the corresponding matrix representation of the initial consistency network.

	I_1	I_2	I_3	I_4	I_5
I_1					
I_2					
I_3					
I_4					
I_5					

time

Fig. 3. Temporal configuration with the matrix representation of its consistency network

Table 5 shows the possible derivations from path consistency that result in crisp results. For example, the derivation of the relation between intervals I_1 and I_2 (r_{12}) is achieved by the intersection of the sets of relations that result from the compositions $r_{13}; r_{32}$, $r_{14}; r_{42}$, and $r_{15}; r_{52}$. Unlike this case, the intersection of the results of the compositions $r_{12}; r_{23}$, $r_{14}; r_{43}$, and $r_{15}; r_{53}$ is equivalent to the set $\{\,\sqcup, \sqcap, \sqsupset\,\}$ (i.e., any of the relations in the set is a possible result) and, therefore, this intersection does not uniquely derive the relation r_{13} ($\sqsupset$). Cases when the intersections of composition paths do not result in crisp results are not included in Table 5.

Table 5 shows that there are seven possible derivations from path consistency; however, it is not possible to eliminate all of these relations. The derivation of r_{12} requires the relation r_{14} and, vice versa, the derivation of r_{14} requires the relation r_{12}. So, we cannot eliminate both relations. The same situation occurs with relations r_{14} and r_{34}, r_{12} and r_{15}, and r_{24} and r_{34}. Note in Table 5 that to derive relation r_{35}, one just needs the composition between relations r_{31} and r_{15}, since only this composition gives a crisp result.

A minimal subgraph will eliminate the maximum number of relations that are derivable, while keeping the minimum number of relations that are needed to completely determine the original consistency network. In this example, there are two possible combinations of derivable relations obtained from node, arc- and path consistency. These combinations create minimal subgraphs with six relations (Figure 4).

A strategy for selecting a minimal subgraph is to add heuristics in the search of minimal subgraphs. A basic heuristics is to choose the subgraph that contains more relevant relations, where relevance is a qualitative order of relations. For example, intervals that are far apart may be considered less important. Such a heuristics follows the classic principle of geography, which establishes a stronger

Table 5. All possible derivations from path-consistency

Relation	Derivation
$r_{12} = $ ⊐	$r_{13}; r_{32} \bigcap r_{14}; r_{42} \bigcap r_{15}; r_{52} = ($ ⊏ $\vee$ ⊐ $\vee$ ⊔$)\bigcap ($ ⌐ $\vee$ ⊐ $\vee$ ⊔ $\vee$ ⌐ $\vee$ ⊏ $)$ $\bigcap ($ ⌐ $\vee$ ⊐ $\vee$ ⊔ $\vee$ ⌐ $\vee$ ⊐ $)$
$r_{14} = $ ⊔	$r_{12}; r_{24} \bigcap r_{13}; r_{34} \bigcap r_{15}; r_{54} = ($ ⊓ $\vee$ ⊔ $\vee$ ⊔ $\vee$ ⊔ $\vee$ ⊓$) \bigcap ($ ⊔ $) \bigcap (ni)$
$r_{15} = $ ⊥	$r_{12}; r_{25} \bigcap r_{13}; r_{35} \bigcap r_{14}; r_{43} = ($ ⊐ $\vee$ ⊥ $\vee$ ⊐$) \bigcap ($ ⊔ $\vee$ ⊥ $\vee$ ⊔$) \bigcap ($ ⊔ $\vee$ ⊐ $\vee$ ⊥ $\vee$ ⊔ $\vee$ ⊐ $\vee$ ⊔ $\vee$ ⊏ $\vee$ ⊓ $\vee$ □ $)$
$r_{24} = $ ⊓	$r_{21}; r_{14} \bigcap r_{23}; r_{34} \bigcap r_{25}; r_{54} = ($ ⊓ $\vee$ ⊔ $\vee$ ⊔ $\vee$ ⊔ $\vee$ ⊓$) \bigcap ($ ⊓ $) \bigcap ($ ⊓ $\vee$ ⊔ $\vee$ ⊔ $\vee$ ⊔ $\vee$ ⊓ $)$
$r_{34} = $ ⊔	$r_{31}; r_{14} \bigcap r_{32}; r_{24} \bigcap r_{35}; r_{54} = ($ ⌐ $\vee$ ⊐ $\vee$ ⌐ $\vee$ ⊔ $\vee$ ⊏$) \bigcap (ni) \bigcap ($ ⊓ $\vee$ ⊔ $\vee$ ⊔ $\vee$ ⊔ $\vee$ ⊓ $)$
$r_{35} = $ ⊥	$r_{31}; r_{15} \bigcap r_{32}; r_{25} \bigcap r_{34}; r_{45} = ($ ⊥ $) \bigcap ($ ⊐ $\vee$ ⊥ $\vee$ ⊐$) \bigcap ($ ⊔ $\vee$ ⊐ $\vee$ □ $\vee$ ⊔ $\vee$ ⊏ $\vee$ ⊥ $\vee$ ⊓ $\vee$ ⊐ $\vee$ ⊓ $)$
$r_{45} = $ ⊥	$r_{41}; r_{15} \bigcap r_{42}; r_{25} \bigcap r_{43}; r_{35} = ($ ⊥ $) \bigcap ($ ⌐ $\vee$ ⊐ $\vee$ ⌐ $\vee$ ⊥ $\vee$ ⊐$) \bigcap ($ ⊥ $)$

	I_1	I_2	I_3	I_4	I_5
I_1			⊓	⊔	⊥
I_2			⊔	⊓	⊔
I_3					
I_4					
I_5					

	I_1	I_2	I_3	I_4	I_5
I_1		⊐	⊓		⊥
I_2			⊔		⊔
I_3				⊔	
I_4					
I_5					

Fig. 4. Consistent minimal subgraphs derived from node-, arc-, and path-consistency

connection between close objects [24]. Another strategy may consider to sort relations based on the information content of the relations in a context. Following the argumentation from information theory [21], information content is defined in terms of the uncertainty that the data eliminate, that is, it is negative to the likelihood or frequency of the data. For example, within a large set of intervals, the use of a less frequent relation reduces the uncertainty of which intervals hold such relation. In this work, we selected the subgraph with a larger number of *non-disjoint* relations, considering that these relations are less frequent in data sets [7].

3.2 Neighborhood-Based Strategy

The Neighborhood-Based Strategy (NBS) is a quantitative technique that concentrates on closely related temporal interval points. Making an analogy to Tobler's First Law of Geography [24] "everything is related to everything else, but nearby things are more related than distant things," this work explores a temporal closeness, that is, "everything is related to everything else, but nearby things in time are more related than distant things in time." Very important here is to say that implicitly one talks about the same location in space. It is not that obvious how applicable neighborhood in time is, especially considering the ontological object of time as temporal intervals; however, to the best of our knowledge there has been not much work that explores a neighboring concept of temporal intervals in configurations.

Neighborhood of temporal intervals: This work characterizes a temporal interval by two parameters: (1) the duration (radius) of an interval and (2) the moment when the interval happens, which is identified by the middle point of the interval [12]. These parameters are mapped onto a 2D space with the x-axis being middle points and y-axis being radii of intervals. Within this vector representation, intervals are considered connected or closely related if they are neighbors. Neighboring points are determined by using the Delaunay Triangulation [17, 18]. The Delaunay Triangulation partitions the Euclidean space, composed of a set of points, into triangles such that no four points of this set are co-circular. The dual of the Delaunay Triangulation, the Voronoi Diagram, represents a partition of space into regions where points of the Delaunay Triangulation are the nuclei of specific areas. These areas are bounded by the perpendicular bisectors of the nucleus and the set of its neighboring points. An example of the determination of neighboring temporal intervals is shown in Figure 5.
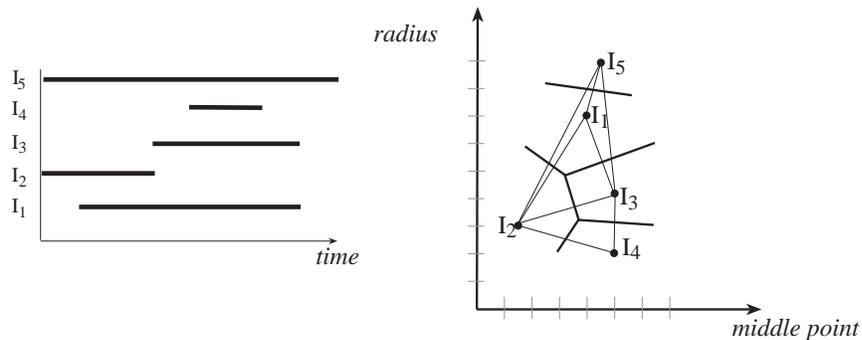
Fig. 5. Representation of temporal intervals and determination of Delaunay Triangulation

From Euler's Equation [18], which is applied for every convex polyhedron with m_n nodes and m_a arcs, one can derive that every node in the Delaunay Triangulation has a degree (i.e., number of neighboring nodes) ≥ 3. If all nodes of a Delaunay Triangulation are substituted with intervals and all arcs with binary relations, we can deduce that for a very large graph, the average number of neighbors (av_ng) of an interval is less than six (Equation 3). Thus, the average total number of arcs must be less than $3n$, that is, it grows linearly by $O(n)$, with n being the number of intervals. This upper bound of the number of relations in the final subgraph contrasts the theoretical bound of $O(n^2)$ of the initial graph.

$$av\_ng = \frac{2m_a}{m_n} \leq 6 - \frac{12}{m_n} \qquad (3)$$

Neighborhood-based simplification: To simplify temporal configurations, these configurations are represented as points in a 2D space. Only the arcs that exist in the resulting graph from the Delaunay Triangulation are considered to represent relations that are part of the minimal subgraph. The graph from the Delaunay Triangulation is undirected; however, we consider the arcs of this graph as binary relations, that is, as directed arcs. For example, consider the temporal configuration and the Delaunay Triangulation in Figure 5, its corresponding minimal configuration graph represented as a matrix is shown in Figure 6.

	I_1	I_2	I_3	I_4	I_5
I_1		⎓	⌐		⊥
I_2			⌐	⌐	⌐
I_3				⊤	⊥
I_4					
I_5					

Fig. 6. Minimal configuration graph that is obtained by applying the neighborhood-based simplification

Unlike the composition-based strategy for simplifying temporal configurations, given a representation of temporal configurations in a 2D space, there exists only one minimal subgraph.

4   Experimental Results of Selecting Temporal Relations

In order to evaluate how much these strategies can simplify configurations of temporal intervals, we use these strategies to simplify configurations of temporal

intervals that are used as queries in a process of content-based retrieval. The idea behind this analysis is to check the effect of eliminating temporal relations of an original query when comparing the results of the retrieval process (i.e., the set of configurations that are solutions) with respect to the desired answer (i.e., the original query). Eliminations will not have a negative impact if, by considering the subset of relations between temporal intervals as constraints in a retrieval process, one obtains results whose temporal intervals satisfy the complete set of constraints of the query. In such a process, there is a basic comparison between the original configuration and the configurations retrieved from the data set.

4.1 Framework for Comparing Temporal-Interval Configurations

We distinguish two levels in the comparison of temporal relations that we have called *consistency* and *equivalence*. Temporal relations are consistent if they are the same within the set of thirteen interval relations defined by Allen (i.e., there is no contradiction in the relation between intervals); temporal relations are equivalent if they are consistent and are the same based on metric characteristics that distinguish relative size, relative overlapping, and relativity of disjointness. For example, in Figure 7 the relation between I_1 and I_2 is consistent with the relation between intervals I_3 and I_4 ($r_{12} = r_{34}$); however, the pair of intervals (I_1, I_2) is not equivalent to (I_3, I_4) in terms of the lengths of intervals and length of overlap.

$$
\begin{array}{cc}
\underline{\text{I}_1} & \underline{\text{I}_3} \\
\underline{\text{I}_2} & \underline{\text{I}_4}
\end{array}
$$

Fig. 7. Consistent but not equivalent temporal relations

In this context, consistency and equivalence are binary decisions; however, equivalence is associated with a *distance function* that depends on metric characteristics of temporal intervals. The comparison of temporal intervals is not concerned with the location of intervals on time and the intervals' absolute lengths. Absolute positions of intervals on time are typically used for *timeslice* queries, which are outside the scope of this work. A relative length is used to make the distance measure less sensitive to scaling of intervals' representations. We define the distance between two pairs of intervals (I_1, I_2) and (I_3, I_4) by:

$$distance((I_1, I_2), (I_3, I_4)) = \begin{cases} |T(I_1, I_2) - T(I_3, I_4)| & \text{if } r_{12} = r_{34} \\ 1 & \text{otherwise} \end{cases} \quad \text{with}$$

$$T(I_j, I_k) = \frac{length(I_j) + length(I_k)}{2length(I_j \bigcup I_k)}$$

$$length(I_j) = I_j^+ - I_j^-$$

$$I_j \bigcup I_k \equiv \{(I^-, I^+) | I^- = min(I_j^-, I_k^-) \land I^+ = max(I_j^+, I_k^+)\} \qquad (4)$$

This function gives values from 0 to 1, with 1 meaning that the relations are inconsistent and, therefore, non equivalent, and with 0 meaning equivalent relations. To relax the equivalence definition, one could consider that two relations are equivalent if *distance* is less than or equal to a given threshold (set in this experiment to 0.05). In this work, we say that a solution is consistent if all relations in the solution are consistent with respect to the corresponding query. Likewise, a solution is said equivalent if the relations in the solution are equivalent to the relations in the query.

4.2   Evaluation of Strategies

For the analysis of composition- versus neighborhood-based strategies, a synthetic set of temporal intervals was created that allows for experimental replication. The synthetic data set was created with all possible intervals from 0 to 10 temporal units, which results in 55 different temporal intervals. Although the data set is small, it has all possible intervals within a temporal range. These intervals, in combination, create a large set of possible configurations (e.g., for configurations with 5 intervals, there exist 3.4×10^6 possible configurations).

We started with a set of 50 configurations that were created by randomly selecting five temporal intervals (i.e., ten temporal relations or constraints) from the data set. These configurations are queries in the retrieval process and are pre-processed to obtain four different cases of simplified queries: original configurations with all constraints, configurations after eliminating two random relations, configurations derived from the composition-based strategy (CBS), and configurations derived from the neighborhood-based strategy (NBS). We used the basic strategy of random eliminations of relations to evaluate whether or not the results with subgraphs were independent of the strategies of elimination. For the comparison of CBS and NBS, and considering the multiple subgraphs obtained with CBS, we selected the minimal subgraph with more *non-disjoint* relations (i.e., *after* and *before* relations). This heuristics was defined after checking that the results of the searches based on these selected subgraphs gave, on average, better results than other subgraphs. This is in agreement with the fact that overlapping intervals are less common and, therefore, the search based on query constraints defined in terms of these relations reduces the search space.

Original and simplified configurations represent query graphs and query subgraphs, respectively. The results of the retrieval process (i.e., the sets of temporal intervals that satisfy the same relations than the relations in query subgraphs) were compared in terms of degree of consistency and degree of equivalence with respect to query graphs (i.e., graphs with all relations). The number of constraint evaluations is used as a measure of performance, which is commonly used in the evaluations of constraint satisfaction problems and makes the evaluation independent of computational resources.

In cases of using the original configurations with all constraints, the process always finds optimal configurations (i.e., configurations where all constraints are satisfied). In the case of random eliminations of constraints, we found that the number of consistent and equivalent configurations were 75% and 66%, respectively. These values less than the number of consistent and equivalent results by using CBS or NBS. A summary of results for CBS and NBS is presented in Table 6. In this table, one has to consider that each configuration may have more than one solution and that each configuration has a number of consistent and equivalent relations (100% of consistent relations implies that a solution is consistent). Figure 8 shows, in detail, the percentage of equivalent relations in the configurations of 50 random queries (i.e., percentage over 10 constraints per query) for CBS and NBS.

Table 6. Summary of results for CBS and NBS (* average with respect to a configuration that includes all relations)

Parameters	CBS	NBS
Average of consistent configurations	100%	89%
Average of equivalent configurations	73%	85%
Average of consistent relations	100%	96%
Average of equivalent relations	90%	94%
Average percentage of constraint evaluations*	76%	60%

As we increased the number of intervals (six and seven intervals) and, therefore, the number of temporal relations to fifteen or twenty-one relations, we observed a tendency to maintain the results of NBS and CBS in terms of the average percentages of relations that are consistent and equivalent in configurations. In particular, the percentages of consistent relations in configurations for NBS and CBS were 94% and 100%, respectively, and the percentages of equivalent relations were 94% and 92%, respectively. The main difference was obtained in the number of equivalent configurations for CBS, which was on average 10% less than the average of equivalent configurations with respect to the initial five intervals.

5 Conclusions and Future Work

This paper has analyzed two different strategies to define a subset of temporal relations that characterizes a set of temporal intervals. The composition-based strategy is based on the composition of temporal intervals, and the neighborhood-based strategy selects relations that connect neighboring intervals represented in a 2D space.

While the composition-based method guarantees that relations can be derived, the neighborhood-based method captures metric characteristics of temporal intervals with respect to relative length, relative overlapping, and relativity
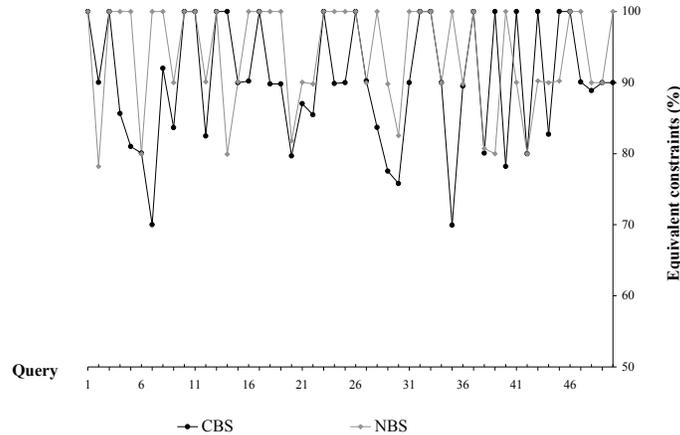
Fig. 8. Percentage of equivalent relations

of disjointness. In most cases, such strategies were able to characterize temporal configurations. From a different perspective, finding minimal subgraphs based on the composition is a complex task that could lead to an intractable problem unless one uses heuristics to reduce the search space. The results from this work indicate the potential of using the algebraic properties of composition and the metric characteristics of intervals for reducing the number of relations considered in a process. The use of fewer relations in a search process was on average computationally less expensive than the use of the complete description of sets of temporal relations. These strategies have a potential use in applications based on the analysis of temporal relations, such as content-based retrieval and data mining.

As future work, we will continue analyzing the efficiency of algorithms to apply these strategies to a larger scale. The results of this work complement the previous results with respect to topological relations between regions in a 2D space. We expect, in the near future, to analyze spatio-temporal relations within the same framework. In particular, a continuation of this work will be to apply the strategies to the simplification of spatio-temporal configurations that are described by a combination of spatial and temporal relations [2].

Acknowledgment

References

1. J. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):823–843, 1983.
2. C. Claramunt and B. Jiang. An integrated representation of spatial and temporal relations between evolving regions. *Geographical Systems*, 3(4):411–428, 2001.
3. E. Clementi, J. Sharma, and M. Egenhofer. Modelling topological spatial relations strategies for query languages. *Computing and Graphics*, 18(6):815–822, 1994.
4. M. Egenhofer. Deriving the composition of binary topological relations. *Journal of Visual Languages and Computing*, 5(2):133–149, 1994.
5. M. Egenhofer. Query preprocessing in spatial query by sketch. *Journal of Visual Computing*, 8(4):403–424, 1997.
6. M. Egenhofer and J. Sharma. Assessing the consistency of complete and incomplete topological information. *Geographical Systems*, 1:47–68, 1993.
7. J. Florence and M. Egenhofer. Distribution of topological relations in geographic databases. In *ACSM/ASPRS*, Baltimore, MD, 1996.
8. C. Freksa. Temporal reasoning based on semi intervals. *Artificial Intelligence*, 54:199–227, 1992.
9. C. Hamblin. Instants and intervals. *Studium Generale*, 24:127–134, 1971.
10. I. Humberstone. Interval semantics for tense logic: Some remarks. *Journal of Philosophical Logic*, 8:171–196, 1979.
11. H. Kriegel and T. Brinkhoff. Efficient spatial query processing. *IEEE Data Engineering Bulletin*, 16(3):10–15, 1993.
12. Z. Kulpa. Diagrammatic representation for a space of intervals. *Machine Graphics*, 6(1):5–24, 1997.
13. G. Langran. Temporal GIS design tradeoffs. *Journal of URISA*, 2(2):16–25, 1990.
14. A. Mackworth. Consistency in networks of relations. *Artificial Intelligence*, 8(1):99–118, 1977.
15. R. Maddux. *Some Algebras and Algorithms for Reasoning about Time and Space*. Technical report, Department of Mathematics, IOWA State University, Ames, IO, 1990.
16. P. Meseguer. Constraint satisfaction problems: an overview. *Artificial Intelligence Communications AICOM*, 2(1):3–17, 1989.
17. J. O'Rouke. *Computational Geometry*. Cmabridge University Press, Cambridge, MA, 1993.
18. F. Preparata and M. Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, Berlin, 1985.
19. B. Ramachandran and F. MacLeod. Modelling temporal changes in GIS using an object-oriented approach. In *Proceedings of the Sixth International Symposium on Spatial Data Handling*, pages 518–537, Edinburgh, Scotland, 1994.
20. A. Rodríguez, M. Egenhofer, and A. Blaser. Query pre-processing of topological constraints: Comparing composition-based with neighborhood-based approach. In T. Hadzilacos, Y. Manolopoulos, J. Roddick, and Y. Theodoridis, editors, *Spatial and Temporal Databases. LNCS 2750*, pages 362–379, Santorini, Greece, 2003. Springer-Verlag.

21. S. Ross. *A First Course in Probability*. Macmillan, 1976.
22. O. Stock. *Spatial and Temporal Reasoning*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.
23. A. Tarski. On the calculus of relations. *Journal of Symbolic Logic*, 6(3):73–89, 1941.
24. W. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46(2):234–240, 1970.