STAIRS 2014
U. Endriss and J. Leite (Eds.)

© 2014 The Authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License.
doi:10.3233/978-1-61499-421-3-91

A Logic of Part and Whole for Buffered Geometries

Heshan DU [a] and Natasha ALECHINA [a]
[a] *University of Nottingham, UK*

Abstract. We propose a new qualitative spatial logic for reasoning about part-whole relations between geometries (sets of points) represented in different geospatial datasets, in particular crowd-sourced datasets. Since geometries in crowd-sourced data can be less inaccurate or precise, we buffer geometries by a margin of error or level of tolerance σ, and define part-whole relation for buffered geometries. The relations between geometries considered in the logic are: buffered part of (BPT), Near and Far. We provide a sound and complete axiomatisation of the logic with respect to metric models and show that its satisfiability problem is NP-complete.

1. MOTIVATION

This work is motivated by our previous work [3] on integrating authoritative geospatial information and crowd-sourced or volunteered geospatial information. Geometry representations of the same location or place in different datasets are usually not exactly the same. Objects are also sometimes represented at different levels of granularity. For example, consider geometries of objects in Nottingham city centre given by the Ordnance Survey of Great Britain (OSGB) [7] and by the OpenStreetMap (OSM) [6] in Figure 1. The position and shape of the Prezzo Ristorante are represented differently in OSGB (dotted) and OSM (solid) (Figure 1.a). The Victoria Shopping Centre is represented as a whole in OSM (Figure 1.b), and as several shops in OSGB (Figure 1.c).

In order to integrate the datasets, we need to determine which objects are the same and sometimes (as in the example of Victoria Shopping Centre) which objects in one dataset are parts of objects in another. One way to produce such matches is to use locations and geometries of objects, although of course we also use any lexical labels associ-

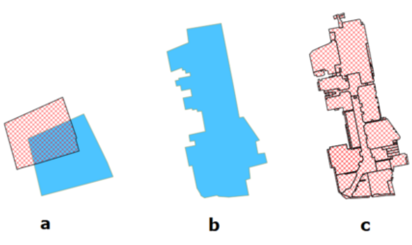

Figure 1. a. Prezzo Ristorante; **b.** Victoria Shopping Centre in OSM; **c.** Victoria Shopping Centre in OSGB

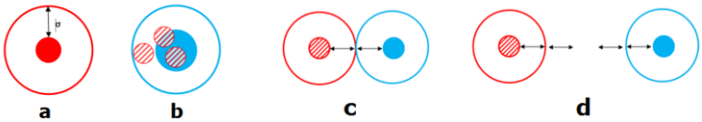

Figure 2. **a.** a buffer; **b.** three dashed circles are buffered part of (BPT) the solid circle; **c.** NEAR; **d.** FAR

ated with the objects, such as names of restaurants etc. The generated matches are seen as assumptions, and are retractable if found incorrect. We check correctness of matches by checking their logical consistency. Some of the checks use ontology reasoning (if an object is classified as a restaurant in one dataset and as a bank in another, together with an axiom stating that the concepts of Restaurant and Bank are disjoint, a contradiction can be derived). Other checks are performed using spatial reasoning. In [4], we proposed a spatial logic LNF that contains relations of being buffered equal (BEQ), Near and Far to validate 'sameAs' matches of objects: if it is conjectured that a_1 is 'sameAs' b_1 and that a_2 is 'sameAs' b_2, then a contradiction can be derived if $NEAR(a_1, a_2)$ and $FAR(b_1, b_2)$. However, LNF is not appropriate for verifying 'partOf' matches. In this paper, we are proposing a logic where we can formalise for example the following argument: if b and c are near, c is part of d, then b cannot be part of a which is far from d. The main concepts of this logic, which we call a Logic of ParT and whole for Buffered geometries (LBPT), are explained in the next section. We also compare it to existing spatial logics.

2. BPT, NEAR, FAR

For the application described in the previous section, we found it difficult to use formalisms such as RCC and other topology or mereology theories [1], since they presuppose accurate geometries or regions with sharp boundaries. Unlike existing models for spatial relations between indeterminate regions or objects with broad boundaries based on rough set theory [8], such as [2] and [9], we could not define a certain inner region, because the same location can be represented using two disconnected polygons from authoritative and crowd-sourced geospatial datasets respectively, which requires that the whole region within the buffer [5] of a geometry is uncertain. We did not adopt probabilistic or fuzzy approaches, such as [12] and [10], because we did not have a good way to define a proper probability function or a membership function for a fuzzy set. The first logic we designed for debugging geometry matches, LNF [4], has the 'buffered equal' relation as a basic relation, which turns out to be less useful when the data is represented at different levels of abstraction (such as a shopping centre in one set and a collection of shops in another). In [4], we gave a complete and sound axiomatisation for LNF, but only with respect to geometries consisting of a single point. In this paper, we start with the 'buffered part of' (BPT) relation as the basic relation, and interpret geometries as sets of points.

As shown in Figure 2.a, by buffering the solid circle c by σ, where σ indicates the margin of error or level of tolerance, we obtain a larger circle, denoted as $buffer(c, \sigma)$, where every point is within σ distance from c. For a geometry c which is possibly represented inaccurately within the margin of error σ, the actual accurate representation is assumed to be somewhere within $buffer(c, \sigma)$. A geometry g is buffered part of a geometry h, if g is within $buffer(h, \sigma)$ (Figure 2.b).

We also have NEAR and FAR relations in the logic LBPT. They formalise concepts of being 'possibly connected' (given a possible displacement by σ) and 'definitely disconnected' (even if displaced by σ) respectively. Two geometries are possibly connected iff their σ buffers are connected. Figure 2.c and Figure 2.d show the boundary case of being NEAR, where $distance(g,h) = 2\sigma$ (their buffers are externally connected) and the case where two geometries are far apart and cannot possibly correspond to connected objects respectively.

3. SYNTAX, SEMANTICS AND AXIOMS OF LBPT

The language $L(LBPT)$ contains a set of individual names, three binary predicates *NEAR*, *FAR* and *BPT*, and logical connectives, $\neg, \wedge, \vee, \rightarrow$.

Applying predicate letters to individual names yields atomic formulas, e.g. $BPT(a,b)$. Every atomic formula is a well-formed formulas (wffs). If α and β are wffs, then $\neg\alpha$, $\alpha \wedge \beta$, $\alpha \vee \beta$, $\alpha \rightarrow \beta$ are wffs.

We interpret the logic over models which are based on a metric space (similar to other spatial logics, such as [13] and [1], and also similar to [11] but for a different logical language).

Definition 1 (Metric Space) *A* metric space *is a pair* (Δ, d)*, where* Δ *is a set and* d *is a metric on* Δ*, i.e., a function* $d : \Delta \times \Delta \longrightarrow \mathbb{R}_{\geq 0}$ *such that for any* $x, y, z \in \Delta$*, the following holds:*

1. $d(x,y) = 0$ *iff* $x = y$*;*
2. $d(x,y) = d(y,x)$*;*
3. $d(x,z) \leq d(x,y) + d(y,z)$*.*

Definition 2 (Metric Model) *A metric model* M *is a tuple* (Δ, d, I, σ)*, where* (Δ, d) *is a metric space,* I *is an interpretation function which maps each constant to a set of elements in* Δ*, and* $\sigma \in \mathbb{R}_{>0}$ *is the margin of error. The notion of* $M \models \phi$ *(*ϕ *is true in model* M*) is defined as follows:*

$M \models BPT(a,b)$ *iff* $\forall p_a \in I(a) \exists p_b \in I(b) : d(p_a, p_b) \in [0, \sigma]$*;*
$M \models NEAR(a,b)$ *iff* $\exists p_a \in I(a) \exists p_b \in I(b) : d(p_a, p_b) \in [0, 2\sigma]$*;*
$M \models FAR(a,b)$ *iff* $\forall p_a \in I(a) \forall p_b \in I(b) : d(p_a, p_b) \in (4\sigma, +\infty)$*;*
$M \models \neg\alpha$ *iff* $M \not\models \alpha$*;* $\quad$ $M \models \alpha \wedge \beta$ *iff* $M \models \alpha$ *and* $M \models \beta$*;*
$M \models \alpha \vee \beta$ *iff* $M \models \alpha$ *or* $M \models \beta$*;* $\quad$ $M \models \alpha \rightarrow \beta$ *iff* $M \not\models \alpha$ *or* $M \models \beta$

where a, b *are individual names,* α, β *are wffs.*

A formula α is valid ($\models \alpha$) if for every metric model M, $M \models \alpha$. The logic LBPT is the set of all valid formulas of $L(LBPT)$.

The following calculus (that we will also refer to as LBPT) will be shown sound and complete for LBPT:

Axiom 0 All tautologies of classical propositional logic;
Axiom 1 $BPT(a,a)$;
Axiom 2 $NEAR(a,b) \rightarrow NEAR(b,a)$;

Axiom 3 $FAR(a,b) \rightarrow FAR(b,a)$;
Axiom 4 $BPT(a,b) \wedge BPT(b,c) \rightarrow NEAR(c,a)$;
Axiom 5 $BPT(b,a) \wedge BPT(b,c) \rightarrow NEAR(c,a)$;
Axiom 6 $BPT(b,a) \wedge NEAR(b,c) \wedge BPT(c,d) \rightarrow \neg FAR(d,a)$;
Axiom 7 $NEAR(a,b) \wedge BPT(b,c) \wedge BPT(c,d) \rightarrow \neg FAR(d,a)$;
MP Modus ponens: $\phi,\ \phi \rightarrow \psi \vdash \psi$.

The notion of derivability $\Gamma \vdash \phi$ in LBPT is standard. A formula ϕ is LBPT-derivable if $\vdash \phi$. A set Γ is (LBPT) inconsistent if for some formula ϕ it derives both ϕ and $\neg\phi$.

4. SOUNDNESS AND COMPLETENESS OF LBPT

In this section we prove that LBPT is sound and complete with respect to metric models, namely that $\vdash \phi \Leftrightarrow \models \phi$. Proofs of some lemmas are omitted due to lack of space. Detailed proofs can be found here: `www.cs.nott.ac.uk/~hxd/report/lbpt.pdf`.

Theorem 1 (Soundness) *Every LBPT derivable formula is valid:* $\vdash \phi \Rightarrow \models \phi$.

Proof. The proof is by an easy induction on the length of the derivation of ϕ. Axioms 1-7 are valid (by the truth definition of *BPT*, *NEAR* and *FAR*) and modus ponens preserves validity. QED.

In the rest of this section, we prove completeness. We will actually prove that given a finite consistent set of formulas, we can build a satisfying model for it. This shows that $\nvdash \phi \Rightarrow \not\models \phi$ and by contraposition we get completeness. The completeness proof is more involved than that for LNF with respect to point geometries [4].

Definition 3 (MCS) *A set of formulas Γ in the language $L(LBPT)$ is maximal consistent, if Γ is consistent, and any set of LBPT formulas over the same set of individual names properly containing Γ is inconsistent. If Γ is a maximal consistent set of formulas, then we call it an MCS.*

Lemma 1 (Lindenbaum's Lemma) *If Σ is a consistent set of formulas in the language $L(LBPT)$, then there is an MCS Σ^+ over the same set of individual names such that $\Sigma \subseteq \Sigma^+$.*

Let $\phi_0, \phi_1, \phi_2, ...$ be an enumeration of LBPT formulas over the same set of individual names as that in Σ. Σ^+ can be defined as follows:

- $\Sigma_0 = \Sigma$;
- $\Sigma_{n+1} = \Sigma_n \cup \{\phi_n\}$, if it is consistent, otherwise, $\Sigma_{n+1} = \Sigma_n \cup \{\neg\phi_n\}$;
- $\Sigma^+ = \bigcup_{n \geq 0} \Sigma_n$.

Given a consistent set of formulas Σ, we construct a metric model satisfying a maximal consistent set Σ^+ containing Σ, following the steps below.

Step 1 We equivalently transform Σ^+ to $B(\Sigma^+)$, a set of basic quantified formulas.
Step 2 We construct a set of distance constraints $D(\Sigma^+)$ from $B(\Sigma^+)$, such that any metric space satisfying $D(\Sigma^+)$ can be extended to a model of $B(\Sigma^+)$.

Step 3 We show that if $D(\Sigma^+)$ is path-consistent, then there is a metric space (Δ, d) satisfying $D(\Sigma^+)$.

Step 4 We show that $D(\Sigma^+)$ is path-consistent, if Σ^+ is consistent.

Since Σ^+ is consistent, then there is a metric space that can be extended to a metric model satisfying $B(\Sigma^+)$, thus, Σ^+, thus, Σ.

In **Step 1**, we equivalently transform Σ^+ to a set of basic quantified formulas defined as follows.

Definition 4 (Basic Quantified Formula) *Observe that atomic LBPT formulas are equi-satisfiable with first order quantified formulas corresponding to their truth conditions in Definition 2:*

- $BPT(a,b)$ *and the formula* $\forall p_a \in a \exists p_b \in b : d(p_a, p_b) \in [0, \sigma]$ *are equi-satisfiable;*
- $NEAR(a,b)$ *and the formula* $\exists p_a \in a \exists p_b \in b : d(p_a, p_b) \in [0, 2\sigma]$ *are equi-satisfiable;*
- $FAR(a,b)$ *and the formula* $\forall p_a \in a \forall p_b \in b : d(p_a, p_b) \in (4\sigma, \infty)$ *are equi-satisfiable.*

We refer to these first order quantified formulas as basic quantified formulas, and use the following abbreviations for them, where g is a non-negative interval.

- $\forall(a,b,g) \equiv (\forall p_a \in a \forall p_b \in b : d(p_a, p_b) \in g)$;
- $\exists(a,b,g) \equiv (\exists p_a \in a \exists p_b \in b : d(p_a, p_b) \in g)$;
- $\chi(a,b,g) \equiv (\forall p_a \in a \exists p_b \in b : d(p_a, p_b) \in g)$;
- $\xi(a,b,g) \equiv (\exists p_a \in a \forall p_b \in b : d(p_a, p_b) \in g)$.

Lemma 2 *For any MCS* Σ^+ *and any pair of individual names* a, b *occurring in* Σ*, exactly one of the following cases holds:*

C1 $BPT(a,b) \wedge BPT(b,a) \in \Sigma^+$;
C2 $BPT(a,b) \wedge \neg BPT(b,a) \in \Sigma^+$;
C3 $\neg BPT(a,b) \wedge BPT(b,a) \in \Sigma^+$;
C4 $\neg BPT(a,b) \wedge \neg BPT(b,a) \wedge NEAR(a,b) \in \Sigma^+$;
C5 $\neg NEAR(a,b) \wedge \neg FAR(a,b) \in \Sigma^+$;
C6 $FAR(a,b) \in \Sigma^+$.

Definition 5 (B(Σ^+)) *Given an MCS* Σ^+*, a corresponding set of basic quantified formulas* $B(\Sigma^+)$ *is constructed as follows. For every pair of individual names* a, b*, we translate the LBPT formulas to basic quantified formulas:*

- $translate(BPT(a,b) \wedge BPT(b,a)) = \{\chi(a,b,[0,\sigma]), \chi(b,a,[0,\sigma])\}$;
- $translate(BPT(a,b) \wedge \neg BPT(b,a)) = \{\chi(a,b,[0,\sigma]), \xi(b,a,(\sigma,\infty))\}$;
- $translate(\neg BPT(a,b) \wedge BPT(b,a)) = \{\xi(a,b,(\sigma,\infty)), \chi(b,a,[0,\sigma])\}$;
- $translate(\neg BPT(a,b) \wedge \neg BPT(b,a) \wedge NEAR(a,b)) =$ $\{\xi(a,b,(\sigma,\infty)), \xi(b,a,(\sigma,\infty)), \exists(a,b,[0,2\sigma]), \exists(b,a,[0,2\sigma])\}$;
- $translate(\neg NEAR(a,b) \wedge \neg FAR(a,b)) =$ $\{\forall(a,b,(2\sigma,\infty)), \forall(b,a,(2\sigma,\infty)), \exists(a,b,[0,4\sigma]), \exists(b,a,[0,4\sigma])\}$;
- $translate(FAR(a,b)) = \{\forall(a,b,(4\sigma,\infty)), \forall(b,a,(4\sigma,\infty))\}$.

Let $names(\Sigma)$ *be the set of individual names occurring in* Σ*. Then,*

$$B(\Sigma^+) = \bigcup_{a \in names(\Sigma), b \in names(\Sigma)} translate(case(a,b))$$

where $case(a,b)$ returns the LBPT formula in the case of a,b specified in Lemma 2.

By the construction above, $B(\Sigma^+)$ contains the same set of individual names as Σ, and any metric model satisfying $B(\Sigma^+)$ satisfies Σ^+.

In **Step 2**, for a set of basic quantified formulas $B(\Sigma^+)$, we construct a set of distance constraints $D(\Sigma^+)$, and then show that if there is a metric space satisfying $D(\Sigma^+)$, then it can be extended to a model of $B(\Sigma^+)$ (hence Σ^+).

Next we turn to producing enough 'points' to populate geometries corresponding to individual names. The next definition specifies the cardinality of the set $points(a)$ (points assigned to an individual name a).

Definition 6 ($num_{B(\Sigma^+)}(a)$) *Let $names(\Sigma)$ be the set of individual names occurring in Σ, $B(\Sigma^+)$ is a corresponding set of basic quantified formulas of Σ^+, an MCS of Σ. For any individual name $a \in names(\Sigma)$,*

$$\begin{aligned} num_{B(\Sigma^+)}(\exists a) &= |\{b \in names(\Sigma) \mid \exists g : \exists(a,b,g) \in B(\Sigma^+)\}| \\ num_{B(\Sigma_+)}(\xi a) &= |\{b \in names(\Sigma) \mid \exists g : \xi(a,b,g) \in B(\Sigma^+)\}| \\ num_{B(\Sigma^+)}(\chi a) &= |\{b \in names(\Sigma) \mid \exists g : \chi(b,a,g) \in B(\Sigma^+)\}| \end{aligned}$$

Then $num_{B(\Sigma^+)}(a) = \max(1, num_{B(\Sigma^+)}(\exists a) + num_{B(\Sigma^+)}(\xi a) + num_{B(\Sigma^+)}(\chi a))$.

We omit subscript $B(\Sigma^+)$ for readability when it is clear from context.

Definition 7 (Witness for a formula) *A witness for a formula $\exists(a,b,g)$ is a pair of constants $p_a \in a$, $p_b \in b$ such that $d(p_a, p_b) \in g$. A witness for a formula $\xi(a,b,g)$ or $\chi(b,a,g)$ is a constant $p_a \in a$, such that $d(p_a, p_b) \in g$, for any constant $p_b \in b$. A constant is clean for a formula, if it is not a witness for any other formula.*

Definition 8 (D(Σ^+)) *Let $B(\Sigma^+)$ be the set of basic quantified formulas of an MCS Σ^+. To every individual name a in Σ, we assign a fixed set of new constants, $points(a) = \{p_a^1, \cdots, p_a^n\}$, where $n = num(a)$. We construct a set of distance constraints $D(\Sigma^+)$ as follows, by iterating through basic quantified formulas in $B(\Sigma^+)$ and eliminating quantifiers on new constants. Initially, $D(\Sigma^+) = \{\}$. For every individual name a in Σ, for every constant $p_a \in points(a)$, we add $d(p_a, p_a) \in \{0\}$ to $D(\Sigma^+)$. Then $\chi(a,a,\{0\})$ always holds. For every pair of different individual names a,b, if*

- *$\exists(a,b,g) \in B(\Sigma^+)$, then we take clean constants $p_a \in points(a)$, $p_b \in points(b)$, and add $d(p_a,p_b) = d(p_b,p_a) \in g$ to $D(\Sigma^+)$ (p_a, p_b become the witness for $\exists(a,b,g)$);*
- *$\xi(a,b,g) \in B(\Sigma^+)$, then we take a clean constant $p_a \in points(a)$, for every $p_b \in points(b)$, we add $d(p_a,p_b) = d(p_b,p_a) \in g$ to $D(\Sigma^+)$;*
- *$\xi(b,a,g) \in B(\Sigma^+)$, then we take a clean constant $p_b \in points(b)$, for every $p_a \in points(a)$, we add $d(p_a,p_b) = d(p_b,p_a) \in g$ to $D(\Sigma^+)$;*
- *$\chi(a,b,g) \in B(\Sigma^+)$, then we take a clean constant $p_b \in points(b)$, for every $p_a \in points(a)$, we add $d(p_a,p_b) = d(p_b,p_a) \in g$ to $D(\Sigma^+)$;*
- *$\chi(b,a,g) \in B(\Sigma^+)$, then we take a clean constant $p_a \in points(a)$, for every $p_b \in points(b)$, we add $d(p_a,p_b) = d(p_b,p_a) \in g$ to $D(\Sigma^+)$;*
- *$\forall(a,b,g) \in B(\Sigma^+)$, then for every pair of constants $p_a \in points(a)$, $p_b \in points(b)$, we add $d(p_a,p_b) = d(p_b,p_a) \in g$ to $D(\Sigma^+)$.*

For every pair of different constants p,q involved in $D(\Sigma^+)$, we add $d(p,q)=d(q,p)\in[0,\infty)$ to $D(\Sigma^+)$, then repeatedly replace $d(p,q)=d(q,p)\in g_1$ and $d(p,q)=d(q,p)\in g_2$ with $d(p,q)=d(q,p)\in(g_1\cap g_2)$, until there is one distance range for each pair of p,q.

Lemma 3 *Any metric space satisfying $D(\Sigma^+)$ can be extended to a model of $B(\Sigma^+)$.*

Proof.(sketch) Suppose S is a metric space satisfying $D(\Sigma^+)$. We extend S to a model M by interpreting every a occurring in Σ as $points(a)$, as specified in Definition 8. We can prove that for any individual name a, $points(a)$ covers all the clean constants needed for constructing $D(\Sigma^+)$. Then by Definition 8, every basic quantified formula has a witness. Therefore M is a model of $B(\Sigma^+)$. QED.

$D(\Sigma^+)$ and $B(\Sigma^+)$ are not equi-satisfiable because of the way we assign witnesses for χ formulas, but we will show that if $B(\Sigma^+)$ is consistent then $D(\Sigma^+)$ can be satisfied in a metric space. The following definitions are essential for **Step 3** and **Step 4**.

Definition 9 (Composition) *For non-negative numbers d_1,d_2, the composition $\{d_1\}\circ\{d_2\}=[|d_1-d_2|,d_1+d_2]$* [1]. *For non-negative intervals g_1,g_2, their composition $g_1\circ g_2=\bigcup_{d_1\in g_1,d_2\in g_2}\{d_1\}\circ\{d_2\}$.*

Definition 10 (Path Consistency) *Given a set of distance constraints D, for every pair of constants a,b, their distance range is strengthened by enforcing path-consistency as follows until a fixed point is reached:*

$$\forall c: g(a,b)\leftarrow g(a,b)\cap(g(a,c)\circ g(c,b))$$

where c is a constant different from a,b, $g(a,b)$ denotes the distance range for a,b (i.e. $d(a,b)\in g(a,b)$). If at the fixed point, for every pair of constants a,b, there exists a valid value for their distance, this is, $g(a,b)\neq\emptyset$, then D is path-consistent.

Definition 11 (Primitive, Composite, Definable Intervals) *Let h be a non-negative interval. h is primitive, if h is one of $[0,\sigma]$, (σ,∞), $[0,2\sigma]$, $(2\sigma,\infty)$, $(2\sigma,4\sigma]$, $(4\sigma,\infty)$, $[0,\infty)$. h is composite, if it can be composed using at least two primitive intervals. h is definable, if it is primitive or composite.*

It is easy to show that if an interval occurs in $D(\Sigma^+)$, then it is an identity interval ($\{0\}$) or a primitive interval.

Definition 12 ($DS(\Sigma^+)$) *We define the set of distance constraints which appear in the process of enforcing path-consistency on $D(\Sigma^+)$, denoted as $DS(\Sigma^+)$, as follows:*

- *Any distance constraint in $D(\Sigma^+)$ is in $DS(\Sigma^+)$;*
- *If distance constraints $d(a,b)\in h$ and $d(b,c)\in g$ are in $DS(\Sigma^+)$, then $d(a,c)\in h\circ g$ is in $DS(\Sigma^+)$;*
- *If distance constraints $d(a,b)\in h$ and $d(a,b)\in g$ are in $DS(\Sigma^+)$, then $d(a,b)\in h\cap g$ is in $DS(\Sigma^+)$*

where a,b,c are constants in $D(\Sigma^+)$.

[1] Based on $d(x,z)\leq d(x,y)+d(y,z)$ (Property 3 of Definition 1).

For a distance constraint $d(a,b) \in h$ in $DS(\Sigma^+)$, we proved that h is a non-negative interval; h is either right-infinite or right-closed; if $lower(h) \neq 0$, then h is left-open.

We are now going to characterise all possible distance constraints occurring in $DS(\Sigma^+)$. Eventually, we will show that all those distance constraints are left and right definable in the sense given below. For an interval h of the form (l,u), $[l,u)$, $(l,u]$ or $[l,u]$, we call l the lower bound of h, represented as $lower(h)$, and u the upper bound of h, represented as $upper(h)$. We allow $lower(h)$ or $upper(h)$ to be ∞. For the lower or upper bound of an interval h, we use $^-$ or $^+$ to denote that h is open or closed respectively.

Definition 13 (Left-Definable) *A distance constraint $d(c_1,c_n) \in h$ $(n > 1)$ is left-definable, iff there exists a sequence of distance constraints $d(c_i,c_{i+1}) \in h_i$ $(0 < i < n)$ in $D(\Sigma^+)$, such that, for $h' = h_1 \circ \ldots \circ h_{n-1}$, the following holds:*

1. *If h is left-open, then h' is left-open, $h \subseteq h'$, and $lower^-(h') = lower^-(h)$;*
2. *If h is left-closed, then h' is left-closed, $h \subseteq h'$, and $lower^+(h') = lower^+(h)$.*

Definition 14 (Right-Definable) *A distance constraint $d(c_1,c_n) \in h$ $(n > 1)$ is right-definable, iff there exists a sequence of distance constraints $d(c_i,c_{i+1}) \in h_i$ $(0 < i < n)$ in $D(\Sigma^+)$, such that, for $h' = h_1 \circ \ldots \circ h_{n-1}$, the following holds:*

1. *If h is right-open, then h' is right-open, $h \subseteq h'$, and $upper^-(h') = upper^-(h)$;*
2. *If h is right-closed, then h' is right-closed, $h \subseteq h'$, and $upper^+(h') = upper^+(h)$.*

Lemma 4 *If a distance constraint $d(a,b) \in h$ is in $DS(\Sigma^+)$, then it is left-definable and right-definable.*

Lemma 4 can be proved by an induction on the number of operations (intersection or composition) applied, to obtain $d(a,b) \in h$ from $D(\Sigma^+)$.

In **Step 3**, following the same way as described in [4], we can construct a metric space satisfying all the constraints in $D(\Sigma^+)$. The main lemmas proved are stated below.

Lemma 5 *Let t be the number of constants in $D(\Sigma^+)$. Enforcing path-consistency on $D(\Sigma^+)$, a fixed point can be reached in $O(t^3)$.*

For any interval h occurring in $D(\Sigma^+)$, $h \subseteq [0,\infty)$. In the worst case, $[0,\infty)$ can be strengthened at most $4t$ times (first strengthen it to $[0,u]$, $u \leq 4\sigma(t-1)$, then strengthen it by σ each time). For t constants, there are $O(t^2)$ distance constraints in $D(\Sigma^+)$. Therefore, the total time of strengthening all the distance constraints is $O(t^3)$.

Lemma 6 *Let t be the number of constants in $D(\Sigma^+)$, $D^f(\Sigma^+)$ be a fixed point of enforcing path consistency on $D(\Sigma^+)$. If $D(\Sigma^+)$ is path-consistent, $D_s(\Sigma^+)$ is obtained from $D^f(\Sigma^+)$ by replacing every right-infinite interval with $\{5t\sigma\}$, every right-closed interval h with $\{upper(h)\}$, then $D_s(\Sigma^+)$is path-consistent.*

Any interval referred in $D^f(\Sigma^+)$ is either right-infinite or right-closed.

Lemma 7 *Let Σ^+ be an MCS. If $D(\Sigma^+)$ is path-consistent, then there is a metric space (Δ,d) such that all the constraints in $D(\Sigma^+)$ are satisfied by d.*

In **Step 4**, we will prove $D(\Sigma^+)$ is path-consistent by contradiction.

Lemma 8 *Let Σ^+ be an MCS. Then its set of distance constraints $D(\Sigma^+)$ is path-consistent.*

Proof.(sketch) Suppose $D(\Sigma^+)$ is not path-consistent. By Definitions 10 and 12, $d(p,q) \in \emptyset$ is in $DS(\Sigma^+)$, for some constants p,q. It is easy to show that for any distance range g occurring in $D(\Sigma^+)$, $g \neq \emptyset$. By Definitions 12, 9, and intersection rules, the last operation to obtain the first $\emptyset$ interval is intersection. By Definition 12, there exist $d(p,q) \in h$ and $d(p,q) \in g$ in $DS(\Sigma^+)$, $h \neq \emptyset$, $g \neq \emptyset$, and $h \cap g = \emptyset$. h,g are non-negative intervals. Without loss of generality, let us suppose $upper(h) \leq lower(g)$.

By Lemma 4, $d(p,q) \in h$ and $d(p,q) \in g$ are left-definable and right-definable. Since $d(p,q) \in h$ is right-definable, then by Definition 14, there exists an h' such that $upper^+(h) = upper^+(h')$ and $h \subseteq h'$. Since $d(p,q) \in g$ is left-definable, then by Definition 13, there exists an g' such that $lower^-(g) = lower^-(g')$ and $g \subseteq g'$. Then h' and g' are identity or definable intervals. By properties of identity or definable intervals, $lower(g') \leq 4\sigma$, thus, $upper(h') \leq 4\sigma$. By properties of intervals in $DS(\Sigma^+)$, h is right-closed; g is left-open, if $lower(g) \neq 0$. Then all the possible cases where $h \cap g = \emptyset$ are listed below:

- $upper(h) = 0$, $lower(g) \in \{\sigma, 2\sigma, 3\sigma, 4\sigma\}$ or $lower^-(g) = 0$;
- $upper(h) = \sigma$, $lower(g) \in \{\sigma, 2\sigma, 3\sigma, 4\sigma\}$;
- $upper(h) = 2\sigma$, $lower(g) \in \{2\sigma, 3\sigma, 4\sigma\}$;
- $upper(h) = 3\sigma$, $lower(g) \in \{3\sigma, 4\sigma\}$;
- $upper(h) = 4\sigma$, $lower(g) = 4\sigma$.

We can show that given an upper bound or a lower bound of a definable interval, there is a limited number of possibilities of it. For example, if $upper(h') = 2\sigma$, then $h' = [0, 2\sigma]$ or $h' = [0, \sigma] \circ [0, \sigma]$. Thus, there are finitely many possibilities for the corresponding sequences of $d(p,q) \in h$ and $d(p,q) \in g$. By Definitions 13 and 14, every distance constraint in the sequences is in $D(\Sigma^+)$. By Definitions 5 and 8, we can know which LBPT formulas in Σ^+ they come from. For example, if $d(p,q) \in [0, 2\sigma]$ is in $D(\Sigma^+)$ and $p \in points(a)$, $q \in points(b)$, then $NEAR(a,b) \in \Sigma^+$. In each case, we can show $\perp$ is derivable using axioms, which contradicts the assumption that Σ^+ is consistent. Therefore, $D(\Sigma^+)$ is path-consistent. QED.

Theorem 2 *If a finite set of formulas Σ is LBPT-consistent, there exists a metric model satisfying it.*

Proof. Given Σ, by Lemma 1, we can construct an *MCS* Σ^+ containing it. If Σ is LBPT-consistent, so is Σ^+, and hence by Lemma 8 and Lemma 7 there is a metric space (Δ, d) such that all constraints in $D(\Sigma^+)$ are satisfied by d. By Lemma 3, the metric space can be extended to a model M of $B(\Sigma^+)$, thus, of Σ^+ (Definition 5). By properties of maximal consistent sets, for every $\phi \in \Sigma^+$, $\phi \in \Sigma^+ \Leftrightarrow M \models \phi$. Hence, since $\Sigma \subseteq \Sigma^+$, M satisfies all formulas in Σ. QED.

5. DECIDABILITY AND COMPLEXITY OF LBPT

From the bound on the size of the satisfying model, we also have the following theorem:

Theorem 3 *The LBPT satisfiability problem is NP-complete.*

Proof.(sketch) NP-hardness of the LBPT satisfiability problem follows from NP-hardness of the satisfiability problem for propositional logic, which is included in LBPT.

To prove that the LBPT satisfiability problem is in NP, we show that given a finite satisfiable set of LBPT formulas Γ, we can guess a model for Γ and verify that this model satisfies Γ, both in time polynomial in the combined size of formulas occurring in Γ.

The completeness proof shows that, if Γ is consistent, it is satisfiable in a metric model M whose size is polynomially bounded by the number of constants in Γ, and distance function has a fixed finite range. We guess a model like this. To check whether it is a proper model, we need to check whether it is a metric space by Definition 1. This can be done in time which is polynomial in the size of M. To check whether M satisfies Γ, we need to check this for each formula in Γ. This can be done in time which is polynomial in the combined size of formulas in Γ and in the size of M. QED.

6. CONCLUSION

We presented a logic LBPT which formalizes the concepts of being 'possibly part of' (BPT), 'possibly connected' (NEAR) and 'definitely disconnected' (FAR). We provided a sound and complete axiomatistion of it with respect to metric models and showed that its satisfiability problem is NP-complete. An LBPT reasoner is under development and testing, for validating 'sameAs' and 'partOf' matches between spatial objects from authoritative and crowd-sourced geospatial datasets.

References

[1] M. Aiello, I. Pratt-Hartmann, and J. Benthem. *Handbook of Spatial Logics*. Springer, 2007.

[2] E. Clementini and P. D. Felice. Approximate Topological Relations. *International Journal of Approximate Reasoning*, 16(2):173–204, 1997.

[3] H. Du, N. Alechina, M.J. Jackson, and G. Hart. Matching Formal and Informal Geospatial Ontologies. In *Geographic Information Science at the Heart of Europe*, Lecture Notes in Geoinformation and Cartography, pages 155–171. Springer, 2013.

[4] H. Du, N. Alechina, K. Stock, and M.J. Jackson. The Logic of NEAR and FAR. In *Conference On Spatial Information Theory, COSIT 2013*, volume 8116 of *LNCS*, pages 475–494. Springer, 2013.

[5] ISO Technical Committe 211. ISO 19107:2003 Geographic information – Spatial schema. Technical report, International Organization for Standardization (TC 211), 2003.

[6] OpenStreetMap. The Free Wiki World Map. http://www.openstreetmap.org, 2012.

[7] Ordnance Survey. Ordnance Survey. http://www.ordnancesurvey.co.uk/oswebsite, 2012.

[8] Z. Pawlak, L. Polkowski, and A. Skowron. Rough Set Theory. In *Wiley Encyclopedia of Computer Science and Engineering*. 2008.

[9] A.J. Roy and J.G. Stell. Spatial Relations between Indeterminate Regions. *International Journal of Approximate Reasoning*, 27(3):205 – 234, 2001.

[10] S. Schockaert, M. D. Cock, C. Cornelis, and E. E. Kerre. Fuzzy region connection calculus: Representing vague topological information. *International Journal of Approximate Reasoning*, 48(1):314–331, 2008.

[11] T. Williamson. *Vagueness*. Routledge, 1996.

[12] S. Winter. Uncertain topological relations between imprecise regions. *International Journal of Geographical Information Science*, 14(5):411–430, 2000.

[13] F. Wolter and M. Zakharyaschev. Reasoning about Distances. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*, pages 1275–1282, 2003.