Maintaining Relational Consistency in a Graph-Based Place Database

Hao Chen
hchen@student.unimelb.edu.au

Maria Vasardani
mvasardani@unimelb.edu.au

Stephan Winter
winter@unimelb.edu.au

Department of Infrastructure Engineering,
The University of Melbourne, Australia

Abstract

People use natural language (NL) descriptions to communicate spatial information, mostly referring to space in qualitative terms. In this research, a graph database is used to store such qualitative spatial information as derived from NL descriptions. It focuses on developing and testing qualitative spatial reasoning mechanisms to maintain relational consistency within the graph database. The study provides a first step into using a graph database to storing and querying qualitative spatial data from NL place descriptions, and provides some insights for the system implementation.

1 Introduction

People represent environment in natural language (NL) descriptions typically in a qualitative way of spatial features and their relationships, such as "Building *A* is to the left side of Building *B*" or "The campus is a little bit north from the city". Such qualitative information could provide an intuitive approach for representing human spatial knowledge. Since this knowledge can be extracted in the form of triplets of locatum-relationship-relatum (as in "campus"-"north of"-"city") a formal model of two nodes and a linking edge comes to mind. Graph databases, which have already proven useful for modelling qualitative relationships in other areas, present a way to store and query these triplets.

This research focuses on building such a graph database that can store triplets while maintaining relational consistency. The research will go back to Qualitative Spatial Reasoning (QSR) for consistency checking. QSR has been used in other contexts, such as robot navigation, e.g. (Moratz, Nebel et al. 2003), general QSR calculi in AI, e.g. (Frank 1991, Freksa 1992, Zimmermann and Freksa 1996), or for systems modelling NL descriptions, e.g. (Belouaer, Brosset et al. 2013, Basiri, Amirian et al. 2014).

The hypothesis of this study is that a graph-based place database can preserve relational consistency in transactions about triplets representing locata, qualitative spatial relationships, and relata. It is anticipated that, as an outcome, a graph-based place database is able to identify and flag if new input is violating the current consistency and causing logical contradictions. For this purpose the paper studies the relationships and calculi in question, suggests reasoning constraints and algorithms for consistency checking, and tests these algorithms.

The rest of this paper is structured as follows: Section 2 discusses related work and tools. Section 3 explains the methodology for data modelling, as well as reasoning configurations to preserve relational consistency. Section 4 explains the experiments and provides a demonstration of the implemented system. The experiment results presented in Section 5 and discussed in Section 6. Section 7 includes conclusions and suggestions of future work.

In: B. Veenendaal and A. Kealy (Eds.): Research@Locate'15, Brisbane, Australia, 10-12 March 2015, published at http://ceur-ws.org

2 Related work

2.1 Graph database as a NoSQL database

The NoSQL database family (NoSQL means 'not only SQL') includes a variety of database management systems that are not restricted to relational SQL models. It is not in contrast to but rather complements relational databases (Basiri, Amirian et al. 2014). Graph databases are one of the major databases in the NoSQL family other than key-value, document and column databases (Tiwari 2011). Graph databases are based on graphs (Biggs, Lloyd et al. 1986) and employ nodes and edges with properties to store data and their relations.

Graph databases have two advantages in modelling rich-connected data, compared to relational databases (Güting 1994, Wiebrock, Wittenburg et al. 2000). First, they provide a natural and efficient way for network-based data modelling using nodes and edges, while relational databases are more suitable for modelling set-based data. Second, querying connected data can be cheaper in a graph database by traversing paths rather than by a join operation (or even recursive join) in a relational database.

Graph databases are today used in modelling information from social networks, such as Facebook, as well as general knowledge, for example the Google Knowledge Graph. In this study, a graph database is used to model place information since it provides a natural, suitable, efficient and flexible way for modelling spatial objects and the qualitative spatial relationships among them, with no need for pre-defining a tabular schema as in a relational database.

2.2 Modelling spatial information in a graph structure

The concept of triplets has been widely utilized (Wiebrock, Wittenburg et al. 2000, Sproat, Coyne et al. 2010, Kordjamshidi, Frasconi et al. 2012). One specific triplet, that of locatum-qualitative spatial relationship-relatum, has been introduced as the *spatial property graph* (Vasardani, Timpf et al. 2013). It consists of the spatial object to be located (the *locatum* L), the reference object (the *relatum* R) and the qualitative spatial relationship between them (r). So far, the spatial property graph has been used for constructing plausible sketch maps (Vasardani, Timpf et al. 2013). Belouaer et al. introduced another rule-based approach to generate a spatial semantic network inferred from verbal descriptions (Belouaer, Brosset et al. 2013). Both approaches focus on identifying semantic similarity among stored nodes. Our study, on the other hand, focuses on the relationships instead of spatial objects.

2.3 Qualitative Spatial Reasoning and Calculi

QSR is a constraint-based process that enables spatial intelligence to reasoning about space (Cohn and Hazarika 2001, Dylla, Frommberger et al. 2006). QSR algorithms can be used to solve constraint-based problems by reasoning on different types of qualitative spatial relations.

In this context, Frank suggested QSR with *cardinal directions* and *qualitative distance* relationships (Frank 1991, Frank 1992). For cardinal directions he suggested a cone model (Figure 1a), half-plane models (Figure 1b), and a neutral zone model (Figure 1c). The neutral zone model with its composition table is applied in this study, with additional rules to increase the flexibility of natural language.

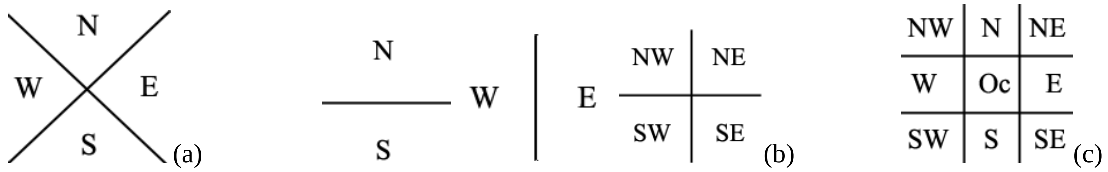

Figure 1: Cardinal direction systems

For qualitative distance relationships Frank discussed two-step (close, far), three-step (close, middle, far) and multi-step systems (Frank 1992). Those systems consider qualitative distance as mappings into intervals that form a partition of the positive real numbers, such as (for a three-step system) close = $[0, 1)$; middle = $[1, 3)$; far = $[3, \infty)$ (Figure 2a). In a linear space such relationships can be added, and dist(A,B) + dist(B,C) $\geq$ dist(A,C). The resulting interval will then be mapped back to the corresponding symbols (close, middle, far) again. In place databases, however, stored place configurations are from two-dimensional space, where nothing else than the usual triangle inequality holds. Figure 2b shows for example a situation where "A is middle to B" and "B is middle to C" means that A and C can be in any relationship, for example near.

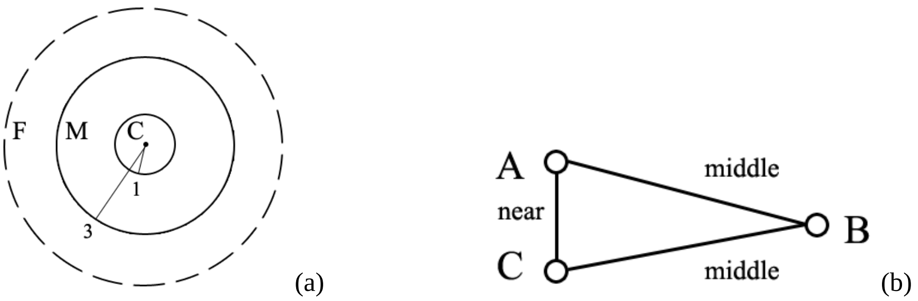

Figure 2: Deriving qualitative distance relationship on a plane

Freksa provided an approach for reasoning with *relative direction* relationships (Freksa 1992). Later, Zimmermann and Freksa discussed composition cases (Zimmermann and Freksa 1996). However, relative direction relations are not considered in this study, for reasons discussed in Section 3.

Egenhofer et al. discussed a methodology for modelling and reasoning with *topological* relations (Egenhofer and Herring 1990, Egenhofer and Franzosa 1991), known as the 4-intersection and 9-intersection models. The two models consider whether the interior ($A°$ and $B°$), the boundary (∂A and ∂B) and the exterior (A^c and B^c) of two spatial objects intersect, and map the different cases to topological relationships. A composition table was also introduced for reasoning (Egenhofer 1991). An alternative model, based on first order logic instead of point-set topology, is the Region Connection Calculus (RCC) introduced by Randell, Cui and Cohn in 1992 (Randell, Cui et al. 1992). The 4-intersection model and the RCC both distinguish the same eight relations between two simple connected spatial objects. In this study, Egenhofer's 4-intersection model and composition table are used.

The different calculi of QSR, each of them applicable on one type of spatial relationships, have been packed into toolboxes. For example, SparQ (Wallgrün, Frommberger et al. 2007) is an integrated system which provides a series of Qualitative Spatial Calculi functionalities and aims at supporting tasks related to constraint-based reasoning and qualitative consistency-checking. In this research, however, the flexibility that comes with the use of natural language requires a critical discussion and extension of such calculi.

3 A relationally consistent graph-based place database

Triplets of place information, as extracted from NL description, can be integrated and stored in a graph database. The graph database should only store relationships that are consistent with each other in order to avoid contradictions in querying or reasoning, even if the information is not necessarily true according to geographic reality. In this section, methods will be provided to maintain relational consistency.

3.1 Modelling place information using graph database

A triplet consisting of {locatum, relation, relatum}, such as {Building A, to the right of, Building B} is represented in a graph database using two distinct nodes for locatum and relatum, and a directed edge from the locatum to the relatum, as shown in Figure 3. In the following sections, L-r->R will be used to represent a triplet, such as "Building A–to the right of–>Building B".

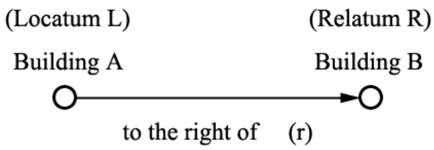

Figure 3: A stored triplet of two nodes and one edge

Each time a new triplet is to be added, the names of the locatum and relatum will be queried to verify whether already nodes of these names exist in the database. If the nodes already exist, the triplet is integrated in the existing graph, otherwise new nodes are created and connected with directed edges. The graph can become a multi-graph when more than one triplet are added with the same locatum and relatum (e.g., A–right of–>B; A–west of –>B).

3.2 Relational consistency

Relational inconsistency occurs in a graph database when two pieces of information can be derived from the stored data that are logically contradicting. For example, if a real world situation as shown in Figure 4 is described by "A-west of->B" and then by "B-west of->A" there is a contradiction. Contradictions can also occur over longer cycles, such as in "A-west of->B", "B-west of->C" and then "C-west of-A".

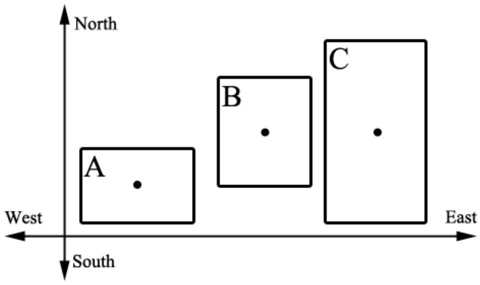

Figure 4: Three buildings in the cardinal reference frame

Definition: Two triplets are relationally-consistent if, without any additional knowledge, no contradicting information can be derived according to the provided reasoning rules.

Definition: A graph database is relationally-consistent if, at any state, it is free from relational inconsistency within the stored data.

3.3 Composition of relationships

A compositional inference is a deduction based on two relationships $\mathrm{R_1(A,B)}$ and $\mathrm{R_2(B,C)}$, which results in $\mathrm{R_3(A,C)}$ (Cohn and Hazarika 2001). Qualitative spatial calculi, with a limited set of expressions, may require $\mathrm{R_3(A,C)}$ to be a set of possible relations $\mathrm{S_3(A,C)}$. Let us denote $\circ$ for the composition operation between two relationships. Composition operation can (but not necessarily) follow the following rules:

Commutative law: $\mathrm{A \circ B = B \circ A}$

Associativity law: $\mathrm{(A \circ B) \circ C = A \circ (B \circ C)}$

Distributive law: $\mathrm{A} = \{\mathrm{A_1..A}_i\}, \mathrm{B} = \{\mathrm{B_1..B}_j\},\ \mathrm{A \circ B} = \sum_{m=1}^{i}\sum_{n=1}^{j} \mathrm{A}_m \circ \mathrm{B}_n$

A table can be constructed to represent the composition results $\mathrm{S_3(A,C)}$. For a calculus of *n* possible relationships, an *n*n* composition table consists of each composition result between any two relationships in *n*.

3.4 General reasoning configuration

For the time being, any new triplet L-r->R, r must be from one of the three categories: cardinal direction, qualitative distance, or topological relationship. All other relationships are not yet covered; they might be stored without consistency checking, or conservatively just rejected.

In order to identify relational inconsistency, any new input triplet L-r->R will be compared to existing knowledge between L and R derived from the stored data.

Definition: Given a triplet L-r->R, an *existing path* (EP) is a sequence of nodes and edges in the graph database that starts from L and ends with R, while all the relationships in the path belong to the category of r.

Definition: An *existing knowledge* (EK) is a set of derived relationships between L and R from an EP. If the length of EP equals to one, then EK is the single relation in the EP. If the length of EP is greater than one, then for all relationships with *1..n* in EP in sequence order, $\mathrm{EK} = r_1 \circ r_2 \circ .. r_j$, which could be either a single relationship or a set of relationships.

A new triplet L-r->R is regarded consistent with a database if it is relationally consistent with all EKs between L and R. If the triplet is consistent, it will be added to the graph database; otherwise an inconsistency will be flagged.

3.5 Reasoning configuration for cardinal direction relationships

Table 1 below shows the implemented cardinal direction relationship set (left), and possible verbalizations (right).

Table 1: Cardinal direction relationships

Cardinal Directions	Example of NL Descriptions
north of	north, to the north of, northern, north from, N, etc.
northeast of	northeast, to the northeast of, northeast from, NE, etc.
east of	east, to the east of, eastern, east from, E, etc.
southeast of	southeast, to the southeast of, southeast from, SE, etc.
south of	south, to the south of, southern, south from, S, etc.
southwest of	southwest, to the southwest of, southwest from, SW, etc.
west of	west, to the west of, western, west from, W, etc.
northwest of	northwest, to the northwest of, northwest from, NW, etc.

For cardinal directions this research adopts the neutral zone system (Figure 1c), as it is assumed that all spatial features have a two-dimensional extend. As for any other system, each cardinal direction can be associated with a region such that these regions form a JEPD partition of the Euclidian plane. The size of the neutral zone is defined by the relatum R, while the locatum L in a triplet L-r->R belongs to one of the eight zones (NW, N, NE, E, SE, S, SW and W).

Table 2 shows Frank's composition table for the neutral zone model. 'Any' in the table means again that any of the eight relationships is possible. However, the composition table is not directly applied in this work. This is because the semantics of cardinal directions in NL descriptions are more vague than a logical model suggests. For example, when people say North, they could mean anywhere within zones of W, NW, N, NE and E—in the neutral zone model—instead of only N, because if the half-planes of cardinal directions intersect in a valid relation, it seems that people consider them as conceptually consistent (i.e., non contradicting), as they can generalize from the more specific relations, to the more general and broader application region of each relation (half-planes). Therefore, without more detailed knowledge, a certain level of uncertainty should be accepted and not be regarded as inconsistency. For instance, in reasoning N and NW should be considered as consistent while N and S should be regarded as inconsistent.

Table 2: Frank's composition table for the neutral zone model

	N	NE	E	SE	S	SW	W	NW
N	N	N	NE	E	Any	W	NW	N
NE	N	NE	E	E	E	Any	N	N
E	NE	E	E	E	SE	S	Any	N
SE	E	E	SE	SE	S	S	S	Any
S	Any	E	SE	S	S	S	SW	W
SW	W	Any	S	S	S	SW	W	W
W	NW	N	Any	S	SW	W	W	W
NW	N	N	N	Any	W	W	N	NW

Definition: An *extended cardinal direction set* (ECD) is a set of consistent cardinal direction relationships given a cardinal direction relationship r in a NL description.

The ECD for any cardinal direction relationship is shown in Table 3. The table provides a set of relationships that are regarded as relationally consistent with r. For instance, "Building A-north of->Building B" and "Building A-east of->Building B" are considered as relationally consistent, and their half-plane intersection exists in the form of a valid relation north-east (NE); while "Building A-north of->Building B" and "Building A-south of->Building B" are relationally inconsistent, as obviously an intersection of their half-planes does not exist.

Table 3: Extended cardinal direction set for r

	ECD set
N	W, NW, N, NE, E
NE	N, NE, E
E	N, NE, E, SE, S
SE	E, SE, S
S	E, SE, S, SW, W
SW	S, SW, W
W	S, SW, W, NW, N
NW	W, NW, N

ECDs provide a 'tolerant' mechanism that first stores triplets which are considered to be relationally consistent. Then, when more triplets with new knowledge between the same two spatial objects are stored, and the number of EP (and EK) increases, the system becomes less tolerant, as a new input triplet will need to be consistent with every EK in order to be consistent with the graph database.

Rule 1: Given an input triplet L-r->R and r is a cardinal direction relationship, if for every EK between L and R, r belongs to the ECD of at least one of the elements in that EK, then the triplet is relationally consistent with the graph database.

3.6 Relative direction relationships

Relative direction relationships require a reference frame for relative directions. The reference frame can be linked to the orientation of the observer (egocentric, to *my* left), the orientation of the relatum (allocentric, to the *left* of the front side), or be a projection of the speaker to the reference frame of the recipient (*your* left). Unfortunately triplets do not provide this spatial reference frame. For instance, when someone says "A is in front of C" then this sentence's ambiguity with respect to the spatial reference frame cannot be resolved without further context (which is not accessible). This is the reason why in this research reasoning mechanisms for relative direction relationships are not provided.

3.7 Reasoning configuration for qualitative distance relationships

Table 4 below shows the implemented qualitative distance relationship set, followed by possible verbalizations.

Table 4: Qualitative distance relationships

Qualitative Distance Relation	Examples of NL descriptions
Near	beside, next to, close to, near, around, etc.
Middle	not far from, middle, etc.
Far	far from, furthest, distance from, very far, end, etc.

The only reasoning rule for the qualitative distance relationship between A and C is the triangle inequality. Frank's three-step intervals are applied: (N = [0, 1); M = [1, 3); F = [3, ∞)). Table 5 shows the corresponding composition table, using the distance computations of Algorithm 1.

Table 5: Composition table of qualitative distance intervals

	N = [0, 1)	M = [1, 3)	F = [3, ∞)
N = [0, 1)	[0, 2) N, M	[0, 4) any	[2, ∞) M, F
M = [1, 3)	[0, 4) any	[0, 6) any	[0, ∞) any
F = [3, ∞)	[2, ∞) M, F	[0, ∞) any	[0, ∞) any

Algorithm 1: Distances for qualitative distance interval composition

a = $[x_a, y_a)$, b = $[x_b, y_b)$ # a ∘ b = $[x_c, y_c)$
$y_c = y_a + y_b$
if $x_b \geq x_a$**:**
  $x_c = |x_b - y_a|$
else:
  $x_c = |x_a - y_b|$
return $[x_c, y_c)$

Rule 2: Given an input triplet L-r->R and r is a qualitative distance relationship, if r is element of EK between L and R, then the triplet is relationally consistent with the graph database.

3.8 Reasoning configuration for topological relationships

Table 6 shows the implemented topological relationship set (left), followed by possible verbalizations (right).

Table 6: Topological relationships

Topological relation	Example of NL descriptions
Disjoint	Building A is away from building B
Meet	Building A is adjacent to building B
Equal	A and B are actually the same building
Inside	The office is inside Building A
coveredBy	The pond is at the north end of the park
Contains	Building A contains the office
Covers	The park covers the pond at its north end
Overlap	Part of Building A is on the neighbour's parcel

Egenhofer' s composition table for the 4-intersection model (Egenhofer 1991) is shown in Table 7. Here, 'any' means no more specific information can be deducted: any topological relationship is possible. The commutative law does not apply for the composition operation between topological relationships. For instance, *disjoint* ∘ *inside* = {d, m, i, cB, o} while *inside* ∘ *disjoint* = d.

Table 7: Composition table of topological relationship

	disjoint	meet	equal	inside	covered By	contains	covers	overlap
disjoint	any	d,m,i, cB,o	d	d,m,i, cB,o	d,m,i, cB,o	d	d	d,m,i, cB,o
meet	d,m,ct, cv,o	d,m,e, cB,cv,o	m	i,cB,o	m,i,cB,o	d	d,m	d,m,i, cB,o
equal	d	m	e	i	cB	ct	cv	o
inside	d	d	i	i	i	any	d,m,i, cB,o	d,m,i, cB,o
covered By	d	d, m	cB	i	i,cB	d,m,ct, cv,o	d,m,e, cB,cv,o	d,m,i, cB,o
contains	d,m,ct, cv,o	ct,cv,o	ct	e,i,cB, ct,cv,o	ct,cv,o	ct	ct	ct,cv,o
covers	d,m,ct, cv,o	m,ct, cv,o	cv	i,cB,o	e,cB,cv,o	ct	ct, cv	ct,cv,o
overlap	d,m,ct, cv,o	d,m,ct, cv,o	o	i,cB,o	i,cB,o	d,m,ct, cv,o	d,m,ct, cv,o	any

Rule 3: Given an input triplet L-r->R and r is a topological relationship, if r is element of EK between L and R, then the triplet is relationally consistent with the graph database.

4 Implementation

The data used to test the algorithms are 731 triplets derived from 42 place descriptions collected by Vasardani et al. (Vasardani, Timpf et al. 2013). The triplets were extracted from NL descriptions of the University of Melbourne's Parkville campus, given by a group of graduate students with different levels of familiarity with the campus. These descriptions had already been pre-processed by NL parsing rules, synonym detection and relationship classification. Of these 731, 325 triplets have a cardinal direction, a qualitative distance, or a topological relationship.

4.1 System overview

The Neo4j community version is used in this research as local host visualization platform. The Neo4j Python-embedded is the main module used to interact with the local database host. The system structure is shown in Figure 5.

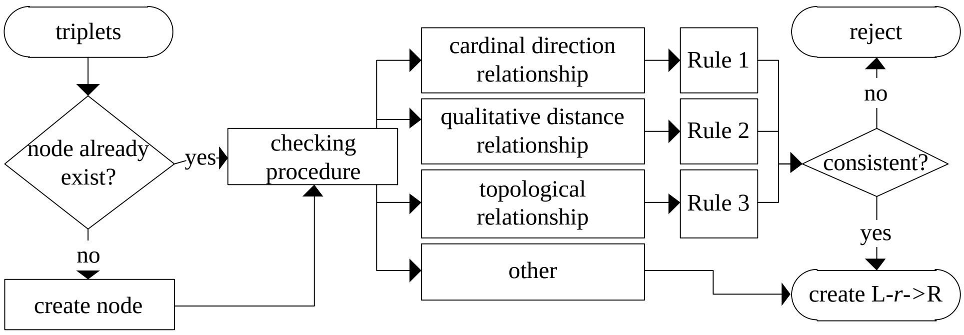

Figure 5: Diagram of reasoning process

4.2 Experiment design

The experiment consists of two parts. In Part 1, during the import stage, the reasoning algorithms were applied to identify inconsistent triplets form the set of 731. Each flagged triplet is then manually verified of whether it really causes relational inconsistency or not. Thus, rate P_1 calculated by Formula 4.2.1, is the precision of the algorithm. For instance, if the system detects 10 inconsistent triplets, and 9 of them are verified, then P_1 equals to 0.9 (maximum 1).

$$P_1 = \frac{\text{number of verified inconsistent triplets}}{\text{number of inconsistent triplets identified}} \tag{4.2.1}$$

In Part 2, after all triplets are imported, an additional set of 60 made-up consistent and inconsistent triplets of the three categories of relationships were input to the database. In this case it was known in advance which of the triplets are consistent or inconsistent. Again it was tested whether the system is able to correctly accept or reject each of them accordingly. The rate P_2 calculated by Formula 4.2.2 provides another insight on the precision of the algorithm.

$$P_2 = \frac{\text{number of correctly acceptedor rejected triplets}}{\text{number of all made−up triplets}} \tag{4.2.2}$$

In line with the hypothesis it is anticipated that both P_1 and P_2 equal to 1, in which case the graph database is proven to be able to maintain relational consistency.

4.3 Maximum query depth

When querying for EPs that start from node L and end with R, a maximum path length has been set, since computing all the paths between two nodes in a graph draws on computation time. This theoretical drawback can only be justified by pragmatics: First, it can be observed that after just a few steps of reasoning, composition tables in many places end with ‘any’ anyway. The second pragmatic argument is based on the First Law of Geography: Near things are more related than far things, which raises the expectation that relevant spatial information is provided in short cycles. The parameter for the maximum path length, Maximum Query Depth (MQD), has been tested in order to observe when the overall computation time becomes unacceptable. This test was done by adding all the 731 triplets into an empty graph database and recording the overall computation time of reasoning for each triplet, over a range of MQD.

5 Observed Results

For Part 1, all 731 triplets were added from the descriptions, and no inconsistent triplet occurred. All triplets were then verified, and P_1=1. For Part 2, 60 triplets were tested. 60 out of 60 were processed correctly—both all consistent and all inconsistent cases were identified—by the graph database, thus P_2=1.

Figure 6a shows the overall number of EP computed in Part 1 for every triplet from the set of 325 in the considered relation categories. Figures 6b, 6c, and 6d show the number of EP for triplets for each respective category (MQD = 3).

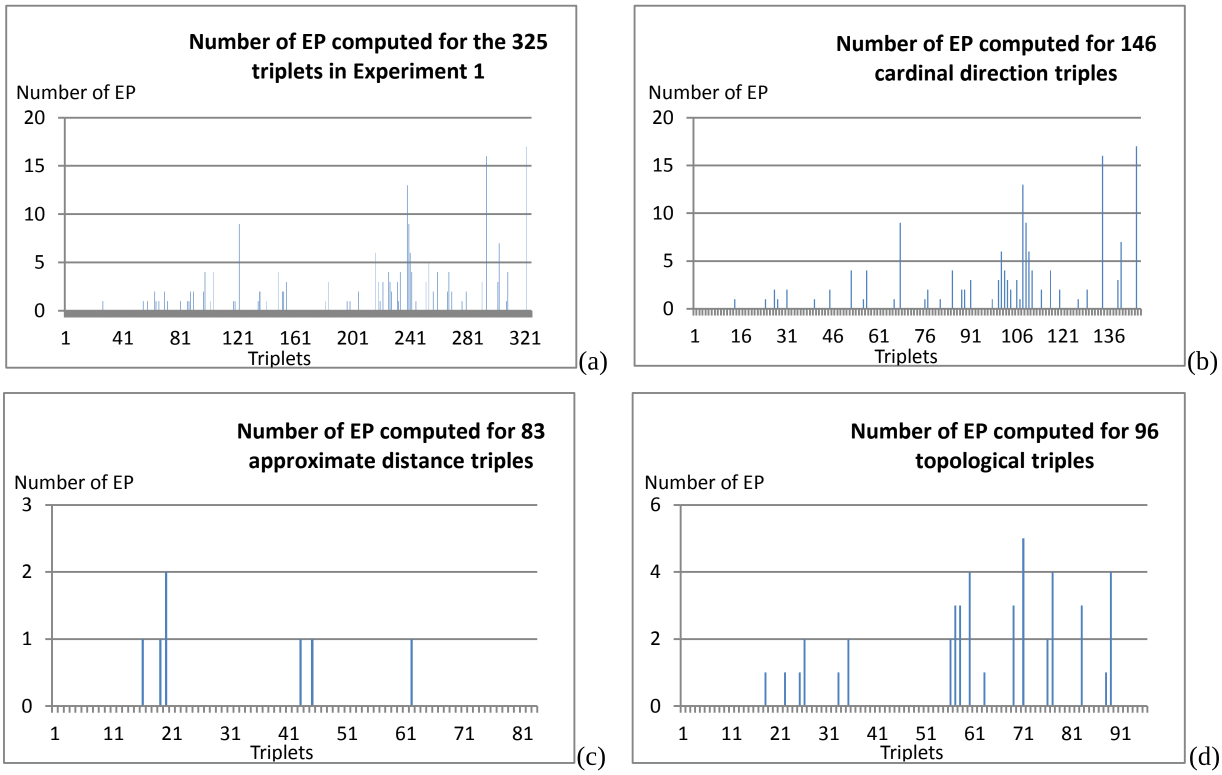

Figure 6: Number of EP computed in Part 1

Figure 7 shows that once a made-up consistent triplet was accepted and added, the following inconsistent triplet has at least one more EP queried than the previous one.

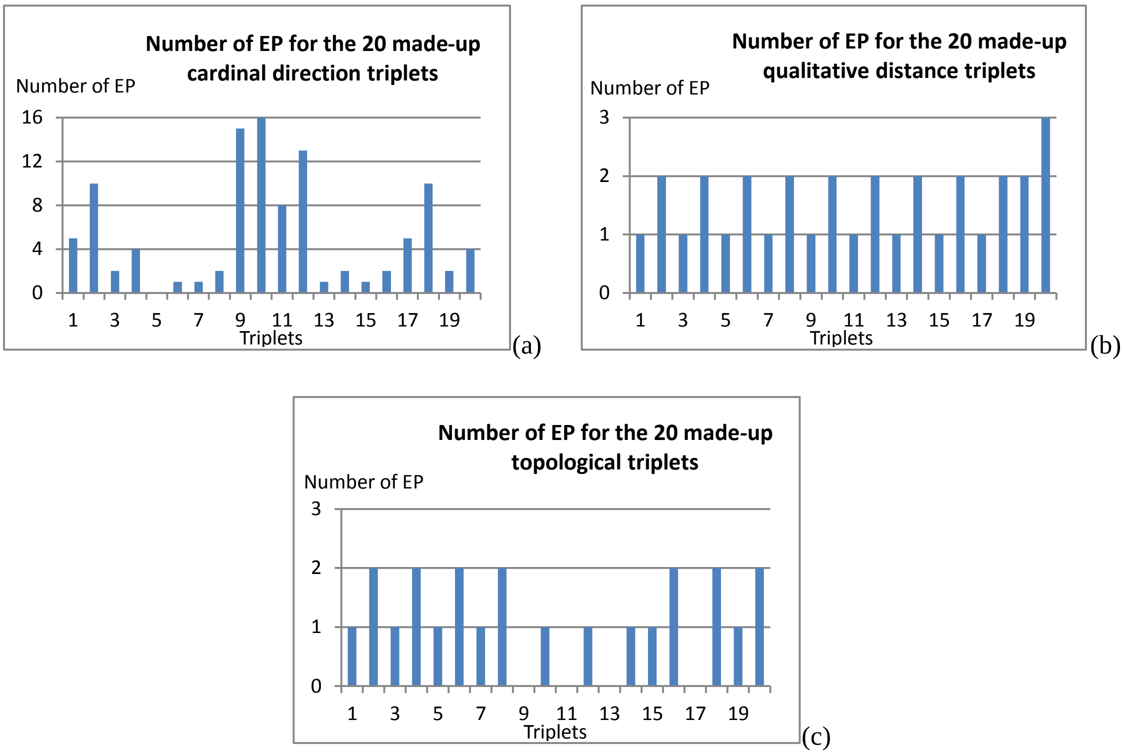

Figure 7: Number of EP computed in Part 2

Computation time for MQD that equals to 1, 2, 3 and 4 respectively is shown in Figure 8. When MQD is set to 1, 2 and 3, the computation time is rather stable (around 0.1s). But when it is set to 4, the computation time suddenly begins to soar as the size of the graph database grows.

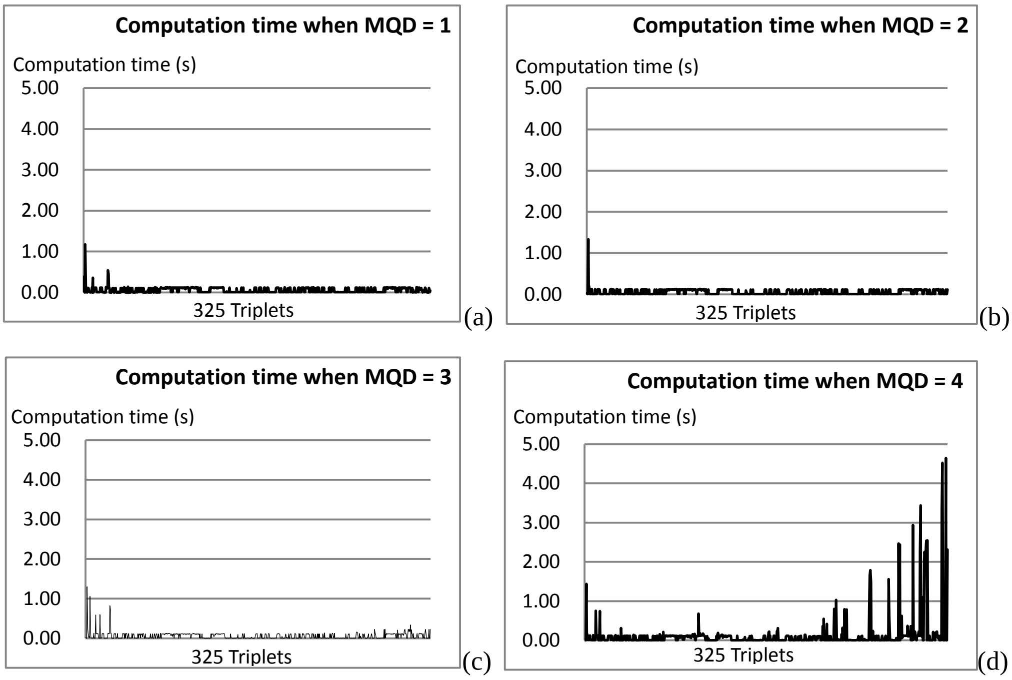

Figure 8: Computation time for MQD equals to 1, 2, 3 and 4

6 Discussion

The implementation of relational consistency rules is the first step of automatically reason with qualitative spatial relationships from NL descriptions in a graph database. The implemented mechanisms are flexible and robust to preserve relational consistency in a graph database.

In Part 1 of the experiment, the 731 triplets extracted from the 42 place descriptions are relationally consistent. This result is not surprising, since the descriptions are from students that are already, to various extents, familiar with the environment of the University of Melbourne's Parkville campus, therefore the possibility of giving misleading descriptions is low (but still possible, such as by mistake, and which makes the consistency checking mechanisms necessary). Alternatively, also malicious users could try to add inconsistent knowledge and should be prevented. However, the system does only flag an inconsistency; it has no mechanisms to decide whether the already stored or the added relations are the truer representation of geographic reality.

According to the result given by Part 2, the system is capable of distinguishing consistent and inconsistent triplets. Additionally, as shown in Figure 6, the numbers of EP found for qualitative distance relationships are overall small. This indicates that there are not as many neighbouring relationships that belong to the qualitative distance relationship category, therefore it is not as rich-connected as cardinal direction and topological relationships.

The system is flexible and accommodates certain degrees of ambiguity and uncertainty. The graph database is robust for maintaining relational consistency with given rules and formalized triplets, even though the reasoning algorithms are flexible.

Among the input spatial objects, 'Campus' is a special one. It is rather a container of most other spatial objects in these (campus) place descriptions. Among all 731 triplets, 144 of them (approximately 19.7%) related to 'Campus'.

Although relative direction relationships were not reasoned, it was observed that some relative direction relationships appear 'inconsistent' in expression when considered without their individual spatial reference frame. For instance, in one triplet, the Engineering Department is "right of" the South Lawn, while in another triplet, the relationship between them is "left of". Triplets of relative directions occupy approximately 25% of the total number of triplets; hence they must be addressed in future work by ways of capturing and maintaining their spatial reference frame.

Additionally, each relationship is only checked for consistency within its own category. Relational consistency checking among different categories is not considered in this research; therefore the system is not yet able to reason relationships from multiple categories. For example, "A-inside->B" and "A-far->B" are contradicting since A and B must be disjoint in order to be far from each other. Due to the fact that 'inside' and 'far' do not belong to the same category, they will not be regarded as relationally-inconsistent by the implemented database. Thus, reasoning rules among different categories of relationship should be developed in future research.

It was anticipated that by setting different MQD, the number of EP queried for any triplet L-r->R would be different. Also, it is expected that the reasoning results would be different since cycles longer than MQD can still be inconsistent with a triplet to be added. When implemented, no difference in reasoning result was detected by setting different MQD; however, the difference of number of EP detected is significant.

Although more EP could be queried by setting larger MQD (and consequently more EK can be used for consistency checking), it is suggested to set MQD to 3 in order to limit the computation time. As shown in Figure 8, it can be predicted that with a larger dataset, the computation time would soon become unacceptable with larger MQD.

7 Conclusions and future work

This research proves that a graph database can preserve relational consistency when modelling qualitative place information derived from NL. The hypothesis is verified by the outcomes of the experiments described in earlier section. The implemented system is overall flexible enough to accommodate NL descriptions and robust for maintaining relational consistency.

Data triplets in this research include spatial relationships from categories such as cardinal directions, qualitative distances, topological relations, relative directions and other (non-categorized) relationships. Reasoning rules were developed for the first three categories. When a new input triplet violates a consistency rule, it will be flagged by the graph database to prevent logical contradiction and relational inconsistency. Currently, the system can only process reasoning based on relationships that belong to the same category. It is not yet able to reason on triplets with relationships from multiple categories.

The path length of reasoning cycles has to be limited for computational reasons. Different MQD were tested for composition reasoning, and an MQD of 3 still delivered acceptable computation times.

This research is the first step in attempting to build a relationally consistent database for place information from NL. Such a graph database, built and maintained in relational consistency, can then be used to solve relevant decision making tasks, such as landmark navigation; or be used for other applications, such as plausible sketch map drawing.

The major limitations of this research are the limited size and context of the test dataset (all campus descriptions of the same campus). More interesting observations are expected if the graph database is used for testing datasets referring to different locations. Second, the robustness of mapping NL descriptions of qualitative distance relationships into intervals needs further research. It is known that qualitative distance judgements change with the context in a conversation, do not follow the commutative law (from “A is near B” does not necessarily follow that, within the same context, “B is near A”), and negation within the three-step model is not addressed. For instance, “not far” could mean either ‘near’ or ‘middle’ in a description, however it is mapped to ‘middle’ in this research. Third, as discussed, the system is not yet able to reason over relationships from different categories, and also it is not able to reason with relative direction relationships.

Apart from the limitations discussed above, other work can also be done in future studies. First, semantic reasoning could be introduced to automatically identify synonyms of relationship labels, thus helping a graph database further to maintain relational consistency over a larger variety of relationship names. Secondly, due to the ambiguity of natural language descriptions, relationships cannot be categorized automatically currently. For instance, ‘in’ and ‘on’ in NL descriptions could stand for not only topological ‘inside’ (i.e. room in the building; clock-tower on the lawn), but also other relations (i.e. station on the road). Mechanisms could be developed by considering the type of the referenced objects as well as the context, in order to categorize relationships automatically.

References

Basiri, A., P. Amirian and A. Winstanley (2014). Use of Graph Databases in Tourist Navigation Application. Computational Science and Its Applications–ICCSA 2014, Springer 8583: 663-677.

Belouaer, L., D. Brosset and C. Claramunt (2013). Modeling spatial knowledge from verbal descriptions. Spatial Information Theory, Springer 8116: 338-357.

Biggs, N., E. K. Lloyd and R. J. Wilson (1986). Graph Theory, 1736-1936, Clarendon Press.

Cohn, A. G. and S. M. Hazarika (2001). "Qualitative spatial representation and reasoning: An overview". Fundamenta informaticae 46(1-2): 1-29.

Dylla, F., L. Frommberger, J. O. Wallgrün and D. Wolter (2006). SparQ: A toolbox for qualitative spatial representation and reasoning. Qualitative Constraint Calculi: Application and Integration, Workshop at KI.

Egenhofer, M. J. (1991). Reasoning about binary topological relations. Advances in Spatial Databases, Springer 525: 141-160

Egenhofer, M. J. and R. D. Franzosa (1991). "Point-set topological spatial relations". International Journal of Geographical Information System 5(2): 161-174.

Egenhofer, M. J. and J. Herring (1991). Categorizing binary topological relations between regions, lines, and points in geographic databases. 9: 94-91.

Frank, A. U. (1991). Qualitative spatial reasoning with cardinal directions. Österreichische Artificial-Intelligence-Tagung/Seventh Austrian Conference on Artificial Intelligence, Springer 287: 157-167.

Frank, A. U. (1992). "Qualitative spatial reasoning about distances and directions in geographic space." Journal of Visual Languages & Computing 3(4): 343-371.

Freksa, C. (1992). Using orientation information for qualitative spatial reasoning, Spatio-Temporal Reasoning, Springer 639: 162-178.

Güting, R. H. (1994). GraphDB: Modeling and Querying Graphs in Databases. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), Morgan Kaufmann: 297-308.

Kordjamshidi, P., P. Frasconi, M. Van Otterlo, M.-F. Moens and L. De Raedt (2012). Relational learning for spatial relation extraction from natural language. Inductive Logic Programming, Springer 7207: 204-220.

Mackaness, W., P. Bartie and C. S.-R. Espeso (2014). Understanding Information Requirements in “Text Only” Pedestrian Wayfinding Systems. Geographic Information Science, Springer 8728: 235-252.

May, A. J., T. Ross and S. H. Bayer (2005). "Incorporating landmarks in driver navigation system design: An overview of results from the REGIONAL project." Journal of Navigation 58(1): 47-65.

Moratz, R., B. Nebel and C. Freksa (2003). Qualitative spatial reasoning about relative position. Spatial cognition III, Springer 2685: 385-400.

Randell, D. A., Cui, Z. and Cohn, A. G. (1992). A Spatial Logic based on Regions and Connection. 3rd Int. Conf. on Knowledge Representation and Reasoning, Morgan Kaufmann: 165-176.

Sproat, R., R. E. Coyne and J. B. Hirschberg (2010). Spatial Relations in Text-to-Scene Conversion, Computational Models of Spatial Language Interpretation Workshop at Spatial Cognition 2010.

Tiwari, S. (2011). Professional NoSQL, John Wiley & Sons.

Vasardani, M., S. Timpf, S. Winter and M. Tomko (2013). From descriptions to depictions: A conceptual framework. Spatial information theory, Springer 8116: 299-319.

Wallgrün, J. O., L. Frommberger, D. Wolter, F. Dylla and C. Freksa (2007). Qualitative spatial representation and reasoning in the SparQ-toolbox. Spatial Cognition V Reasoning, Action, Interaction, Springer 4387: 39-58.

Wiebrock, S., L. Wittenburg, U. Schmid and F. Wysotzki (2000). Inference and visualization of spatial relations. Spatial Cognition II, Springer 1849: 212-224.

Zimmermann, K. and C. Freksa (1996). "Qualitative spatial reasoning using orientation, distance, and path knowledge." Applied Intelligence 6(1): 49-58.