Exploiting Qualitative Spatial Reasoning for Topological Adjustment of Spatial Data

Jan Oliver Wallgrün
Department of Geography, GeoVISTA center
The Pennsylvania State University
University Park, State College, PA 16802, USA
wallgrun@psu.edu

ABSTRACT

Formal models of spatial relations such as the 9-Intersection model or RCC-8 have become omnipresent in the spatial information sciences and play an important role to formulate constraints in many applications of spatial data processing. A fundamental problem in such applications is to adapt geometric data to satisfy certain relational constraints while minimizing the changes that need to be made to the data. We address the problem of adjusting geometric objects to meet the spatial relations from a qualitative spatial calculus, forming a bridge between the areas of qualitative spatial representation and reasoning (QSR) and of geometric adjustment using optimization approaches. In particular, we explore how constraint-based QSR techniques can be beneficially employed to improve the optimization process. We discuss three different ways in which QSR can be utilized and then focus on its application to reduce the complexity of the optimization problem in terms of variables and equations needed. We propose two constraint-based problem simplification algorithms and evaluate them experimentally. Our results demonstrate that exploiting QSR techniques indeed leads to a significant performance improvement.

Categories and Subject Descriptors

I.2.4 [**Knowledge Representation Formalisms and Methods**]: Relation Systems; G.1.6 [**Optimization**]: Constrained optimization

General Terms

ALGORITHMS, THEORY

Keywords

topological relations, adjustment, qualitative spatial reasoning, constrained optimization, conflation, data cleaning

1. INTRODUCTION

ACM SIGSPATIAL GIS '12, November 6-9, 2012. Redondo Beach, CA, USA

Formal models of spatial relations play an important role in many spatial information processing applications, for instance as a basis to formulate spatial integrity constraints. Over the last three decades, the research field of qualitative spatial reasoning (QSR) [5, 18] has proposed and investigated a multitude of different relational formalisms, commonly referred to as qualitative spatial calculi, with a focus on logical reasoning and consistency checking. Arguably the two spatial calculi most prevalent in the area of geographic information science are the 9-Intersection model by Egenhofer [7] and the RCC-8 calculus by Randell, Cui, and Cohn [17]. Both approaches define topological relations between two spatial regions and, in the case of simple regions in the plane, define the exact same eight binary relations (see Fig. 1). The basic relations of the 9-Intersection model and RCC-8 have, for instance, been utilized to describe spatial relationships in query and retrieval scenarios [4, 1], to formalize (geo)spatial concepts, change, and processes [3, 12, 10, 6], and to specify spatial knowledge and integrity constraints in the context of spatial and spatio-temporal database applications [8, 11, 19].

While detecting spatial inconsistencies and violations of integrity constraints is an important problem of spatial information processing, procedures that are furthermore capable of resolving these inconsistencies or constraint violations automatically are often required or desirable. For instance, when integrity constraints demand that the 2D geometries of certain objects may not overlap (meaning their topological relation has to be disconnected (DC) in terms of RCC-8), one would want to automatically displace or deform the objects as much as needed to ensure disconnectedness without violating any other spatial background constraints. Such constrained optimization problems arise naturally in the area of conflation and data cleaning, in map generalization scenarios, and in the context of layout problems in general. Research in these areas has focused on particular topological constraints, in particular non-overlap [16, 21], using for instance optimization techniques and adjustment theory. General approaches able to deal with the problem of adjusting geometric data to the relations from a topological or other qualitative calculus are, however, still lacking. In addition, to our knowledge there exists no research explicitly trying to connect the areas of QSR and geometric adjustment, investigating, for instance, how constraint-based QSR techniques can be employed to improve the optimization process.

In our work, we aim at developing a flexible framework for spatial adjustment with qualitative spatial calculi and at

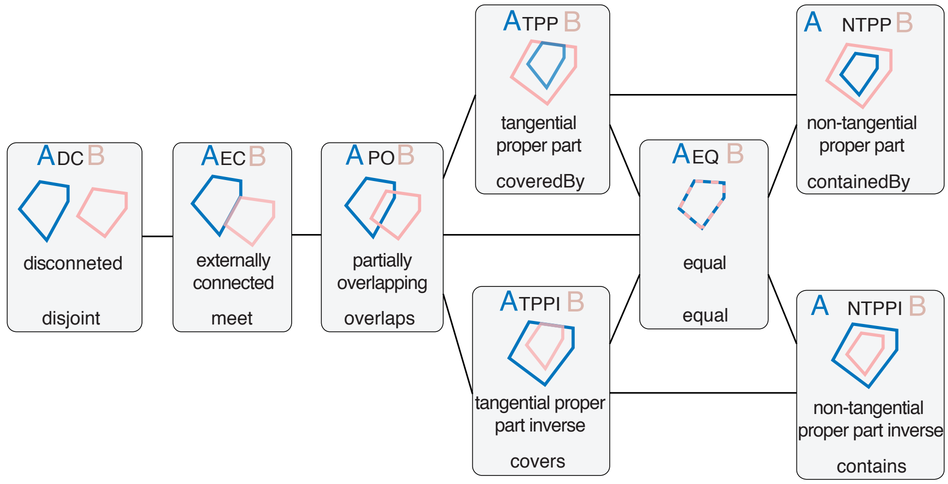


Figure 1: Eight basic topological relations distinguished by the RCC-8 and 9-Intersection models.

exploring the combined application of QSR and optimization techniques to solve spatial conflation, data cleaning, and layout problems. In this paper, we address the concrete problem of adjusting polygons in 2D space through translation / displacement to meet topological constraints given in the form of relations from the 9-Intersection and RCC-8 models. With the goal of extending our recently proposed formalization of these relations in terms of sets of (in)equations using Minkowski sums and an adjustment approach based on mixed-integer programming (MIP) [23] described further in Sec. 3, we focus on the question of how in turn constraint-based QSR techniques can be employed to improve this general approach. In Sec. 4, we propose and discuss three different ways in which QSR can be utilized in this context. We then focus on one of these options, namely the application of QSR to reduce the complexity of the optimization problem in terms of variables and equations needed (Sec. 5). We present two QSR based algorithms, a basic version and an extended one, for modifying a set of spatial constraints such that the complexity of the resulting MIP problem is reduced. In Sec. 6, we report on our experimental evaluation of these two algorithms using randomly generated problem instances. The experiments confirm that the exploitation of QSR techniques has the potential to significantly simplify the MIP problems and lead to a performance increase of the overall adjustment approach.

2. RELATED WORK

The problem of adapting spatial data to meet certain spatial constraints has received significant attention in several application domains of spatial data processing such as data integration and cleaning and map generalization. Spatial constraints discussed in this context can be graphical / metrical, topological, structural, etc. (see for instance [22, 13]). The rich spectrum of employed approaches ranges from local search [24], over global optimization and least square adjustment techniques [21, 9], to agent-based frameworks [14].

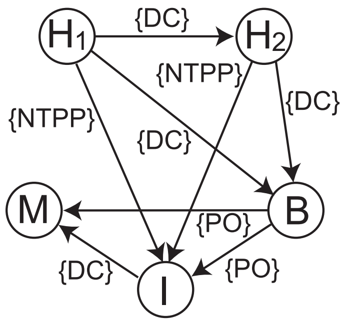
The approach closest to our adjustment method is the work described in [16]. It also employs Minkowski polygons to formalize non-overlap constraints. The adjustment approach used in this paper extends this line of research to allow for describing all topological relations defined in the previously mentioned qualitative calculi. A formalization of RCC-8 relations in terms of systems of (in)equations has already been proposed in [2]. However, our approach has the advantage that it is much more efficient in terms of the required number of (in)equations (linear vs. quadratic).

Employing qualitative spatial reasoning techniques [5, 18] for preprocessing relational information has been discussed in the context of spatial query processing (see for instance [20]). However, the methods we develop in the main part of this paper are different in that they aim at deriving a constraint network with a minimal number of constraints that are not the universal relation in contrast to reducing the overall number of base relations.

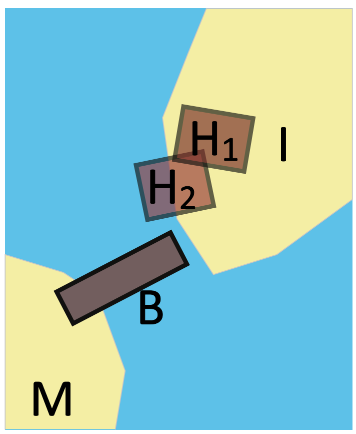
3. QUALITATIVE ADJUSTMENT

3.1 Adjustment with qualitative constraints

Formal models for the representation of and reasoning with spatial relations referred to as qualitative spatial calculi define a set $\mathcal{B}$ of *base relations* (e.g., the eight topological relations from Fig. 1) over a domain D of spatial objects (e.g., points or regions in 2D) together with a set of operations that enable elementary reasoning. Spatial knowledge is then expressed in the form of relational statements, e.g., $A\{\mathrm{DC}\}B$ for describing the fact that object A is disconnected from object B. Incomplete knowledge is expressed by disjunctions of base relations written as sets, e.g., $A\{\mathrm{DC}, \mathrm{EC}\}B$. The universal relation U is the disjunction of all basic relations and can express complete ignorance or unconstrainedness. A qualitative knowledge base or a set of relational conditions (for instance spatial integrity constraints) can be illustrated as a *qualitative constraint network* (QCN) (see Fig. 2(a)) with nodes standing for the objects and edges labeled by the respective qualitative relations (no connecting edge means the relation is U). A QCN is said to be *consistent* if it has at least one *solution*, meaning we can find a set of spatial objects that satisfy the relations in the network. Each solution corresponds to a refined version of the network in which each constraint is a single base relation, referred to as an *atomic* QCN.



(a) constraint network



(b) geometric data

Figure 2: Example of a qualitative adjustment problem: Given (a) a qualitative constraint network over a spatial calculus $\mathcal{C}$ (here RCC-8) and (b) geometric data for the involved objects, adjust the geometric data until all relations from the network are satisfied while changing the geometries as little as possible.

The general problem we are concerned with, referred to as the *qualitative adjustment problem*, is to adapt the geometric data for a set of objects $\mathcal{O} = \{O_1, \ldots O_n\}$ such that it satisfies the qualitative spatial relations given in a QCN N over calculus $\mathcal{C}$ and $\mathcal{O}$, while minimizing the changes as specified by a cost function c. A simple example of such a problem is shown in Fig. 2(b). It depicts a set of polygonal objects resulting from integrating erroneous data from different sources. It violates several integrity constraints such as the two buildings H_1 and H_2 overlapping each other, H_2 not being completely located on the island I, and the bridge B not connecting the mainland M with the island I. The QCN in Fig. 2(a) formalizes the constraints that the data is supposed to satisfy: H_1 and H_2 should be disconnected (RCC-8 relation DC), I should contain both H_1 and H_2 (RCC-8 relation NTPP), and the bridge should overlap both M and I (RCC-8 relation PO).

The qualitative adjustment problem falls into the general category of constrained optimization problems. While we here restrict ourselves to relations from a single qualitative calculus, the definition can easily be extended to allow for relations from different spatial calculi to deal with the general problem of finding a geometric configuration that satisfies a set of qualitative constraints.

Algorithm 1 Qualitative adjustment algorithm

function ADJUST($\mathcal{O}, N$)

Input:

$\mathcal{O}$ set of polygons O_i

N QCN over calculus $\mathcal{C}$ and objects $\mathcal{O}$

Output:

$\mathcal{O}'$ new set of polygons O'_i with O'_i being a displaced version of O_i

Choose reference vertex p_i for each O_i in $\mathcal{O}$
Compute $M^+(p_i, O_i, O_j)$ and $M^-(p_i, O_i, O_j)$
Translate relations in N into systems of (in)equations
Translation into an equivalent MIP problem
Run MIP solver to compute new ref. points p'_i
Generate $\mathcal{O}'$ by translating each O by $p'_i - p_i$
Return $\mathcal{O}'$

end function

In the following, we describe our approach to solve the qualitative adjustment problem for topological constraints from the 9-Intersection and RCC-8 models between polygonal objects in 2D and the case that only translation / displacement of the objects is allowed, preserving their shape and size. A more detailed description of the underlying formalization of topological constraints can be found in [23].

3.2 A Topological Adjustment Approach Using Minkowski Sums and MIP

Our general approach is illustrated in Alg. 1 and is based on a translation of the qualitative relations occurring in the input QCN into systems of (in)equations. These are then further translated into a mixed-integer programming (MIP) problem (either a mixed-integer linear programming problem (MLP) or a mixed-integer *non*-linear programming problem (MNLP)). A dedicated MIP solver takes the MIP formulation and cost function as input and solves the problem. If the solver is able to find a solution, new geometries for the input objects that satisfy all spatial constraints can be directly derived from this solution. An optimal solution found by the MIP solver is guaranteed to be an optimal solution to the original qualitative adjustment problem. However, due to the heuristic nature of many MIP solving approaches, the result may not always be globally optimal.

While this general approach can be used to realize all kinds of qualitative spatial constraints (e.g., also direction constraints) and allows for other kinds of transformations (e.g., deformations), Alg. 1 describes a specific realization for topological constraints and displacement only, which allows for a more compact and efficient formalization and forms the basis for the analysis performed later in this paper. The approach employs Minkoswki sums of the input polygons to keep the complexity of the formalization in terms of variables and (in)equations needed low. In the following description, we will use the abbreviated relation names of RCC-8, e.g., DC for disconnected, etc. (see again Fig. 1).

3.2.1 Minkowski-based Formalization

In principle, topological relations between simple polygons in the plane can be expressed using the x, y coordinates of all vertices of the two involved objects. However, while this approach has the advantage of being general enough to allow for all kinds of transformations to the objects during the optimization process, it requires a quadratic number of (in)equations. More precisely, for two convex polygons with m and n vertices, respectively, it requires $2 \times (m+n)$ real variables and $O(m*n)$ (in)equations. In contrast, our approach tailored to displacement scenarios employs Minkowski sums of the involved polygons reducing the number of variables required per relation to only a constant number of four real variables and $O(m+n)$ (in)equations.

Fig. 3 illustrates the Minkowski based formalization approach for two polygons P and Q. For both objects, reference vertices p and q are chosen arbitrarily and their coordinates will become the four real variables in the formalization. Then new polygonal objects are computed using two different Minkowski sums. The first $M^+(p,P,Q) = \{-(a-p)+b \mid a \in P \wedge b \in Q\}$ is constructed by shrinking P to point p, while growing Q accordingly (an operation for instance used in motion planning [15] and layouting [16] to avoid overlap / collision between objects). As illustrated in Fig. 3, we get a grown version of Q shown as the dark shaded polygon. We extended this idea by introducing a second Minkowski sum $M^-(p,P,Q) = cl(Q \setminus \{-(a-p)+b \mid a \in P \wedge b \in \partial Q\})$ shown by the light shaded polygon contained in Q in Fig. 3. If p falls into the area represented by $M^+(p,P,Q)$, there will be an overlap between P and Q. If not P and Q will be disconnected. If p falls into the area represented by $M^-(p,P,Q)$, P will be contained by Q. Taken together, these two Minkowski sums are sufficient to formalize all eight topological base relations of RCC-8.

For simplicity, we equate the point sets $M^-(p,P,Q)$ and $M^+(p,P,Q)$ with the respective polygonal objects that describe their boundaries. Both $M^-(p,P,Q)$ and $M^+(p,P,Q)$ can be computed using the general operation of convoluting polygons as for instance provided by the CGAL computational geometry library[1]. In the general case where both P and Q may be concave, $M^-(p,P,Q)$ and $M^+(p,P,Q)$ may have multiple components and have holes which are themselves concave. We here restrict ourselves to the case where both P and Q (and, hence, also $M^-(p,P,Q)$ and $M^+(p,P,Q)$) are simple convex polygons. In addition, we only provide one example of the formalization, namely how the relation PO (partially overlap) can be realized based on a combination of both $M^-(p,P,Q)$ and $M^+(p,P,Q)$. For the complete formalization and a generalization of the approach to non-convex objects, we refer again to [23].

Assuming P and Q are both convex polygons, the relation PO holds between P and Q if point p is at the same time inside $M^+(p,P,Q)$ and outside of $M^-(p,P,Q)$. The inside condition can be described by the requirement that p is *right of* all bounding edges of $M^+(p,P,Q)$ when assuming a clockwise orientation of the edges. For each edge $e_i^+ \in E^+$ (the edges of $M^+(p,P,Q)$) with start point q_i and end point $q_{i\oplus 1}$, *right_of* can be formalized based on the crossproduct $cross(p,(q_i,q_{i\oplus 1})) = (x_p - x_{q_i})(y_{q_{i\oplus 1}} - y_{q_i}) + (y_p - y_{q_i})(x_{q_{i\oplus 1}} - x_{q_i})$ as $right\_of(p,(q_i,q_2)) \Leftrightarrow cross(p,(q_1,q_2)) > 0$. The result is a conjunction of $|E^+|$ inequations. The outside condition, on the other hand, can be described by the requirement that p is *left of* at least one of the edges $e_i^- \in E^-$ (the edges of $M^-(p,P,Q)$). *left_of* is analogously defined as $left\_of(p,(q_i,q_{i\oplus 1})) \Leftrightarrow cross(p,(q_i,q_{i\oplus 1})) < 0$. This leads to a disjunction of $|E^-|$ inequations and the following overall

[1] http://www.cgal.org/

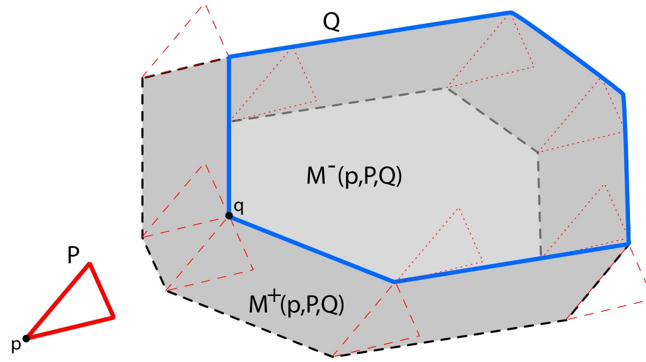


Figure 3: Minkowski sums $M^+(p,P,Q)$, $M^-(p,P,Q)$.

definition of PO_cvx for relation PO:

$$PO\_cvx(P,Q) \Leftrightarrow \bigwedge_{i=1}^{|E^+|} right\_of(p,e_i^+) \wedge \left(M^-(p,P,Q) = \emptyset \vee \bigvee_{i=1}^{|E^-|} left\_of(p,e_i^-) \right)$$

The other topological relations can be formalized similarly. Given the systems of (in)equations resulting from the input QCN, the problem can be transformed into an equivalent non-linear programming problem. However, a *mixed-integer* programming (MIP) approach is needed to represent the disjunctions occurring in the formalizations of the different topological relations.

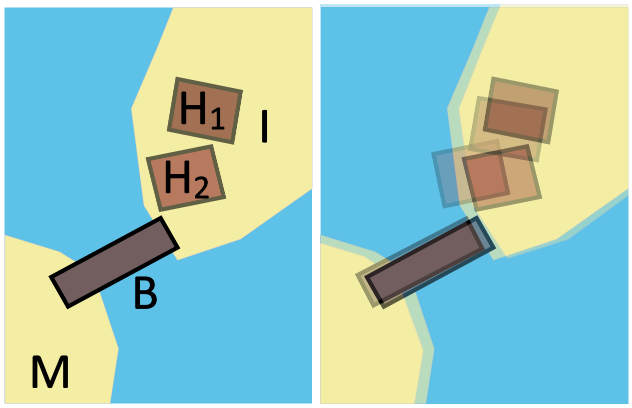
3.2.2 Translation into an MIP Problem

MIP problems are defined in the following way: Given

1. a vector $\vec{x} = (x_i)$ of variables over the reals,
2. a vector $\vec{y} = (y_j)$ of integer variables,
3. a cost function $c(\vec{x},\vec{y})$,
4. and a constraint function $g(\vec{x},\vec{y})$,

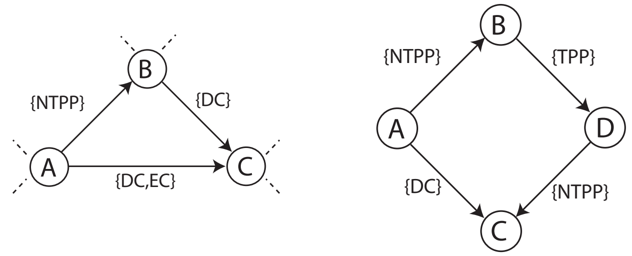
minimize $c(\vec{x},\vec{y})$ subject to $g(\vec{x},\vec{y}) \le 0$.

When the systems of (in)equations are translated into an MIP problem, the coordinates of the reference vertices become the real variables x_i. All other vertices occurring in the Minkowski polygons are expressed relatively to the chosen reference vertex. Binary variables and additional inequations are then used to formulate the problem as a set of conjunctively connected inequations forming the constraint matrix $g(\vec{x},\vec{y})$. Disjunctions are resolved by introducing a binary integer variable $y_i \in \{0,1\}$ into each of the n (in)equations part of the disjunction such that it is always satisfied if y_i is 1. This can be achieved by inserting additive or subtractive terms consisting of a multiplication of y_i with a large constant C. The actual disjunction is then realized by enforcing that at least one of the y_i for the given disjunction is 0 by demanding that their sum is smaller than n. This approach adds n binary variables and one additional inequation. The number of (in)equations per base relation thus stays linear wrt. the number of vertices occurring in the



(a) (b)

Figure 4: (a) Computed solution for the qualitative adjustment problem from Fig. 2 and (b) an illustration of the displacements performed.



(a) (b)

Figure 5: Examples of QCNs: In (a) the base relation EC can be removed based on the composition operation. The network in (b) is inconsistent.

Minkowski polygons. If the input QCN contains disjunctions of relations as constraints (other than the universal relation U), formalizing these may require a hierarchical organization of binary variables.

The final component needed for the MIP specification is the cost function over the real variables x_i. Taking a least square adjustment approach, we use the basic cost function $c(\vec{x}) = \sum_{i=1}^{n}(x_i - \bar{x_i})^2$ which minimizes the squared displacement distance over all objects. However, other kinds of cost functions can be used, for instance expressing preferences over which objects should be moved.

Fig. 4(a) shows the result of applying the approach described in this section to the simple example from Fig. 2. An ϵ parameter has been introduced into the equations to ensure a minimal gap between the boundaries for relations DC and NTPP as well as a minimal overlap for PO (i.e., the bridge overlapping both the mainland M and the island I). The problem description comprised 10 real variables, 38 binary variables, and 83 (in)equations. Computing the solution took 2.3 seconds on a 3GHz i5 computer using the Bonmin[2] MIP solver. Fig. 4(b) illustrates the displacement of the objects that was needed to satisfy all the constraints.

4. EXPLOITING QSR

The formalization of qualitative spatial relations in terms of systems of (in)equations forms a bridge between the research field of qualitative spatial (and temporal) reasoning and geometric adjustment approaches, allowing us to construct geometric instances for qualitatively given information. We here address the question how in turn traditional constraint-based qualitative spatial reasoning techniques can be employed to improve the overall adjustment procedure. Constraint-based qualitative spatial reasoning employs constraint propagation based on a so-called composition table to infer new information by ruling out certain base relations from the disjunctions appearing in a QCN. The cells of the composition table state which base relations can hold between two objects A and C given the relation holding between A and another object B and the relation between B and C. Standard procedures for checking the consistency of a QCN (i.e., whether the relations in the network can be satisfied by a set of geometric objects) [18] are like-wise based ultimately on the composition operation ($\circ$). Given the information $A\{\mathrm{DC}, \mathrm{EC}\}C$, $A\{\mathrm{NTPP}\}B$, and $B\{\mathrm{DC}\}C$ (see Fig. 5(a)), employing the composition operation allows us to refine the disjunction $\{\mathrm{DC}, \mathrm{EC}\}$ between A and C to just $\{\mathrm{DC}\}$ because the table states that $\{\mathrm{NTPP}\} \circ \{\mathrm{DC}\} = \{\mathrm{DC}\}$.

The so-called *algebraic closure* procedure [18] performs the operation $C_{ik} = C_{ik} \cap (C_{ij} \circ C_{jk})$ for triples of nodes N_i, N_j, N_k in a QCN until a fixpoint is reached (C_{ij} denotes the relational constraint between N_i and N_j in the network). It has a run time complexity of $O(n^3)$ for QCNs with n nodes. The composition results $C_{ij} \circ C_{jk}$ are looked up in the composition table, which can be seen as providing precompiled geometric information in the form of all consistent atomic networks of three objects. If the result of the intersection is the empty relation, no solution to the network exists, meaning it is inconsistent. For most calculi, the algebraic closure procedure provides only an incomplete consistency check that may not discover all inconsistencies. To get a sound and complete consistency check, the procedure has to be embedded into a backtracking search in which constraints are split either down to the level of single base relations or to relations from a subset for which algebraic closure is sufficient to decide consistency.

In the following, we propose and discuss three general ways in which QSR techniques such as composition-based reasoning can contribute to improve the performance of the MIP based adjustment approach in terms of quality and efficiency. Quality here is measured in terms of the costs of the final solution computed, while efficiency concerns the time needed to find this solution. We will then focus on one of the proposed approaches, namely the utilization of composition-based reasoning to simplify the MIP problem descriptions, and investigate it in detail in Sec. 5 and 6. The other proposed approaches will require a similar analysis as part of future research.

4.1 Preceding Consistency Check

In application scenarios in which it cannot be excluded that the derived set of constraints forming the input QCN is contradictory, performing a consistency check as a preprocessing step can be expected to increase the performance of the overall adjustment approach. If an inconsistency is discov-

[2] `http://www.coin-or.org/Bonmin/`

ered employing QSR methods, translating the problem into an MIP problem and running the MIP solver can be avoided. This is particularly beneficial as in many cases using just the MIP solver, inconsistency of the input constraints can only be concluded from the fact that no solution was found within a given time threshold.

However, one has to take into account that for many qualitative calculi consistency checking is NP-hard. Hence, a sound and complete qualitative consistency checking procedure may be prohibitively expensive and may actually decrease the average run time performance, in particular when inconsistencies only occur rarely. We therefore consider the algebraic closure procedure a promising candidate for an approximate but polynomial consistency checking method that may provide a good trade-off. First tests indicate that this is indeed the case. To give one example, algebraic closure was able to determine inconsistency of the simple network shown in Fig. 5(b) within less than 100ms. Feeding problem instances based on this QCN directly into an MIP solver resulted in running times around 5 seconds on the same computer. While the benefits of a preceding consistency check clearly depend on the possibility or probability of input QCNs being inconsistent, performing algebraic closure has the additional benefit of potentially removing base relations from the QCN which cannot be satisfied and in doing so simplifying the MIP problem as we will discuss in the next section.

4.2 Problem Simplification

In addition to discovering inconsistencies and avoiding the optimization procedure completely, using composition and algebraic closure can be used to simplify the complexity of the resulting MIP problem in terms of the number of variables and equations needed. This can be achieved by a combination of two different strategies. On the one hand, the algebraic closure procedure can be used to remove base relations from disjunctions occurring in the input network which cannot lead to a solution, potentially leading to a smaller number of equations and auxiliary binary variables required. On the other hand, if it is possible to set one or more constraints in the network to the universal relation U without altering the possible solutions of the network, the optimization problem will also be simplified because U does not constrain the spatial configuration of the involved objects. Replacement by U is possible if the remaining relations in the network imply the original relation. The overall goal would then be to build an equivalent QCN (one with the same set of solutions) with as many occurrences of U as possible. Since no (in)equations are needed to formalize the universal relation U, this seems a particularly promising approach, which is the reason why we will study it in detail in Sec. 5 and the experimental evaluation in Sec. 6.

4.3 Problem Decomposition

Even with a simplification of the problem description as proposed in the previous section, computation times may become unacceptable for a large numbers of objects. However, problems encountered in practice often can be decomposable into smaller subproblems which can be solved independently, followed by a combination of the individual solutions into an overall solution. Taking again topological adjustment as an example, the containment relations NTPP, TPP, and their respective inverses can induce a hierarchical structure on the objects which can be used to decompose the problem. Let us say, we have a set of objects A_i for which the constraints demand that they are all in relation NTPP with object B which in turn is supposed to be in relations DC or EC with the remaining objects C_i. We then can solve the problem of arranging objects A_i within B and the problem of arranging B with respect to all C_i individually and put the solutions together. While this looks similar to what was discussed for the problem simplification approach, the difference here is that smaller subproblems are solved individually which ultimately allows for a parallelization of the adjustment procedure.

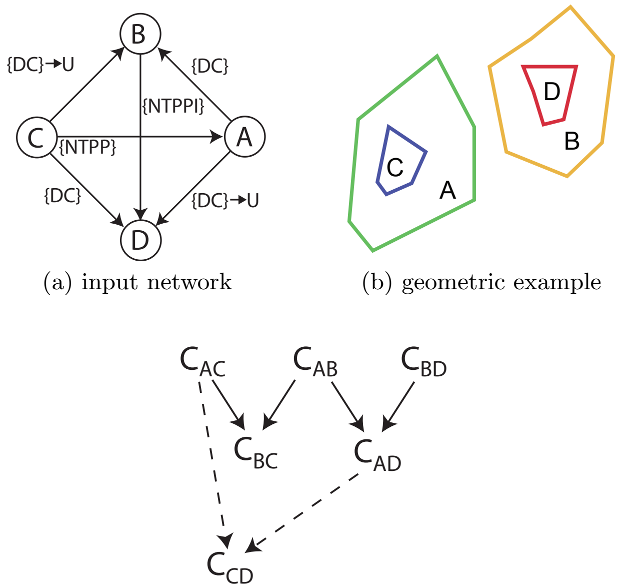
This proposal raises the question of how the actual decomposition can be performed. Generally speaking, when we can find a subgraph in a QCN of which only one node is connected to nodes outside of this subgraph via constraints that are not the universal relation U, the adjustment problem for this subgraph can be solved individually. This notion can be combined with the previously discussed idea of replacing constraints by U when this is possible without changing the solutions of the input QCN, followed by the identification of such subgraphs in the modified network.

5. SIMPLIFICATION APPROACHES

As we discussed in Sec. 4.2, using algebraic closure can remove base relations from disjunctions in the input QCN that cannot lead to a solution. If the network contains the disjunction $A\{\mathrm{NTPP}, \mathrm{NTPPI}\}B$ but no solution exists in which the relation between A and B is NTPPI, the NTPPI can be safely removed; and if there exists a composition for A and B whose result does not contain NTPPI, the algebraic closure operation will do so. This can simplify the MIP formulation but only for certain disjunctions as in this example. If the disjunction for instance is {DC,EC}, this disjunction can be formalized with the same number of (in)equations as just {DC} (see [23]). Hence, there is no benefit in removing the EC if algebraic closure allows to do so. Moreover, removing base relations from the universal relation U, the disjunction of all base relations, is counterproductive as U does not lead to any (in)equations at all. In contrast, trying to remove constraints that are implied by other constraints in the network and replace them by U seems more promising and has the additional advantage that it can simplify the MIP description even when all constraints in the input network are either single base relations or U, a case that is very common in practice (see for instance the example from Fig. 2). We therefore focus on this aspect by proposing two algorithms following this approach and evaluating them in Sec. 6.

Fig. 6(a) contains a small example illustrating the idea we just described. The composition of $C\{\mathrm{NTPP}\}A$ and $A\{\mathrm{DC}\}B$, for instance, implies that the relation between C and B can only be DC. Hence, this relation can safely be replaced by U without changing the solutions of the network. In general, given a QCN N, the problem can be described as finding a QCN N' over the same set of objects and having the same set of solutions (meaning N' and N are equivalent), while maximizing the number of constraints that are the universal relation U.

The two algorithms that we describe in the following are both greedy algorithms not guaranteed to provide an optimal simplification but have the advantage of being quite efficient, while still being very effective in reducing the complexity of the problem descriptions, as our evaluation will



(a) input network

(b) geometric example

(c) auxiliary graph structure used by the extended algorithm

Figure 6: Example in which the basic simplification algorithm will not find the best solution.

show. Both algorithms are based on reverting the composition operation based on the following insight: *Whenever* $C_{ij} \circ C_{jk} \subseteq C_{ik}$ *holds, replacing* C_{ik} *with* U *will result in an equivalent network.* The truth of this statement is obvious from the fact that the equivalence preserving algebraic closure operation will perform the opposite operation of replacing U with the composition $C_{ij} \circ C_{jk}$.

Our first algorithm is the most basic variant of performing this reversed composition operation by taking a QCN N as input and looking at all triples of nodes to check whether the constraints between them satisfy the condition stated above. A pseudocode version of the approach is shown in Alg. 2. Whenever the condition is satisfied, the constraint C_{ik} is replaced by U and the approach continues with the next triple. We leave the order in which the triples are considered unspecified as our tests have shown that, at least in the case of RCC-8 relations, not much could be gained using different heuristics compared to a random approach that simply loops over the three indices i, j, k. Our implementation of the algorithm contains a few optimizations not shown in the pseudocode, which aim at avoiding unnecessary lookup operations in the composition table. After termination, the algorithm will have transformed the input QCN into an equivalent one with a potentially increased number of appearances of U in the network.

While our evaluation will show that our basic simplification algorithm achieves very good results in practice, it is rather easy to construct an example in which it will not result in the highest possible number of universal relations in the output. The example from Fig. 6(a) is such a case: Let us assume that both the relation between B and C and the one between A and D have been replaced by U using the composition triples CAB and ABD, respectively. At this point no further replacements are possible. This leaves four relations that are not U. However, when looking at an example of the described configuration shown in Fig. 6(b), it becomes clear that it actually can be unambiguously described by just three statements, namely $C\{\text{NTPP}\}A$, $D\{\text{NTTP}\}B$, and $A\{\text{DC}\}B$. In principle, it would be possible to use the composition of $C\{\text{NTTP}\}A$ and $A\{\text{DC}\}D$ to replace the relation between C and D by U but since the relation between A and D has already been replaced by U, the basic algorithm will not do so.

Algorithm 2 Basic Simplification Algorithm

procedure SIMPLIFY(N)
  Input:
  N QCN over n objects O_i with C_{ij} being the constraint between O_i and O_j
  Result:
  modified N equivalent to the original network
    for all $i, j, k \in \{1, ..., n\}$ **do**
      if $C_{ij} \circ C_{jk} \subseteq C_{ik}$ **then** $C_{ik} \leftarrow U$ **end if**
    end for
end procedure

This leads to the idea underlying our second, extended simplification algorithm. Instead of immediately replacing a relation with U, it only stores the information that this replacement should happen but uses the original information for further processing. In the pseudocode version in Alg. 3, this is achieved by setting a previously initialized copy C'_{ij} of C_{ij} to U. The actual replacement will take place in a final loop at the end. However, following this approach care must be taken to ensure that the resulting network will still be equivalent to the original input network. Equivalence can be violated when we directly or indirectly use a constraint b that has been noted down for being replaced based on the composition of constraints c and d to decide that either c or d may be replaced as well. In other words, we need to make sure the directed dependency graph which states which constraints have been used to determine replacement of which other constraints remains acyclic. To achieve this, our algorithm builds up such an auxiliary graph structure in which the nodes represent the constraints. When it has been decided that the triple C_{ij}, C_{jk}, C_{ik} can be used to replace C_{ik} by U, two directed arcs are introduced into the graph connecting both C_{ij} and C_{jk} to C_{ik}. In the pseudocode of the algorithm, this is realized by the variables $Parents_{ij}$ which store the parents of constraint C_{ij} in the graph. To decide if a replacement is possible, we then get a second condition, namely that C_{ik} is neither an ancestor of C_{ij} nor of C_{jk} in the graph. In the pseudocode version, the auxiliary function $Ancestor(C_{ik}, C_{jk})$ is used to verify this, using a recursive search through the nodes above C_{jk} in the graph. For the example from Fig. 6(a), the current graph structure is shown in Fig. 6(c) by the fully drawn arcs. Since no cycle would be introduced, the relation between A and D can be noted down for replacement leading to the new connections shown by the dashed arrows. The final result will indeed be a network with only three relations different from U.

Both algorithms loop over all $O(n^3)$ triples of objects exactly once. The extended version has the additional overhead of having to traverse the ancestors in the auxiliary graph in order to check whether a replacement can take place. Given these two different simplification approaches, we were now interested whether they would indeed be able

Algorithm 3 Extended Simplification Algorithm

procedure SIMPLIFY(N)

Input:

N QCN over n objects O_i with C_{ij} being the constraint between O_i and O_j

Result:

modified N equivalent to the original network

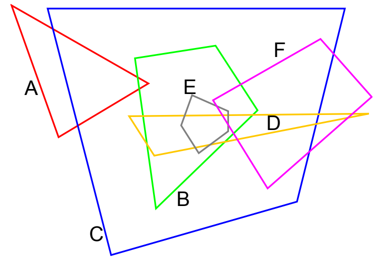
```
;; initialization
for all i, j ∈ {1, ..., n} do
    C'_ij ← C_ij
    Parents_ij ← ∅
end for
;; main loop with acyclic graph construction
for all i, j, k ∈ {1, ..., n} do
    if C_ij ∘ C_jk ⊆ C_ik and Parents_ik = ∅
      and not Ancestor(C_ik, C_ij)
      and not Ancestor(C_ik, C_jk) then
        C'_ik ← U
        Parents_ik ← {C_ij} ∪ {C_jk}
    end if
end for
;; performing actual replacement
for all i, j ∈ {1, ..., n} do C_ij ← C'_ij end for
end procedure
```

to improve the performance of our adjustment approach and how they would compare with each other. In the next section, we report on the results from our experiments.

6. EXPERIMENTAL EVALUATION

We implemented the two simplification approaches and incorporated them into the qualitative adjustment framework described in Sec. 3.2. In order to perform an experimental evaluation and collect statistical data about the benefits of employing the QSR based simplification as a preprocessing step, we then set up an instance generator able to generate random adjustment problems (see Fig. 7 for an example). The generator produces instances by first generating a random geometric configuration $G_{original}$ of m convex polygons. Then, the qualitative constraint network $N_{original}$ for this geometric configuration is derived. Finally, the geometric configuration is varied by randomly displacing the objects resulting in a new geometric configuration $G_{displaced}$. The configuration $G_{displaced}$ and the constraint network $N_{original}$ form the input to the adjustment algorithm. Since $G_{displaced}$ has been derived from the geometric configuration corresponding to $N_{original}$, we know that a solution must exist, although in general we cannot expect that $G_{original}$ is the optimal solution for the generated instance.

We generated random instances with 3 to 23 objects. It has to be noted that the generated problem instances are at the most challenging end of the spectrum as they specify a base relation between each pair of objects leading to $O(m^2)$ relational constraints. The problem instances were then fed into our adjustment algorithm (a) without simplification, (b) with the basic simplification algorithm applied, and (c) with the extended version applied to simplify the input constraint network. The resulting MIPs were then solved using Bonmin on a 3GHz i5 computer. We recorded several statistics including the complexity of the MIP program in terms of the number of variables and the number of (in)equations, as well as the computation times for the preprocessing step doing the simplification and the running times of the MIP solver. Moreover, we recorded the value of the cost function for the computed solutions. The results are illustrated in Figs. 8 to 10 for all three settings and in dependence of the number of objects.



A PO B, A PO C,
A DC D, A **NTPPI** E,
A DC F;
B NTPP C, B PO D,
B NTPPI E, B **DC** F;
C PO D, C NTPPI E,
C **DC** F;
D **DC** E, D PO F;
E **DC** F

Figure 7: Example of a randomly generated problem instance. The bolded relations are not satisfied in the geometric configuration and need to be adjusted.

First, Fig. 8 shows the complexity of the MIP problem in terms of the number of variables and (in)equations required. We see that both have been significantly reduced when employing one of the simplification algorithms compared to the setup without QSR based simplification. When comparing the two simplification variants with each other, there is, somewhat surprisingly, no visible difference between the two approaches. Looking at the exact numbers, the extended simplification approach performs marginally better than the basic version but instances in which the two approaches produce different results occurred very rarely in the experiment. More precisely, this was the case only for 1.66% of all instances. The total reduction in required variables compared to the setup without simplification was 34.40% (34.45% for the extended version) when averaged over all runs. While the diagram shows the overall number of variables, both real and binary ones, the reduction exclusively concerns the number of binary variables needed to formalize disjunctions of (in)equations. Regarding the number of (in)equations, the average reduction over all runs was 23.70% (23.74% for the extended version).

The question now was whether this significant simplification of the problem descriptions leads to an increased performance of the MIP solving process. As shown in Fig. 9(a), this is indeed the case as the computation times are significantly lower for both variants. Averaged over all runs the reduction in computation times of the MIP solver was 52.33% for the basic algorithm and 52.40% for the extended variant. Hence, our conclusion at this point is that exploiting QSR based reasoning to simplify the MIP problem descriptions can lead to a drastic improvement of the optimization procedure. Fig. 9(b) answers the question at which costs this improvement is achieved by showing the computation times for the simplification procedures. Over the entire experiment these times remained below three seconds which is rather negligible compared to the time saved during the MIP solving. Comparing the two simplification variants, we see that as expected the computation times for the more involved extended variant are higher than those of the basic version and show a quicker increase with higher object numbers. Together with the discussed results on the perfor-

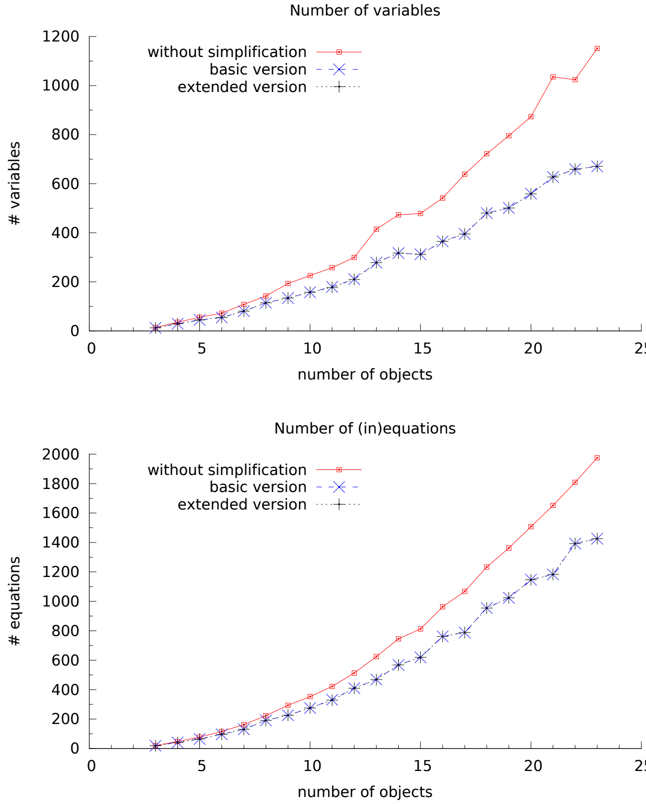


Figure 8: Results regarding the reduction in the number of variables and (in)equations.

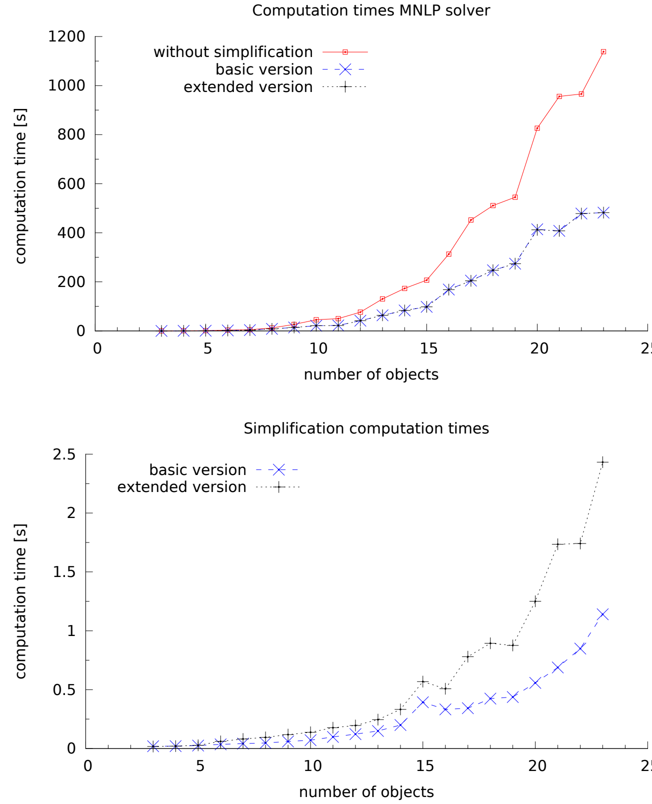


Figure 9: Results regarding the computation times of the MIP solver and the preprocessing.

mance increase, this indicates that the basic simplification algorithm constitutes an excellent trade-off between simplification effectiveness achieved and the effort required to do so, and that more involved approaches trying to find the optimal simplification may not be worth the extra effort.

Finally, Fig. 10 shows us that the simplification approaches did not result in a meaningful improvement regarding the cost values of the computed solutions. The overall reduction in costs was just 1.31%. Overall, we conclude that while the more complex unsimplified problem descriptions do not lead to a decreased quality of the computed solutions, the merit of employing the QSR based simplification approach is that the time to compute these solutions is significantly reduced.

7. CONCLUSIONS

We addressed the problem of adapting geometric data to satisfy a set of qualitative spatial integrity constraints and described a general approach to the problem based on a formalization of qualitative relations as systems of (in)equations and a translation of these equational systems into an MIP problem. We focused on the question whether and how constraint-based qualitative spatial reasoning techniques can be beneficially employed to improve the performance within this general approach and discussed several options, in particular the simplification of the problem by using QSR to reduce the number of required variables and (in)equations. We presented two different simplification algorithms and performed an experimental evaluation by embedding them into the concrete system for adjusting polygonal objects in the plane to satisfy topological constraints of the RCC-8 and 9-Intersection models considering only displacement. The results demonstrate that indeed a significant performance increase can be achieved by employing the QSR based simplification approaches. While we did not record a noticeable improvement in the quality of the solutions, i.e., no improvement in the values of the cost function, the simplification approaches resulted in a significantly improved run time performance. Comparing the two variants with each other, it turned out that the extended version only improved performance marginally over the basic version. This raises doubts whether even more involved approaches to find the optimal simplification will be worth the extra effort.

Our future research will aim at a thorough investigation of the other proposed ways of exploiting QSR techniques in the context of our qualitative adjustment framework. Furthermore, we plan to extend this framework in several ways: by incorporating other spatial calculi, by allowing for other types of transformations, by facilitating different types of objects (e.g., polygonal and line objects), and by increasing the overall expressivity, for instance by providing means for defining constraints about newly constructed entities, e.g., the intersection or union of input objects, or parts of entities.

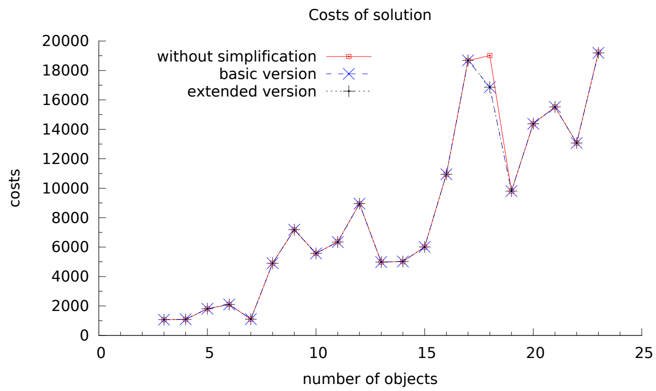


Figure 10: Results regarding the cost values show only a marginal improvement in solution quality.

8. REFERENCES

[1] A. I. Abdelmoty, P. Smart, B. El-Geresy, and C. Jones. Supporting frameworks for the geospatial semantic web. In *Proceedings of the 11th International Symposium on Advances in Spatial and Temporal Databases*, pages 355 – 372, 2009.

[2] M. Bhatt, J. H. Lee, and C. Schultz. CLP(QS): A declarative spatial reasoning framework. In M. J. Egenhofer, N. A. Giudice, R. Moratz, and M. F. Worboys, editors, *Spatial Information Theory (COSIT'11)*, pages 210–230. Springer, 2011.

[3] C. Claramunt, M. Thériault, and C. Parent. A qualitative representation of evolving spatial entities in two-dimensional spaces. *Innovations in GIS V, Carver, S. Ed.*, pages 119–129, 1997.

[4] E. Clementini, J. Sharma, and M. J. Egenhofer. Modeling topological spatial relations: Strategies for query processing. *Computers and Graphics*, 18(92):815–822, 1994.

[5] A. G. Cohn and S. M. Hazarika. Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, 46(1-2):1–29, 2001.

[6] M. Duckham, M.-H. Jeong, S. Li, and J. Renz. Decentralized querying of topological relations between regions without using localization. In *Proc. 18th ACM SIGSPATIAL GIS*, pages 414–417, 2010.

[7] M. J. Egenhofer. Reasoning about binary topological relations. In O. Gunther and H.-J. Schek, editors, *Advances in Spatial Databases, Second Symposium on Large Spatial Databases*, pages 143–160. 1991.

[8] V. Haarslev and R. Möller. Spatioterminological reasoning: Subsumption based on geometrical inferences. In R. J. Brachman, F. M. Donini, E. Franconi, I. Horrocks, A. Y. Levy, and M.-C. Rousset, editors, *Description Logics*, 1997.

[9] L. Harrie. The constraint method for solving spatial conflicts in cartographic generalization. *Cartography and Geographic Information Science*, 26(1), 1999.

[10] J. Jiang and M. Worboys. Detecting basic topological changes in sensor networks by local aggregation. In *Proc. 16th ACM International Conference on Advances in Geographic Information Systems*, pages 1–10, 2008.

[11] A. Khan and M. Schneider. Topological reasoning between complex regions in databases with frequent updates. In *18th ACM SIGSPATIAL Int. Conf. on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS)*, pages 380–389, 2010.

[12] A. Klippel, M. Worboys, and M. Duckham. Identifying factors of geographic event conceptualisation. *International Journal of Geographical Information Science*, 22(2):183–204, 2008.

[13] L. Kulik, M. Duckham, and M. J. Egenhofer. Ontology-driven map generalization. *J. Vis. Lang. Comput.*, 16(3):245–267, 2005.

[14] S. Lamy, A. Ruas, Y. Demazeau, M. Jackson, W. Mackaness, and R. Weibel. The application of agents in automated map generalisation. In *19th Int. Cartographic Conference*, pages 160–169, 1999.

[15] J.-C. Latombe. *Robot Motion Planning*. Kluwer Academic Publishers, 1991.

[16] K. Marriott, P. Moulder, P. Stuckey, and A. Borning. Solving disjunctive constraints for interactive graphical applications. In T. Walsh, editor, *Principles and Practice of Constraint Programming*, volume 2239, pages 361–376. Springer, 2001.

[17] D. A. Randell, Z. Cui, and A. Cohn. A spatial logic based on regions and connection. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 165–176. Morgan Kaufmann, 1992.

[18] J. Renz and B. Nebel. Qualitative spatial reasoning using constraint calculi. In M. Aiello, I. E. Pratt-Hartmann, and J. F. van Benthem, editors, *Handbook of Spatial Logics*. Springer, 2007.

[19] M. A. Rodríguez, N. Brisaboa, J. Meza, and M. R. Luaces. Measuring consistency with respect to topological dependency constraints. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 182–191, 2010.

[20] M. A. Rodríguez, M. J. Egenhofer, and A. D. Blaser. Query pre-processing of topological constraints: Comparing a composition-based with neighborhood-based approach. In T. Hadzilacos, Y. Manolopoulos, J. F. Roddick, and Y. Theodoridis, editors, *Advances in Spatial and Temporal Databases, SSTD 2003*, pages 362–379, 2003.

[21] M. Sester. Generalization based on least squares adjustment. In *International Archives of Photogrammetry and Remote Sensing*, volume 33, 2000.

[22] S. Steiniger and R. Weibel. Relations among map objects in cartographic generalization. *Cartography and Geographic Information Science*, 34(3), 2007.

[23] J. O. Wallgrün. Topological adjustment of polygonal data. In *Advances in Spatial Data Handling*. Springer, to appear.

[24] J. M. Ware and C. B. Jones. Conflict reduction in map generalization using iterative improvement. *GeoInformatica*, 2(4):383–407, 1998.