"© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works."

# On Finding Approximate Solutions of Qualitative Constraint Networks

Jason Jingshi Li † Artificial Intelligence Group The Australian National University Email: jason.li@gmx.ch

Abstract-Qualitative Spatial and Temporal Reasoning (QSTR) represents spatial and temporal information in terms of human comprehensible qualitative predicates and reasons about qualitative information by solving qualitative constraint networks (QCNs). Despite significant progress in the past three decades, more and more evidence has shown that it is inherently hard to find exact solutions for expressive qualitative constraints. In many applications, however, we are often required to make decisions in a very limited time. In these cases, finding a good approximate solution in seconds is much more desirable than waiting days for an exact solution. In this paper, we will exploit the algebraic structure of qualitative calculi (e.g. Interval Algebra and RCC8) as well as their conceptual neighbourhood graphs to develop approximate methods for consistency checking in QSTR. Moreover, we propose and empirically compare four independent methods to serve as tools for finding good approximate solutions for the given qualitative calculi.

# I. INTRODUCTION

Spatial and temporal information is pervasive and increasingly involved in our everyday life. Many tasks in the real or virtual world require sophisticated spatial reasoning abilities. Qualitative Spatial and Temporal Reasoning (QSTR) is the AI subfield that represents spatial and temporal information in terms of human comprehensible qualitative predicates.

In the past three decades, QSTR has made significant progress in relation modelling. Since Allen's seminal work on the Interval Algebra (IA) [1], dozens of qualitative calculi (i.e. relation models) have been proposed. Moreover, prominent relation models such as Interval Algebra and the Region Connection Calculus RCC8 [22] have been applied in areas including natural language processing, geographical information systems, robotics, content-based image retrieval.

QSTR reduces qualitative reasoning (e.g. consistency checking and entailment) to solving qualitative constraint networks. The *consistency problem*, i.e. the problem of deciding the consistency of qualitative constraint networks, is one central task in qualitative reasoning. In the past three decades, the consistency problem has been studied for many different calculi individually and jointly. For example, Nebel and Bürckert show that reasoning with IA is NP-complete in general but also identify the unique maximal tractable subclass of IA which contains all basic IA relations [19]. Renz identifies all three Sanjiang Li QCIS, FEIT University of Technology Sydney Email: sanjiang.li@uts.edu.au



Fig. 1. An inconsistent IA constraint network (left) and one of its consistent approximation (middle) and a solution to the approximation (right)

maximal tractable subclasses of RCC8 that contain all basic RCC8 relations [24], [23].

The consistency problem becomes much harder if we extend for example RCC8 to more expressive topological languages or combine other spatial aspects (e.g. connectivity, Boolean constraints, directions, convexity) with RCC8 constraints. For example, the consistency problem becomes PSPACE hard if RCC8 constraints are augmented with Boolean constraints [28]. When combining RCC8 with the DRM (direction relation matrix) model, the consistency problem becomes NP-hard even when only basic constraints are considered [16].

NP-hardness is, however, not the end of the study of qualitative constraint solving. Several different tools have been develops to decide consistency efficiently [27], [9], [14], [12]. Approximate methods for consistency checking are more desirable in applications where efficiency is vital. In these cases, finding a good approximate solution in seconds is much more desirable than waiting days for an exact solution, where by a good approximate solution we mean a qualitative configuration that satisfies or almost satisfies each constraint.

Tools for finding approximate solutions can be useful in at least two ways. First, if the given qualitative constraint problem is inconsistent, this suggests that the modelling of the problem is improper and some constraints are too hard. In this case, we can use the 'good' approximate solution computed by our methods to remodel the constraint problem. An example in IA is shown in Figure 1, where  $v_i$  (i = 1, 2, 3) are three temporal events. On the left, it specifies that event  $v_1$  should *overlap*  $v_2$ ,  $v_2$  occurs *during*  $v_3$  and  $v_3$  *finishes* with  $v_1$ . One can easily check the composition table of IA to verify that such a constraint network is inconsistent. However, we can quickly modify the relation from  $v_1$  to  $v_3$  to *overlap*, which is within the conceptual neighborhood of the original relation. Then the second constraint network is consistent, and on the right we show one such consistent instantiation of this network. It is

<sup>&</sup>lt;sup>†</sup> This work was carried out when the author was working at the Artificial Intelligence Laboratory, EPFL in Lausanne, Switzerland.

reasonable to regard the second constraint network as a 'good' approximate solution of the first network as 'overlap' is a conceptual neighbour of 'finished\_by' [7].

Second, when the consistency of the given qualitative constraint problem is hard to determine, then we first determine if the problem has a 'good' approximate solution, which assumedly can be efficiently decided. If it has no 'good' approximate solution, it has no solution at all; if it has a 'good' approximate solution, then, starting from the 'good' approximate solution and combined with local search, we may have a great chance to find an exact solution if the problem is consistent.

Approximate methods have been considered in QSTR more than 20 years ago [6], and remain an interesting research topic in QSTR (see e.g. [20], [3], [2], [26]). Some of these works, e.g. [20], approximate a spatial object by a point or a box. Therefore, the shape/structure of the object is ignored. Most other works deal with the impreciseness or uncertainty in spatial/temporal phenomena (see e.g. [3], [2], [25], [26].) That is, the spatial/temporal objects and/or predicates are regarded as fuzzy or vague.

In classical constraint solving, an "approximate consistency solving" technique (e.g. arc or path consistency) usually means a decision procedure that can *partially* decide the consistency problem. That is, the consistency of a set of constraints may remain undetermined after applying the approximation.

Unlike the approximate methods described in the two preceding paragraphs, this paper will exploit the *algebraic structure* of the qualitative calculus as well as its *conceptual neighbourhood graph* (CNG) [7] to develop approximate but efficient methods for consistency checking in QSTR. The idea has been slightly touched in the combination of topological and directional spatial information [15].

The remainder of this paper is organised as follows. Section 2 recalls IA and RCC8 and their conceptual neighbourhood graphs. Section 3 introduces two distances for measuring the "goodness" of an approximate solution and examine the computational complexities of finding good approximate solutions. In Section 4 we give four methods for finding approximate solutions to a qualitative constraint network and these methods are then empirically compared in Section 5. The last section concludes the paper and outlines future work.

# **II. PRELIMINARIES**

The qualitative approach to spatial and temporal knowledge representation and reasoning is mainly based on qualitative calculi. Suppose U is the universe of spatial or temporal entities. Write **Rel**(U) for the Boolean algebra of binary relations on U. A *qualitative calculus* on U is a finite Boolean subalgebra of **Rel**(U). For a relation  $\alpha$  in a qualitative calculus A, we call  $\alpha$  a *basic* relation if it is an atom in A. We note that the set B of basic relations form a *jointly exhaustive and pairwise disjoint* (JEPD) set, i.e. any two elements in U are related by a unique relation R by the set of basic relations it contains.

Relation	Sym.	Conv.	Meaning
before	b	bi	$x^- < x^+ < y^- < y^+$
meets	m	mi	$x^- < x^+ = y^- < y^+$
overlaps	0	oi	$x^{-} < y^{-} < x^{+} < y^{+}$
starts	S	si	$x^- = y^- < x^+ < y^+$
during	d	di	$y^- < x^- < x^+ < y^+$
finishes	f	fi	$y^- < x^- < x^+ = y^+$
equals	eq	eq	$x^- = y^- < x^+ = y^+$

#### TABLE I

Basic IA relations and their converses, where  $x = [x^-, x^+], y = [y^-, y^+]$  are two intervals.

Relation	Meaning
DC	$a \cap b = \varnothing$
EC	$a \cap b \neq \varnothing, a^{\circ} \cap b^{\circ} = \varnothing$
PO	$a \not\subseteq b, b \not\subseteq a, a^{\circ} \cap b^{\circ} \neq \emptyset$
TPP	$a \subset b, a \not\subset b^{\circ}$
NTPP	$a\sub{b^{\circ}}$
EQ	a = b
-	1

TABLE II

Topological interpretation of basic RCC8 relations in a topological space X, where a, b are regions in X, and  $a^{\circ}, b^{\circ}$  are the interiors of a, b, respectively.

In the following subsection, we recall two most important qualitative calculi.

#### A. Interval Algebra and RCC8

Definition 1 (Interval Algebra [1]): Let U be the set of closed intervals on the real line. Thirteen binary relations between two intervals  $x = [x^-, x^+]$  and  $y = [y^-, y^+]$  are defined by the order of the four endpoints of x and y, see Table I. The Interval Algebra (IA) is generated by these JEPD relations.

The RCC8 algebra has interpretations in arbitrary topological spaces.

Definition 2 (RCC8 Algebra): Let U be the set of nonempty regular closed sets, or *regions*, in a topological space X. The RCC8 algebra is generated by the topological relations

# $DC, EC, PO, EQ, TPP, NTPP, TPP^{\sim}, NTPP^{\sim},$

where DC, EC, PO, TPP, NTPP and EQ are defined in Table II, and  $TPP^{\sim}$  and  $NTPP^{\sim}$  are the converses of TPP and NTPP respectively.

# B. Conceptual Neighbourhood Graph

Two basic relations in a qualitative calculus  $\mathbb{A}$  are called *conceptual neighbours* if there exists a smooth transition from one to another directly [7]. For example, in RCC8, **DC** is a conceptual neighbour of **EC**, but not a conceptual neighbour of **PO** (see Table 2). The conceptual neighbourhood graph (CNG) of RCC8 or IA is obtained by connecting basic relations that are conceptual neighbours.

Definition 3: For two basic relations  $\alpha$ ,  $\beta$ , the *conceptual distance* of  $\alpha$  to  $\beta$ , written dist( $\alpha$ ,  $\beta$ ), is the length of a shortest path from  $\alpha$  to  $\beta$  in the conceptual neighbour graph.

Note that dist(PO,  $\alpha$ )  $\leq 2$  for any RCC8 basic relation  $\alpha$  and dist(eq,  $\beta$ )  $\leq 3$  for any IA basic relation  $\beta$ .



Fig. 2. Conceptual neighbourhood graphs of (a) IA and (b) RCC8

For a basic relation  $\alpha$  and a nonempty set R of basic relations, the conceptual distance of  $\alpha$  to R is defined as

$$\mathsf{dist}(\alpha, R) = \min\{\mathsf{dist}(\alpha, \beta) : \beta \in R\}.$$
(1)

#### C. Consistency Problem

Let  $\mathbb{A}$  be a qualitative calculus with universe U and S be a subset of  $\mathbb{A}$ . A qualitative constraint over S has the form (xRy), where R is a relation in S and x, y are spatial/temporal variables taking values from U. A qualitative constraint network (QCN) over S is a set of qualitative constraints over S. Without loss of generality, we assume that  $\mathcal{N}$  has the form  $\{v_i R_{ij} v_j : 1 \le i, j \le n\}$ , where  $R_{ii}$  is the identity relation for each variable  $v_i$ , and, for any two variables  $v_i, v_j$  with  $i \ne j$ ,  $R_{ij}$  is the unique constraint from  $v_i$  to  $v_j$  and  $R_{ji} = R_{ij}^{\sim}$ , the converse of  $R_{ij}$ . We say a QCN  $\mathcal{N}$  over S is *consistent* or *satisfiable* if it has a solution in U. The *consistency problem* over S, written CSPSAT(S), is the problem of deciding the consistency of qualitative constraint networks over S. It is well-known that both CSPSAT(IA) and CSPSAT(RCC8) are NP-complete [19], [23].

A constraint network  $\mathcal{N} = \{v_i R_{ij} v_j : 1 \leq i, j \leq n\}$  is called *basic* if each  $R_{ij}$  is a basic relation. In this paper, we also call a (consistent) basic constraint network as a (consistent) *scenario*. For a consistent scenario  $\mathbf{s} = \{v_i \sigma_{ij} v_j :$  $1 \leq i, j \leq n\}$  and a network  $\mathcal{N} = \{v_i R_{ij} v_j : 1 \leq i, j \leq n\}$ , we say  $\mathbf{s}$  is a consistent scenario (or an *exact solution*) of  $\mathcal{N}$ if  $\sigma_{ij}$  is a basic relation contained in  $R_{ij}$  for any i, j. It is clear that a network is consistent iff it has an exact solution in the above sense.

We say a constraint network  $\mathcal{N} = \{v_i R_{ij} v_j : 1 \le i, j \le n\}$ is *path-consistent* if, for every three variables  $v_i, v_j, v_k$   $(1 \le i, j, k \le n)$ , we have <sup>1</sup>

$$R_{ij} = R_{ji}^{\sim} \quad \text{and} \quad R_{ij} \subseteq R_{ik} \circ_w R_{kj}, \tag{2}$$

where  $R \circ_w S$  is the *weak composition* [5] of R and S, i.e. the smallest relation in  $\mathbb{A}$  that contains the usual composition of Rand S. As a local property, path-consistency can be enforced in cubic time. Recall that IA has a unique maximal tractable subclass that contains all basic relations [19], called ORD-Horn and written  $\mathcal{H}$ , and RCC8 has three maximal tractable subclasses  $\widehat{\mathcal{H}}_8, \mathcal{C}_8, \mathcal{Q}_8$  [23]. Let S be one of the above maximal tractable subclasses. The consistency of any network over Scan be determined by enforcing path-consistency [19], [23], and hence CSPSAT(S) is in P.

<sup>1</sup>This definition of path-consistency is different from that for finite CSPs [18], [17] as we use weak composition instead of composition.

We say a constraint network  $\mathcal{N}$  is *pseudo-consistent* if no empty constraint occurs when enforcing path-consistency. It is clear that every consistent network is pseudo-consistent, but not the vice versa.

# III. APPROXIMATE SOLUTION: DEFINITION AND COMPLEXITIES

Let  $\mathbb{A}$  be a qualitative calculus with universe U, and let  $\mathcal{B}$  be the set of its basic relations. We assume path-consistency implies consistency for basic constraint networks over  $\mathbb{A}$ . Several important qualitative calculi, e.g. the point algebra, the interval algebra, RCC5 and RCC8, enjoy this property.

As the consistency problem is NP-complete for IA and RCC8, it could be very hard to find an exact solution. In this section we discuss what will happen if our task is to find a 'good' approximate solution.

Definition 4 (approximate solution): For a network  $\mathcal{N}$  of constraints over  $\mathbb{A}$ , an approximate solution of  $\mathcal{N}$  is a consistent scenario s that is over the same set of variables  $\{v_1, \dots, v_n\}$  as  $\mathcal{N}$ .

Note that this is a very general notion. In the following we provide measures for evaluating the quality of approximate solutions.

# A. When Is an Approximate Solution Good?

It is clear that two approximate solutions of the same network  $\mathcal{N}$  are not equally good. To measure the "goodness" of an approximate solution s to  $\mathcal{N}$ , we introduce two distances that are intuitive and easy to compute. The first distance is obtained by adding up the conceptual distances between corresponding relations in s and  $\mathcal{N}$  (1).

Definition 5 (sum conceptual distance): Suppose  $\mathbf{s} = \{v_i \sigma_{ij} v_j : 1 \leq i, j \leq n\}$  is a consistent scenario and  $\mathcal{N} = \{v_i R_{ij} v_j : 1 \leq i, j \leq n\}$  is a QCN. The sum conceptual distance of  $\mathbf{s}$  to  $\mathcal{N}$  is defined as

$$\mathsf{dist}_{\Sigma}(\mathbf{s}, \mathcal{N}) = \sum \{ \mathsf{dist}(\sigma_{ij}, R_{ij}) : 1 \le i < j \le n \}.$$
(3)

A consistent scenario s is called an *optimal approximate* solution of  $\mathcal{N}$  if dist(s,  $\mathcal{N}$ )  $\leq$  dist(s',  $\mathcal{N}$ ) for any other consistent scenario s'.

It is clear that the smaller the distance is the better the approximate solution is. In particular, if  $\mathcal{N}$  is consistent, then every optimal approximate solution is an exact solution of  $\mathcal{N}$  and has distance 0 to  $\mathcal{N}$ .

*Remark 1:* The above notion of sum conceptual distance was first proposed for measuring the likeliness of *two consistent scenarios* in [4], where the authors also define the distance between a consistent scenario s and a network  $\mathcal{N}$  as:

$$\mathsf{dist}_{\Sigma}^{+}(\mathbf{s}, \mathcal{N}) = \begin{cases} \min\{\mathsf{dist}(\mathbf{s}, \mathbf{s}') : \mathbf{s}' \in [\mathcal{N}]\}, & \text{if } \mathcal{N} \text{ is consistent} \\ 0, & \text{otherwise;} \end{cases}$$

where [N] is the set of consistent scenarios of N. This extension is different from (3) even when N is consistent.<sup>2</sup>

<sup>&</sup>lt;sup>2</sup>Let  $\mathcal{N}$  and **s** be specified as  $\{v_1 \mathbf{TPP} v_2, v_2 \mathbf{NTPP} v_3, v_1 \star v_3\}$  and, respectively,  $\{v_1 \mathbf{TPP} v_2, v_2 \mathbf{TPP} v_3, v_1 \mathbf{TPP} v_3\}$ , where  $\star$  is the universal relation. Then  $\mathsf{dist}_{\Sigma}(\mathbf{s}, \mathcal{N}) = 1$  but  $\mathsf{dist}_{\Sigma}^+(\mathbf{s}, \mathcal{N}) = 2$ .

Moreover, it does not suit our purpose for characterising the "goodness" of approximate solutions.

Another way to measure the goodness of an approximate solution s to a network  $\mathcal{N}$  is by computing the maximum conceptual distance between corresponding relations in s and  $\mathcal{N}$ .

Definition 6 (max conceptual distance and Cn-solution): Suppose  $\mathbf{s} = \{v_i \sigma_{ij} v_j : 1 \le i, j \le n\}$  is a consistent scenario and  $\mathcal{N} = \{v_i R_{ij} v_j : 1 \le i, j \le n\}$  is a QCN. The max conceptual distance of  $\mathbf{s}$  to  $\mathcal{N}$  is defined as

$$\mathsf{dist}_{\max}(\mathbf{s}, \mathcal{N}) = \max\{\mathsf{dist}(\sigma_{ij}, R_{ij}) : 1 \le i < j \le n\}.$$
(4)

For an integer  $l \geq 0$ , we say a consistent scenario s is an *l*-th order conceptual solution of  $\mathcal{N}$  if dist(s, $\mathcal{N}$ ) = *l*. In the following, we call each first order conceptual solution a conceptual neighbourhood solution (Cn-solution for short) of  $\mathcal{N}$ . If  $\mathcal{N}$  has a Cn-solution, then we also say that  $\mathcal{N}$  is cn-consistent.

Note that 0-th order conceptual solutions are exact solutions of  $\mathcal{N}$ .

Similar to propositional merging [10], there exist many other ways to characterise the "goodness" of approximate solutions. For example, we may consider the sum of the squares of the conceptual distances between corresponding relations in the consistent scenario and the network. The sum and max conceptual distances are in a sense the two limits of the classes of distances proposed in [10]. In this paper, we will also combine the sum and max conceptual distances to obtain a finer measure of the fitness of an approximate solution.

#### B. Complexities

Recall that finding an exact solution (i.e. a consistent scenario) for an IA or RCC8 network is in general NP-complete. It is straightforward to show that finding an optimal approximate solution is also NP-complete.

In the following, we consider the computational complexity of determining when a network has an *l*-th order Cn-solution. We call this problem the *l*-th order Cn-consistency problem. Clearly, the consistency problem and the Cn-consistency problem are actually the zero order and, respectively, the first order cn-consistency problems.

For a basic relation  $\alpha$ , an arbitrary relation R, and an integer  $l \ge 0$ , we call

$$\operatorname{cn}^{l}(\alpha) = \{\beta \in \mathcal{B} : \operatorname{dist}(\alpha, \beta) \le l\}$$
(5)

the *l*-th order conceptual neighbourhood of  $\alpha$ , and call

$$\operatorname{cn}^{l}(R) = \{\beta \in \mathcal{B} : \operatorname{dist}(\beta, R) \le l\} = \bigcup \{\operatorname{cn}^{l}(\sigma) : \sigma \in R\}$$
(6)

the l-th order conceptual neighbourhood of R.

For a network  $\mathcal{N} = \{v_i R_{ij} v_j\}$ , we write

$$\mathsf{cn}^{l}(\mathcal{N}) = \{ v_{i} \mathsf{cn}^{l}(R_{ij}) v_{j} \},\tag{7}$$

and call  $\operatorname{cn}^{l}(\mathcal{N})$  the *l*-th order conceptual neighbourhood network of  $\mathcal{N}$ . If l = 1, then we omit the superscripts in  $\operatorname{cn}^{l}(\alpha)$ ,  $\operatorname{cn}^{l}(R)$ , and  $\operatorname{cn}^{l}(\mathcal{N})$ .



EXAMPLES OF CONCEPTUAL NEIGHBOURHOODS OF RCC8 RELATIONS, WHERE T=**TPP**, N=**NTPP**, TI=**TPP** $^{\sim}$ , NI=**NTPP** $^{\sim}$ , AND, FOR EXAMPLE, CN({**NTPP**}) = {**TPP**, **NTPP**, **EQ**}.

Proposition 1: A network  $\mathcal{N}$  is Cn-consistent if and only if  $cn(\mathcal{N})$  is consistent.

Write  $CN_{RCC8}$  for the subclass of RCC8 that is comprised of all conceptual neighbourhoods. That is, each relation in  $CN_{RCC8}$  is the union of a set of conceptual neighbourhoods of basic relations. Then the cn-consistency problem of RCC8 is exactly the consistency problem over the class  $CN_{RCC8}$ , i.e. CSPSAT( $CN_{RCC8}$ ).

*Theorem 1:* The cn-consistency problem of RCC8 is NP-complete.

*Proof:* We first compute the set  $CN_{RCC8}$ . There are all together 21 such relations (shown in Table III). Then we compute  $\widehat{CN}_{RCC8}$ , the closure of  $CN_{RCC8}$  in RCC8 under converse, intersection, and weak composition. It is comprised of 107 relations, with 24 relations belonging to none of the three maximal tractable subsets. Moreover, the whole RCC5 algebra is contained in  $\widehat{CN}_{RCC8}$ , i.e.  $RCC5 \subset \widehat{CN}_{RCC8} \subset RCC8$ . By the NP-completeness of CSPSAT(RCC5) and CSPSAT(RCC8), we know  $\widehat{CN}_{RCC8}$  is also NP-complete.

Similarly, we can show

*Theorem 2:* The cn-consistency problem and the second order cn-consistency problem of IA are both NP-complete.

**Proof:** For each IA relation R, it can be proved that  $\operatorname{cn}^2(R) = \{\alpha : \operatorname{dist}(\alpha, R) \leq 2\} = \bigcup \{\operatorname{cn}(\beta) : \beta \in \operatorname{cn}(R)\} = \operatorname{cn}(\operatorname{cn}(R))$ . That is, each second order conceptual neighbourhood is also a first order conceptual neighbourhood. Let  $\mathcal{CN}_{IA}$  and  $\mathcal{CN}_{IA}^2$  be the subclasses of IA that are comprised of, respectively, all first order conceptual neighbourhoods. Since  $\mathcal{CN}_{IA} \supseteq \mathcal{CN}_{IA}^2$ , we need only show that the second order conceptual order conceptual neighbourhoods. Since  $\mathcal{CN}_{IA} \supseteq \mathcal{CN}_{IA}^2$ , we need only show that the second order conceptual order conceptual neighbourhoods. Since the subclasses of IA that are composition of  $\mathcal{CN}_{IA}^2$  under converse, intersection, and weak composition happens to be IA itself. Because CSPSAT(IA) is NP-complete, the second order cn-consistency problem of IA is also NP-complete.

Recall that dist(PO,  $\alpha$ )  $\leq 2$  for every RCC8 basic relation  $\alpha$  and dist(eq,  $\beta$ )  $\leq 3$  for every IA basic relation  $\beta$ . This implies that each RCC8 network has a second (and *l*-th for  $l \geq 2$ ) order cn-solution, where every two different variables are related by PO; and each IA network has a third (and *l*-th for  $l \geq 3$ ) order cn-solution, where every two different variables are related by PO; and each IA network has a third (and *l*-th for  $l \geq 3$ ) order cn-solution, where every two different variables are related by eq.

# **IV. FINDING APPROXIMATIONS**

In this section we look at different ways of finding approximations. One fairly obvious idea is to transform the existing SAT encodings [21], [9] to a MaxSAT problem. However, in our first attempt we have found that this could be very inefficient, as the previous SAT encoding only has to encode the weak composition constraints of the existing relations, whereas finding approximation would require encoding composition constraint for all possible relations. Instead, we propose four efficient methods for finding approximations to a qualitative constraint network.

#### A. Approximate with cn-solutions

The notion of cn-solutions provides one good characterisation of "approximate solutions." A cn-solution can be found by using standard constraint search. As a constraint network may have many cn-solutions, it is reasonable to measure the 'goodness' of cn-solutions in terms of conceptual distance (3).

Let  $s = \{v_i \sigma_{ij} v_j : 1 \le i < j \le n\}$  be a consistent scenario and  $\mathcal{N} = \{v_i R_{ij} v_j : 1 \le i < j \le n\}$  be a QCN. Note that if s is a Cn-solution of  $\mathcal{N}$ , then s satisfies m dist(s,  $\mathcal{N}$ ) constraints in  $\mathcal{N}$  and 'almost' satisfies each of the remaining constraints in  $\mathcal{N}$ , where m is the number of nontrivial constraints in s. It is clear that the smaller the distance is the better the approximate solution is. In particular, if  $\mathcal{N}$ is consistent, then each optimal approximate Cn-solution has distance 0 to  $\mathcal{N}$  and, hence, is an exact solution.

Given that we have already shown that deciding cnconsistency is NP-hard and finding a cn-solution may still involve modifying a large number of constraints, we will instead turn and look for polynomial methods that provide 'good' approximate solutions, where the 'goodness' of an approximate solution s is measured by the conceptual distance dist(s,  $\mathcal{N}$ ) of s to  $\mathcal{N}$  (3).

#### B. Approximate with Tractable Subsets

One approach is to make use of the identified tractable subsets of the given qualitative calculus, as they allow consistency of the constraint networks containing only the tractable relations to be decided in polynomial time.

Above we have seen that computing a Cn-solution is hard in general. In the following, however, we show that a Cn-solution can be found in polynomial time if the network is pseudoconsistent, which is a very weak condition. To this end, we introduce the following notion.

Definition 7: Let  $\mathcal{D}$  be a subclass of  $\mathbb{A}$ . Suppose  $\mathcal{N}$  is a network. Write

$$\mathcal{N}_d = \{ x R_{ij}^d y : 1 \le i, j \le n \},\tag{8}$$

where  $R_{ij}^d$ , called the  $\mathcal{D}$ -closure of  $R_{ij}$ , is the smallest relation in  $\mathcal{D}$  that contains  $R_{ij}$ . We call  $\mathcal{N}_d$  the  $\mathcal{D}$ -coarse network of  $\mathcal{N}$ .

*Lemma 1:* Let  $\mathcal{D}$  be a subclass of  $\mathbb{A}$  that is closed under converse, intersection, and weak composition. Suppose  $\mathcal{N}$  is a path-consistent network. Then the  $\mathcal{D}$ -coarse network of  $\mathcal{N}$  is also path-consistent.

*Proof:* For every three variables  $v_1, v_2, v_3$  in  $\mathcal{N}$ , we need to show that  $R_{13}^d$  is contained in  $R_{12}^d \circ_w R_{23}^d$ , the weak composition of  $R_{12}^d$  and  $R_{23}^d$ . By the path-consistency of  $\mathcal{N}$ , we have  $R_{13} \subseteq R_{12} \circ_w R_{23}$ . Because  $\mathcal{D}$  is closed under weak composition,  $R_{12}^d \circ_w R_{23}^d$  is a relation in  $\mathcal{D}$ . It therefore contains  $R_{13}^d$  since it already contains  $R_{13}$ .

Proposition 2: Let  $\mathcal{D}$  be a subclass of  $\mathbb{A}$  that is closed under converse, intersection, and weak composition. Suppose each path-consistent network over  $\mathcal{D}$  is consistent. Then the  $\mathcal{D}$ -coarse network of each pseudo-consistent network  $\mathcal{N}$  over  $\mathbb{A}$  is consistent.

*Proof:* Suppose  $\mathcal{N}$  is a pseudo-consistent network over  $\mathbb{A}$ . Let  $\mathcal{N}^p$  be the path-consistent network obtained by enforcing path-consistency to  $\mathcal{N}$ . Note that  $\mathcal{N}^p$  refines  $\mathcal{N}$ . By Lemma 1, the  $\mathcal{D}$ -coarse network of  $\mathcal{N}^p$ , written  $\mathcal{N}^p_d$ , is also path-consistent. Since  $\mathcal{N}^p_d$  is over  $\mathcal{D}$ , by assumption, we know that  $\mathcal{N}^p_d$  is consistent. Because  $\mathcal{N}^p_d$  refines  $\mathcal{N}_d$ , we know  $\mathcal{N}_d$  is consistent.

By Proposition 2 we know that  $\mathcal{N}_d$  is consistent if  $\mathcal{N}$  is pseudo-consistent. On the other hand, it is not difficult to construct examples to show that there are inconsistent networks whose  $\mathcal{D}$ -coarse network are consistent.

Recall that the ORD-Horn subset  $\mathcal{H}$  of IA, and the three maximal tractable subsets  $\hat{\mathcal{H}}_8$ ,  $\mathcal{Q}_8$ , and  $\mathcal{C}_8$  of RCC8 are all closed under converse, intersection, and weak composition. Moreover, path-consistency implies consistency for networks over any of these subclasses. As a consequence of Proposition 2 we have

Corollary 1: Suppose  $\mathcal{N}$  is a pseudo-consistent IA (or RCC8) network. Let  $\mathcal{D}$  be the (or a) maximal tractable subset of IA (or RCC8). Then the  $\mathcal{D}$ -coarse network of  $\mathcal{N}$  is consistent.

Moreover, using the refinement method for  $\widehat{\mathcal{H}}_8$  [23], we have

*Theorem 3:* Every pseudo-consistent RCC8 network is cn-consistent.

**Proof:** Let  $\mathcal{N}$  be a pseudo-consistent RCC8 network. Write  $\mathcal{N}^p$  for the path-consistent network obtained by enforcing path-consistency to  $\mathcal{N}$ . Then the  $\hat{\mathcal{H}}_8$ -coarse network of  $\mathcal{N}^p$ is path-consistent and a consistent scenario can be constructed by using the refinement method of [23]. For any RCC8 relation R, write R' for its  $\hat{\mathcal{H}}_8$ -closure. It is straightforward to check that the conceptual distance  $\operatorname{dist}(\sigma, R)$  (1) of each basic relation  $\sigma$  in R' to R is either 0 or 1. As a consequence, we know that any consistent scenario of the  $\hat{\mathcal{H}}_8$ -coarse network of  $\mathcal{N}^p$  is a Cn-solution of  $\mathcal{N}^p$ . In particular, the one constructed by using the refinement method of [23] is a Cn-solution of  $\mathcal{N}$ .

The above theorem shows that, despite that it is in general

NP-hard, the cn-consistency problem can be partially determined in polynomial time (by checking if a network is pseudoconsistent) for RCC8.

# C. Approximate with Consistency Subsets

Let  $\mathcal{B}$  be the set of basic relations of  $\mathbb{A}$ . A subset S of  $\mathcal{B}$  is called a *consistency subset* if any assignment  $\sigma$  of basic relations in S to edges in  $\{(v_i, v_j) : 1 \leq i < j \leq n\}$  is a consistent scenario, where  $V = \{v_1, ..., v_n\}$   $(n \geq 1)$  is a set of ordered variables.

For example  $\{eq\}$  is a consistency subset of IA. In fact, IA has altogether 11 consistency subsets, all are singletons. RCC8 has 12 consistency subsets, 8 are singletons and the other 4 are subsets of  $\{DC, EC, PO\}$ .

A consistency subset can be used in finding an approximate consistent scenario for arbitrary networks in linear time. Let  $\mathcal{N} = \{v_i R_{ij} v_j\}$  be an RCC8 network over  $V = \{v_1, ..., v_n\}$ . The following scenario  $\mathbf{s} = \{v_i \sigma_{ij} v_j\}$  provides an approximation to  $\mathcal{N}$ :

$$\sigma_{ij} = \begin{cases} \mathbf{DC}, & \text{if } \mathbf{DC} \in R_{ij}; \\ \mathbf{EC}, & \text{else if } \mathbf{EC} \in R_{ij}; \\ \mathbf{PO}, & \text{else.} \end{cases}$$

Note that s is path-consistent and, hence, consistent. This is because any triple of relations from  $\{DC, EC, PO\}$  is a composition triad of the RCC8 composition table. As the conceptual distance of **PO** to any other basic RCC8 relation is less than or equal to 2, this provides a reasonable approximation to  $\mathcal{N}$ . In fact, s is a second order cn-solution of  $\mathcal{N}$ .

#### D. Approximate with Constraint Search

A fourth possible approximation approach is to modify the constraints when performing the standard constraint search. Whenever an inconsistency is detected, instead of backtracking, we simply relax the violated constraint. In practice, this entails modifying the constraint solver such that when enforcing path-consistency, instead of refining the relation on an edge to empty relation when inconsistency is detected, the solver modifies the relation to the weak composition of the two other relations. For example, if when enforcing pathconsistency we found that  $R_{02} \cap R_{01} \circ_w R_{12} = \emptyset$ , instead of return inconsistent, we simply "repair"  $R_{02}$  to  $R_{01} \circ_w R_{12}$ . Note that this is not exactly the same as removing the original constraint of  $R_{02}$  completely, as the said relation  $R_{02}$  may have already been refined during search, and the product of the weak composition of the two other relations may still have a non-empty intersection of the original relation. This allows us to avoid backtracking, and hence the complexity of the procedure is the same as enforcing path-consistency at  $O(n^3)$ , where n is the number of variables in the constraint network.

## V. EMPIRICAL EVALUATIONS

In this section, we empirically evaluate, on both IA and RCC8 benchmarks, the four approximation methods mentioned in the previous section, viz. approximate with tractable subsets, approximate with consistency subsets, repairing inconsistent constraints with a modified constraint search, and



Fig. 3. Average approximation distance of random RCC8 networks of 100 nodes (upper), and average CPU time of approximating random RCC8 networks of 100 nodes (below), 100 instances per data point.

finding a Cn-solution using standard constraint search. The approaches are compared by the average approximation distance, which is defined as the average ratio between the sum of the conceptual distance of the approximating scenario to the original constraint network (3) and the number of specified constraints of the original constraint network.<sup>3</sup> The constraint search solver is GQR-1418 [27], and the experiments were run on 2.4 GHz processors with a 2GB memory limit.

# A. RCC8

First we evaluate approximating constraint networks of RCC8. We randomly generate two sets of RCC8 networks. In the first set, we allow any RCC8 constraint network size 100 with average degree between 4 to 20. In the second set, we allow only pseudo-consistent networks size 110 with only the relations outside the three maximal tractable subsets of RCC8, with average degree falling in the phase transition region. These are standard benchmarks in literature for deciding consistency in RCC8.

Figure 3 shows the average approximation distance and the CPU solving time for the four different approaches in the first set. We have found that despite CN-search is proved to be NP-complete, while the other three approaches are polynomial, in practice the CPU time used for the approaches are fairly similar. The tractable-subset approach which uses  $\hat{\mathcal{H}}_8$  (denoted by horn in the figure) failed to find any approximations for random networks with average degree more than 12, as none of those networks are pseudo-consistent. We observe that the

<sup>&</sup>lt;sup>3</sup>Recall that we assume  $R_{ji} = R_{ij}^{\sim}$  and hence only consider those constraints  $R_{ij}$  such that i < j and  $R_{ij}$  is not the universe relation.



Fig. 4. Average approximation distance of difficult RCC8 networks of 110 nodes (upper), and average CPU time of approximating difficult RCC8 networks of 110 nodes (below), 100 instances per data point.

constraint-search based approximation (c-search) performed best in most cases, where only very few edges are modified, while the approach using tractable subsets failed to find any approximations for average degree greater than 12. Furthermore, as the number of constraint increases, the c-search approach would require to modify more constraints, but the solving time decreases. Figure 4 shows a similar picture for the difficult networks in the phase transition region, with the approximation based on the tractable subset  $\hat{\mathcal{H}}_8$  become the most expensive approach.

# B. Interval Algebra

Second, we evaluate the approaches for approximating randomly generated constraint networks of Interval Algebra. Since there are no non-trivial consistency subsets for IA, we omit it from this evaluation. Similar to RCC8, we randomly generate two sets of IA networks with 100 nodes, with the first set containing random IA constraint networks with average degree between 4 and 20 (Fig. 5), and the second containing only pseudo-consistent networks with only relations outside the maximal tractable subset ORD-Horn with the average degree in the phase transition region (Fig. 6).

Here we see similar patterns to the random and difficult instances in RCC8, in that in practice the CN-search are not more expensive in CPU time than other polynomial approaches. Also, c-search yields the best performance for networks of low average degree, and that the CPU time decreases as the average degree of the networks are increased. However, in Figure 5 we observe a noticeable difference for random IA networks compared to random networks in RCC8, in that c-search became more expensive in terms of



Fig. 5. Average approximation distance of random IA networks of 100 nodes (upper), and average CPU time of approximating random IA networks with 100 nodes (below), 100 instances per data point.



Fig. 6. Average approximation distance of difficult IA networks of 100 nodes (upper), and average CPU time of approximating difficult IA network of 100 nodes (below), 100 instances per data point.

average distance than approximation using tractable subset and conceptual-neighbourhood search.

#### VI. CONCLUSION AND FUTURE WORK

In this paper we studied finding approximate solutions of qualitative constraint networks. We have determined that the tasks of finding a conceptual neighbourhood solution for RCC8 and IA is NP-complete, and proposed and empirically compared four different methods for finding approximate solutions. Our experiments have shown that approximations can be found efficiently, and the approximation based on constraint search dominates the other approaches in terms of the averaged approximation distance when the average degree of the instances is low.

There are several directions for future research. First, it may be possible to combine the methods proposed here with local search or greedy search for better solutions. Second, the proposed approaches do not give estimations of how close the approximate solutions founded are to the optimal approximate solutions. This may be determined by computing the or a most non-redundant subnetwork of a given network.

Except the ORD-Horn class, there are 17 maximal tractable subclasses of IA [11], which do not contain all basic IA relations. Although path-consistency does not imply consistency for these subclasses, we could also use these tractable subclasses for finding approximate solutions (as described in Section 4.2). Given an IA network, we may first choose the closest (or the most appropriate) tractable subclass and this may improve the efficiency of the approximate method.

Furthermore, in this paper we only discussed qualitative calculi in which path-consistency imply consistency of basic constraint networks. There are other important calculi, e.g. the Direction Relation Matrix (DRM) model [8] for representing cardinal directions between extended objects, which does not enjoy this property but the consistency problem over basic constraint networks is still tractable [29]. Future work will also consider extending approximations to these calculi.

Lastly, we note that our methods can be applied to find approximate solutions to both consistent networks and inconsistent networks. This suggest that these methods could also be used to repairing or merging inconsistent qualitative constraint networks [4], [13].

#### **ACKNOWLEDGEMENTS**

This work was partially supported by Australian Research Council (DP120104159, DP120103758, FT0990811), National Natural Science Foundation of China (61228305), and the OpenSense project funded by the NanoTera.ch program. We thank Jochen Renz and Tony Cohn for fruitful discussions on our joint ARC Project "Approximate Reasoning with Qualitative Spatial Constraints Involving Landmarks."

#### REFERENCES

- [1] J. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26:832–843, 1983.
- [2] E. Clementini and P. Di Felice. Approximate topological relations. International Journal of Approximate Reasoning, 16(2):173–204, 1997.
- [3] A. Cohn and N. Gotts. The 'egg-yolk' representation of regions with indeterminate boundaries. In P. Burrough and A. Frank, editors, *Proceedings, GISDATA Specialist Meeting on Geographical Objects* with Undetermined Boundaries. Francis Taylor, 1996.
- [4] J.-F. Condotta, S. Kaci, and N. Schwind. A framework for merging qualitative constraints networks. In D. Wilson and H. C. Lane, editors, *FLAIRS Conference*, pages 586–591. AAAI Press, 2008.
- [5] I. Düntsch, H. Wang, and S. McCloskey. A relation-algebraic approach to the region connection calculus. *Theoretical Computer Science*, 255:63-83, 2001.

- [6] S. Dutta. Qualitative spatial reasoning: A semi-quantitative approach using fuzzy logic. In A. P. Buchmann, O. Günther, T. R. Smith, and Y.-F. Wang, editors, SSD, volume 409 of *Lecture Notes in Computer Science*, pages 345–364. Springer, 1989.
- [7] C. Freksa. Temporal reasoning based on semi-intervals. Artificial Intelligence, 54(1):199–227, 1992.
- [8] Goyal, R. and Egenhofer, M., 1997. The direction-relation matrix: A representation for directions relations between extended spatial objects. *In: The Annual Assembly and the Summer Retreat of University Consortium for Geographic Information Systems Science.*
- [9] J. Huang, J. J. Li, and J. Renz. Decomposition and tractability in qualitative spatial and temporal reasoning. *Artificial Intelligence*, 195: 140–164, 2013.
- [10] S. Konieczny and R. P. Pérez. On the frontier between arbitration and majority. In D. Fensel, F. Giunchiglia, D. L. McGuinness, and M.-A. Williams, editors, *KR*, pages 109–120. Morgan Kaufmann, 2002.
- [11] A. Krokhin, P. Jeavons and P. Jonsson. Reasoning about temporal relations: The tractable subalgebras of allen's interval algebra. *Journal* of the ACM, 50(5):591–640, 2003.
- [12] J. J. Li Qualitative Spatial and Temporal Reasoning with Answer Set Programming. In *Proceedings of the IEEE 24th International Conference on Tools with Artificial Intelligence (ICTAI'12)*, 2012.
- [13] J. J. Li, T. Kowalski, J. Renz and S. Li Combining Binary Constraint Networks in Qualitative Reasoning. in *Proceedings of the 18th European Conference in Artificial Intelligence (ECAI'08)*, p.515–519, 2008
- [14] J. J. Li, and J. Renz. In defense of large qualitative calculi. In Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10), 2010.
- [15] S. Li and A. G. Cohn. Reasoning with topological and directional spatial information. *Computational Intelligence*, 28(4):579–616, 2012.
- [16] W. Liu, S. Li, and J. Renz. Combining RCC-8 with qualitative direction calculi: Algorithms and complexity. In *Proceedings of the* 21th International Joint Conference on Artificial Intelligence (IJCAI'09). AAAI, 2009.
- [17] A. Mackworth. Consistency in networks of relations. Artificial Intelligence, 8:99–118, 1977.
- [18] U. Montanari. Networks of constraints: fundamental properties and applications to picture processing. *Information Science*, 7:95–132, 1974.
- [19] B. Nebel and H.-J. Bürckert. Reasoning about temporal relations: A maximal tractable subclass of Allen's interval algebra. *Journal of the* ACM, 42(1):43–66, 1995.
- [20] D. Papadias and M. J. Egenhofer. Algorithms for hierarchical spatial reasoning. *Geoinformatica*, 1(3):251–273, 1997.
- [21] D. N. Pham, J. Thornton and A. Sattar Modelling and solving temporal reasoning as propositional satisfiability. *Artificial Intelligence*, 172(15), pages 1752-1782, 2008.
- [22] D. Randell, Z. Cui, and A. Cohn. A spatial logic based on regions and connection. In *KR-92*, pages 165–176, 1992.
- [23] J. Renz. Maximal tractable fragments of the Region Connection Calculus: A complete analysis. In D. Dean, editor, *Proceedings of* the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99), pages 448–454. Morgan Kaufmann, 1999.
- [24] J. Renz and B. Nebel. On the complexity of qualitative spatial reasoning: A maximal tractable fragment of the Region Connection Calculus. In *IJCAI*, pages 522–527, 1997.
- [25] S. Schockaert, M. D. Cock, and E. E. Kerre. Fuzzifying Allen's temporal interval relations. *IEEE T. Fuzzy Systems*, 16(2):517–533, 2008.
- [26] S. Schockaert, M. De Cock, and E. E. Kerre. Spatial reasoning in a fuzzy region connection calculus. *Artificial Intelligence*, 173(2):258– 298, 2009.
- [27] M. Westphal, S. Wölfl, and Z. Gantner. GQR: A fast solver for binary qualitative constraint networks. In AAAI Spring Symposium: Benchmarking of Qualitative Spatial and Temporal Reasoning Systems, pages 51–52. AAAI, 2009.
- [28] F. Wolter and M. Zakharyaschev. Spatial reasoning in RCC-8 with Boolean region terms. In *Proceedings of the 14th European Conference* on Artificial Intelligence (ECAI'00), pages 244–250, 2000.
- [29] X. Zhang, W. Liu, S. Li, and M. Ying. Reasoning with cardinal directions: An efficient algorithm. In D. Fox and C. Gomes, editors, *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelli*gence (AAAI'08), 2008.