The University of Maine Digital Commons @UMaine

Electronic Theses and Dissertations

Fogler Library

12-2009

Qualitative Spatial Reasoning with Holed Regions

Maria Vasardani

Follow this and additional works at: http://digitalcommons.library.umaine.edu/etd
Part of the Geographic Information Sciences Commons

Recommended Citation

Vasardani, Maria, "Qualitative Spatial Reasoning with Holed Regions" (2009). *Electronic Theses and Dissertations*. 553. http://digitalcommons.library.umaine.edu/etd/553

This Open-Access Dissertation is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DigitalCommons@UMaine.

QUALITATIVE SPATIAL REASONING WITH HOLED REGIONS

By

Maria Vasardani

B.S. University of Athens, 2000

M.S. University College London, 2003

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Spatial Information Science and Engineering)

The Graduate School

The University of Maine

December 2009

Advisory Committee:

Max J. Egenhofer, Professor in Spatial Information Science and Engineering, Advisor

M. Kate Beard-Tisdale, Professor in Spatial Information Science and Engineering

Anthony G. Cohn, Professor of Automated Reasoning, University of Leeds

Robert Franzosa, Professor in Mathematics and Statistics

Michael F. Worboys, Professor in Spatial Information Science and Engineering

THESIS

ACCEPTANCE STATEMENT

On behalf of the Graduate Committee for Maria Vasardani, I affirm that this manuscript is the final and accepted thesis. Signatures of all committee members are on file with the Graduate School at the University of Maine, 42 Stodder Hall, Orono Maine.

Max J. Egenhofer, Professor in Spatial Information Science and Engineering

© 2009 Maria Vasardani

All Rights Reserved

LIBRARY RIGHTS STATEMENT

In presenting this thesis in partial fulfillment of the requirements for an advanced degree

at the University of Maine, I agree that the Library shall make it freely available for

inspection. I further agree that permission for 'fair use' copying of this thesis for

scholarly purposes may be granted by the Librarian. It is understood that any copying of

publication of this thesis for financial gain shall not be allowed without my written

permission.

Signature:

Maria Vasardani

Date: 12/10/2009

QUALITATIVE SPATIAL REASONING WITH HOLED REGIONS

By Maria Vasardani

Thesis Advisor: Dr. Max J. Egenhofer

An Abstract of the Thesis Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (in Spatial Information Science and Engineering) December, 2009

The intricacies of real-world and constructed spatial entities call for versatile spatial data types to model complex spatial objects, often characterized by the presence of holes. To date, however, relations of simple, hole-free regions have been the prevailing approaches for spatial qualitative reasoning. Even though such relations may be applied to holed regions, they do not take into consideration the consequences of the existence of the holes, limiting the ability to query and compare more complex spatial configurations. To overcome such limitations, this thesis develops a formal framework for spatial reasoning with topological relations over two-dimensional holed regions, called the *Holed Regions* Model (HRM), and a similarity evaluation method for comparing relations featuring a multi-holed region, called the *Frequency Distribution Method* (FDM).

The HRM comprises a set of 23 relations between a hole-free and a single-holed region, a set of 152 relations between two single-holed regions, as well as the composition inferences enabled from both sets of relations. The inference results reveal that the fine-grained topological relations over holed regions provide more refined composition results in over 50% of the cases when compared with the results of hole-free regions relations. The HRM also accommodates the relations between a hole-free region and a multi-holed region. Each such relation is called a *multi-element* relation, as it can be deconstructed into a number of elements—relations between a hole-free and a single-holed region—that is equal to the number of holes, regarding each hole as if it were the only one.

FDM facilitates the similarity assessment among multi-element relations. The similarity is evaluated by comparing the frequency summaries of the single-holed region relations. The multi-holed regions of the relations under comparison may differ in the number of holes. In order to assess the similarity of such relations, one multi-holed region is considered as the result of dropping from or adding holes to the other region. Therefore, the effect that two concurrent changes have on the similarity of the relations is evaluated. The first is the change in the topological relation between the regions, and the second is the change in a region's topology brought upon by elimination or addition of holes. The results from the similarity evaluations examined in this thesis show that the topological placement of the holes in relation to the hole-free region influences relation similarity as much as the relation between the hole-free region and the host of the holes. When the relations under comparison have fewer characteristics in common, the placement of the holes is the determining factor for the similarity rankings among relations.

The distilled and more correct composition and similarity evaluation results enabled by the relations over holed regions indicate that spatial reasoning over such regions differs from the prevailing reasoning over hole-free regions. Insights from such results are expected to contribute to the design of future geographic information systems that more adequately process complex spatial phenomena, and are better equipped for advanced database query answering.

ACKNOWLEDGMENTS

Writing this thesis would have not been possible without the endless support and love of my family—my father Aris, my mother Matina and my sister Matoula. They respected and went along with my decision to study far away from home and were there to welcome and take care of me during my visits. I want to thank them for supporting this, as well as all the important decisions in my life.

I would like to express my deepest gratitude to my advisor, Dr. Max Egenhofer for supporting and believing in this work. He was always an inspirational, encouraging and excellent judge of my progress. Our meetings gave me the motivation to continue even through difficult times and our academic discussions always fired up my research interests. His guidance and experience made sure that I was always on track, without suppressing my scientific curiosity. I am also grateful to the other members of my committee: Mike Worboys, for his acute remarks and the interesting discussions over coffee from time to time, Kate Beard for her perceptive comments and her restless willingness to help, Robert Franzosa for giving me a different outlook on things and clearing up confusions I may had, Anthony Cohn for being very patient and cooperative with our long-distance communication and for providing with helpful comments and suggestions. I would also like to express my gratitude to Eliseo Clementini, the external examiner from the University L'Aquila, Italy, for his thorough review of the final thesis draft and his prompt response.

I wish to thank my former advisors Paul Densham and Douglas Fleweling for strongly supporting my goal to obtain a PhD and suggested UMaine as the place to achieve it. From the University of Athens, I would like to thank Dr. Nikolaos Skarpelis for advising and encouraging me through the early years of my studies.

On a more personal tone, I would like to thank all my friends from back home who supported my decision to study abroad and all the new friends in UMaine that made sure I had a good time here. Special thanks to Kostas T. and Georgina, for always being at the other end of the long-distance call line, day or night, for laughs or not, making the geographical distance seem smaller. Kostas N. and I became close friends since the first day I moved to Maine and for that I am grateful. Thank you to everyone else I met here, especially Marcus, Katie, Arda, Harris and many thanks to my roommate and very good friend Aleksandra, for the many funny Greek moments.

This work was partially supported by the National Geospatial-Intelligence Agency under grant number NMA201-01-1-2003 (Principal Investigator: Max Egenhofer), and by the University on Maine through a Provost Fellowship of the University of Maine, a Graduate Research (Chase) Award, a Maine Economic Improvement Fund Fellowship, and a College of Engineering Graduate Research Assistant Award.

TABLE OF CONTENTS

ACKNO	WLEDGMENTS	iv
LIST OF	F TABLES	xiii
LIST OF	F FIGURES	xiv
Chapter	1 INTRODUCTION	1
1.1	Qualitative Spatial Representation and Reasoning about Holed Regions	3
1.1.	1 Topological Spatial Relations	4
1.1.	2 Composition Inferences	5
1.1.	.3 Consistency of Topological Relations	6
1.1.	4 Relation Similarity	7
1.2	Motivation	8
1.3	Need for Explicit Accounting for Holes	10
1.4	Scope of the Thesis	12
1.5	Hypothesis	14
1.6	Research Approach	15
1.7	Major Results	17
1.8	Intended Audience	18
1.9	Thesis Organization	18
Chapter	2 SPATIAL OBJECTS WITH HOLES	22
2.1	Theory of Holes in Three-Dimensional Objects	23
2.2	Deciphering Holes from Figures in the Plane	25

2.3 Topological Relations Between Two-Dimensional Regions In the Plane	26
2.3.1 Models of Topological Relations for Hole-Free Regions	27
2.3.1.1 Basic Topological Concepts	27
2.3.1.2 9-Intersection	29
2.3.1.3 Region Connection Calculus	31
2.3.2 Models for Representing Complex Regions	33
2.3.3 Models of Topological Relations for Complex Objects	34
2.3.3.1 Building on Relations between Hole-Free Regions	36
2.3.3.2 Models Based on A Smaller Group of Relations	37
2.3.3.3 Enhanced 9-Intersection	38
2.4 Look-A-Likes of Regions with Holes	42
2.5 Change in the Structure of Dynamic Areal Phenomena	49
2.6 Change in Topological Relations between Regions	51
2.6.1 Conceptual Neighborhood Graphs for Binary Topological Relations	51
2.6.2 Similarity of Binary Relations	53
2.7 Summary	54
Chapter 3 RELATIONS BETWEEN A HOLE-FREE AND A SINGLE-HOLED	
REGION	56
3.1 Qualitative Model of a Single-Holed Region	57
3.2 Deriving the Topological Relations between a Hole-Free and a	
Single-Holed Region	60
3.3 Twenty-Three Relations between a Hole-Free and a Single-Holed Region	61

3.4 Pro	operties of the Twenty-Three Relations	62
3.4.1	Converse Relations.	64
3.4.2	Implied Relations	65
3.5 Co	nceptual Neighborhood Graph of the 23-t _{RRh}	66
3.5.1	Construction of the Graph	67
3.5.2	Distances between the 23- <i>t_{RRh}</i> on the CNG	69
3.6 Co	mpositions over a Single-Holed Region	70
3.7 An	alysis of Compositions	76
3.7.1	Cardinality and Crispness	77
3.7.2	Unambiguous Composition Results	80
3.8 Su	mmary	81
Chapter 4	RELATIONS BETWEEN TWO SINGLE-HOLED REGIONS	83
4.1 To	pological Relations between Two Single-Holed Regions	84
4.2 Pro	operties of the Single-Holed Region Relations	90
4.3 Co	nceptual Neighborhood Graph of the 152-t _{RhRh}	93
4.4 Co	mpositions Involving Single-Holed Region Relations	94
4.5 An	alysis of Compositions	97
4.5.1	Absolute Frequencies of Composition Cardinalities.	98
4.5.2	Cumulative Frequencies of Composition Cardinalities	99
4.5.	2.1 Average Increase of Cumulative Cardinality Frequencies	101
4.5.	2.2 Saturation and Accumulation Pace of Cumulative Cardinality	
	Frequencies	103
4.6 Su	mmary	104

Chapter 5	COMPARING RELATIONS WITH A MULTI-HOLED REGION	106
5.1 A l	Multi-Holed Region and the Relation between a Hole-Free and a	
Mu	ılti-Holed Region	108
5.2 Fre	equency Distribution Method	109
5.2.1	Topological Dissimilarity and Relation Difference	110
5.2.2	Computing the Minimum Cost of Relation Transformation with the	
	Transportation Algorithm	112
5.2.	2.1 Balanced Transportation Problem	113
5.2.	2.2 Similarity Values Obtained from the Transportation Algorithm	114
5.3 Exa	ample of Relation Similarity Evaluation	115
5.4 Pro	operties of the Frequency Distribution Method	118
5.4.1	Effect of the Principal Relation on the Similarity Evaluation	118
5.4.2	Ramifications of the Weight Constant	120
5.4.3	Analysis of Random Query Experiments	123
5.5 Su	mmary	124
Chapter 6	SIMILARITY ASSESSMENT WITH VARYING NUMBERS	
	OF HOLES.	126
6.1 Lir	mitations of the FDM	129
6.1.1	Out-of-Balance System	129
6.1.2	Rebalancing the System	131

6.2 Br	idging Different Topological Changes	132
6.2.1	Changes in the Relation between Regions	132
6.2.2	Changes in the Topological Structure of a Region Between Relations	135
6.2.3	The Hybrid Conceptual Neighborhood Graph	136
6.2.4	Cost of the Change in the Topological Structure of the Region	137
6.2	.4.1 Cost of Hole Modification: Cost Model A	138
6.2	.4.2 The Weighted Hybrid CNG and the Table of Costs from CMA	139
6.3 Al	ternative Cost Assigning Methods	141
6.3.1	Balanced Hybrid CNG: Cost Model B	142
6.3.2	Grouped Hybrid CNG: Cost Model C	147
6.3.3	Modified Grouped Hybrid CNG – Longest Paths: Cost Model D	150
6.4 Ad	lapted FDM for Relations of Regions with Different Numbers of Holes	152
6.4.1	Frequency Units that Rebalance the System	152
6.4.2	Example of Similarity Evaluation with the Modified FDM	155
6.5 Su	mmary	159
Chapter 7	EVALUATION OF SIMILARITY RESULTS	161
7.1 Ex	perimental Setup	162
7.1.1	Evaluation Criteria	162
7.1.2	Datasets and Queries	164
7 1 3	Evaluation Procedure	165

7.2 Analysis of Similarity Rankings	166
7.2.1 First Criterion: Influence of the Number of Holes	167
7.2.2 Second Criterion: Influence of the Principal Relation	169
7.2.3 Third Criterion: Influence of the Refining Relations	170
7.2.4 Additional Observations On the Similarity Rankings	172
7.3 Assessment of the Similarity Evaluation Method for Relations with	
Different Numbers of Holes	173
7.4 Summary	179
Chapter 8 CONCLUSIONS	180
8.1 Summary of the Thesis	181
8.2 Contributions and Major Findings	183
8.2.1 Contributions	183
8.2.2 Major Findings	186
8.3 Future Work	188
8.3.1 Alternatives to the S-HRM	188
8.3.1.1 Relaxing the Definition of the Single-Holed Region	189
8.3.1.2 Relaxing the Scaling Transformation Requirements	191
8.3.2 Relations of Holed Regions on the Sphere	193
8.3.3 Extending the FDM	195
8.3.3.1 M: N-Holed Regions Relations	195
8.3.3.2 Holes with Different Roles	197
8.3.3.3 Hole Generalization	198
8.3.4 Introducing Human Language Expressions	199

REFERENCES	201
BIOGRAPHY OF THE AUTHOR	211

LIST OF TABLES

Table 2.1 Topological interpretation of the eight base relations of RCC-8	32
Table 3.1 Each of the 23 - t_{RRh} 's neighboring relations according to each of the	
constituent relations.	68
Table 3.2 Distances between relations for the 23- <i>t_{RRh}</i>	69
Table 4.1 Neighbor cardinalities for the 152- <i>t</i> _{RhRh} CNG.	93
Table 6.1 Distances between relations of different type on the hybrid CNG,	
assigned using CMA.	141
Table 6.2 Schematic depiction of the complete table of costs for Cost Model A	141
Table 6.3 Distances between relations of different type on the hybrid CNG,	
assigned using CMB.	146
Table 6.4 Schematic depiction of the complete table of costs for Cost Model B	146
Table 6.5 Distances between relations of different type on the hybrid CNG,	
assigned using CMC.	150
Table 6.6 Schematic depiction of the complete table of costs for Cost Model C	150
Table 6.7 Distances between relations of different type on the hybrid CNG,	
assigned using CMD.	151
Table 6.8 Schematic depiction of the complete table of costs for Cost Model D	152
Table 7.1 Refining relations for queries and multi-refinement dataset relations	165

LIST OF FIGURES

Figure 1.1	Examples of regions with holes.	2
Figure 1.2	Coverage holes in networks	3
Figure 1.3	Example of regions with a hole due to coverage holes in a buoy	
	network in the Gulf of Maine	9
Figure 1.4	Example sketched geological application with multi-holed oil-	
	baring region D and hole-free excavating regions A, B, and C	10
Figure 1.5	The eight possible configurations if region A overlaps region B and	
	B overlaps region C	11
Figure 1.6	The three possible configurations if B overlaps with A such that B's	
	hole is inside A, and B also overlaps with C such that B's hole	
	meets C.	12
Figure 2.1	The hole in surface y encloses object x and its hole, however, x is	
	not part of the hole.	24
Figure 2.2	Various hole fillers.	25
Figure 2.3	Holes vs. figures	26
Figure 2.4	The eight binary topological relations as defined in the 9-	
	intersection and their matrices.	30
Figure 2.5	A complex region with five faces.	34
Figure 2.6	Three topologically distinct grouped together by RCC	35
Figure 2.7	Topological configurations realized by the Schneider-Behr model	40
Figure 2.8	A vague region with kernels (k), boundary (b) and a hole (blank	
	area).	43
Figure 2.9	Example of a vague region modeled by the egg-yolk representation	44

Figure 2.10	The 46 possible relations between two egg-yolk pairs	44
Figure 2.11	Two topologically distinct scenarios of a relation between two	
	regions, each with a hole, that are grouped together under the egg-	
	yolk theory	45
Figure 2.12	A vague region with a single-component core.	46
Figure 2.13	Different broad boundary regions.	47
Figure 2.14	A region with broad boundary.	48
Figure 2.15	An areal object	51
Figure 2.16	Conceptual neighborhood graph of the 9-intersection for regions	53
Figure 2.17	Conceptual neighborhood graph for RCC-8	53
Figure 3.1	A single-holed region <i>B</i> .	58
Figure 3.2	Single-holed B's five topologically distinct and mutually exclusive	
	parts.	58
Figure 3.3	Topological relation between a hole-free A and a single-holed	
	region B.	59
Figure 3.4	A three-region spatial scene that captures the constituent relations	
	of a binary topological relation between a region A and a single-	
	holed region B.	61
Figure 3.5	Graphical depictions and specifications of the 23 topological	
	relations between a hole-free and a single-holed region.	62
Figure 3.6	Topologically different configurations for the [overlap meet] t_{RRh}	63
Figure 3.7	Constraints imposed by a specified $t(A, B^*)$ on $t(A, B_H)$	65
Figure 3.8	Constraints imposed by a specified $t(A, B_H)$ on $t(A, B^*)$	65
Figure 3.9	Conceptual Neighborhood Graph for the 23-t _{RRh}	68

Figure 3.10	The 16 combinations of compositions of binary relations with hole-	
	free regions (R) and single-holed regions (Rh).	71
Figure 3.11	The spatial scene over four regions used for the derivation of the	
	composition $t(A, B)$; $t(B, C)$	72
Figure 3.12	Iconic representation of compositions and relations on the 8 - t_{RRh}	
	CNG	72
Figure 3.13	Correspondence between the iconic representations of t_{RRh} s on the	
	23- <i>t_{RRh}</i> CNG and on the 8- <i>t_{RR}</i> CNG	73
Figure 3.14	Composition table t_{RRh} ; t_{RhR} .	74
Figure 3.15	Comparison of the frequencies of compositions results	78
Figure 3.16	Crispness improvements (in absolute counts) for t_{RRh} ; t_{RhR} vs.	
	t_{RR} ; t_{RR}	79
Figure 3.17	Composition of relations inside and contains over a single-holed	
	region B using different sets of relations.	79
Figure 4.1	Representation of the topological relation between two single-holed	
	regions, A and B.	84
Figure 4.2	Hierarchy of the constituent relations (κ) for a t_{RRh} and a t_{RhRh}	85
Figure 4.3	Example configurations of the 152 t_{RhRh}	86
Figure 4.4	The specifications of the 152 topological relations between two	
	holed regions, captured by their constituent relations	89
Figure 4.5	Counts (#) of t_{RhRh} s by their principal relation $\pi(t_{RhRh})$, inter-hole	
	relation $\omega(t_{RhRh})$, minor relation $\psi(t_{RhRh})$, and reverse-minor relation	
	$\psi^{-1}(t_{RhRh})$	90
Figure 4.6	The Conceptual Neighborhood Graph for the 152- <i>t</i> _{RhRh} .	94

Figure 4.7	The spatial scene for deriving the composition $t_{RhR}(A, B)$; $t_{RRh}(B, C)$	96
Figure 4.8	The spatial scene for deriving the composition	
	$t_{RhRh}(A, B)$; $t_{RhRh}(B, C)$	96
Figure 4.9	Normalized cardinality frequencies for compositions that result in	
	t_{RR} and t_{RhRh} .	98
Figure 4.10	Cumulative normalized frequencies over normalized composition	
	cardinalities for t_{RR} ; t_{RR} and t_{RhR} ; t_{RRh}	100
Figure 4.11	Cumulative normalized frequencies over normalized composition	
	cardinalities for compositions resulting in t_{RR} and t_{RhRh}	100
Figure 4.12	Approximations for cumulative frequency graphs	102
Figure 5.1	Topological relations between a multi-holed region B and a region	
	A	106
Figure 5.2	A multi-holed region n-B.	108
Figure 5.3	Depiction of the different $t(A, Bi)$ that together make up the four-	
	element relation $t(A, 4-B)$ between hole-free region A and multi-	
	holed region 4-B.	109
Figure 5.4	Example topological multi-element relations	116
Figure 5.5	The five indistinguishable multi-element relations	119
Figure 5.6	Paths connecting different elements	120
Figure 5.7	Moving from relation [cB cv] to any of its four neighboring	
	relations causes the same amount of topological change.	121
Figure 5.8	Similarity evaluation of multi-element relations.	121
Figure 5.9	The ranking of multi-element relations according to their similarity	
	with the query relation.	122

Figure 5.10	Randomly generated archived topological scenarios.	123
Figure 6.1	Relations between a hole-free and a multi-holed region with	
	different numbers of holes.	126
Figure 6.2	Sketched geological scenarios of multi-holed oil baring reservoirs	128
Figure 6.3	Example multi-element relations with different numbers of holes	129
Figure 6.4	Example of transforming a multi-element relation into another,	
	when the number of holes between them is different	132
Figure 6.5	The Conceptual Neighborhood Graphs for $8-t_{RR}$ and $23-t_{RRh}$ with	
	edge labels for possible topological changes from one relation to the	
	other	134
Figure 6.6	The alignment of the $8-t_{RR}$ with the $23-t_{RRh}$ CNG	136
Figure 6.7	Different hole elimination procedures	139
Figure 6.8	Hybrid CNG with costs of relation transformations assigned to the	
	edges for CMA	140
Figure 6.9	Different relation transformation paths	143
Figure 6.10	The reference edge $E([d\ d]-disjoint)$ for Cost Model B, assigned	
	the unit cost after hole disappearance.	144
Figure 6.11	Schematic example of calculating the cost of an intra-level link	145
Figure 6.12	Hybrid CNG with costs of relation transformations assigned to the	
	edges from Cost Model B.	147
Figure 6.13	The grouped hybrid CNG.	148
Figure 6.14	Grouped hybrid CNG reasoning.	149
Figure 6.15	Hybrid CNG with costs of relation transformations assigned from	
	Cost Model C	149

Figure 6.16	Multi-element relation comparison.	155
Figure 6.17	Ranking results against reference relation $t(A, 5-B)$	159
Figure 7.1	Instances depiction of queries QO, QI_2,QI_5 and datasets D_I and D_{II}	165
Figure 7.2	The left (L), right (R), and bottom (B) continua of the principal	
	relations on the 8- <i>t_{RR}</i>	167
Figure 7.3	Similarity ranking of dataset D _{II} against QI ₅ , with cost model C	168
Figure 7.4	Similarity rankings of dataset D_{I} against QI_{5} , with cost model $D.$	169
Figure 7.5	Similarity rankings of dataset D_I , against QI_2 , with cost model $A.\dots$	170
Figure 7.6	Similarity values after the first evaluation round of cost model B	171
Figure 7.7	Similarity values after the first evaluation round of cost model C	171
Figure 7.8	Similarity rankings of dataset D_{II} against QI_2 , with cost model B	172
Figure 7.9	Similarity rankings of dataset D_I against QO, with cost model D	176
Figure 7.10	Hybrid CNG highlighting the overlap, coveredBy, and inside	
	relations that participate in dataset D _I .	176
Figure 7.11	Highlighted elements.	177
Figure 8.1	Single-holed region B in different relations with the hole B_H	189
Figure 8.2	The five new t_{RRh} relations when the hole is $coveredBy$ the	
	generalized region.	190
Figure 8.3	The 23- t_{RRh} CNG for which B_H coveredBy B^* .	190
Figure 8.4	The 23- <i>t_{RRh}</i> CNG featuring the C-neighbors.	192
Figure 8.5	An ambiguous depiction of the hole in a region that sits on the	
	sphere	194

Figure 8.6	The 2-holed region to 2-holed region relation between A and B as	
	the union of two multi-element relations, each of which may be	
	broken down to two elements	196
Figure 8.7	The two t_{RhRh} s that form the 2-holed to 2-holed region relation	
	between A and B.	196
Figure 8.8	Hole generalization	199

Chapter 1 Introduction

In an effort to model and reason about geographic reality, Geographic Information Systems (GISs) rely on geometric representations of natural and constructed geographic entities and the spatial relations between them. Geographic entities have different spatial dimensions and may be represented by point, linear, areal features, or combinations of these. Among the relations between spatial entities that are essential to GISs are the *topological* relations, which are derived by such topological information as connectivity, adjacency, continuity, and containment of the geometric representations.

Frequently, more complex spatial objects, such as two-dimensional regions with holes, are necessary for modeling the plethora of geographic phenomena that extend over a specific area and are characterized by cavities or other internal discontinuities. Examples of holes in spatial entities include the space left out after imposing a buffer zone around an entity (Fig. 1.1a), an island in a lake (Fig. 1.1b), or Lesotho, which is completely surrounded by South Africa (Fig. 1.1c). While geometric models of spatial data have matured enough to capture complex spatial objects such as regions with holes (Frank and Kuhn 1986; OGC 1999, 2005; Worboys and Bofakos 1993, Clementini *et al.* 1995, Schneider and Behr 2006), there are no relation models that offer explicit support for querying and reasoning about topological relations between holed regions.

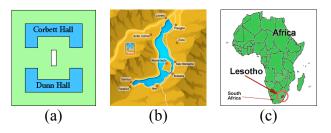


Figure 1.1 Examples of regions with holes: (a) the smoke free zone around two buildings on the UMaine campus implied by a 30ft buffer zone, (b) Lago D'Iseo with island Monte Isola, and (c) South Africa surrounding Lesotho.

The recent proliferation of geosensor networks (Stefanidis and Nittel 2004) for monitoring spatially distributed phenomena further challenges the current spatial querying and reasoning methods of information systems. Sensor networks often capture information about continuous properties of a phenomenon, such as air temperature or water contamination, and that information is displayed in the form of regions of a certain quality within the area covered by the network. In that setting, holes occur frequently, for example, in the form of *coverage holes* (Fig. 1.2a-b)—areas that cannot be reached by a certain number of sensors (Ahmed *et al.* 2005)—due to technical failures, resulting also in information displayed in the form of regions with holes.

Most of the underlying collected data are in the form of quantitative measures (e.g., water contamination is recorded when the quantity of the sensed chemical elements reaches or exceeds a certain threshold). However, most of the reasoning is qualitative in nature, since it is often the areal objects formed by the collected data (e.g., regions of high water contamination) and the objects' properties and relations that are of importance for further reasoning and analysis. Representing continuous properties by discrete systems of symbols and providing calculi for reasoning with spatial entities is the essence of qualitative reasoning (Cohn and Hazarika 2001).

Topology offers a useful toolkit for abstracting and generating qualitative descriptions of spatial data. In the case of sensor data, for example, models that address explicitly the topological relations between regions with holes due to missing data provide the opportunity for enhancing qualitative reasoning. In general, topological information is an important component of any geographic database and GIS, permitting users to derive spatial relationships between entities. This thesis provides a framework for qualitative spatial reasoning using topological relations involving holed regions and investigates the results of inference mechanisms over such relations such as composition and similarity assessment.

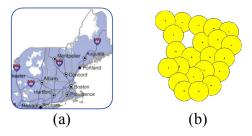


Figure 1.2 Coverage holes in networks: (a) A cellular telephony provider's coverage area (darker) with holes (white parts) in New England and neighbors, and (b) a coverage hole in a sensor network (disks represent the coverage range around each sensor).

1.1 Qualitative Spatial Representation and Reasoning about Holed Regions

The two main approaches to representing spatial information are quantitative and qualitative in nature. In the quantitative approach, it is common to have a coordinate system in which all spatial information is maintained. The exact position and extent of all entities is required and reasoning is usually performed with geometrical or numerical methods such as computing the distances between objects. In the alternative approach, spatial information is represented using a finite vocabulary—a set of symbols—that describes the spatial configurations by specifying the properties of and the relations between spatial entities (Hernández 1994). These sets of symbols have to be relevant to

the kind of reasoning performed and the resulting group of qualitative values is termed a *qualitative quantity space* (Hayes 1979; Forbus 1984), in which indistinguishable values have been grouped into equivalence classes (Cohn and Hazarika 2001).

People conceptualize geographic space differently from manipulable, table-top spaces (Zubin 1989; Mark 1993; Montello 1993; Mark and Freundshuh 1995). When interacting with the world, people's cognitive mechanisms handle complex situations by storing and processing qualitative information, even if it is originally obtained in quantitative form through perception (Freksa and Röhrig 1993; Kuipers 1994).

Qualitative spatial representations (Hernández 1994; Cohn 1997) appear to resemble closer how humans represent and communicate spatial knowledge, namely by specifying the relationships of and between spatial entities. Qualitative spatial reasoning is challenged to provide calculi and appropriate reasoning techniques that allow GISs to represent and reason with spatial entities, make predictions and help in decision making about physical systems without employing traditional quantitative techniques, so prevalent in computer graphics and computer vision (Cohn and Renz 2007).

1.1.1 Topological Spatial Relations

It is common in qualitative spatial reasoning to consider a specific aspect of space, such as topology, orientation, distance, shape, or size, with the exception of a few cases of combined information (Worboys 1996; Gerevini and Renz 2002; Li 2007), and to develop a system of qualitative relations between spatial entities (Hernández 1994). The focus here is on topology. Topological distinctions are inherently qualitative and topology offers a rich theory of space by categorizing it into different kinds, called topological spaces, according to different properties (Renz 2002). Topological relations

are invariant under continuous transformations of the underlying space and are related to the notions of connectivity, adjacency, inclusion and continuity.

In this thesis we consider the topological relations that involve two-dimensional holed regions embedded in \mathbb{R}^2 with the usual topology. To determine such relations we build on the set of eight binary topological relations {disjoint, meet, overlap, coveredBy, inside, covers, contains, equal} between hole-free regions, as defined by the 9-intersection (Egenhofer and Herring 1990). In order to reason about the relations involving holed regions we employ the inference mechanisms of relation composition and similarity and we assess how the presence of holes in the regions influences the inference results.

1.1.2 Composition Inferences

One of the most important inference mechanisms in qualitative spatial reasoning is the *composition* of relations. The composition over a common region B is the deduction of the topological relation t_k between regions A and C from the knowledge of the two topological relations, $t_i(A, B)$ and $t_j(B, C)$, and it is denoted by t_i ; t_j . The composition inference is actually a set of relations. If the outcome is the *empty* set, then a contradiction is implied, while the *universal* relation—the union of all possible relations—indicates that no information can be obtained by the inference. For the rest of the cases, the smaller the inference set, the more precise, refined, and less ambiguous is the composition result, with the singleton being optimum, implying a completely determined inference. The composition of pairs of n relations can be stored in an $n \times n$ composition table for reference, a technique that is useful given a fixed set of relations (Cohn and Hazarika 2001). Based on the composition results available in the composition

table for the eight binary relations between hole-free regions (Egenhofer 1994), this thesis determines the composition tables for sets of relations involving one or two single-holed regions.

1.1.3 Consistency of Topological Relations

When determining a new set of binary relations, it is essential to verify their *topological consistency*, that is, to ensure that the relations experience no internal contradictions due to their properties and can be realized in a particular space (Egenhofer and Sharma 1993). Topological consistency problems may be expressed as *constraint satisfaction problems* (CSPs) (Ladkin and Maddux 1994) over a network of binary topological relations (Egenhofer and Sharma 1993): each relation is interpreted as a constraint restricting possible values of its arguments. Topological relations can be formulated as constraints operating on a *relation algebra* (Tarski 1941; Ladkin and Maddux 1994).

A relation algebra on a set A comprises a set of binary relations \mathcal{R} , and each binary relation is a subset of the Cartesian product A × A. The set of relations is closed under the operations of *union*, *intersection*, *composition*, *conversion*, and *complement* (Renz 2002). The *jointly exhaustive and pairwise disjoint* (JEPD) relations defined in the 9-*intersection* (Egenhofer and Herring 1990) form a relation algebra with *equal* being the *identity relation* and the disjunction of all relations in the set being the *universal relation* (Egenhofer and Sharma 1993). To solve CSPs, a special purpose inference procedure, which can be shown to be complete for checking consistency of a set of relations is necessary (Bennett 1998). For binary topological relations this inference procedure is the composition of relations, which tests the consistency of any triple of relations. The

compositions of the base 9-intersection relations are computed using the semantics of the relations and their composition table has already been defined in (Egenhofer 1994).

Relying on the fact that the 9-intersection relations are closed under intersection, converse, and composition, constraint propagation can be applied in order to eliminate from the domain those values that are not consistent with the constraints (Cohn and Renz 2007). A well-known constraint propagation method for spatial CSPs is the *path-consistency algorithm* (Mackworth 1977). Path consistency is enforced by making sure that for any consistent instantiation of any two variables x and y, there exists an instantiation of any third variable z, such that the three variables together are consistent for the relations r(x, y), s(x, z) and t(z, y) (Eqn. 1.1).

$$\forall x, y, z := r(x, y) \cap (s(x, z); t(z, y))$$
 (1.1)

The path-consistency algorithm refines the resulting relations until either a fixed point is reached or one constraint is refined to the empty relation. If the propagation reaches the empty relation then the constraint set is inconsistent, otherwise the resulting set is *path-consistent*. By verifying path consistency of relations for holed regions that are combinations of the 9-intersection relations, it is possible to determine which of these relations and their compositions are topologically consistent.

1.1.4 Relation Similarity

Frequently, spatial entities with a certain area have more than one internal discontinuity, thus, their geometric representations are multi-holed regions. The most fine-grained model for topological relations between regions with holes explicitly enumerates the realizable 9-intersection relations for pairs of regions, considering each hole to be a region as well (Egenhofer *et al.* 1994). The resulting relations are dependent on the

number of the holes, making this approach inefficient and painstaking as the number of holes increases. It is often the case that cognitively, the details of a topological scenario get absorbed and give way to the more general, dominant topological features. When querying, for instance, a database of different archived scenarios that feature relations involving multi-holed regions, the user is more likely to pay less attention to the exact number and individual relations of the holes. Focus is rather on the quest for closely related scenarios that could either substitute the query relation with minimum consequences or provide useful information due to their similarity with the query. It is desirable, therefore, to look for a way of comparing different relations that feature multi-holed regions rather than continuing the enumeration of plausible relations.

1.2 Motivation

A model of topological relations that addresses holed regions explicitly provides additional opportunities for spatial querying and analysis. Figure 1.3 shows a sensor network of buoys deployed in the Gulf of Maine to record such variables as wind speed, wave height, air and water temperatures, and atmospheric pressure (http://www.gomoos.com). The two snapshots in Figures 1.3a and 1.3b refer to the available data on two consecutive days, where the highlighted regions depict two zones with data unavailability, each around a malfunctioning buoy, yielding holes in the coverage regions. With the movement of the buoys on the ocean surface, these zones of data availability shift, and so do the holes of missing data. The comparison of both regions over the 2-day period reveals that for the first buoy, both the region and its hole have moved on the second day (Figure 1.3c), while the second buoy's region has changed only slightly in size but its hole has moved (Figure 1.3d). In order to perform such

analyses at a more abstract level, so that it could be achieved conveniently through a spatial query language, one needs an account of the possible relations between such regions with holes. Beyond the enumeration of all possible cases, however, the inferences that can be drawn from such relations reveal new insights. For instance, what could be derived if one had further information available that on day 3 the regions have moved such that they are now disjoint from their positions on day 2?

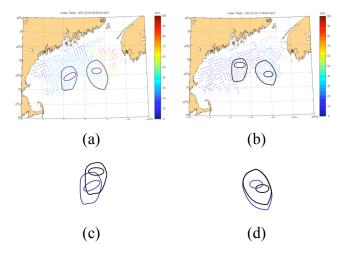


Figure 1.3 Example of regions with a hole due to coverage holes in a buoy network in the Gulf of Maine: (a) two regions, A (left) and B (right), with missing data on 02/04/2005, (b) two regions, A' (left) and B' (right), with missing data on 02/05/2005, (c) overlay regions A and A', and (d) overlay of regions B and B'.

Figure 1.4 shows a sketched topological scenario that requires comparisons among different multi-holed region relations. Region D represents an underground sedimentary rock formation with concentrations of oil residing in the holes H1 through H8. Regions A, B and C are ground, hole-free regions that have been deemed appropriate for excavation in order to reach the oil deposits. Given that all holes are of equal profit importance, how different is the topological relation between regions A and D from that between B and D and the relation between regions C and D? If all three relations are

topologically similar, choosing any excavation region is comparably beneficial, but if they are very dissimilar, further oil-drilling profit estimations are needed.

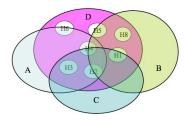


Figure 1.4 Example sketched geological application with multi-holed oil-baring region D and hole-free excavating regions A, B and C.

To answer such and similar questions, this thesis develops a reasoning framework that applies to two-dimensional regions with holes. This framework comprises sets of topological relations, which enable new information inferences using the relation composition mechanism and a similarity evaluation model that quantifies the similarity of topological relations of holed regions against a query relation.

1.3 Need for Explicit Accounting for Holes

Qualitative spatial reasoning has mostly been performed either using models that explicitly exclude regions with holes, or that implicitly consider hole-free and holed regions in the same manner. The 9-intersection (Egenhofer and Herring 1990), based on point set topology, and the Region Connection Calculus (RCC) (Randell et al. 1992b), which employs spatial logic, are two fundamental approaches to defining topological relations between two-dimensional spatial entities. The 9-intersection—as defined in (Egenhofer and Herring 1990)—specifically excludes regions with holes by applying only to two-dimensional point sets that are topologically equivalent to a closed disc, while RCC treats regions with holes the same as hole-free, homogeneous regions.

The following example, however, highlights the need for an explicit distinction of relations over hole-free regions from relations over hole-free regions. If a simple region B overlaps with simple region A and also overlaps with simple region C, then for the composition A overlaps B and B overlaps C, one can deduce from the composition table in (Egenhofer 1994) the set of all the possible relations between A and C. In this case, the composition yields the universal relation U₈ (i.e., all the eight binary topological relations are possible between A and C) (Fig. 1.5a-h).

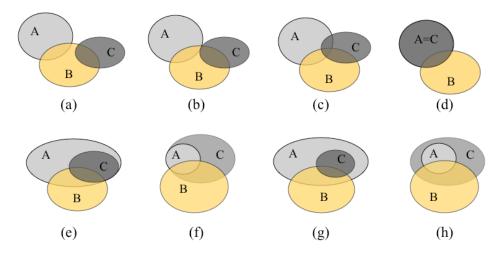


Figure 1.5 The eight possible configurations if region A *overlaps* region B and B *overlaps* region C: (a) A *disjoint* C, (b) A *meet* C, (c) A *overlaps* C, (d) A *equal* C, (e) A *covers* C, (f) A *coveredBy* C, (g) A *contains* C, and (h) A *inside* C.

On the other hand, if we take into consideration that the common region B has a hole and it *overlaps* with the two hole-free regions, A and C, then the existence of the hole changes the composition scenario. For instance, let A *overlap* with B and completely *contain* B's hole, and let B *overlap* with C such that B's hole *meets* C. This composition results in three possible relations between A and C: they could *overlap* (Fig. 1.6a), A could *cover* C (Fig. 1.6b), or A could *contain* C (Fig. 1.6c).

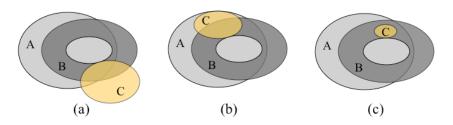


Figure 1.6 The three possible configurations if B *overlaps* with A such that B's hole is *inside* A, and B also overlaps with C such that B's hole *meets* C: (a) A *overlaps* C, (b) A *covers* C, and (c) A *contains* C.

The insertion of a hole into the common region of a composition scenario changes the results, in this case, into a more refined set (i.e., three vs. eight possible relations). Treating both cases with the same model for hole-free regions would neglect the constraints that the hole imposes, leading to incorrect and more ambiguous results for the end user who has to make the choices.

1.4 Scope of the Thesis

This thesis focuses on spatial relations between two-dimensional regions with holes, which are the geometric representations of natural or constructed holed entities found in geographic space. Such entities may be lakes with islands, temporary buffer zones with uncovered central areas or coverage holes created by the absence of sensor nodes in a network. The holed regions are embedded in the Euclidean plane \mathbb{R}^2 only, since they are represented by continuous point-sets; thus regions in the discrete plane \mathbb{Z}^2 are not included. The relations we consider are between regions in the same plane, therefore, regions cannot be embedded in \mathbb{R}^3 . In cases where examples from the three-dimensional physical world are used, it is implied that we consider the projection of the holed regions on the plane, a two-dimensional 'slice' of the objects. Finally, regions embedded on the surface of the sphere (\mathbb{S}^2) are also excluded, because even though the surface of the sphere resembles the Euclidean plane, the set of basic relations between hole-free regions

on the sphere is different from that on the plane—eleven vs. eight basic relations (Egenhofer 2005). Since the sets of relations derived in this thesis are based on the eight binary relations as these are defined for the plane, they do not cover all possibilities for two-dimensional regions on the sphere.

The types of relations that we are concerned with here are topological, as topology provides appropriate calculi for qualitative spatial reasoning. Relations of the *metric* type, expressed in terms of distances and directions (Peuquet and Ci-Xiang 1987; Hernández 1991; Frank 1992, 1995; Goyal 2000) or *partial* and *total order relations* (Kainz *et al.* 1993) such as *in front of* or *above* (Freeman 1975; Chang *et al.* 1989; Hernández 1991) are not objects of concern in this work. The goal for the thesis is to identify where and how holes matter for reasoning with topological relations and not necessarily to obtain shortcuts for the fast implementation of such relations. To achieve a higher-level reasoning, the sets of relations derived in this work explicitly for holed regions are based on the region-region relations, rather than the intersections or connections of the regions for supplementing relations that apply to a wider range of two-dimensional abstractions of spatial entities.

The topological relations considered are only coarse relations based on the empty or non-empty intersections of boundaries, interiors and exteriors of regions. Details about the dimensions of these three components or the dimension of their intersections (Egenhofer and Franzosa 1995) are not considered. The relations are also restricted between objects of co-dimension zero (the difference between the dimension of the embedding space and the dimension of the object). As a result, relations between regions and lines, or regions with points are not examined. Finally, the regions may not comprise different components due to a separated interior (i.e., multi-component regions).

The higher-level models of relations developed here apply to abstract geometric representations of regions and do not carry with them any particular semantic or ontological restrictions for either the host regions or the holes themselves. The domain of the regions and what they stand for depends on the specific application for which the representations are constructed and is not of focus in this work. A region with holes is a two-dimensional entity with separated exterior and boundary. The holes are disjoint from each other and they do not touch the region's boundary. If the holes are filled with some material, they may resemble distinguished parts of the same region. However, the set of relations for regions with distinguished parts would be different than the sets of relations derived for holed regions in this thesis. This is because different relations may hold between parts of the same entity than the relation between holes. Therefore, relations for regions with distinguished parts are not considered in this thesis. The sets of relations for holed regions could only apply for regions with not only distinguished, but also disjoint parts.

1.5 Hypothesis

GISs have so far used models of topological relations that have either explicitly ignored two-dimensional regions with holes, or have treated them in the same manner as simple, homogeneous regions. The hypothesis of this thesis is: "Taking into consideration the holes in two-dimensional regions that represent real-world spatial entities with cavities, imposes new constraints on the topological relations that can hold between such regions, differentiating these relations from the relations that hold between hole-free regions, and affects the inferences of the reasoning process towards more refined results." The hypothesis is verified by (1) building a model of topological relations that comprises

different sets of relations when mixed scenarios of hole-free and single-holed regions are involved, and (2) studying the composition inferences of such relations, which allows for the derivation of new quantitative measures that verify the hypothesis. Furthermore, the similarity assessment among relations featuring a multi-holed region against a query relation, show that taking into consideration the holes produces more exact similarity rankings than the rankings produced by ignoring the holes and only comparing the relations that the host region is involved in.

1.6 Research Approach

In order to build a new formalism for reasoning with two-dimensional single-holed regions, new sets of relations are derived. Their basis is the 9-intersection (Egenhofer and Herring 1990), in which the values empty or non-empty of the nine intersections among interiors, boundaries, and exteriors specify the topological relations between two hole-free regions. A *spatial scene*, which captures the binary relations between host-regions, between holes, and between host-regions and holes, serves as the conceptual model for deriving the new sets (Egenhofer and Vasardani 2007; Vasardani and Egenhofer 2008). A customized consistency checker ensures that each new relation is node-, arc- and path-consistent (Mackworth 1977).

Composition tables provide benchmarks for the inferences that are made possible using the new sets of relations. In order to make inference comparisons, we complement the qualitative analysis with new quantitative measures such as the composition *crispness* (i.e., the difference in the cardinality of the inference result when a hole is present in one of the regions) and the *cumulative frequencies* (i.e., the consecutive summing of the frequencies in the composition results) that provide numerical evidence for hypothesis'

evaluation. Much of the analysis is based on a set of graphs created with the quantitative data from the numerical examination of the composition tables (Vasardani and Egenhofer 2008).

To compare relations featuring a multi-holed region, a similarity evaluation method is built using the transportation algorithm (Murty 1976; Strayer 1989). The comparison of relations is translated into a balanced transportation problem of calculating the cost for transforming one relation into the other. By decomposing relations between a hole-free and a multi-holed region into elements—relations between the hole-free region and single-holed regions, as if each hole was unique—transformation of relations is interpreted as transforming the elements of one relation into those of another (Vasardani and Egenhofer 2009). The cost of such transformations is a direct indication of the dissimilarity of relations, therefore, an indirect indication of their similarity

The costs of element transformations are identified as the distances of the elements on the conceptual neighborhood graph of the relations between a hole-free and a single-holed region. Modification of that graph into a hybrid one that contains distances between relations with a single-holed region and relations of hole-free regions facilitates similarity comparisons of relations with different numbers of holes. The assumption is that the different number of holes is the result of eliminating holes and changing the relations between a hole-free and a single-holed region into relations between two hole-free regions. There also exist different reasoning processes for evaluating the change in the topological structure of a region attributed to the elimination of the hole and therefore, different possibilities are discussed.

1.7 Major Results

This thesis develops the *Holed Regions Model* (*HRM*), a formal framework for qualitative spatial reasoning using topological relations. *HRM* differs from existing models in that it explicitly incorporates relations with holed regions in the reasoning process. The *HRM* has two parts: (1) the Single-Holed Regions Model (S-HRM) and (2) the Multi-Holed Region Model (M-HRM). The S-HRM analytically examines the effects of single holes in the reasoning process and it comprises:

- 23 relations between a hole-free and a single-holed region;
- 152 relations between two single-holed regions;
- composition tables for pairs of relations between hole-free and/or single-holed
 regions that enable the derivation of composition inferences; and
- conceptual neighborhood graphs of the new sets of relations.

Along with the S-HRM, a number of quantitative measures that enable the comparison of composition tables over different domains is developed. Systematic examination using such measures shows that compositions of relations with single-holed regions are more refined and yield more unique results than compositions over hole-free regions, while reducing the ambiguities for close-to-unique references.

The M-HRM examines relations involving a multi-holed region, providing a similarity evaluation method for such relations and answering questions as to which relations may be surrogate to each other in decision-making cases. The importance of the method lies in its independence from the number of holes involved and the flexibility to accommodate various models for comparing relations of regions with different numbers of holes. The similarity results prove as well, that holes impose additional constraints for

qualitative reasoning, making similarity comparisons based only on the relation of the hosts of the holes less adequate for decision-making.

1.8 Intended Audience

This thesis is intended for people working in qualitative spatial reasoning in general and the use of spatial relations to model and reason about reality in particular. It is of interest to designers of spatial databases and spatial query languages that seek to built models based on the relations between entities and to researches from the field of geographic information science that study spatial-relation algebras. It may be of interest to experts from the fields of artificial intelligence, cognitive science, and computer science as well, as it relates to the representation of spatial information, which is necessary to any intelligent system. This thesis is of particular interest to designers of future GISs who are concerned with incorporating commonsense geographic knowledge in the systems and to researchers in cartographic generalization. Finally, it may be of interest to scientists working on various environmental subjects that require the representation and the analysis of information in the form of areas with missing data due to a sensor network's failures

1.9 Thesis Organization

The remainder of this thesis is organized into seven chapters:

Chapter 2 reviews the different aspects of a complete theory about holes and a number of existing models of topological relations. The focus is specifically on the *9-intersection* that serves as the basis for the development of the *HRM*, and other extensions of it or different models, such as RCC, that can possibly deal with two-

dimensional regions with holes. Models of geometric arrangements that resemble regions with holes are also examined.

Chapter 3 introduces the methodology used to derive the new sets of relations and the composition tables. The first step in the development of the *HRM* is the derivation of the set of relations between a hole-free and a single-holed region and their ordering in a conceptual neighborhood graph. Initially, the conceptual model of a single-holed region and the conceptual framework for deriving new relations, as well the derivation of the compositions of such relations, are introduced. The examination of the relation's properties is followed by the numerical analysis of the relations' composition inferences. New quantitative measures and comparisons with previous work provide insights about the influence of the presence of the holes.

The complementary step of the reasoning framework is the determination and analysis of the relations between two single-holed regions, which is the subject of Chapter 4. Similarly, the relations' properties are analyzed, their conceptual neighborhood graph is constructed, and their composition tables are derived. The composition inferences are compared with compositions of the same, as well as different domains. Measurements such as the *absolute* and the *cumulative composition frequencies*, accompanied by a set of graphs, complete the quantitative analysis of the comparisons.

The focus of Chapter 5 is the development of a method that assesses the similarity among relations between a hole-free and a multi-holed region when the multi-holed regions have the same number of holes. Such relations are decomposed into elements whose frequencies are stored in vectors. The chapter introduces the concept of associating the relation similarity with the cost of transforming one relation to the other,

by transforming one relation's frequency vector into that of the other. The vectors' transformation is interpreted as a case of the balanced transportation problem and the transportation algorithm is applied for its solution. This chapter also analyzes how the basis for calculating the cost of the overall relation transformation are the relation distances on the conceptual neighborhood graph of the 23 relations between a hole-free and a single-holed region.

Chapter 6 expands the similarity evaluation method to cover relations between a hole-free and a multi-holed region when the multi-holed regions have different numbers of holes. To achieve this the method is altered based on the assumption that the difference in the number of holes is due to the elimination of holes from one relation to the other. To enable the calculation of the cost of relation transformation when holes are dropped, the conceptual graph is replaced by a hybrid graph, which provides the distances between relations with a single-holed region and relations with hole-free regions that result from dropping holes. Different models that evaluate such distances are discussed.

In order to asses the similarity evaluation method and the different relation distance models developed in the previous chapter, Chapter 7 systematically analyzes the results of ranking two synthetic datasets for their similarity against specific queries. By using certain evaluation criteria, the different models are compared for their results and the similarity rankings are analyzed with the visual aid of graphs designed to facilitate interpretation. The chapter discusses the various patterns recognized in the analysis of the results and the main conclusions from comparing relations with different numbers of holes

Chapter 8 concludes this thesis with a summary, a review of the major contributions and findings, as well as the opportunities for future research that either complement or are made possible through this study.

Chapter 2 Spatial Objects with Holes

The existence of holed objects and holed regions has been acknowledged and discussed in the literature not only from the spatial (Egenhofer and Herring 1990a; Clementini *et al.* 1993), but also from a more philosophical point of view (Casati and Varzi 1994; Lewis and Lewis 1970; 1996). Hole-free regions have been the prevailing data types for spatial analysis within GISs (Egenhofer 1989; Egenhofer and Franzosa 1995; Randell *et al.* 1992b), therefore the first attempts to reason with holed regions have evolved from models for hole-free regions (Egenhofer *et al.* 1994; Clementini *et al.* 1996b). Certain research efforts, however, have focused on representing more complex objects (Worboys and Bofakos 1993; Clementini *et al.* 1995; Schneider and Behr 2006), among them regions with holes as well, and modeling relations between them (Cohn *et al.* 1997; McKenney *et al.* 2007). In addition, models also exist for objects, the geometric representations of which visually resemble regions with a hole (Cohn and Gotts 1996a; Erwig and Schneider 1997; Roy and Stell 2001; Bejaoui *et al.* 2008).

This chapter reviews such research efforts and builds the necessary knowledge for examining exclusively two-dimensional regions with a hole in the plane, the relations among them, as well as tools that may be used for inferring new knowledge from comparisons between topological relations.

The chapter is structured as follows: Section 2.1 provides a brief review of a theory about holes in three-dimensional objects, while the discussion about holes in two-dimensional objects in the plane and how to distinguish them from figures in the plane starts in Section 2.2. Section 2.3 deals explicitly with the representation and topological relations between regions in the plane, examining first models for hole-free regions and then expanding to models for more complex regions. Models of objects that visually resemble a region with a hole are studied in Section 2.4, while the dynamic changes that occur in the structure of a region with the appearance or disappearance of a hole are discussed in Section 2.5. Section 2.6 examines how changes in the topological relations between regions are modeled, and the chapter is summarized in Section 2.7.

2.1 Theory of Holes in Three-Dimensional Objects

The nature of holes can be perceived in a variety of ways. Holes can be either considered synonymous to *perforations* on a material (Lewis and Lewis 1970), or from a materialistic viewpoint, holes are equated with the hole-linings of their material host-object. In that sense, holes are that part of their host's material that surrounds them (Lewis and Lewis 1996). Holes are dependent entities: they cannot exist alone unless some host provides a surface in which they exist. This ontological dependency suggests that *holehood* is a relational property—there are no holes *simpliciter* (Casati and Varzi 1994). However, holes are interpenetrated by other entities—including other holes (Fig. 2.1). This property implies that something can be spatially enclosed in a hole without being a part thereof (Varzi 1996).

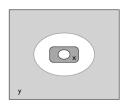


Figure 2.1 The hole in surface y encloses object x and its hole, however, x is not part of the hole.

Part-whole relations for holes can be formulated in the framework of *mereology*, the formal theory of parthood and relations (Varzi 1996). The basic ontological premise H(x, y), interpreted as "x is a hole in (or through) y," is eventually supplemented with hole-specific axioms. The primitive mereological relation is P(x, y), which reads "x is a (possibly improper) part of y" and is given the basic axiom that being a part of means overlapping, where O(x, y) reads "x overlaps y" and is usually defined as having some part in common.

Topology can be used to differentiate three main kinds of holes in three-dimensional objects. Superficial holes (or hollows) correspond to simple depressions of various depths or indentations in the surface of the host. They could, in principle, be eliminated by elastic deformation. Perforating holes (tunnels) introduce non-eliminable topological discontinuities. Internal holes (cavities) are completely hidden inside the host so that they mark a splitting in the host's complement (Casati and Varzi 1994).

Finally, an explicit *morphological* (shape-oriented) analysis relies on a characteristic property of holes: they are *fillable* because they are spacious and involve concavities (Varzi 1996). The relation considered is F(x, y) which reads "x fills y" (Casati and Varzi 1994) and it can be taken to indicate various types of fillers (Fig. 2.2) (Varzi 1996).

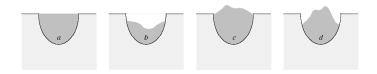


Figure 2.2 Various hole fillers: (a) *exact* filler, (b) *proper*, but not complete, (c) *complete* but not proper, and (d) neither complete nor proper (after Varzi 1996).

2.2 Deciphering Holes from Figures in the Plane

Defining holes in the three-dimensional context is a complex procedure that involves hollows, perforations, and cavities. In a geographic setting, however, the embedding in a space that resembles the bird's-eye-view is most popular. In the plane, a visual hole can be defined as a two-dimensional region on a surface, perceived as an aperture (i.e., a missing piece of the surface) through which the background surface is visible. In this context, holes in the plane are background regions that are surrounded by a foreground figure (Bertamini 2006). The letter O provides such an example. In mathematical terms this configuration implies that a hole in an object is a topological structure that prevents the object from being continuously shrunk to a point (Weisstein 2007).

However, the perceptual interpretation of a uniform, connected region that is surrounded by another such region can be visually ambiguous. Holes may be confused for smaller objects in front of larger ones, especially if the shape of their boundary resembles a well-identified object. Since background regions only exist relative to some foreground, enclosure is a defining criterion for a hole to be perceived as such, and not as an object (Bertamini and Croucher 2003; Bertamini 2006). Apart from enclosure, three additional factors are critical for identifying a hole: (1) depth factors that indicate that the enclosed region sits behind its surrounding (Fig. 2.3a), (2) grouping factors that relate the

enclosed region to its surrounding, and (3) the absence of figural factors, such as convexity, symmetry, or familiarity (Fig. 2.3b) (Nelson and Palmer 2001).



Figure 2.3 Holes vs. figures: (a) a hole is recognized due to depth factor, and (b) an hourglass figure is distinguished due to familiarity with the object (Nelson and Palmer 2001).

2.3 Topological Relations Between Two-dimensional Regions In the Plane

In qualitative spatial representations it is commonplace to describe properties of and relations between such geometric entities as points, lines, and regions of a certain space (i.e., the two-dimensional or three-dimensional Euclidean space). Relations between such entities capture various aspects of space, such as topology, orientation, distance, size, and shape. In general, relations can be grouped into three different categories: (1) topological relations, which remain invariant with respect to continuous transformations of the underlying space, such as translation, scaling and rotation (Egenhofer 1989; Egenhofer and Herring 1990; Randell *et al.* 1992b; Clementini *et al.* 1993); (2) *metric* relations in terms of distances and directions (Peuquet and Ci-Xiang 1987; Hernández 1991; Frank 1992, 1995; Goyal 2000), and (3) *partial* and *total order* relations of spatial objects (Kainz *et al.* 1993) as described by prepositions such as *in front of*, *behind*, *above*, and *below* (Freeman 1975; Chang *et al.* 1989; Hernández 1991).

The focus of this thesis is on the topological relations that may hold between two regions if at least one of them has a hole. However, topological relations have first been examined for simple regions; therefore, the most prevalent among the models for relations between simple regions are discussed.

2.3.1 Models of Topological Relations for Hole-Free Regions

The 9-intersection (Egenhofer and Herring 1990), which forms relations between hole-free regions, and the Region Connection Calculus (RCC) (Randell *et al.* 1992b), which defines relations for both hole-free and holed regions without an explicit differentiation, are the most popular theories of topological relations in geographic information science. The 9-intersection uses concepts from point-set topology (Alexandroff 1961; Munkres 1966; Spanier 1966), while RCC employs spatial logic. The two models have been developed using different definitions of a region. Therefore, before examining them, it is appropriate to refer to a few basic concepts of point-set topology with neighborhoods, open and closed sets that help distinguish the differences in the regions as used by the two models and by extensions of the models that cover more complex objects, or by alternative theories.

2.3.1.1 Basic Topological Concepts

Let S be a topological space and X be a subset of points of S. The *interior* of X, denoted by X^o , is the union of all open sets contained in X, that is, the interior of X is the largest open set contained in X. A point $x \in X^o$, if and only if there is an open set U such that $x \in U \subset X$, which means that there is a neighborhood of x contained in X.

The *closure* of X, denoted by \overline{X} , is the intersection of all closed sets that contain X, i.e., the closure of X is the smallest closed set containing X. A point $x \in \overline{X}$, if and only if $U \cap X \neq \emptyset$ for every open set U containing x. The *exterior* of X then, denoted by X^- , is the complement of the closure of X, i.e., $X^-=S-\overline{X}$. If C is a closed set, then $\overline{C}=C$.

The *boundary* of X, denoted by ∂X , is the intersection of the closure of X and the closure of the complement of X, i.e., $\partial X = \overline{X} \cap \overline{S-X}$. The boundary is then a closed set and a point $x \in \partial X$, if and only $U \cap X \neq \emptyset$ and $U \cap (S-X) \neq \emptyset$ for every open set U containing x. The relations between interior, closure and boundary are given in Equations 2.1a and 2.1b.

$$X^{\circ} \cap \partial X = \emptyset \tag{2.1a}$$

$$X^{\circ} \cup \partial X = \overline{X} \tag{2.1b}$$

A separation of X is a pair of non-empt subsets, A and B, such that $A \cup B = X$, $\overline{A} \cap B = \emptyset$ and $A \cap \overline{B} = \emptyset$. If there exists a separation of X, then X is said to be disconnected, otherwise X is connected. If C is connected, then \overline{C} is also connected.

The *regularization* of X is the closure of the interior of X, that is, $reg(X) = \overline{X}^{\circ}$. The regularization process eliminates from a set any pathological features such as punctures or arcs and any non-areal points such a mixtures of points, lines and areas. An object on which regularization has no effect is termed regular closed, i.e., X is *regular closed* if and only if $\overline{X}^{\circ} = X$ (Worboys and Duckham 2004).

The Euclidean plane with the usual topology is the most important example of a topological space for the purposes of GIS. A *homeomorphism* (or topological transformation) of \mathbb{R}^2 is a bijection of the plane that transforms each neighborhood in the domain to a neighborhood in the image. Any neighborhood in the image must be the result of the application of the transformation to a neighborhood in the domain. If Y is the result of applying a homeomorphism to a point set X, then X and Y are *topologically equivalent*. Examples of homeomorphisms include the notions of translation, rotation scaling and skew in the Euclidean plane. Properties that are preserved under homeomorphisms, such as connectedness, are called *topological invariants*.

2.3.1.2 9-Intersection

The 9-intersection (Egenhofer and Herring 1990b) is a generalization of the 4-intersection (Egenhofer 1989; Egenhofer and Herring 1990a; Egenhofer and Franzosa 1991). Even though the intersection method can be applied to geometric features of different dimensions, such as points, lines, and polygonal areas in the plane, the focus here is on polygonal areas alone. Specifically, the sets of interest are *spatial regions*, as these are defined in Egenhofer and Franzosa (1991): Let S be a connected topological space. A *spatial region* A in S is a non-empty proper subset of S such that A^o is connected, and A is regular closed, that is, $A = \overline{A^o}$. It follows from the definition that the interior of a spatial region is non-empty and that a spatial region is closed and connected since it is the closure of a connected set. The boundary of such a spatial region A is always non-empty, that is, $\partial A \neq \emptyset$ (Egenhofer and Franzosa 1991).

The definition of binary topological relations between two spatial regions, A and B, is based on the intersections of the interiors, boundaries and, exteriors of A and B. A 3x3 matrix, M, called the 9-intersection, concisely represents these intersections (Eqn. 2.2). The 4-intersection is the 2×2 subset of M—the intersections of interiors and boundaries, in this standardized sequence.

$$M = \begin{pmatrix} A^{\circ} \cap B^{\circ} & A^{\circ} \cap \partial B & A^{\circ} \cap B^{-} \\ \partial A \cap B^{\circ} & \partial A \cap \partial B & \partial A \cap B^{-} \\ A^{-} \cap B^{\circ} & A^{-} \cap \partial B & A^{-} \cap B^{-} \end{pmatrix}$$

$$(2.2)$$

Each intersection in M is examined against topologically invariant criteria under homeomorphisms. By considering the property of *content* (i.e., *empty* and *non-empty* values)—which is set-theoretic and, therefore, topologically invariant—one distinguishes 2^9 =512 possible matrices. For each such matrix, there exists a corresponding binary topological relation. The relations that can actually be realized in a particular space are a

subset of the 512 ones, depending on the particular properties of the objects, the embedding space (i.e., \mathbb{R}^n , \mathbb{Z}^n), the surface of an n-sphere) and their co-dimensions, that is, the difference between the dimension of the embedding space and the objects. For two spatial regions embedded in the plane, a set of 18 binary relations can be defined according to the content criterion (Egenhofer and Herring 1990b). These relations apply to spatial regions with either connected or disconnected boundaries an example of the latter case being a region with holes. However, if only regions with connected boundaries, or homeomorphic to a two-dimensional disk are considered, then there are only eight *jointly exhaustive and pairwise disjoint* (JEPD) relations between two regions—*disjoint, meet, overlap, equal, covers, coveredBy, contains, inside* (Fig. 2.4) (Egenhofer and Herring 1990a).

$ \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} $ disjoint	$ \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} $ contains	$ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} $ inside	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ equal
$ \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} $	$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$
meet	covers	coveredBy	overlap

Figure 2.4 The eight binary topological relations for hole-free regions as defined in the 9-intersection and their matrices.

This set is a closed set of relations with complete coverage. The *converse* relation \bar{t} is defined for each topological relation t such that $\overline{t(A, B)} = t(B, A)$, and the disjunction of all eight topological relations forms the *universal* relation, denoted by U. All compositions in the 9-intersection are extensional and therefore form the basis for a

strong relation algebra that can be used to reason about topological relations (Egenhofer 1991; Smith and Park 1992).

The 9-intersection on simple regions has been refined in a number of ways, especially using additional topological invariants, such as the dimension of the intersection components, their types (touching or crossing), or the number of the components, to express more details about the topological relations (Egenhofer 1989; 1993; Clementini *et al.* 1993; Egenhofer and Franzosa 1995)

2.3.1.3 Region Connection Calculus

The Region Connection Calculus (RCC) is a fully axiomatized first-order theory of topological relations (Randell *et al.* 1992b; Cohn *et al.* 1997). It is founded on Clarke's (1981, 1985) calculus of individuals based on connection. Unlike the 9-intersection, which models relations between different spatial entities (i.e., points, lines, regions), RCC applies to a single domain, that is, *spatial regions*, which for RCC are defined as non-empty, regular closed subsets of some topological space *U*. The interiors and exteriors of such regions need not be connected and can therefore, have holes or disjoints parts of arbitrary, but all of the same, dimension.

The basic primitive relation of RCC is the binary relation C(x, y) between two regions x and y, which reads "x connects with y." The interpretation of C(x, y) in RCC is that spatial regions x and y are connected if their topological closures share a common point and the only requirement is that the relation C is reflexive and symmetric, enforced by axioms included in the theory. Using C(x, y), a large number of different relations can be defined (Gotts 1994; Gotts 1996), such as $P(x, y) = \forall z [C(z, x) \rightarrow C(z, y)]$, which reads "x is part of y", or $O(x, y) = \exists z [P(z, x) \land P(z, y)]$, which reads "x overlaps y." A set of eight

JEPD base relations, denoted as RCC-8 (Table 2.1), allows for topological distinctions rather than just mereological (Randell *et al.* 1992b). Most other relations definable in RCC are refinements of these relations and other primitive functions can be added, such as the *convex-hull* function (Randell *et al.* 1992b; Davis *et al.* 1999).

Table 2.1 Topological interpretation of the eight base relations of RCC-8 (Bennett 1998).

RCC-8 Relation	Topological Constraints	
$\langle A, B \rangle \in DC$ (A is disconnected from B)	$A \cap B = \emptyset$	
$\langle A, B \rangle \in EC$ (A is externally connected with B)	$A^{\circ} \cap B^{\circ} = \emptyset, A \cap B = \emptyset$	
$\langle A, B \rangle \in PO$ (A partially overlaps B)	$A^{\circ} \cap B^{\circ} \neq \emptyset, A \not\subset B, B \not\subset A$	
$\langle A, B \rangle \in TPP$ (A is a tangential proper part of B) $\langle A, B \rangle \in TPP^{-1}$ (B is a tangential proper part of A)	$A \subseteq B, A \not\subset B^{\circ}$ $B \subseteq A, B \not\subset A^{\circ}$	
$\langle A, B \rangle \in \mathbf{NTPP}$ (A is a non-tangential proper part of B) $\langle A, B \rangle \in \mathbf{NTPP}^{-1}$ (B is a non-tangential proper part of A)	A⊂B° B⊂A°	
$\langle A, B \rangle \in EQ$ (A is equal to B)	A = B	

RCC-8 is the constraint language formed by the eight JEPD base relations and by all possible unions of the base relations. The latter represent indefinite knowledge. Using first-order definitions, it can be verified that exactly one of the eight base relations holds between two spatial regions. Since they are pairwise disjoint, there are $2^8 = 256$ different RCC-8 relations altogether (including the empty and the universal relation). The spatial regions in RCC-8, without loss of generality, due to the intended interpretation of the C relation, are *regular* closed (equivalent to the closure of their interior) subsets of a topological space, not necessarily internally or externally connected, and with no particular dimension. By applying the appropriate set-theoretic operations, the *converse*, *intersection*, and *union* of relations can be defined. The eight base relations are closed under composition and their composition table can be computed using their formal definitions (Randell *et al.* 1992a).

If given a straightforward topological interpretation in terms of point-set topology, the RCC-8 relations correspond to the eight topological relations of the 9-intersection (Li 2006). This matching seems to be a natural agreement of what topological distinctions are important (Renz 2002). However, the two models have different domains of discourse: the 9-intersection applies to simple regions, which are two-dimensional objects that are embedded in \mathbb{R}^2 and homeomorphic to the unit disk, while RCC takes the general definition of a region as a non-empty regular closed set. This difference explains also why the 9-intersection is applicable to hole-free, one-piece regions, while RCC is implicitly applicable to regions with possible disconnected interiors or exteriors.

2.3.2 Models for Representing Complex Regions

Complex regions, a sub-category of complex spatial objects, mainly refer to regions with separations of the interior (multiple components) and of the exterior (holes). For example, Italy has multiple components (the mainland and the islands) and two holes (the Vatican City and San Marino). A few models have been developed for representing complex spatial objects in general. The focus here is on complex regions.

Worboys and Bofakos (1993) introduced the concept of a *generic area* based on a labeled-tree representation. Areal objects with holes and islands nested to any finite level are modeled by describing each level with a set of nodes. The nodes are topologically equivalent to the unit disk, pairwise disjoint or with a finite intersection, and may spatially contain their child nodes. The root node represents the whole region object. Clementini *et al.* (1995) define *composite regions* as closed (non-empty) two-dimensional subsets of \mathbb{R}^2 that have closed components. Each component is a simple region (with no holes). The components' interiors are pairwise *disjoint* and their

boundaries are either pairwise *disjoint* or intersect at a finite set of points. A *complex region*, however, has components that may either be simple regions or regions with holes (Clementini and Di Felice 1996).

The OpenGIS Consortium (OGC) has informally described geometric features, called *simple features*, in the OGC Abstract Specification (OGC 1999) and in the Geography Markup Language (GML) (OGC 2005), an XML encoding for the transport and storage of geographic information. Among these features, *MultiPolygons* refer to complex regions that may have multiple components and holes. An informal definition of such objects that permit multi-components and holes has also been given in ESRI's Spatial Database Engine (SDE) (ESRI 1995).

A complex region (Schneider and Behr 2006) comprises one or several regular sets, called *faces* that may have holes and other faces as islands in the holes. A hole within a face can touch the boundary of the face or of another hole, at a single point at most. Each face is atomic and cannot be further decomposed. A complex region can comprise multiple faces which can be *disjoint*, *meet* at one or many single boundary points, or lie *inside* a hole of another face and possibly share one or several single boundary points with the hole (Fig. 2.5).

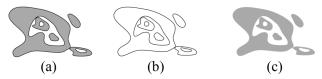


Figure 2.5 A complex region: (a) with five faces, (b) its boundary, and (c) its interior.

2.3.3 Models of Topological Relations for Complex Objects

Models of relations for hole-free regions have been typically either applied directly to more complex regions or formed the basis for developing models for such complex regions. An example of the former approach is RCC-8, which considers for regions with holes the same relation definitions as for regions without holes (Randell *et al.* 1992b). While compact and compatible across regions with an arbitrary number of holes, this approach implies, for instance, that the three external-connection configurations—B is externally connected to A from the outside (Fig. 2.6a), through A's hole (Fig. 2.6b), or by filling A's hole completely (Fig. 2.6c)—are categorized by the same RCC-8 relation. In order to distinguish these cases would either have to count the components of the complement or use different kinds of connection of increasing strength identified in a later study (Cohn and Varzi 2003). The problem is partially addressed with RCC-23, a larger number of relations based on the *conv* primitive, a function such that *conv*(A) denotes the *convex hull* of region A. While RCC-23 differentiates externally connected (Fig. 2.6a) from connected from the inside (Fig. 2.6b), it groups together the latter case with that of B filling A's hole (Fig. 2.6c). This problematic grouping of topologically different configurations under the same relation is a limitation that can be remedied if the model of relations is explicitly designed to account for the holes in the regions.

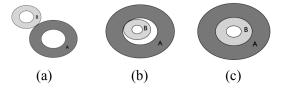


Figure 2.6 Three topologically distinct grouped together by RCC. RCC-8 groups all three together under relation EC(A, B). RCC-23 groups (b) and (c) together under INSIDEi EC(A, B).

Models for complex objects have been systematically derived, either directly or indirectly, from the intersections between interiors, boundaries and possibly exteriors of two objects, in three main ways: (1) building on the atomic relations between hole-free regions (Egenhofer *et al.* 1994), (2) forming a smaller group of relations and adding

boundary operators (Clementini *et al.* 1995; Clementini and Di Felice 1996), and (3) recognizing different components and exteriors of the complex objects and imposing additional intersection constraints to differentiate between enriched intersection matrices (Schneider and Behr 2006; McKenney *et al.* 2007; McKenney *et al.* 2008; Kurata 2008).

2.3.3.1 Building on Relations between Hole-Free Regions

A holed region has a separated exterior, unlike hole-free regions. In the case of a separated exterior, there exists an unbounded set that is the *outer exterior* (or the set of points in the plane that surrounds the region and its holes) and a number n > 0 of bounded sets that form the *inner exterior(s)*. The union of the outer and the inner exterior(s) makes up the entire exterior of a holed region. A hole in a region A corresponds to the closure of an inner exterior and it is pairwise disjoint from any other hole. The generalized region A^* (Egenhofer *et al.* 1994) is defined as the union of A and all the holes that are contained in A, essentially *exactly filling* all holes, similar to the filling between holes and material bodies (Section 2.4). Both the generalized region and each of the holes are then simple regions, homeomorphic to the unit-disk, and the generalized region always *contains* each of the holes.

Using the relations defined in the 9-intersection, the topological relationship between two holed regions is characterized as the conjunction of the relations between the underlying hole-free regions—the generalized regions and the holes. For two holed regions A and B with n and m holes respectively, a matrix of (n+1)(m+1) elements represents the topological relation between A and B (Egenhofer et al. 1994). A relaxation on the restrictions of the model allows the holes to be at the region's boundary or along the boundary of another hole. This relaxation changes the types of topological relations that may hold between the generalized region and each hole to

(contains \lor covers \lor equal), and between any two holes to (disjoint \lor meet). In an extension of the model, the dimension of the relation meet between two touching holes can vary, depending on whether the holes meet at a common point or along a set of points that form a line (Paiva 1998).

In this model, the number of topological relationships between two holed regions is dependent on the number of holes in each region, resulting in an arbitrary number of relations. To avoid production of an infinite set of valid topological relations, this thesis identifies the finite set of relations that can hold in topological configurations involving a certain number of single-holed regions, even if the underlying concept of using the relations between the simple regions in such configurations is the same.

2.3.3.2 Models Based on A Smaller Group of Relations

Initially applied to composite regions (Clementini *et al.* 1995), the relations between hole-free regions defined in the Calculus-Based Method (CBM) (Clementini *et al.* 1993) are extended to cover complex geometric features (Clementini and Di Felice 1996). The CBM uses the four intersections between the interiors and boundaries of simple features (points, lines or areas) to define five topological relations—*touch, in, cross, overlap* and *disjoint*. To enhance the use of the relations, operators that extract boundaries from areas and lines are defined. For a simple region A, the *boundary* operator (*b*) returns the closed curve ∂A (Eqn. 2.3).

$$(A, b) = \partial A \tag{2.3}$$

Based on the relations of the CBM for hole-free regions, the Topological Relations for Composite Regions (TRCR) are defined at two levels of granularity called the *course level* and the *detailed level*, respectively (Clementini *et al.* 1995). At the course level, the

general relationship between two composite regions is given, while at the detailed level, the single components are considered. A set of rules links the two levels of description. When dealing with a composite region A made up of n components, the boundary is made up of n closed curves that may be disjoint or touch at a finite set of points (Eqn. 2.4). With TRCR, however, the relations between composite regions are defined in an $ad\ hoc$ manner and not systematically derived from the underlying model.

$$(\mathbf{A}, b) = \bigcup_{i=1}^{n} \partial \mathbf{A}_{i}$$
 (2.4)

For a model of relations between complex regions that may have holes, the boundary operator returns a complex line, which is the union of several (either *disjoint* or intersecting at a finite set of points) circular lines. Additionally, the *cross* relationship needs to be modified appropriately (Clementini and Di Felice 1996). This approach allows self-intersecting features and the objects of representation resulting from its definitions are not necessarily unique.

2.3.3.3 Enhanced 9-Intersection

The relations defined in the 9-intersection, combined with a collection of topological *constraint* rules enable the definition of topological relations between complex regions with holes in a number of different models.

In the Schneider and Behr (2006) relation model topological relations are derived in a manner independent from the number of components that make up the participating complex regions (faces in this case), by extending the 9-intersection to point sets that belong to complex rather than to simple spatial objects. The philosophy is that if two regions intersect for example, according to a given definition, then the number of intersecting face pairs (as long as it is greater than zero) is irrelevant, since it does not

influence the fact of intersection. In order to exclude impossible configurations, a number of topological constraints (conditions) are imposed in each of the 2^9 =521 possible matrices of the 9-intersection. The ones that pass that initial test are also subject to validation through drawing, to determine whether they represent realizable configurations in the plane.

For determining the topological relations between two complex regions, nine constraint rules are applied to all 512 9-intersection matrices and this process reveals that 33 matrices satisfy the rules. All 33 configurations proved realizable in the plane. Among them, the eight topological relations between simple regions (Egenhofer 1989; Egenhofer and Franzosa 1991) can also be identified.

Much like RCC-8, this model provides a single relation for configurations with or without holes distinguishing, however, some additional details that are grouped together in RCC-8. These distinctions come, nevertheless, at the premium of causing some anomalies when holes are introduced into or removed from regions, leading at times to perplexing reclassifications of relations. For instance, configurations in Figures 2.7a and 2.7b have the same 9-intersection matrices and the difference between the two is that in 2.7b, B's hole is filled. The same 9-intersection is shared by Figure 2.7c, in which B fills A's hole, as well as Figures 2.6a and 2.6b (Fig. 2.6c is the same as 2.7c). However, when B's hole is filled, going from Figure 2.7c to 2.7d, the vanilla 9-intersection of Figure 2.7d is a different one, classifying it as a new relation. What would be expected is the grouping of figures 2.7a-b (and probably 2.6a-b) under the same relation, allowing Figures 2.7c-d to share another relation, since they involve the different configuration of having one region filling the other's hole. RCC-8, on the other hand, offers a more consistent treatment of such scenarios.

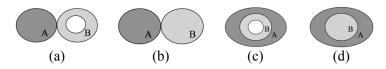


Figure 2.7 Topological configurations realized by the Schneider-Behr model (2006). Relations (a), (b), and (c) have the same vanilla 9-intersection, whereas (d) has a different one.

If the 9-inersection is applied to complex spatial objects, it considers only the interior, exterior and boundary point sets of the whole object, providing the relation for the coarse or *global* level (Clementini *et al.* 1995). As a result, local topological information regarding relations between individual and separate components of the object is lost. By defining global topological relations based on the existence of local topological relations between the various components, a *hybrid* model for topological relations between complex regions is developed (McKenney *et al.* 2007).

Given that the set of simple regions with holes (SRHs) is a subset of the set of complex regions, it follows that the set of topological relations that hold between two SRHs is a subset of the set of topological relations that can hold between two complex regions. Applying a new constraint to the 33 identified topological relations between complex regions (Schneider and Behr 2006), 15 relations are eliminated, allowing 18 relations between two SRHs (McKenney *et al.* 2007). A *localized topological predicate* (LTP) characterizes the topological relations between two complex regions A and B, by asserting which of the 18 relations between two SRHs hold between the components of the regions. Each LTP then is a conjunctive Boolean expression with exactly 18 clauses, denoted by an 18-bit vector. With the use of 17 constraints, all the invalid bit vectors are eliminated. The remaining set of LTP for complex regions consists of 137,209 18-bit vectors. If four additional bits are added to each LTP, which reveal the global relation of the objects, the number of valid 22-bit vectors is much higher. These extremely large

numbers of relations defined between two complex regions and their non-intuitive representation as Boolean 22-bit vectors are some of the drawbacks of this model, making its use cumbersome.

Problematic is also the fact that these models are based on topological relations between SRHs, a data type which is not typically implemented in current spatial database systems. A solution to this problem was sought in a later approach, where predicates as the relations that can hold between the simple regions that are included in a SRH—namely the hosts of the holes and the holes themselves—are first identified (McKenney et al. 2008). The topological relation between two SRHs is characterized as a component based topological relationship (CBTR). Each CBTR consists of four sets of topological predicates, containing all predicates that exist between the simple regions in the two SRHs, and two Boolean values that indicate whether a special condition holds. However, even this approach does not offer a manageable set of intuitively named topological relations between regions with holes that could facilitate the recognition of such relations by the user of a system.

As an extension of the 9-intersection, the 9⁺-intersection distinguishes the intersections between the subparts of the objects' topological parts (i.e., the subdivisions of the object's interior, boundary or exterior) (Kurata and Egenhofer 2006). In this model, the topological relations between A and B are characterized by the 9⁺-intersection matrix in Equation 2.5, whose nine bracketed elements are matrices by themselves, each representing the intersection between the subparts of A's one topological part and those of B's one topological part. $A^{\circ i}$, $\partial_i A$ and A^{-i} are the the i^{th} supart of A's interior, boundary and exterior, while $B^{\circ j}$, $\partial_j B$ and B^{-j} are the j^{th} subpart of B's interior, boundary and exterior, respectively. Just like the 9-intersection, topological relations

between A and B are distinguished by the presence or absence of all intersections listed in the matrix.

$$M^{+}(A,B) = \begin{bmatrix} \begin{bmatrix} A^{\circ i} \cap B^{\circ j} \end{bmatrix} \begin{bmatrix} A^{\circ i} \cap \partial_{j}B \end{bmatrix} \begin{bmatrix} A^{\circ i} \cap B^{-j} \end{bmatrix} \\ \begin{bmatrix} \partial_{i}A \cap B^{\circ j} \end{bmatrix} \begin{bmatrix} \partial_{i}A \cap \partial_{j}B \end{bmatrix} \begin{bmatrix} \partial_{i}A \cap B^{-j} \end{bmatrix} \\ \begin{bmatrix} A^{-i} \cap B^{\circ j} \end{bmatrix} \begin{bmatrix} A^{-i} \cap \partial_{j}B \end{bmatrix} \begin{bmatrix} A^{-i} \cap B^{-j} \end{bmatrix} \end{bmatrix}$$

$$(2.5)$$

The 9⁺-intersection is extended to derive the candidates for the topological relations between various pairs of objects embedded in the various spaces (Kurata 2008). The difference with the previous models based on the 9-intersection extensions, is that the constraints applied on the 9⁺-intersection are nine *universal constraints*, instead of specific constraints to each pair of objects in each space. Using this model and the 9⁺-intersection matrix in Equation 2.5, the same set of relations between a hole-free region A and a single-holed region B that are derived in this thesis, is produced by different 9⁺-intersection matrices. In these matrices $\partial_1 B$, $\partial_2 B$, B^{-1} and B^{-2} are B's outer boundary, hole boundary, outer exterior and inner exterior (the hole). The advantage of this thesis' approach over the 9⁺-intersection is that the relations defined in this work are based on the relations identified by the original 9-intersection, whose definitions are readily available in current geographic information systems for the simple regions data type and need not be reconstructed by applying the universal constraints for the subparts on the fly.

2.4 Look-A-Likes of Regions with Holes

Apart from crisp regions of space that represent two-dimensional geographic phenomena with well-determined boundaries (i.e., the area occupied by a shopping mall), data stored in a GIS are frequently about areas of space with vague or indeterminate boundaries

(Burrough and Frank 1996), such as urban areas or areas occupied by particular species or habitats of plants. Such vague regions may visually resemble regions with a hole, if one depicts them graphically by only drawing the regions' outlines.

A vague region in Erwig and Schneider (1997) is a pair of disjoint regions: the *kernel*, which describes the area that belongs to the vague region and the *boundary*, which is the area or parts of it that may or may not belong to the vague area or it is *unknown* whether they belong to the vague area or not. The boundary need not necessarily be one-dimensional but can be a region as well. Kernels and boundaries may be adjacent, and may have holes which themselves may contain kernels and boundaries with holes (Fig. 2.8) (Erwig and Schneider 1997).

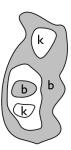


Figure 2.8 A vague region with kernels (k), boundary (b) and a hole (blank area).

While RCC (Section 2.3.1.3) deals with crisp regions (i.e., regions with determined boundaries and certain extent), the egg-yolk model specifies relations between vague regions. It is possible to use two or more concentric *subregions* that indicate degrees of membership in a vague region (Lehmann and Cohn 1994). In the simplest case with two subregions, the inner one is referred to as the *yolk*, the outer as the *white*. Taken together, they define the *egg* (Fig. 2.9). The following constraint conservatively defines limits on the possible complete crisping or precise versions of the pair of vague regions that represent the egg and the yolk of an egg-yolk configuration: any acceptable complete crisping must lie between the inner and outer limits defined by the egg and yolk (Cohn

and Gotts 1996a). This is a way to show that the entire non-crisp region's penumbra of vagueness lies between these limits.

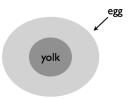


Figure 2.9 Example of a vague region modeled by the egg-yolk representation.

The egg-yolk formalism allows for the RCC-5 relations (i.e., a subset of the RCC-8 where PP and TPP are grouped under PP, their converses under PP⁻¹, and DC and EC are grouped together under DC) between any egg-egg, yolk-yolk, or any egg and yolk each belonging to different configurations, with the constraint that a yolk is always a proper part of its own egg. There are 46 possible configurations recognized between two egg-yolk pairs (Fig. 2.10). Each of these configurations may result in a subset of the RCC-5 set of relations, according to the various complete crispings of the vague regions that can take place in each occasion (Cohn and Gotts 1996a).

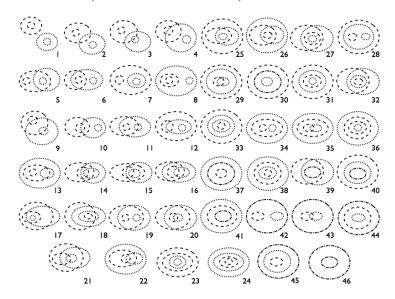


Figure 2.10 The 46 possible relations between two egg-yolk pairs (numbering after Cohn and Gotts 1996a).

The egg-yolk approach may be considered comparable to the regions with holes approach (Egenhofer *et al.* 1994). A correspondence between a region with a hole and an egg-yolk pair gives almost the same analysis. Each relation between two vague regions in RCC-5 is found as a scene representation that shows the relations according to the 9-intersection between each pair of simple regions that exist in the scene (Li 2006). However, eight base relations are considered in (Egenhofer *et al.* 1994), in contrast to five base relations in the egg-yolk theory. Since *disjoint* and *meet* are grouped under the DR relation, and TPP and NTPP under PP, the egg-yolk theory leaves out many more relations between regions with holes than the 9-intersection method does. For example, if two regions *meet* and they each *cover* their single hole as shown in Fig. 2.11a, then according to the egg-yolk model their relation would be classified the same as the topologically different relation shown in Fig. 2.11b. In a different approach, where regions may or may not be crisp, referred to as *OCregions* (for 'Optionally Crisp'), more relations than the original 46 between two egg-yolk pairs are possible (Cohn and Cotts 1996b).



Figure 2.11 Two topologically distinct scenarios of a relation between two single-holed regions that are grouped together under the egg-yolk theory. Both (a) *meet* and (b) *disjoint* relations are grouped together under case (b).

In a generalization of the egg-yolk model, a vague region is modeled by a pair of RCC regions representing the lower and upper approximations (Roy and Stell 2001). When the lower and upper approximations are equal, the region is crisp. Alternatively, rough sets (Pawlak 1994) are used to model vague regions (Bittner and Stell 2000). In

this case, the lower and upper approximations are a pair of classical sets that together represent the rough set. As expected, the lower approximation consists of all the elements that certainly belong to the set, whereas the upper approximation consists of all elements that possibly belong to the set. Finally, vague regions that visually resemble regions with holes may also be represented by as a single-component fuzzy set that does not have irregularities such as isolated vague points and vague lines, or punctures and cuts (i.e. removed vague points and vague lines, respectively) (Dilo *et al.* 2007). A vague region is then a broad boundary region, such that points in the broad boundary have different membership values that gradually change between neighboring points. If the membership values change abruptly along a line making a stepwise jump, a core area is more distinguishable, resembling more in this case a region with a hole (Fig. 2.12). However, such abrupt value changes from both sides of the line are not allowed (Dilo *et al.* 2007).

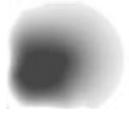


Figure 2.12 A vague region with a single-component core.

Another model for spatial objects with indeterminate boundaries includes, in many occasions, geometric representations that resemble those of regions with a hole (Clementini and Di Felice 1996b). The broad-boundary model describes the uncertainty in the boundaries as a two-dimensional exact zone surrounding the object, which acts as a separation of the space that surely belongs to the region from the space that is surely outside. It uses as basis the 9-intersection, only now the boundary of each region may be at parts or completely two-dimensional. The following definitions, different from the 9-intersection, apply:

- A region with a broad boundary A is made up of two simple regions, A_1 and A_2 , with $A_1 \subseteq A_2$, where ∂A_1 is the inner boundary of A and ∂A_2 is the outer boundary of A.
- The *broad boundary* ΔA of a region with a broad boundary is a closed connected subset of \mathbb{R}^2 with a hole. ΔA comprises the area between the inner boundary and the outer boundary of A, such that $\Delta A = \overline{A_2 A_1}$. The interior of region A is defined as: $A^\circ = A_2 \Delta A$ (Fig. 2.13) (Clementini and Di Felice 1996b).

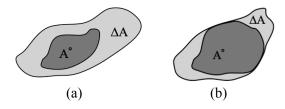


Figure 2.13 Different broad boundary regions: (a) a region with a completely two-dimensional broad boundary and (b) a region with a broad boundary that is only at parts two-dimensional.

Applying geometric conditions less restrictive than for simple regions results in 44 out of the 2⁹ matrices that correspond to possible geometric realizations between regions with broad boundaries, instead of 8 between simple regions (Clementini and Di Felice 1996b). In case the boundary maintains its 2-dimensionality, the geometric models visually resemble regions with a hole (Fig. 2.13a).

This approach considers a broad boundary region (BBR) as a whole and uses the 9-intersections among interiors, broad boundaries and exteriors for determining topological relations between two BBRs A and B made up of two crisp regions each. A different approach, the 4-tuple method, represents such relations as the combination of four individual relations: $t(A_1, B_1)$, $t(A_1, B_2)$, $t(A_2, B_1)$, and $t(A_2, B_2)$ (Du *et al.* 2007). Relations $t(A_1, B_1)$ and $t(A_2, B_2)$ can be any of the set of five relations {*disjoint, overlap, contain, inside, equal*} while $t(A_1, B_2)$ and $t(A_2, B_1)$ can be any of the same set, excluding

equal. The 4-tuple method makes finer distinctions and is more expressive than the 44 BBRs relations in (Clementini and Di Felice 1996b).

In a similar fashion, in the Qualitative Min-Max (QMM) model a region with a broad boundary is composed by two crisp subregions: (1) *a maximal extent* A_{max} (i.e., the representation of the region when the boundary is considered as far as possible) and (2) a *minimal extent* A_{min} (i.e., the representation of the region when the boundary is considered as close as possible) (Bejaoui *et al.* 2008). A_{max} and A_{min} are related by one of the following relations: $equal(A_{max}, A_{min})$, $covers(A_{max}, A_{min})$, or $contains(A_{max}, A_{min})$ and the broad boundary is then the difference between the two extents (Fig. 2.14).



Figure 2.14 Minimal and maximal extends of a region with a broad boundary.

Topological relations between two regions with broad boundaries are determined by specifying the subrelations between the minimal and maximal extents of the regions involved, in a 2×2 matrix (Eqn. 2.6) (Bejaui *et al.* 2008). Considering the restriction to three possible relations between the two extents of each region, only 242 out of the 8^4 =4096 possible matrices are valid relations between two regions with broad boundaries.

$$\begin{array}{c|cccc}
B_{\min} & B_{\max} \\
A_{\min} & t_1(A_{\min}, B_{\min}) & t_2(A_{\min}, B_{\max}) \\
A_{\max} & t_3(A_{\max}, B_{\min}) & t_4(A_{\max}, B_{\max})
\end{array} \tag{2.6}$$

Vague or indeterminate regions may frequently resemble holed regions; they are, however, conceptually and semantically dissimilar. Pairs of egg-yolk regions for example conceptualize the initial possible configurations that delimit the crispening of regions. The actual relation is realized only after one or both egg-yolk regions have been replaced

with a crisp region, resulting possibly in the relation that holds between two simple regions. Semantically, an egg-yolk region is an area wherein somewhere the actual entity lies, therefore, presenting no topological discontinuity.

Broad-boundary regions conceptualize an entity and a distinct surrounding area that is considered the possible boundary extend. Therefore, the inner border separates parts of the same entity. Holed regions, however, conceptualize real-world holed entities with a distinct topological discontinuity—the boundary of the hole—separating the actual region from the hole (empty space) inside it. Consequently, relations specified between broad-boundary regions cannot properly capture the semantics of relations between regions that represent actual holed entities. Furthermore, some of the conditions acknowledged in (Clementini and Di Felice 1996b), which prevent the regions from having thick boundaries for example, exclude a number of relations between holed regions where the thickness of the generalized region is not an issue. A more complete set of relations defined explicitly to capture unique configurations between two holed regions (or between hole-free and holed regions) can prove more valuable.

2.5 Change in the Structure of Dynamic Areal Phenomena

Apart from models that represent the topology of static regions with holes, current research has also focused on the changes that may occur in the topological structure of evolving complex regions. Users are often interested in the changes that occur over time in dynamic geographic phenomena such as floods and wildfires, which are monitored by wireless sensor networks (Jixiang and Worboys 2008). The locations occupied by a phenomenon at a particular time are abstracted into areal objects (Jixiang and

Worboys 2006), the evolution of which through time represents the spatial evolution of the geographic phenomenon.

An areal object is defined as a set of points R in 2D space, such that R is regular closed, bounded (contained in the 'inside' of a Jordan curve) and both R and its complement have a finite number of connected components. It is not assumed that R is simply connected or even connected, so holes and disconnected components are allowed. Holes themselves may contain islands with holes, recursively (Jixiang and Worboys 2009).

The topological relationship between the areal objects, holes and islands at a particular snapshot are represented by a rooted tree structure (Worboys and Bofakos 1993). In the tree, each node represents a region and the direct successors of each node represent the holes or islands contained in it. A region is contained within another region if and only if there is a directed path from the node representing the host region, to the node representing the contained region. The root is double circled (Fig. 2.15) (Jixiang and Worboys 2009). As the areal objects evolve through time, their topological structure may undergo changes such as the appearance and disappearance of holes. Such topological changes are called *topological events* and result in changes in the corresponding tree structure. With the use of different *tree morphisms*, which are structure-preserving functions from one tree to another, different kinds of topological changes are distinguished (Jixian and Worboys 2006; 2009).

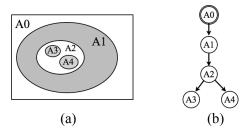


Figure 2.15 An areal object: (a) the object and (b) its tree representation.

2.6 Change in Topological Relations between Regions

The model discussed in Section 2.5 represents the dynamic topology of areal objects, which evolve due to topological changes in the structure of the objects such as the appearance or disappearance of holes. Another approach to topological change is the analysis of the transitions between binary topological relations as the result of continuous changes that preserve the topological structure of the objects involved. Therefore, the study of transitions between relations differs from the study of changes in the topological structure of objects. The analysis of transitions aims at the construction of *conceptual neighborhood graphs* (CNGs) that describe all immediate transitions among members of a set of relations (Egenhofer and Al Taha 1992; Egenhofer and Mark 1995), and serve as frameworks for the determination of relations' similarity (Bruns and Egenhofer 1996).

2.6.1 Conceptual Neighborhood Graphs for Binary Topological Relations

In closed sets of binary topological relations whose members are mutually exclusive, a *deformation* changes the relation between two objects into a new one that must be one of the remaining members in the set. Both sets of binary relations defined from the 9-intersection and the RCC-8 present these properties. Deformations are gradual changes that may occur by typical *translation*, *scaling* (expansion or reduction), or *rotation* of either object in the embedded space. Even though the differences in the various paths

through the neighborhood graph may not be a sufficient factor in human conceptualization of moving geographic regions for example, they are a necessary one, along with others, such as the identity of the moving region, or the concept of a referent region (Klippel *et al.* 2008). Each of the relations has as set of "closest relations" to which it is more likely to deform to when the geometry of one of the objects is altered. The changes are gradual and each relation has to deform first to its closest neighbors before subsequently changing to more distant relations—no jumps are expected.

In the case of the 9-intersection, each relation corresponds to a different intersection matrix. By evaluating the different entries in different matrices—empty or non empty—the topological difference for all pairs of the eight binary relations may be calculated. If transferred to a graph where nodes represent the topological relations and the edges stand for the possible changes, the topological difference reveals for each node, which of the edges link it with its topologically closest relations. Each edge represents the smallest necessary topological change to transform one relation to a different one, with no intermediate cases. The formed graph is called the Conceptual Neighborhood Graph (CNG) (Fig. 2.16) and it makes evident which of the transitions can occur gradually when changing the geometry of one object (Egenhofer and Al Taha 1992).

Accordingly, in RCC, different sets of base relations can be expressed in the form of axioms which stipulate direct transitions that are allowed between pairs of objects over time. A pictorial representation of the RCC-8 CNG and their direct topological transitions produces their conceptual neighborhood graph (Fig. 2.17) (Randell *et al.* 1992b).

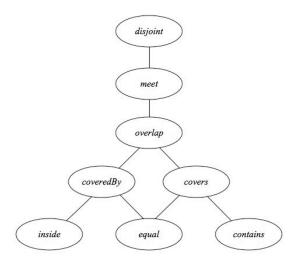


Figure 2.16 Conceptual neighborhood graph of the 9-intersection for regions (Egenhofer and Al-Taha 1992).

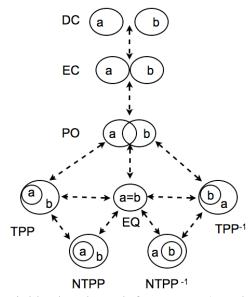


Figure 2.17 Conceptual neighborhood graph for RCC-8 (Randell *et al.* 1992b). The graph is rotated by 90° to be aligned with Fig. 1.16.

2.6.2 Similarity of Binary Relations

Similarity, as the assessment of deviation from equivalence, is an organizing mechanism by which individuals classify objects, form concepts, and make generalizations (Tversky 1977). People frequently and casually judge the similarity of spatial configurations in everyday life. While the common principle is enabling the

quantification of deviations from equivalence for spatial configurations, the methods employed may differ. One of the methods is applying the concept of gradual change as this was described in Section 2.6.1, to impose an order on sets of spatial relations. Two spatial relations that require little change to deform one into the other are more similar than relations that require more change (Bruns and Egenhofer 1996).

Topological constraints are attractive for assessing similarity as they don't respond to subtle geometric variations and they only get changed when significant alterations occur. Based on the concept of gradual change, topological relations are organized on conceptual neighborhood graphs, which facilitate the ordering of topological relations, and support the determination of similar relations. For example, on the CNG for the 9-intersection for simple regions, statements like "meet is similar to overlap" or "meet is more similar to overlap than to contains" are enabled. The similarity ordering of sets of relations was first introduced for Allen's 1-dimensional interval relations (Freksa 1992), while conceptual neighbors of distance and direction relations between regions (Bruns and Egenhofer 1996) and of topological line-regions relations have also been described (Egenhofer and Mark 1995).

2.7 Summary

Many real world entities have holes of some sort and therefore when reasoning about objects in the three dimensions or in the plane, it is desirable if not necessary that holed objects are also accounted for. In the three-dimensional environment, holes have mainly been studied in philosophy and in visual cognition. In qualitative spatial reasoning, which is the focus of this thesis, there have been efforts to incorporate two-dimensional holed regions in models of topological relations between spatial entities.

The basis of the initial models have been models of topological relations between simple regions, such as the 9-intersection, which has been modified and extended to include regions with holes, or RCC, which implicitly applies uniformly to simple, holed or multi-component regions. Different models focus on more complex objects, among them regions with holes. Furthermore, the use of models for objects with indeterminate boundaries that resemble geometrically regions with holes has been suggested.

We use the 9-intersection model to develop our approach for reasoning about twodimensional regions with holes as well, but we offer a more thorough examination of the sets of possible relations between specific configurations involving regions with a single hole. The following chapter describes an important first step for our reasoning framework: the complete set of topological relations between a simple region and a region with a single hole.

Chapter 3 Relations between a Hole-Free and a Single-Holed Region

A first step in the development of the Holed Regions Model (*HRM*) is the derivation of the set of relations between a hole-free and single-holed region, when both regions are two-dimensional and embedded in the plane. This chapter defines and analyzes the new set of 23 relations between a hole-free and a single-holed region and examines quantitatively the relation composition inferences. As the foundation, this chapter introduces the qualitative model and the conceptual tools necessary for deriving sets of relations and their composition tables, as well as analysis techniques. Quantitative measures, such as *crispness* and *complete crispness*, which complement the qualitative analysis of the compositions are employed. Overall, this chapter exposes the reader to the methodology applied to developing the Single-Holed Regions Model (S-HRM), which incorporates single-holed regions in the reasoning process.

Section 3.1 specifies the qualitative conceptualization of a region with a hole and the canonical model used for modeling a single-holed region as well as such a region's topological relation with another region. Section 3.2 presents a method to derive the binary relations that are feasible between a hole-free and a single-holed region. Section 3.3 presents the 23 relations that can be found between such a pair of regions, followed

by an analysis of these relations' algebraic properties in Section 3.4. Section 3.5 presents the conceptual neighborhood graph for the 23 relations while Section 3.6 derives the qualitative inferences that can be made with the new set of relations, focusing on compositions over a common single-holed region. Section 3.7 analyzes these compositions, comparing their reasoning power with the compositions of topological relations between hole-free regions. The chapter is summarized in Section 3.8.

3.1 Qualitative Model of a Single-Holed Region

The S-HRM is based on the relations between a spatial region and a single-holed region, or between two single-holed regions so it is important to specify what spatial regions, holes, and single-holed regions are. The definition of a spatial region is adopted from Egenhofer and Herring (1990a): A *spatial region* A is a connected point-set with non-empty interior and connected boundary, homeomorphic to a 2-disk.

A single-holed region B (Fig. 3.1) is a spatial region with a disconnected boundary and separated exterior. There exists a semi-bounded set, which is the *outer exterior*, separated from the interior of the region by the *outer boundary*, and an *inner exterior*, which fills the region's hole. The hole is bounded by the *inner boundary* of the holed region. In order to get a single-holed region, we subtract from a hole-free region A the interior of another spatial region H, which is fully contained in A (Eqn. 3.1a). The hole is then defined as the closure of the interior of H (Eqn. 3.1b)

$$B = A \setminus H^{\circ}$$
, with A contains H (3.1a)

$$B_{\rm H} = \overline{H^{\circ}} \tag{3.1b}$$

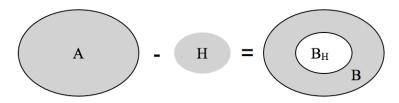


Figure 3.1 A single-holed region B.

A single-holed region B is further analyzed in five topologically distinct and mutually exclusive parts (B°, $\partial_0 B$, $\partial_1 B$, B⁻⁰, B⁻¹) (Fig. 3.2), which establish four neighborhood relations— $neighbor(B^{-1}, \partial_1 B)$, $neighbor(\partial_1 B, B^{\circ})$, neighbor($\partial_0 B, B^{-0}$).

B° is B's interior

B⁻¹ is the inner exterior of B, which fills B's hole

B⁻⁰ is the outer exterior of B

 $\partial_1 B$ is the inner boundary of B, which separates B° from B^{-1}

 $\partial_0 B$ is the outer boundary of B, which separates B° from B^{-0}

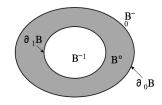


Figure 3.2 Single-holed region B's five topologically distinct and mutually exclusive parts.

There are two elements for the qualitative description of a region with a hole: (1) the hole B_H ($B^{-1} \cup \partial_1 B$) and (2) the generalized region B^* , formed by the union of the hole and the single-holed region ($B_H \cup B$). The definitions of B^* and B_H imply that they are each a spatial region and, therefore, homeomorphic to a 2-disk so that the eight binary relations defined from the 9-intersection (Egenhofer and Herring 1990a) apply to B^* and B_H (but not to B, because B has a hole and is, therefore, not homeomorphic to a 2-disk). The topological relation between B^* and B_H is always *contains*. This is a more restrictive model than the generic region-with-holes model (Egenhofer *et al.* 1994), where B_H could have also been *coveredBy* or even *equal* to B^* , thereby leading to somewhat different semantics of a region with a hole.

The topological relation between a region A and a region with a hole B is modeled as a spatial scene (Egenhofer 1997), comprising A, B*, and B_H together with the nine binary topological relations among these three regions (Fig. 3.3).

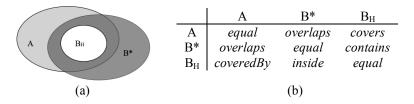


Figure 3.3 Topological relation between a hole-free A and a single-holed region B: (a) a graphical depiction of the configuration and (b) the corresponding symbolic description as a spatial scene.

The notation t_{RR} is used for the relation between two hole-free regions, whereas t_{RRh} and t_{RhR} are used for the relation between a hole-free and a single-holed region and its converse, respectively. In such a spatial scene, five of the nine binary topological relations are implied for any configuration between a hole-free and single-holed region: each region is *equal* to itself, B* *contains* B_H, conversely B_H is *inside* B*, and for the two relations between A and B* and A and B_H, their converse relations (from B* to A and from B_H to A) are implied by the arc consistency constraint (Macworth 1977); therefore, a model of such a spatial scene only requires the explicit specification of the two relations between A and B* and A and B_H to denote t_{RRh} . These relations between A and B* and A and B_H are called the *constituent relations* κ_i of a topological relation between a hole-free and a single-holed region. Their horizontal 1×2 matrix is a direct projection of the top elements in the two right-most columns of the spatial scene description (Eqn. 3.1).

$$t_{RRh}(A, B) = [\kappa_1 \kappa_2] = [t(A, B^*) t(A, B_H)]$$
 (3.1)

The first constituent relation is called the *principal relation* $\pi(t_{RRh})$ (Eqn. 3.2a), and the second constituent relation is called the *refining relation* $\rho(t_{RRh})$ (Eqn. 3.2b). In some cases, only the principal relation suffices to specify a t_{RRh} completely (Section 3.2).

However, the choice of the terminology for the constituent relations does not imply a more significant weight for the semantics of the principal relation, over that of the refining.

$$\pi(t_{RRh}(A, B)) = t(A, B^*) \tag{3.2a}$$

$$\rho(t_{RRh}(A, B)) = t(A, B_H) \tag{3.2b}$$

3.2 Deriving the Topological Relations between a Hole-Free and a Single-Holed Region

The spatial scene can also be used for the derivation of the topological relations that may actually exist between a hole-free and a single-holed region. Since two of the scene's nine topological relations are subject to variations (the relations between A and B* and A and B_H), a total of 8^2 =64 t_{RRh} could be specified. But only a subset of these 64 relations is feasible. For example, t_{RRh} = [contains disjoint] is infeasible, because A cannot contain B* and at the same time be disjoint from B_H (which is inside B*). Therefore, a topological relation between a hole-free and a single-holed region is feasible if (1) the relation's scene representation is consistent and (2) there exists a corresponding graphical depiction.

The binary topological relation between a region A and a single-holed region B is established as a 3-region scene comprising A, B*, and B_H with the constraint that B* *contains* B_H (Fig. 3.4). The topological relation between a hole-free and a single-holed region holds if this 3-region scene is node-consistent, arc-consistent, and path-consistent (Macworth 1977) for the four values $t(A, B^*)$, $t(A, B_H)$ and their corresponding converse relations $t(B^*, A)$ and $t(B_H, A)$.

	A	B*	B_{H}
A	equal	$t(A, B^*)$	$t(A, B_H)$
B^*	$t(B^*, A)$	equal	contains
B_{H}	$t(B_H, A)$	inside	equal

Figure 3.4 A three-region spatial scene. The scene captures the constituent relations of a binary topological relation between a region A and a single-holed region B.

The range of these four relations is the set of the eight t_{RR} . With four variables over this domain, a total of $8^4 = 4,096$ configurations could be described for the topological relations between a hole-free and a single-holed region. Only a subset of them is feasible, however. The feasible configurations are those whose 3-region scenes are consistent. Since in the feasible configurations, $t(A, B^*)$ must be equal to the converse of $t(B^*, A)$, the enumeration of the relations in the feasible configuration can be reduced. The same converseness constraint also holds for $t(A, B_H)$ and $t(B_H, A)$; therefore, for a feasible t_{RRh} , two of the four relations are implied. Thus, only two of the four unknown relations are necessary to completely describe a feasible t_{RRh} , reducing the number of possible configurations to $8^2 = 64$.

3.3 Twenty-Three Relations between a Hole-Free and a Single-Holed Region

In order to determine systematically the feasible t_{RRh} s, a scene consistency checker has been implemented, which iterates for each unknown (i.e., universal) relation over the eight possible relations and determines whether that spatial scene is node-consistent, arc-consistent, and path-consistent (Macworth 1977). Only those configurations that fulfill all three consistency constraints are candidates for a valid t_{RRh} . Twenty-three spatial scenes representing a hole-free and a single-holed region have been found to be consistent and examples of their graphic depictions have been identified (Fig. 3.5). The remaining 64–23=41 candidate configurations for t_{RRh} have been found to be inconsistent.

The set of the 23- t_{RRh} consists of jointly exhaustive and pairwise disjoint (JEPD) relations—there is always one of the 23 relations that applies between any possible pair of a hole-free and a single-holed region, and for any configuration no more than one of the 23 relations applies.

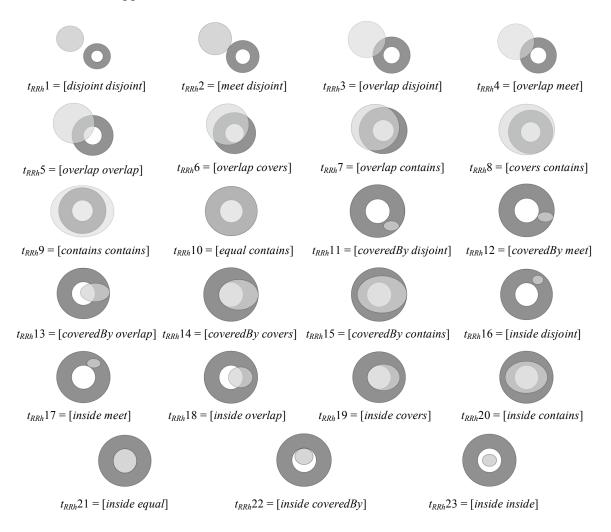


Figure 3.5 Graphical depictions and specifications of the 23 topological relations between a hole-free and a single-holed region.

3.4 Properties of the Twenty-Three Relations

These 23- t_{RRh} can be viewed as refinements of the eight t_{RR} . Five of the eight t_{RR} —disjoint, meet, covers, contains, equal—do not reveal further details if region B has a

hole, because in each of these cases the relation between A and B* is so strongly constrained that only a single relation is possible between A and B's hole, B_H . The remaining three t_{RR} —overlap, coveredBy, inside—are less constraining as each offers multiple variations for the topological relations between A and B_H : overlap and coveredBy each have five refinements, while inside has a total of eight refinements.

Without a specification of the relation between A and B_H , the configurations A overlap B^* and A coveredBy B^* are underdetermined, that is, one can only exclude for each case the three relations A equal B_H , A coveredBy B_H , and A inside B_H , but cannot pin down which of the remaining five choices—A disjoint B_H , A meet B_H , A overlap B_H , A covers B_H , A contains B_H —actually holds. Likewise, the configuration A inside B^* is undetermined without a specification of the relation between A and B_H , because any of the eight t_{RR} could hold between A and B_H .

The 23 relations between a hole-free and a single-holed region are *coarse* topological relations, which means that different topological configurations may be attributed by the same t_{RRh} , even if the topological relations they represent are not equivalent. Figure 3.6 shows three examples of distinct topological configurations that are characterized by the same t_{RRh} , in this case [*overlap meet*]. To allow for *topological distinguishability* in these cases, knowledge of additional topological invariants about the constituent relations, such as the dimension of the boundary intersection for the *meet* relation (Fig. 3.6a-b), or the number of the components of the boundary intersection for the *overlap* relation (Fig. 3.6a and 3.6c), is necessary (Egenhofer and Franzosa 1995).

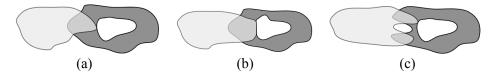


Figure 3.6 Topologically different configurations for the [*overlap meet*] t_{RRh} .

3.4.1 Converse Relations

Since the domain and co-domain of a t_{RRh} refer to different types—a hole-free region and a single-holed region—there are no symmetric t_{RRh} s. There is neither an identity relation, nor are there reflexive, or transitive t_{RRh} s. The concept of a converse relation (i.e., the relation between a single-holed and a hole-free region) still exists, however. The relation converse to t_{RRh} is implied through the converse property of the constituent relations (Eqn. 3.2)— $t(B^*,A)=\overline{t(A,B^*)}$ and $t(B_H,A)=\overline{t(A,B_H)}$ —which is captured in a transposed matrix of the constituent relations (Eqn. 3.3).

$$t_{RhR} = \begin{bmatrix} t(B^*, A) \\ t(B_H, A) \end{bmatrix} = \begin{bmatrix} \overline{t(A, B^*)} & \overline{t(A, B_H)} \end{bmatrix}^T$$
(3.3)

The converse property leads immediately to 23- t_{RhR} . Their names are chosen systematically so that all pairs of converse relations have the same index (Eqn. 3.4).

$$\forall x:1...23: t_{RhR}x = \overline{t_{RRh}x}$$
 (3.4)

From among the 23 pairs of converse relations, five pairs have identical constituent relations (Eqn. 3.5a-e), because each element of these five pairs has a symmetric converse relation, that is, $t_{RhR}x = (t_{RRh}x)^T$. This type of symmetry refers to the specification of the constituent relations and differs from the symmetry at the level of the domains.

$$\begin{bmatrix} disjoint \\ disjoint \end{bmatrix}^{T} = \begin{bmatrix} disjoint \ disjoint \end{bmatrix}$$
 (3.5a)

$$\begin{bmatrix} meet \\ disjoint \end{bmatrix}^{T} = \begin{bmatrix} meet \ disjoint \end{bmatrix}$$
 (3.5b)

$$\begin{bmatrix} overlap \\ disjoint \end{bmatrix}^{T} = \begin{bmatrix} overlap \ disjoint \end{bmatrix}$$
 (3.5c)

$$\begin{bmatrix} overlap \\ meet \end{bmatrix}^{T} = \begin{bmatrix} overlap \ meet \end{bmatrix}$$
 (3.5d)

$$\begin{bmatrix} overlap \\ overlap \end{bmatrix}^{T} = \begin{bmatrix} overlap \ overlap \end{bmatrix}$$
 (3.5e)

3.4.2 Implied Relations

The dependencies among a hole-free region's relations to the generalized region and the hole reveal various levels of constraints (Fig. 3.7). While five $t(A, B^*)$ imply a unique relation for $t(A, B_H)$, two other $t(A, B^*)$ restrict $t(A, B_H)$ to five choices. Only one $t(A, B^*)$ —inside—yields the universal relation U_8 , imposing no constraints on $t(A, B_H)$.

Known Relation <i>t</i> (A , B *)	Implied Relation $t(A, B_H)$
disjoint	disjoint
meet	disjoint
covers	contains
contains	contains
equal	contains
overlap	<pre>not {equal, coveredBy, inside}</pre>
coveredBy	<pre>not {equal, coveredBy, inside}</pre>
inside	U_8

Figure 3.7 Constraints imposed by a specified $t(A, B^*)$ on $t(A, B_H)$.

Conversely, knowledge of the relation $t(A, B_H)$ constrains $t(A, B^*)$ in three cases—if $t(A, B_H)$ is *equal*, *coveredBy*, or *inside*—to a unique relation; there are three choices for three relations between A and B* (if $t(A, B_H)$ is *meet*, *overlap*, or *covers*); five choices in one case (if $t(A, B_H)$ is *disjoint*); and six choices if $t(A, B_H)$ is *contains* (Fig. 3.8).

Known Relation $t(A, B_H)$	Implied Relation t(A, B*)
equal	inside
coveredBy	inside
inside	inside
disjoint	<pre>not {equal, covers,contains}</pre>
meet	{overlap, coveredBy, inside}
overlap	{overlap, coveredBy, inside}
covers	{overlap, coveredBy, inside}
contains	not {disjoint, meet}

Figure 3.8 Constraints imposed by a specified $t(A, B_H)$ on $t(A, B^*)$.

The dependencies may be seen as an opportunity for minimizing the number of relations that are recorded. For example, if knowledge of either one constituent relation implied a unique relation, then it would be sufficient to record only the known relation, thereby cutting into half the amount of relations to be stored for each t_{RRh} . Such a simple choice does not apply, however. Since five $t(A, B_H)$ are implied uniquely by $t(A, B^*)$, $t(A, B_H)$ needs to be recorded only in three cases to fix a complete t_{RRh} specification. Conversely only three $t(A, B^*)$ are implied uniquely by their $t(A, B_H)$. Therefore, the common-sense choice of favoring the relation with respect to the generalized region over the relation to the hole gets further support.

3.5 Conceptual Neighborhood Graph of the 23- t_{RRh}

With closed sets of JEPD relations, such as the 23- t_{RRh} , it is possible to track the topological changes that transform one relation into another of the set (Section 2.6.1). Since more than one relation may be the immediate result of a certain topological change, a linear ordering is usually impossible; therefore, linking together relations that can be immediately transformed into each other results in the conceptual neighborhood graph (CNG), which corresponds to a partially ordered set. The ordering maintained by the CNG also facilitates the determination of similar relations, by considering more similar the relations that are separated by fewer changes. Relations that belong to the same level of the partially ordered set are equally similar to the reference relation (Bruns and Egenhofer 1996). The sum of the steps of topological change between two relations determines their distance on the graph.

3.5.1 Construction of the Graph

In order to assemble the 23- t_{RRh} CNG it is essential to specify the closest neighbor(s) for each relation. The immediate neighboring relations will differ in either the principal relation (i.e., between the hole-free and the generalized region) or the refining relation (i.e., between the hole-free region and the hole), but not both. Each of these two constituent relations is one of the 8- t_{RR} between two simple, hole-free regions. Thus, each such relation is separated by one edge from the relation(s) with which it shares one of its constituent relations and their remaining constituent relations are immediate neighbors on the 8- t_{RR} CNG (Egenhofer and Al Taha 1992).

Table 5.1 shows the results of analyzing each of the 23- t_{RRh} for determining their closest neighbors by examining the constituent relations on the 8- t_{RR} CNG. The closest neighbors in the first column of Table 3.1 share the same refining but differ in their principal relations. In the second column the closest neighbors share the same principal and differ in their refining relations. Using an edge to link the closest neighbors, one derives the 23- t_{RRh} CNG (Fig. 3.9).

Examination of the graph reveals certain distinct local neighborhoods with a certain number of members each: two singleton neighborhoods for those t_{RRh} with *disjoint, meet, equals, covers* and *contains* as their principal relation, five-member neighborhoods for principal relations *overlap* and *coveredBy*, and an eight-member neighborhood for principal relation *inside*.

Table 3.1 Each of the 23- t_{RRh} 's neighboring relations according to the constituent relations

relation	S.	
t_{RRh}	Principal Relation Closest Neighbors	Refining Relation Closest Neighbors
[d d]	[m d]	-
[m d]	[d d], [o d]	-
[o d]	[m d], [cB d]	[o m]
[o m]	[cB m]	[o d], [o o]
$[o\ o]$	$[cB\ o]$	[o m], [o cv]
[<i>o cv</i>]	$[cB\ cv]$	[o o], [o ct]
[<i>o ct</i>]	$[cB\ ct], [cv\ ct]$	[o cv]
$[cv \ ct]$	[<i>e ct</i>], [<i>ct ct</i>]	-
$[ct \ ct]$	$[cv\ ct]$	-
[<i>e ct</i>]	$[cv\ ct], [cB\ ct]$	-
[cBd]	$[o\ d],[i\ d]$	[cB m]
[cB m]	[o m], [i m]	[cBd], [cBo]
$[cB\ o]$	$[o\ o],[i\ o]$	[cB m], [cB cv]
$[cB\ cv]$	$[o\ cv], [i\ cv]$	$[cB\ o], [cB\ ct]$
$[cB\ ct]$	[o ct], [i ct], [e ct]	$[cB\ cv]$
[<i>i d</i>]	[cB d]	[<i>i m</i>]
[<i>i m</i>]	[cB m]	[<i>i d</i>], [<i>i o</i>]
$[i \ o]$	$[cB\ o]$	[i m], [i cB], [i cv]
[<i>i cv</i>]	$[cB\ cv]$	[i o], [i ct], [i e]
[<i>i ct</i>]	$[cB\ ct]$	[i cv]
[<i>i e</i>]	-	$[i\ cv], [i\ cB]$
[i cB]	-	[i o], [i e], [i i]
$[i \ i]$	-	[i cB]
	$[d\ d]$	
	1	
	ر ا ا	
	[m d]	
	[o d]	
	/ 1	
	$\begin{bmatrix} cB d \end{bmatrix}$ $\begin{bmatrix} om \end{bmatrix}$	
	[i d]	
	[00]	

Figure 3.9 Conceptual Neighborhood Graph for the 23-t_{RRh}.

[cB ct]

[ct ct]

3.5.2 Distances between the $23-t_{RRh}$ on the CNG

The distance between two relations is the length of the shortest path between them. Two identical relations have zero distance between them, which is the shortest of all distances. The next shortest distance is the path of length one between two immediate neighbors and it is the minimum change that occurs in order to have one relation transform into another. The maximum distance, d_{max23} , of the 23- t_{RRh} CNG is the shortest path of length eight between four pairs of relations: $t_{RRh} = [d \ d] - t_{RRh} = [ct \ ct]$, $t_{RRh} = [d \ d] - t_{RRh} = [i \ i]$, $t_{RRh} = [i \ d] - t_{RRh} = [ct \ ct]$ and $t_{RRh} = [i \ i] - t_{RRh} = [ct \ ct]$, and it is the biggest change that can happen to one t_{RRh} relation to transform it into another. Table 3.2 shows the distances for all possible pairs of the 23- t_{RRh} relations.

Table 3.2 Distances between relations for the 23- t_{RRh} (for pairs t_{RRh} i- t_{RRh} j, i=1..23 and j=1..12).

t_{RRhi} t_{RRhj}	[<i>d d</i>]	[m d]	[o d]	[o m]	[<i>o o</i>]	[<i>o cv</i>]	[<i>o ct</i>]	[cv ct]	[<i>ct ct</i>]	[<i>e ct</i>]	$[cB\ d]$	$[cB\ m]$
dd	0	1	2	3	4	5	6	7	8	8	3	4
$[m \ d]$	1	0	1	2	3	4	5	6	7	7	2	3
[o d]	2	1	0	1	2	3	4	5	6	6	1	2
[o m]	3	2	1	0	1	2	3	4	5	5	2	1
$[o\ o]$	4	3	2	1	0	1	2	3	4	4	3	2
$[o\ cv]$	5	4	3	2	1	0	1	2	3	3	4	3
[<i>o ct</i>]	6	5	4	3	2	1	0	1	2	2	5	4
$[cv \ ct]$	7	6	5	4	3	2	1	0	1	1	6	5
$[ct \ ct]$	8	7	6	5	4	3	2	1	0	2	7	6
[<i>e ct</i>]	8	7	6	5	4	3	2	1	2	0	5	4
[cBd]	3	2	1	2	3	4	5	6	7	5	0	1
[cB m]	4	3	2	1	2	3	4	5	6	4	1	0
$[cB\ o]$	5	4	3	2	1	2	3	4	5	3	2	1
$[cB\ cv]$	6	5	4	3	2	1	2	3	4	2	3	2
$[cB\ ct]$	7	6	5	4	3	2	1	2	3	1	4	3
[i d]	4	3	2	3	4	5	6	7	8	6	1	2
[i m]	5	4	3	2	3	4	5	6	7	5	2	1
$[i \ o]$	6	5	4	3	2	3	4	5	6	4	3	2
[i cv]	7	6	5	4	3	2	3	4	5	3	4	3
[i ct]	8	7	6	5	4	3	2	3	4	2	5	4
[<i>i e</i>]	8	7	6	5	4	3	4	5	6	4	5	4
[i cB]	7	6	5	4	3	4	5	6	7	5	4	3
$[i \ i]$	8	7	6	5	4	5	6	7	8	6	5	4

Table 3.2 Continued (for pairs t_{RRh} i- t_{RRh} j, i=1..23 and j=13..23).

t_{RRhi} t_{RRhj}	[cB o]	[cB cv]	[cB ct]	[i d]	[i m]	[i o]	[i cv]	[i ct]	[i e]	[i cB]	[i i]
[d d]	5	6	7	4	5	6	7	8	8	7	8
[m d]	4	5	6	3	4	5	6	7	7	6	7
[o d]	3	4	5	2	3	4	5	6	6	5	6
[o m]	2	3	4	3	2	3	4	5	5	4	5
$[o\ o]$	1	2	3	4	3	2	3	4	4	3	4
$[o\ cv]$	2	1	2	5	4	3	2	3	3	4	5
[<i>o ct</i>]	3	2	1	6	5	4	3	2	4	5	6
$[cv\ ct]$	4	3	2	7	6	5	4	3	5	6	7
$[ct \ ct]$	5	4	3	8	7	6	5	4	6	7	8
[<i>e ct</i>]	3	2	1	6	5	4	3	2	4	5	6
[cB d]	2	3	4	1	2	3	4	5	5	4	5
[cB m]	1	2	3	2	1	2	3	4	4	3	4
$[cB\ o]$	0	1	2	3	2	1	2	3	3	2	3
$[cB\ cv]$	1	0	1	4	3	2	1	2	2	3	4
$[cB\ ct]$	2	1	0	5	4	3	2	1	3	4	5
[i d]	3	4	5	0	1	2	3	4	4	3	4
[i m]	2	3	4	1	0	1	2	3	3	2	3
$[i \ o]$	1	2	3	2	1	0	1	2	2	1	2
[i cv]	2	1	2	3	2	1	0	1	1	2	3
[i ct]	3	2	1	4	3	2	1	0	2	3	4
[i e]	3	2	3	4	3	2	1	2	0	1	2
[i cB]	2	3	4	3	2	1	2	3	1	0	1
$[i \ i]$	3	4	5	4	3	2	3	4	2	1	0

3.6 Compositions over a Single-Holed Region

A key inference mechanism for relations is their composition, that is, the derivation of the relation t(A, C) from the knowledge of the two *relations* t(A, B) and t(B, C). A complete account of all relevant compositions considers first all combinatorial compositions of relations with hole-free regions (R) and single-holed regions (Rh). Since all compositions involve two binary relations (i.e., _ _ ; _ _), each over a pair of R and Rh, there are $2^4 = 16$ possible combinations (Fig. 3.10). Eight of these sixteen combinations specify invalid compositions (C 3–6 and C 11–14), because the domain and co-domain of the composing relations' common argument are of different types (i.e., trying to form a composition over a region and a region with a hole). Among the remaining eight combinations, C 1 is the

well-known composition of region-region relations. Two pairs of combinations capture converse compositions—C 2 and C 9, as well as C 8 and C 15—while three combinations capture symmetric compositions—C 7, C 10, and C 16.

Figure 3.10 The 16 combinations of compositions of binary relations with hole-free regions (R) and single-holed regions (Rh).

From among these combinations of compositions involving a single-holed region, we focus here on C 7, the inferences from t_{RRh} ; t_{RhR} . With this specific combination it is possible to evaluate the influence that the hole bears when inserted in the common region of the composed relations. The range of the composition results is the set of eight t_{RRS} between two hole-free regions, the same as the range of the t_{RR} ; t_{RR} composition results, which are known from earlier work (Egenhofer 1994). By comparing the two composition tables, one can distinguish how the resulting t_{RRS} of the corresponding compositions differ when a hole is inserted in the common region, from when it is absent. For combinations C 2, C 8, C 9 and C 15 the range of the composition results is the set of the 23- t_{RRh} or their converse t_{RhRS} , therefore, the composition tables of these combinations cannot be examined against hole-free composition results. Finally, the range of the compositions results for combinations C 10 and C 16 is the set of t_{RhRhS} between two single-holed regions (Rh), which will be examined in the next chapter.

A spatial scene serves again as the framework for a computational derivation of all compositions for C 7. Objects A and C are two regions without a hole, whereas object B is a single-holed region. The corresponding spatial scene has four regions (A, B*, B_H, and C) with their sixteen region-region relations (Fig. 3.11). The pair of relations $t(A, B^*)$ -

 $t(A, B_H)$ must be a subset of the 23 valid t_{RRh} , while the pair of relations $t(B^*, C)$ - $t(B_H, C)$ must be a subset of the 23 valid t_{RhR} . Furthermore, $t(B^*, A)$ and $t(B_H, A)$ must be the respective converse relations of $t(A, B^*)$ and $t(A, B_H)$. The same converse property must hold for the pair $t(C, B^*)$ - $t(C, B_H)$ with respect to $t(B^*, C)$ - $t(B_H, C)$. With 23 relations for each t_{RRh} and t_{RhR} , there are 529 compositions. The range of the inferred relation t(A, C) is the set of the eight t_{RR} . This composition t(A, B); t(B, C) is specified for any spatial scene that is node-consistent, arc-consistent, and path-consistent. To determine systematically all consistent compositions, we have developed a software prototype of a consistency checker that evaluates a spatial scene for the three consistencies. All compositions were found to be valid (i.e., none of the compositions resulted in the empty relation).

	A	B*	B_{H}	C
A	equal	U	U	U
B*	U	equal	contains	U
B_{H}	U	inside	equal	U
C	U	$oldsymbol{U}$	\overline{U}	equal

Figure 3.11 The spatial scene over four regions used for the derivation of the composition t(A, B); t(B, C).

To summarize the composition results graphically, the iconic representation of the region-region relations based on their conceptual neighborhood graph is used (Hernández 1994). A highlighted relation in the graph indicates that this relation is part of the particular composition result. The universal relation U_8 is an icon with all relations highlighted (Fig. 3.12a), and a unique inference has one relation highlighted (Fig. 3.12b).



Figure 3.12 Iconic representation of compositions and relations on the 8- t_{RRh} CNG: (a) the universal relation, and (b) a unique composition result (*inside*).

The composing relations range over the 23 relations between a hole-free and a single-holed region t_{RRh} , and their converses t_{RhR} . Therefore, their iconic representation would be captured on different neighborhood graphs, namely the 23- t_{RRh} CNG and the 23- t_{RhR} CNG. However, in order to allow for a more concise, less crowded, and consistent representation of the composition table the composing relations are also depicted on the 8- t_{RR} CNG. Correspondence between the representations on different graphs is achieved by using two different symbols for depicting the two constituent relations of a t_{RRh} or a t_{RRh} on the 8- t_{RR} CNG. In particular, a large empty circle highlights the principal relation, while the black disc emphasizes the refining relation of either a t_{RRh} or a t_{RRh} . Figure 3.13 shows examples of corresponding representations of a t_{RRh} on the 23- t_{RRh} CNG and on the 8- t_{RR} CNG when the two constituent relations are different (Fig. 3.13a) and when the two constituent relations are the same (Fig. 3.13b). Using the iconic representations of composing relations and results on the 8- t_{RR} , the complete composition table for t_{RRh} ; t_{RhR} is given in Figure 3.14.

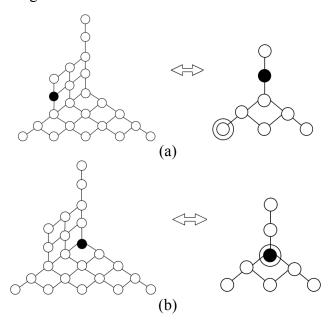


Figure 3.13 Correspondence between the iconic representations of t_{RRh} s on the 23- t_{RRh} CNG and on the 8- t_{RR} CNG: (a) for $t_{RRh} = [inside\ meet]$ and (b) for $t_{RRh} = [overlap\ overlap]$.

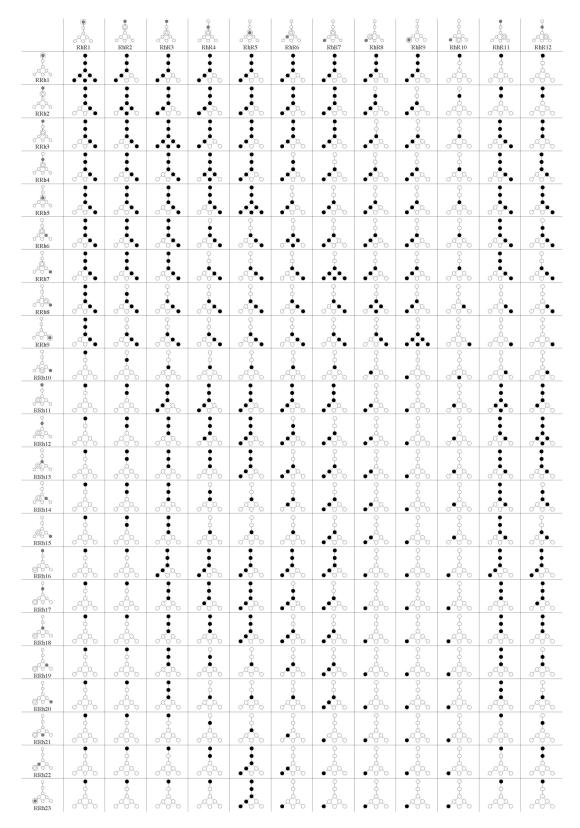


Figure 3.14 Composition table t_{RRh} ; t_{RhR} (for $t_{RhR}=[t_{RhR}1...t_{RhR}12]$).

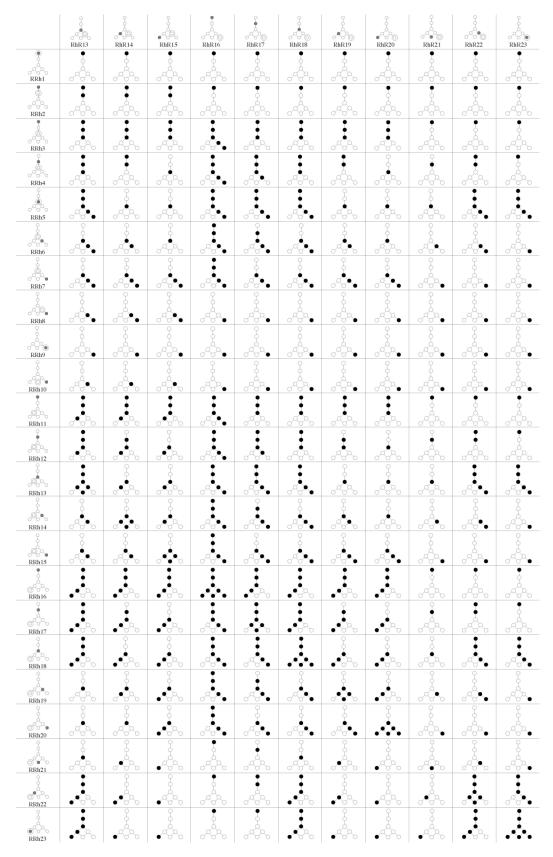


Figure 3.14 Continued (for t_{RhR} =[t_{RhR} 13... t_{RhR} 23]).

3.7 Analysis of Compositions

The examination of the 529 t_{RRh} ; t_{RhR} compositions and the comparison against the 64 compositions of t_{RR} ; t_{RR} (Egenhofer 1994), which form the benchmark for the assessment of the reasoning power of compositions involving regions with holes reveal the following:

The t_{RRh} ; t_{RhR} composition table (Fig. 3.14a and 3.14b) shows that all 529 compositions are valid (i.e., there is no empty relation as the result of any of the compositions). This means none of the 529 4-object scenes considered to calculate the compositions (Fig. 3.11) is inconsistent. The same level of consistency was also found for the t_{RR} ; t_{RR} composition table. Furthermore, all compositions are compatible with the composition results of their principal relations (Eqn. 3.6), that is, the inferences from the principal relations provide an upper bound for the reasoning over regions with a hole.

$$\forall a, b: 1... 23: \quad t_{RRh}a \; ; \; t_{Rh}b \subseteq \pi(t_{RRh}a) \; ; \; \pi(t_{Rh}b) \tag{3.6}$$

Among the 529 compositions there are 263 (49.7%) whose results are identical to the compositions of the relations' principal relations (Eqn. 3.7). Therefore, for slightly less than half of the inferences the hole is of no importance, while it matters for the remaining 266 inferences. Among the 266 compositions whose results are more refined than the compositions of their principal relations, 95 compositions are refined to uniqueness (Eqn. 3.8). If one were to resort in these cases to the compositions of their principal relations, one would incorrectly infer that these compositions are underdetermined.

$$\exists a,b \mid a \neq b: \quad t_{RRh}a ; t_{Rh}b = \pi(t_{RRh}a) ; \pi(t_{Rh}B)$$
(3.7)

$$\exists a,b \mid a \neq b$$
: $t_{RRh}a$; $t_{Rh}ab \subset \pi(t_{RRh}a)$; $\pi(t_{Rh}ab) \land \#(t_{RRh}a; t_{Rh}ab) = 1$ (3.8)

3.7.1 Cardinality and Crispness

To further assess the inference power of the compositions, we use the *composition's* cardinality and the *composition table's cardinality*. The composition cardinality card $_{23}^y$ is the count of relations in that composition result (Eqn. 3.9a). The composition cardinality of a unique composition is 1, a composition with two choices has composition cardinality 2, etc. Therefore, the upper bound of the composition cardinality applies to compositions that result in the universal relation, yielding a cardinality equal to the number of relations in that set. The composition table's cardinality γ_{23} is the sum of the cardinalities of all compositions in the table (Eqn. 3.9b). This yields the *composition table's normalized* crispness $\overline{\Gamma_{23}}$ (Eqn. 3.9c), whose lowest value of 0 stands for compositions that result in the universal relation and whose value increases linearly for composition results with fewer choices. The latter measure also applies to subsets of a composition table to assess and compare the inferences of particular groups of relations. The corresponding measures for t_{RR} ; t_{RR} can be defined accordingly.

$$\operatorname{card}_{23}^{ij} = \#(t_{RRh}i ; t_{RhR}j) \tag{3.9a}$$

$$\gamma_{23} = \sum_{i=1...\#(U_{23})}^{j=1...\#(U_{23})} \operatorname{card}_{23}^{ij}$$
(3.9b)

$$\overline{\Gamma_{23}} = 1 - \frac{\gamma_{23}}{\#(U_8) * \#(U_{23}) * \#(U_{23})}$$
(3.9c)

While the cardinality of the t_{RRh} ; t_{RhR} composition table is over seven times higher than that of the t_{RR} ; t_{RR} composition table (γ_{23} =1389 vs. γ_{8} =193), the overall inferences from t_{RRh} ; t_{RhR} are crisper, because the average composition cardinality of all t_{RRh} ; t_{RhR} is approximately 8% higher than that of t_{RR} ; t_{RR} ($\overline{\Gamma_{23}}$ =0.67 vs. $\overline{\Gamma_{8}}$ =0.62).

The increase in crispness is primarily due to a decrease in the relative number of compositions with a cardinality of 5 (and to a lesser degree cardinalities 6 and 8), while simultaneously the relative numbers of compositions with cardinalities 3, 2, and 4 (and to a miniscule amount those of compositions with cardinality 1) increase (Fig. 3.15). Overall, 239 ambiguities of pure topological reasoning are reduced but not fully eliminated, when considering the holes in the regions.

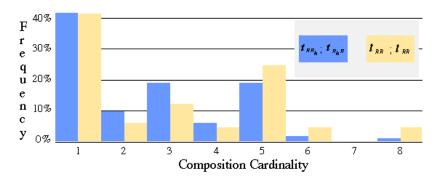


Figure 3.15 Comparison of the frequencies of compositions results. The cardinalities range between 1 (unique inference) and 8 (universal relation) and the composition tables being compared are t_{RRh} ; t_{RhR} and t_{RR} ; t_{RR} .

Compositions $t_{RRh}a$; $t_{RhR}b$ are only subject to crispening if $\pi(t_{RRh}a) \in \{overlap, coveredBy, inside\}$ and $\pi(t_{RhR}b) \in \{overlap, covers, contains\}$, yielding nine groups of compositions that feature crispenings (Figure 3.16). In these groups, the compositions with $\pi(t_{RRh}a)$ =inside and $\pi(t_{RhR}b)$ =contains have the highest crispness improvements, both in absolute counts (319) as well as per composition (i.e., 5.23, which corresponds to an average crispness improvement of 75%).

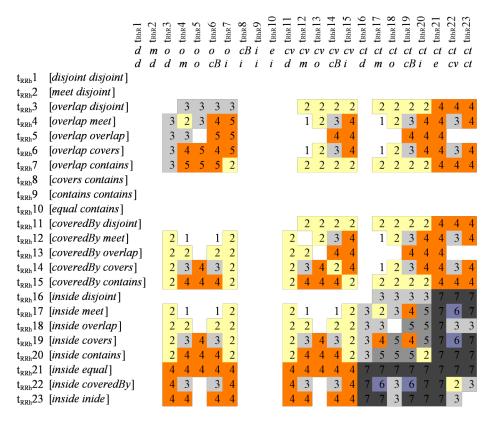


Figure 3.16 Crispness improvements (in absolute counts) for t_{RRh} ; t_{RhR} vs. t_{RR} ; t_{RR} (compositions without improvement left out; darker shading indicates stronger improvement).

The crispness improvements displayed in Figure 3.16 demonstrate also the advantage of using the 23- t_{RRh} for reasoning with single-holed region relations instead of sets of relations whose domain includes both holed and hole-free regions with no distinction. Such a set is the RCC-8 (Randell *et al.* 1992b), which implicitly applies to regions with holes as well. The RCC-8 composition table shares the same number of relations in each composition result as the 9-intersection composition table for hole-free regions (Randell *et al.* 1992a). However, even though holed regions are not excluded from its domain, the restrictions imposed by the presence of holes on the relation compositions are ignored, resulting many times in composition results that cannot be realized. Figure 3.17 shows an example of a composition over a single-holed region which results in a single relation (Fig. 3.17a), but using RCC-8, seven more *invalid*

relations (Fig. 3.17b) are allowed in the composition set. In the specific example, if hole-free region A is *inside* single-holed region B and it contains its hole, and region B and its hole both *contain* hole-free region C, then the only possible relation between A and C is *contains* (Fig. 3.17a). Using RCC-8, the relation between B's hole and region C is ignored, therefore, if relation *inside/*NTPP is composed with relation *contains/*NTPPi, any of the eight basic relations is possible between A and C. However, it is clear that given that C is *contained* in B's hole, only relation *contains*—NTTPi in RCC terminology—is possible, while the rest of the composition relations are invalid (Fig. 3.17b).

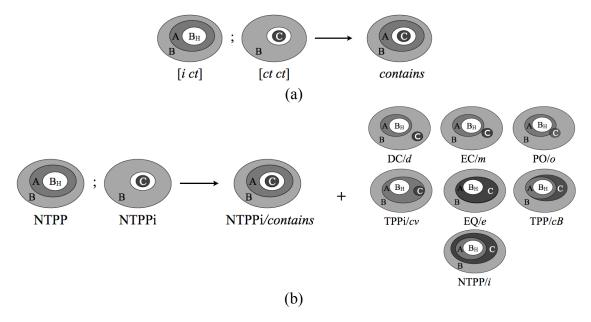


Figure 3.17 Composition of relations *inside* and *contains* over a single-holed region B using different sets of relations: (a) using the 23- t_{RRh} and (b) using RCC-8, which produces seven additional invalid composition relations, because of ignoring the relation between region C and region B's hole.

3.7.2 Unambiguous Composition Results

In absolute numbers the count of compositions with unique results goes up from 27 in t_{RR} ; t_{RR} to 224 in t_{RRh} ; t_{RhR} . Since—for a different set of relations, though—people have

been found to make composition inferences more correctly if the result is unique (Rodríguez and Egenhofer 2000), this increase augurs well for people's performance when reasoning over relations with holes.

From among the 266 compositions with crisper results, 27 (i.e., 10.2%) yield a *complete crispening*, that is a conversion from a universal composition to a unique composition. Complete crispenings occur only for compositions $t_{RRh}a$; $t_{RhR}b$ with $\pi(t_{RRh}a)$ =inside and $\pi(t_{RhR}b)$ =contains (Fig. 3.16). Resorting in these cases to the composition of their principal relations would incorrectly imply that these inferences are undetermined. For all 266 compositions whose results are crisper, on average the crispness of each of these 266 compositions improves by 3.5 counts. Given that the highest possible improvement is seven (for a complete crispening), the average crispness improvement is 50%.

3.8 Summary

This chapter studied systematically the topological relations between a hole-free and a single-holed region, offering new insights for spatial reasoning over such relations. While the 9-intersection captures eight topological relations between two regions, this number increases by 187.5% to 23 when one of the regions has a hole, yielding refinements of the eight region-region relations. Knowing the relation between a hole-free region and the generalized region implies a 63% chance (5 out of 8 relations) of uniquely identifying the complete relation between the two objects without any explicit reference to the relation with the hole.

The 23 relations' compositions over a common region with a hole show that these compositions form subsets—although not necessarily true subsets—of the results

obtained from the compositions of hole-free regions. In 36% of the true subsets, the result is unique (i.e., a single relation). Approximately half of the compositions over a region with a hole yield fewer possible relations, with an 8% increase in the average crispness when compared to the results of compositions over a hole-free region. This decrease in the number of relations in the composition results is due to a general trend of fewer results comprising five or more possibilities, in combination with an increase of the occurrence of results of fewer possibilities (four or less) and by a 10% increase of complete crispness—yielding a unique relation—among these improved results. This leads to an average crispness improvement of 50% for those results. These insights relate to people's reasoning, because relations that include a single-holed region lead to a higher relative number of unique possible results.

Chapter 4 Relations between Two Single-Holed Regions

This chapter develops the second part of the S-HRM, with the focus being on the definition, analysis, and quantitative examination of the composition inferences of the relations between two single-holed regions. The same techniques that were introduced in Chapter 3 are used to derive the set of relations, the conceptual neighborhood graph, as well as the composition tables. However, the composition inferences are now compared with compositions over the same, as well as different domains. This analysis is enabled by using quantitative measurements, such as the *absolute* and the *cumulative composition cardinality frequencies*, which are based on a set of graphs that the numerical analysis of the composition tables yields.

The remainder of chapter is structured as follows: in Section 4.1 a set of JEPD topological relations between two single-holed regions is systematically derived and analyzed for their properties (Section 4.2). Section 4.3 presents the CNG for the set of relations between two single-holed regions and Section 4.4 determines the inferences that one obtains from the composition of two relations: (1) with one single-holed region each, and (2) with two single-holed regions each, and examines the influence of introducing a hole to the common region. Section 4.5 introduces a novel set of quantitative measures to

compare composition tables with different domains and applies them to the composition tables derived in Chapters 3 and 4. The chapter is summarized in Section 4.6.

4.1 Topological Relations between Two Single-Holed Regions

The topological relation between two single-holed regions, A and B, denoted $t_{RhRh}(A, B)$, is modeled in a spatial scene comprising the generalized regions, A* and B*, and their corresponding holes, A_H and B_H , together with the sixteen binary topological relations among these four regions (Fig. 4.1).

Figure 4.1 Representation of the topological relation between two single-holed regions, A and B. The spatial scene comprises eight binary relations and their converses (the remaining eight relations are fixed for two single-holed regions).

At most the four relations $t(A^*, B^*)$, $t(A^*, B_H)$, $t(A_H, B^*)$, and $t(A_H, B_H)$ are required to specify any t_{RhRh} (Eqn. 4.1), because their converse relations are implied by the arc consistency constraint (Mackworth 1977). These four relations are called the constituent relations κ_j of a topological relation between two single-holed regions. Their 2×2 matrix is a direct projection of the top-right elements of the spatial scene (Fig. 4.1).

$$t_{RhRh}(A, B) = \begin{bmatrix} \kappa_1 & \kappa_2 \\ \kappa_3 & \kappa_4 \end{bmatrix} = \begin{bmatrix} t(A^*, B^*) & t(A^*, B_H) \\ t(A_H, B^*) & t(A_H, B_H) \end{bmatrix}$$
(4.1)

The principal relation between two single-holed regions, $\pi(t_{RhRh})$, captures the relation between the two generalized regions. It is the top-left constituent relation (Eqn. 4.2a). The other three constituent relations are referred to as the *inter-hole relation* $\omega(t_{RhRh})$ (Eqn. 4.2b), the minor relation $\psi(t_{RhRh})$ (Eqn. 4.2c), and the reverse minor

relation $\psi^{-1}(t_{RhRh})$ (Eqn. 4.2d) and they are the refining relations of a t_{RhRh} . Figure 4.2 shows the hierarchy of the constituent relations for a t_{RRh} and a t_{RhRh} .

$$\pi(t_{RhRh}(A, B)) = t(A^*, B^*)$$
 (4.2a)

$$\omega(t_{RhRh}(A, B) = t(A_H, B_H) \tag{4.2b}$$

$$\psi(t_{RhRh}(A, B)) = t(A^*, B_H) \tag{4.2c}$$

$$\psi^{-1}(t_{RhRh}(A, B)) = t(A_H, B^*)$$
(4.2d)

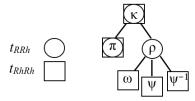


Figure 4.2 Hierarchy of the constituent relations (κ) for a t_{RRh} and a t_{RhRh} . The different symbols highlight which constituent relations are featured in each type of relation: circle for t_{RRh} s, square for t_{RhRh} s.

While the domain of the constituent relations is the set of eight region-region relations, only a subset of the $8^4 = 4,096$ is feasible, given the constraint that each hole must be included in its corresponding generalized region. For instance, the configuration with the principal relation *meet* and all other constituent relations *overlap* is impossible, whereas the configuration with the principal relation *meet* and all other constituent relations *disjoint* is feasible.

The complete set of feasible topological relations between two single-holed regions A and B is derived from the 4-region spatial scene (Fig. 4.1). To specify a t_{RhRh} , each of the four constituent relations is replaced in the spatial scene by a single t_{RR} . The t_{RhRh} is then feasible if (1) its 4-region scene is node-, arc-, and path consistent (Mackworth 1977) and (2) there exists a planar graphical depiction of that spatial scene.

By taking into account the dependencies known from the 23- t_{RRh} (Egenhofer and Vasardani 2007), the upper bound of 4,096 tests can be reduced. Splitting the *constituent*

relations into two 1×2 matrices generates two t_{RRh} —one between simple region A* and a single-holed region B (Eqn. 4.3a) and another between a simple region A_H and the single-holed region B (Eqn. 4.3b). Instead of four variables that range over the 8- t_{RR} , this approach leaves two variables that each range over the 23- t_{RRh} , yielding a total of 23²=529 possible combinations.

$$t_{RRh}(A^*, B) = [\kappa_1 \kappa_2] = [t(A^*, B^*) t(A^*, B_H)]$$
 (4.3a)

$$t_{RRh}(A_H, B) = [\kappa_3 \kappa_4] = [t(A_H, B^*) t(A_H, B_H)]$$
 (4.3b)

The set of all consistent configurations is derived with the scene consistency checker that iterated over the 23 feasible relations for each of the unknown t_{RRh} and determined computationally whether that scene was node-consistent, arc-consistent, and path-consistent. Out of the 529 candidates, 152 combinations fulfilled these criteria. Their feasibility was confirmed by drawing for each, an example configuration (Fig. 4.3). Figure 4.4 provides a numbering scheme to enable the mapping from the graphical domain onto their symbolic representations. This scheme is purely symbolic as the numbers in the symbolic names do not imply an ordering. Like other sets of topological relations, the 152 t_{RhRh} are JEPD.

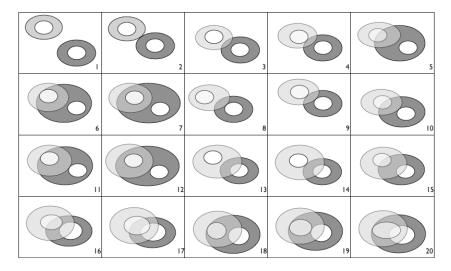


Figure 4.3 Example configurations of the 152 t_{RhRh} (for $t_{RhRh}1...t_{RhRh}20$)

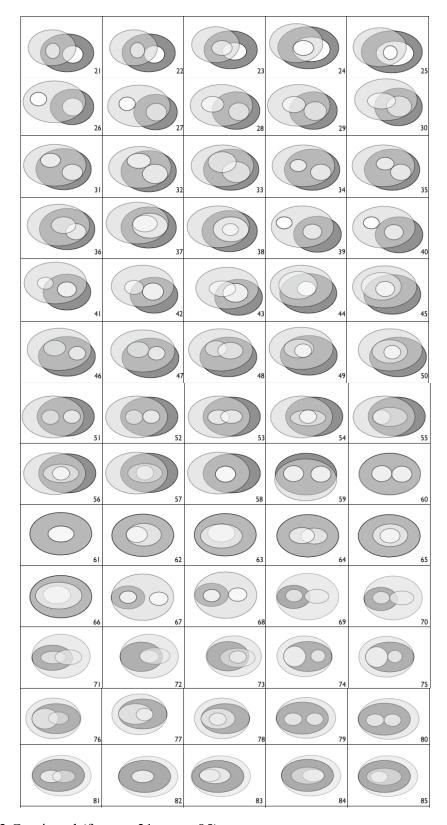


Figure 4.3 Continued (for $t_{RhRh}21...t_{RhRh}85$)

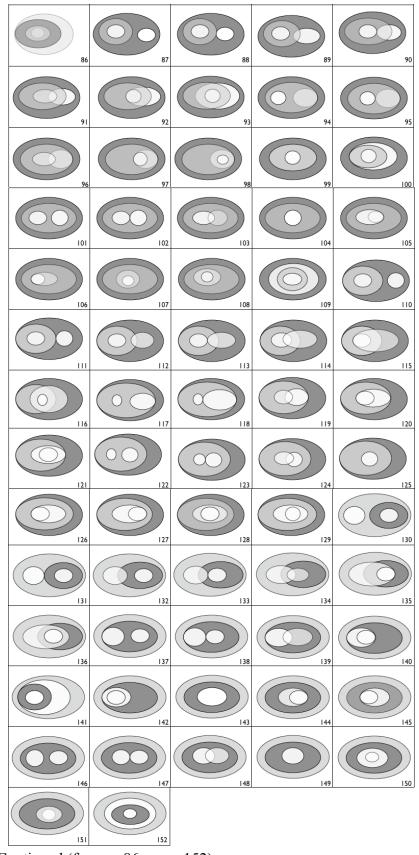


Figure 4.3 Continued (for $t_{RhRh}86...t_{RhRh}152$)

Figure 4.4 The specifications of the 152 topological relations between two holed regions, captured by their constituent relations (d=disjoint, m=meet, o=overlap, e=equal, cB=coveredBy, i=inside, cv=covers, ct=contains).

4.2 Properties of the Single-Holed Region Relations

Are all four constituent relations needed to capture the 152 relations? To answer this question we examine the occurrence of the 152 relations for each constituent relation. Figure 4.5 summarizes the occurrences of all t_{RhRh} by their principal relation $\pi(t_{RhRh})$, inter-hole relation $\omega(t_{RhRh})$, minor relation $\psi(t_{RhRh})$, and reverse-minor relation $\psi^{-1}(t_{RhRh})$. The counts of the latter two show the expected converse behavior, but otherwise the distributions of these counts differ widely.

rel	$\#(\pi(t_{RhRh})=rel)$	$\#(\omega(t_{RhRh})=rel)$	$\#(\psi(t_{RhRh})=rel)$	$\#(\psi^{-1}(t_{RhRh})=\text{rel})$
disjoint	1	48	9	9
meet	1	22	7	7
overlap	56	23	23	23
equal	8	5	1	1
covers	20	12	23	1
inside	23	15	1	87
coveredBy	20	12	1	23
contains	23	15	87	1

Figure 4.5 Counts (#) of t_{RhRh} s by their principal relation $\pi(t_{RhRh})$, inter-hole relation $\omega(t_{RhRh})$, minor relation $\psi(t_{RhRh})$, and reverse-minor relation $\psi^{-1}(t_{RhRh})$.

The summary reveals that the most constraining principal relations are *disjoint* and *meet*, which each yield only a single possible relation for two single-holed regions. In both cases the relation between the generalized regions is so strong that the two holes must be *disjoint* and each hole must be *disjoint* from the other generalized region as well. For the remaining 150 relations, the mere knowledge of the principal relation leaves the t_{RhRh} underdetermined, that is, one can exclude some—but not all—options for the three constituent relations with the holes. Among the underdetermined relations, those with the principal relation *equal* are the most constraining ones yielding eight variations, which is the same count as the cardinality of different region-region relations. This coincidence is not accidental, however, because when two generalized regions are *equal*, then they

necessarily *contain* each other's holes, which may assume any of the eight region-region relations between them. So any underdetermined t_{RhRh} with an *equal* principal relation is completed with the specification of the inter-hole relation $\omega(t_{RhRh})$. Less restrictive is the principal relation *covers* (and its converse *coveredBy*). When A* *covers* B* then it necessarily contains B_H, which means all but three of the 23- t_{RRh} between A_H and a region with a hole B apply (A_H cannot be *equal* to, *cover* or *contain* B*). Even less restrictive is the principal relation *contains* (and its converse *inside*). Since A* *contains* B* also implies A* *contains* B_H, these t_{RhRh} are completed with the information about which of the 23- t_{RRh} holds between A_H and the region with hole B. The least constraining principal relation is *overlap*, for which 56 different cases apply.

The occurrences of t_{RhRh} according to the inter-hole relation $\omega(t_{RhRh})$ show that no t_{RhRh} is fully determined by its $\omega(t_{RhRh})$ alone, while the remaining two constituent relations—the minor $\psi(t_{RhRh})$ and the reverse-minor $\psi^{-1}(t_{RhRh})$ with the same distributions for pairs of converse relations—each has three uniquely defined relations: (1) when one generalized region is *equal* to the other region's hole and (2) when the generalized region is somehow contained in the region's hole, or converse the region's hole is somehow contained in the generalized region.

When considering combinations of the constituent relations, two of these duals exhibit good increases in uniquely defined relations: $\pi(t_{RhRh})$ and $\omega(t_{RhRh})$ together specify another 21 relations that are not yet covered by them individually, while $\psi(t_{RhRh})$ and $\psi^{-1}(t_{RhRh})$ together yield another 11 unique relations. The best result from considering triples— $\pi(t_{RhRh})$, $\omega(t_{RhRh})$, and $\psi(t_{RhRh})$ or $\psi^{-1}(t_{RhRh})$ —yields altogether 66 uniquely specified relations. The analysis of the triples also shows that 20 of the 152 relations (i.e.,

13%) are not covered by the union of all three-combination triples, therefore, requiring all four constituent relations.

The relation converse to t_{RhRh} , denoted by $\overline{t_{RhRh}}$, is implied through the converse property of each constituent relation (Eqns. 4.4a-d). For the matrix representation of t_{RhRh} (Eqn. 4.1) this means that $\overline{t_{RhRh}}$ is the transposed matrix of the corresponding converse constituent relations (Eqn. 4.4e).

$$t(\mathbf{B}^*, \mathbf{A}^*) = \overline{t(\mathbf{A}^*, \mathbf{B}^*)} \tag{4.4a}$$

$$t(B^*, A_H) = \overline{t(A_H, B^*)}$$
(4.4b)

$$t(B_{H}, A^{*}) = \overline{t(A^{*}, B_{H})}$$
 (4.4c)

$$t(B_{H}, A_{H}) = \overline{t(A_{H}, B_{H})} \tag{4.4d}$$

$$t_{RhRh}(B,A) = \overline{t_{RhRh}(A,B)} = \left[\overline{t(A^*,B^*)} \frac{\overline{t(A^*,B_H)}}{t(A_H,B^*)} \frac{\overline{t(A^*,B_H)}}{t(A_H,B_H)} \right]^T = \left[t(B^*,A^*) t(B^*,A_H) \right]$$
(4.4e)

This dependency allows us to derive for each of the 152- t_{RhRh} (Fig. 4.2) its converse relation. For instance, t_{97} and t_{140} form a pair of converse relations, because the constituent relations along the main diagonal form converse pairs—inside/contains and coveredBy/covers—and the converses of the constituent relations off the diagonal map onto each other—covers/coveredBy and inside/contains.

A special role is assumed by those t_{RhRh} whose constituent relations are identical to their own converse relations, which identifies these t_{RhRh} as *symmetric* relations. For the t_{RhRh} , symmetry is implied both at the level of the domain and co-domain—both regions are single-holed—and at the level of relation specifications, in contrast to the t_{RRh} which have symmetric relations only at the level of relation specifications (Section 3.4.1). An obvious symmetric relation is t_1 with each constituent relation *disjoint*, because the converses of all four constituent relations are *disjoint* again, whereas a less obvious case

is t_{51} (with *overlap* and *disjoint* along the main diagonal being identical to their converses, while the two elements off the diagonal—*contains* and *inside*—map onto each other). Among the 152- t_{RhRh} , eighteen relations are symmetric (t_1 through t_3 , t_9 , t_{15} through t_{17} , t_{31} through t_{33} , t_{51} through t_{53} , t_{58} through t_{61} , and t_{64}), while the remaining 134 relations form 67 pairs of converse relations.

4.3 Conceptual Neighborhood Graph of the $152-t_{RhRh}$

As with the set of 23- t_{RRh} , it is possible to construct the CNG for the set of JEPD relations between two single-holed regions. Two $t_{RhRh}s$ are neighbors if they share three constituent relations (κ) and the fourth κ of one relation is an immediate neighbor of the other relation's corresponding κ on the 8- t_{RR} CNG. After analyzing each of the 152- t_{RhRh} for determining the closest neighbors by examining the constituent relations and using an edge to link the closest neighbors, one derives the 152- t_{RhRh} CNG (Fig. 4.6).

The numbering scheme of the graph corresponds to that of Figure 4.3. The 152- t_{RhRh} CNG is symmetric, in contrast to the 23- t_{RRh} CNG. The symmetry axis comprises the eighteen symmetric relations, while the remaining 134 relations form pairs of converses at mirroring locations on both sides of the symmetry axis. The relations on the CNG have between one and seven neighbors each. Table 4.1 shows the number of the relations that are characterized by a certain neighbor cardinality.

Table 4.1 Neighbor cardinalities for the 152- t_{RhRh} CNG.

Neighbor Cardinality	1	2	3	4	5	6	7
Relation Frequency	3	7	29	41	51	18	3

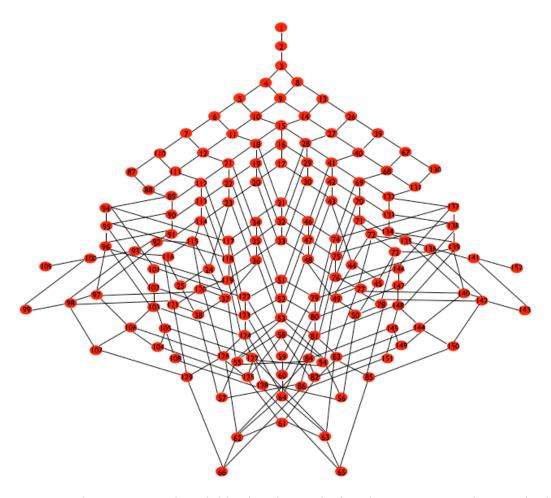


Figure 4.6 The Conceptual Neighborhood Graph for the 152- t_{RhRh} . The numbering scheme corresponds to the one in Figures 4.3 and 4.4.

4.4 Compositions Involving Single-Holed Region Relations

This section examines the influence of the hole on the inferences that result from a composition. It is suggested that when the hole of the common region of the two composed relations is taken into account, then the composition results comprise fewer relations, reducing ambiguities.

We consider the two composition cases that result in a t_{RhRh} : (1) the compositions t_{RhR} ; t_{RRh} —that is, the composition of a relation between a single-holed and a hole-free region, with a relation between a hole-free and a single-holed region—and (2) the

compositions t_{RhRh} ; t_{RhRh} —that is, the composition of two relations each between two single-holed regions. The results are also compared with the composition cases that result in a t_{RR} —the compositions t_{RR} ; t_{RR} and t_{RRh} ; t_{RhR} , both available from earlier investigations (Egenhofer 1994, Egenhofer and Vasardani 2007). The goal is to examine the influence of adding a hole to the common region of the composed relations, when the rest of the involved regions remain the same.

It might be tempting to derive the compositions t_{RhRh} ; t_{RhRh} simply from their constituent relations by applying region-region composition (Egenhofer 1994) for each corresponding pair of relations and mapping these constituent compositions back onto t_{RhRh} . For instance, the composition of t_{109} ; t_{109} could be determined over each pair of constituent relations, deriving four times that inside; inside = inside, so that t_{109} ; t_{109} is t_{109} . While this result is correct in this particular case, the approach is incomplete and too simplistic, leading at times to incorrect composition inferences. It would, for instance, derive for the composition of t_{99} ; t_{99} with three compositions of inside; inside = inside and one composition of equal; equal = equal the incorrect result t_{99} . Instead, the derivation of the compositions of t_{RhR} ; t_{RRh} and t_{RhRh} ; t_{RhRh} requires the more involved model of a spatial scene with the complete set of relations between generalized regions and their holes.

To derive the compositions of t_{RhR} ; t_{RRh} , a spatial scene over five regions (A*, A_H, B, C* and C_H) is needed as the framework for deriving the t_{RhR} ; t_{RRh} composition table. Regions A and C have a hole, whereas the common region B is hole-free, yielding 25 region-region relations (Fig. 4.7). The four derived relations $t(A^*, C^*)$, $t(A^*, C_H)$, $t(A_H, C^*)$, and $t(A_H, C_H)$ are the constituent relations of the inferred t_{RhRh} .

	A*	A_{H}	В	C*	C_{H}
A *	equal	contains	<i>t</i> (A*, B*)	<i>t</i> (A*, C*)	$t(A^*, C_H)$
A_{H}	inside	equal	$t(A_H, B^*)$	$t(A_H, C^*)$	$t(A_H, C_H)$
В	<i>t</i> (B, A*)	$t(B, A_H)$	equal	<i>t</i> (B, C*)	$t(B, C_H)$
C^*	$t(C^*, A^*)$	$t(C^*, A_H)$	<i>t</i> (C*, B)	equal	contains
C_{H}	$t(C_H, A^*)$	$t(C_H, A_H)$	$t(C_H, B)$	inside	equal

Figure 4.7 The spatial scene for deriving the composition $t_{RhR}(A, B)$; $t_{RRh}(B, C)$.

The range of t(A, B) and t(C, B) is the set of $23-t_{RhR}$, therefore, there are $23^2 = 529$ compositions. The range of the inferred t(A, C) is the set of $152-t_{RhRh}$. The scene consistency checker found all compositions to be valid (i.e., node-, arc- and path-consistent) so that no composition resulted in an empty relation. The two extreme scenarios—compositions with unique and universal results—reveal 44 unique cases (i.e., 8.3% of the 529 compositions) and six universal cases (i.e., 1.1%).

The analogous approach was used to derive the t_{RhRh} ; t_{RhRh} composition table, starting with a spatial scene over six regions for three single-holed regions (Fig. 4.8).

	A*	${ m A_H}$	B*	B_{H}	C*	C_{H}
A*	equal	contains	<i>t</i> (A*, B*)	$t(A^*, B_H)$	<i>t</i> (A*, C*)	$t(A^*, C_H)$
A_{H}	inside	equal	$t(A_H, B^*)$	$t(A_H, B_H)$	$t(A_H, C^*)$	$t(A_H, C_H)$
B*	$t(B^*, A^*)$	$t(B^*, A_H)$	equal	contains	$t(B^*, C^*)$	$t(B^*, C_H)$
B_{H}	$t(B_H, A^*)$	$t(B_H, A_H)$	inside	equal	$t(B_H, C^*)$	$t(B_H, C_H)$
C^*	$t(C^*, A^*)$	$t(C^*, A_H)$	$t(C^*, B^*)$	$t(C^*, B_H)$	equal	contains
C_{H}	$t(C_H, A^*)$	$t(C_H, A_H)$	$t(C_H, B^*)$	$t(C_H, B_H)$	inside	equal

Figure 4.8 The spatial scene for deriving the composition $t_{RhRh}(A, B)$; $t_{RhRh}(B, C)$.

Since the range of each t(A, B) and t(B, C) is the 152 t_{RhRh} , there are $152^2 = 23,104$ compositions. The scene consistency checker found all of them valid—that is, each relation inferred with the three network consistency constraints has no empty relation as its composition result. Among the 23,104 compositions, 2,239 inferences (i.e., 9.7%) are unique and six (i.e., 0.03%) are universal.

The t_{RhRh} ; t_{RhRh} composition table identifies that three t_{RhRh} relations are transitive (i.e., the composition of such relations results in the same relation). These are t_{61} , t_{109} and t_{152} . Among them t_{61} is the *identity relation* for t_{RhRh} , because it is symmetric, transitive, and reflexive. For t_{61} , the pair of generalized regions A* and B*, as well as the pair of holes A_H and B_H, is *equal*—the two constituent relations along t_{61} 's main diagonal. These two constraints imply that A* *contains* B_H and A_H is *inside* B*, which are t_{61} 's two constituent relations off the main diagonal.

The typical non-transitive composition with a unique inference composes two different relations and implies a single relation of a different category. For instance, t_2 ; t_{152} implies t_1 . Among the non-transitive compositions with unique inferences is, however, an interesting case that deviates from this pattern, providing a scenario that does not occur with region-region relations. The composition of t_{143} ; t_{143} —a single-holed region whose hole is filled by another single-holed region, which in turn has its hole filled with yet another single-holed region—is unique, but unlike a transitive relation, it does not result in t_{143} . Instead, this composition results in t_{152} (i.e., the most-inner nested region is fully contained in the hole of the outer most region).

4.5 Analysis of Compositions

In order to examine a hole's influence on the inferences, we compare some properties of the composition tables involving holed regions. The results offer an insight into how the reasoning changes when holes are taken into account and justify the need for a separate qualitative model that acknowledges holed regions explicitly.

Four composition tables are available for this analysis. The first two composition tables (t_{RR} ; t_{RR} and t_{RRh} ; t_{RhR}) both yield t_{RR} , while the second two composition tables

 $(t_{RhR}; t_{RRh} \text{ and } t_{RhRh}; t_{RhRh})$ both yield t_{RhRh} . Therefore, each pair of comparisons has compatible domains. They also feature the same pattern of inserting a hole into the previously hole-free common region.

4.5.1 Absolute Frequencies of Composition Cardinalities

The result of each composition consists of some subset of the complete set of relations over which the resulting type of relations ranges. The number of the relations in each subset is the cardinality of each composition result. For the compositions resulting in $t_{RR}-t_{RR}$; t_{RR} and t_{RRh} ; t_{RhR} —the results are subsets of the eight t_{RR} , while for compositions resulting in $t_{RhRh}-t_{RhR}$; t_{RRh} and t_{RhRh} ; t_{RhRh} —the results are subsets of the 152 t_{RhRh} . The focus here is on the frequencies of the composition cardinalities in each composition table. In order to make the results comparable, the cardinality frequencies are normalized (Fig. 4.9).

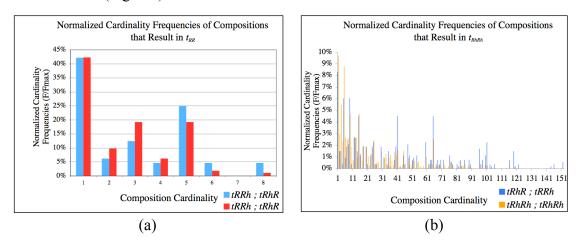


Figure 4.9 Normalized cardinality frequencies for compositions that result in (a) t_{RR} and (b) t_{RhRh} .

Examination of both pairs of composition table frequencies shows that inserting a hole into the common region increases the frequencies of crisper compositions. This increase is due to the additional constraints that the hole imposes on the relation with

other regions. The frequencies of t_{RR} inferences increase for composition cardinalities 1–4 (Fig. 4.9a), while for t_{RhRh} inferences the frequencies increase, with a few exceptions, for composition cardinalities 1 to approximately 38 (Fig. 4.9b).

The frequencies also reveal that compositions of relations with more holes make more fully unambiguous inferences. For both pairs of composition tables summarized in Figure 4.9, the compositions with the highest normalized frequencies are those with a singleton, that is, cardinality 1. For the two composition tables with t_{RR} inferences, the normalized frequencies of the singleton results are very similar—42.2% for t_{RR} ; t_{RR} and 42.3% for t_{RRh} ; t_{RhR} —whereas for the two composition tables with t_{RhRh} inferences, the normalized frequencies increase—from 8.3% for t_{RhR} ; t_{RRh} to 9.7% for t_{RhRh} ; t_{RhRh} . These quantified relative increases are smaller due to the larger range of the composition results (i.e., 152 vs. 8 possible relations). However, the larger number of single-holed regions involved in the compositions with t_{RhRh} imposes more constraints on the reasoning, thereby increasing the number of fully unambiguous inferences.

4.5.2 Cumulative Frequencies of Composition Cardinalities

The cumulative frequency of the composition cardinalities provides another measure for a hole's influence on the inferences. Each composition cardinality cc_i has a corresponding normalized frequency \tilde{f}_i . Composition cardinalities are ordered, such that $cc_i < cc_{i+1}$, with $cc_{min} = 1$ and cc_{max} being the cardinality of the set's universal relation. Taken over the ordered sequence j of all composition cardinalities cc_j , the *cumulative frequency* of the composition cardinalities captures the sum of all normalized frequencies $\tilde{f}_1...\tilde{f}_j$. Since the composition cardinalities typically differ depending on the set of relations considered, they are also normalized onto a scale of 0 to 1 (by the cardinality of the set's universal

relation), yielding *normalized composition cardinalities*. Figure 4.10 shows the cumulative normalized frequencies over normalized composition cardinalities for the two compositions t_{RR} ; t_{RR} and t_{RhR} ; t_{RRh} . Both are compositions over a hole-free region. The differences in the two graphs stem from properties of the underlying domains of the composition results—8 vs. 152 composition cardinalities—as well as the inference power of the different relation sets.

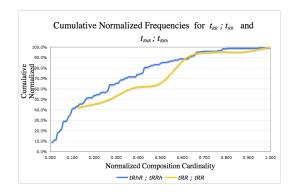


Figure 4.10 Cumulative normalized frequencies over normalized composition cardinalities for t_{RR} ; t_{RR} and t_{RhR} ; t_{RRh} .

Better comparisons can be made between pairs of compositions with equal inference ranges (Fig. 4.11). Both pairs of corresponding curves have approximately the same origin and lead to the identical culmination point (at 100% accumulation for all composition cardinalities).

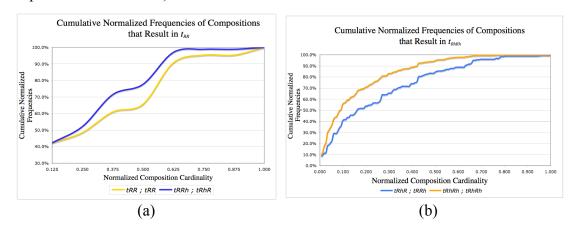


Figure 4.11 Cumulative normalized frequencies over normalized composition cardinalities for compositions resulting in (a) t_{RR} and (b) t_{RhRh} .

Examination of the graphs shows that the insertion of a hole increases monotonically the cumulative frequencies of the compositions, preserving the change in the accumulation pace. The pairs of cumulative frequency graphs for t_{RR} ; t_{RR} and t_{RRh} ; t_{RhR} (Fig. 4.11a) as well as t_{RhR} ; t_{RRh} and t_{RhRh} ; t_{RhRh} (Fig. 4.11b) have similar shapes. In both cases, the curves increase strictly monotonically, except at their tail ends where they remain almost constant. The composition cardinalities over a single-holed region always have higher cumulative frequencies than their respective composition cardinalities over a hole-free region. The same shape and the higher cumulative frequencies, motivate further investigations into the properties of the cumulative normalized frequencies.

4.5.2.1 Average Increase of Cumulative Cardinality Frequencies

Since the curves of corresponding composition pairs form essentially closed areas, it is feasible to quantify the difference in the cumulative frequencies by considering the increase in the area from the composition curve without a hole to the composition curve with a hole (Eqn. 4.5a,b).

$$\Delta H_{RR} = \frac{\int_{\min RR,R}^{\max RR,R} RR_h R(x) dx}{\int_{\min RR,R}^{\max RR,R} RRR(x) dx} - 1$$
 (4.5a)

$$\Delta H_{R_h R_h} = \frac{\int_{\min R_s R_s R_s}^{\max R_s R_s R_s} R_h R_h R_h (x) dx}{\int_{\min R_s R_s}^{\max R_s R_h R_h} R_h (x) dx} - 1$$
 (4.5b)

To calculate the area increases, 6th degree polynomial trend lines were fitted to the curves (Fig. 4.12). Since the single trend lines for t_{RR} results did not generate the desired fit (Fig. 4.12a), another approximation with the curves split in half at their apparent break points was used as well (Fig. 4.12b). The increase calculated from these two approximations was then averaged. The quantitative values obtained are $\Delta H_{RR} = 8\%$ and $\Delta H_{RhRh} = 12\%$.

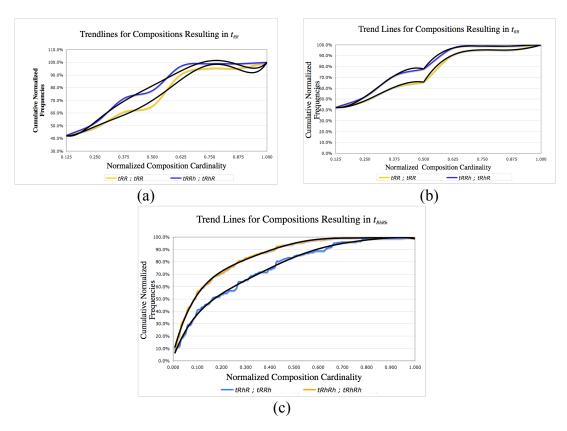


Figure 4.12 Approximations for cumulative frequency graphs: (a)-(b) trendlines for compositions resulting in t_{RR} , and (c) single trendline for compositions resulting in t_{RR} .

Diversions from the average increase in cumulative frequencies reflect an increase of the frequencies of crisper composition results when a hole is inserted. For the pair of compositions resulting in t_{RR} , the increase in the cumulative frequencies was assessed at the eight normalized composition cardinalities separately, in order to compare them with the average increase of 8%. For up to 0.5 normalized composition cardinalities, the increase in cumulative frequencies is above average—18.39% for up to 0.5 normalized composition cardinalities and 17.25% for up to 0.375 normalized composition cardinalities (which translates to 3 and 4 out of the eight t_{RR}) (Fig. 4.11a). These numbers indicate that overall, when the composition's common region is a single-holed region, it increases the occurrence of crisper results, and specifically for composition results with up to 3 or 4 relations.

This increase of crisper composition results is also verified for compositions that result in t_{RhRh} . Samples of cumulative frequencies, taken at the normalized composition cardinalities of 0.125, 0.250, 0.375, 0.5, 0.625, 0.750, 0.875, and 1, show that the biggest increase in cumulative frequencies occurs for up to 0.125 composition cardinalities—which translates to 19 out of 152 relations for the t_{RhRh} domain. This increase is 36%, in contrast to 12%, which was the average increase calculated by the difference in the areas under the trendlines. The increase in the cumulative frequencies continues to be higher than the average for up to a normalized composition cardinality of 0.5 (i.e., compositions with 76 relations in their results) (Fig. 4.11b). These observations verify the hypothesis that when the composition's common region is a single-holed region, crisper results are anticipated than for compositions over a region without a hole.

4.5.2.2 Saturation and Accumulation Pace of Cumulative Cardinality Frequencies

It is also observed that saturation of the cumulative frequency is reached earlier when the common region of the composition has a hole. For t_{RRh} ; t_{RhR} , for a composition cardinality of up to 0.5 relations, the accumulation reaches 78%, compared to 66% reached at the same point for t_{RR} ; t_{RR} (Fig. 4.11a). Therefore, up through a composition cardinality of 4 relations, the accumulation of the t_{RRh} ; t_{RhR} is reached earlier than that of t_{RR} ; t_{RR} . For composition cardinalities of more than 0.5 normalized relations, the accumulation slows down for t_{RRh} ; t_{RhR} , but increases for t_{RR} ; t_{RR} , confirming that the hole in the common region affects the reasoning by decreasing ambiguity of the inferences. Similarly due to the higher frequencies of crisper results, the cumulative frequencies for t_{RhRh} ; t_{RhRh} increase faster up to 0.5 normalized relations, after which they slow down. At this point, however, the pace increases for t_{RhR} ; t_{RRh} (Fig. 4.11b).

Finally, examining the cumulative frequency graphs reveals that the accumulation pace is even faster when more holes are included in the pair of composed relations. For the crisper composition results, the accumulation pace in t_{RhRh} ; t_{RhRh} is the highest among all composition scenarios considered. For up to 50% of the composition cardinalities, the cumulative normalized frequency has already reached 95%, in contrast to 84% for t_{RhR} ; t_{RRh} , 78% for t_{RRh} ; t_{RhR} , and 66% for t_{RR} ; t_{RR} (Fig. 4.11). When compared with the composition of the same domain, approximately 60% of the accumulation has been reached already for a composition cardinality of up to 0.125 normalized relations, which translates to up to 19 relations per composition (Fig. 4.11b). The same percentage is reached only after the number of relations has more than doubled in the case of t_{RhR} ; t_{RRh} , where the cumulative frequency reaches 60% after the composition cardinality has reached 0.263 normalized relations (which translates to up to 40 relations per composition results). This faster pace continues until the accumulation is approximately 90%, after which the pace is smaller for t_{RhRh} ; t_{RhRh} . The additional constraints that the higher number of holes overall impose in this composition scenario (i.e., each pair of composed relations comprises two single-holed regions whereas all the rest of the compositions happen between relations that comprise at most one single-holed region) are responsible for crisper composition results.

4.6 Summary

This chapter completed the S-HRM by providing the definition of the set of 152 relations between two single-holed regions, their neighborhood graph, and the analysis of the relations' composition tables. While the number of relations increases to 152 when both regions in the relation are single-holed, overall the frequencies of composition results

with lower cardinalities are the highest for relations with two single-holed regions, than for relations with one or with no single-holed region.

The compositions that result in two single-holed region relations, t_{RRR} , were compared against compositions that result in hole-free region relations, t_{RR} , using not only the absolute, but also the cumulative cardinality frequencies. The frequency graphs that correspond to the composition tables revealed that inserting a hole in the common region of a pair of composed relations increases monotonically the cumulative cardinality frequencies, independently of the range of the resulting relations (i.e., t_{RR} or t_{RhRh}). Furthermore, lower cardinalities presented increases much higher than the average increase in the cumulative cardinality frequencies, while frequency saturation is achieved earlier for relations with a hole in the common region, than for relations with a hole-free common region. The composition cardinalities of relations with two single-holed regions experience the fastest accumulation pace, reaching 95% of the total accumulation for up to 76 relations (50% cardinality) in the composition results. The results verify the assumption that taking into consideration the holes in the reasoning process leads to more refined inferences, implying less ambiguous choices for the end user.

Chapter 5 Comparing Relations with A Multi-Holed Region

Frequently, topological scenarios involving holed regions extend beyond regions with a single hole. Relations between a hole-free and a single-holed region and between two single-holed regions have been examined in Chapters 3 and 4, respectively. Three types of relations involve multi-holed regions: (1) relations between a hole-free and a multi-holed region (Fig. 5.1a), (2) relations between a single-holed and a multi-holed region (Fig. 5.1b), and (3) relations between two multi-holed regions (Fig. 5.1c).

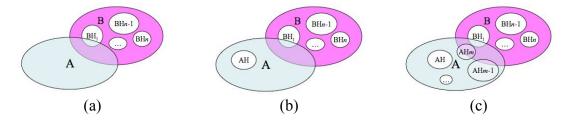


Figure 5.1 Topological relations between a multi-holed region B and a region A: (a) region A is hole-free, (b) A is single-holed, and (c) A is a multi-holed region, where both regions may have the same or different number of holes.

The relation model discussed in this chapter considers relations of the first type—between a hole-free and a multi-holed region. It is called the Multi-Holed Region Model (M-HRM) and it complements the S-HRM. The approach is to summarize for each relation, which of the 23 topological relations hold between the hole-free and a single-holed region (Egenhofer and Vasardani 2007), selecting one hole at a time. This model

also allows for relations between a hole-free and a multi-holed region to be compared for their topological similarity by applying a balanced algorithm that uses their summaries to transform one relation into the other. The cost associated with the transformation expresses a measure of the dissimilarity between the two relations, which is subsequently converted into their similarity value (Nedas 2006). Quantification of the relations' similarity provides a measure with which relations can be compared and ranked against a query, aiding users with their database quests for resembling topological scenarios. Similarity values are yet another inference mechanism made available using the sets of relations derived in this thesis.

In Section 5.1 the formal definitions of a multi-holed region and of the relation between a hole-free and a multi-holed region are given. Section 5.2 describes in detail the similarity evaluation method. First, the computation of the dissimilarity between relations is translated into evaluating the cost of transforming one relation into the other, followed by descriptions of how to compute the minimum cost for this transformation using the transportation algorithm and how to convert it into a similarity value. Section 5.3 provides an example of how to determine the topological similarity among relations between a hole-free and a multi-holed region. Section 5.4 studies the properties of the similarity assessment method and the results of applying it on a set of randomly generated relations for ranking them against a specific query, with the help of a software prototype. Section 5.5 concludes the chapter.

5.1 A Multi-Holed Region and the Relation between a Hole-Free and a Multi-Holed Region

A multi-holed region with n holes, denoted by n-B, is derived by subtracting from a hole-free region B, the union of the interiors of n simple regions Hi, i=1...n (Fig. 5.2). Region n-B comprises the generalized region B* and n holes such that B*contains each of the holes and all holes are pairwise *disjoint*. The generalized region B* is defined as the union of the multi-holed region n-B and the regions Hi, which fill its n holes.

Multi-holed region n-B with holes Hi, i=1.. n:

- n-B=B-ÜHi°
- $B^* = n B^{i=1} \cup H1 ... \cup Hn 1 \cup Hn$
- $\forall i: i=1...n, B^* contains Hi$
- $\forall i, j: i=1...n, j=1...n, i \neq j: Hi disjoint Hj$

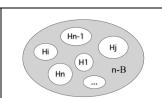


Figure 5.2 A multi-holed region n-B.

The topological relation between a hole-free region A and a multi-holed region n-B, denoted by t(A, n-B), is modeled as the union of the n relations between A and each of the single-holed regions $\mathcal{B}i$, i=1...n, as if each hole Hi was unique (Eqn. 5.1a). Because of this trait, the relation is named *multi-element* and each of the n *elements*, denoted by $t(A, \mathcal{B}i)$, is one of the 23- t_{RRh} between a hole-free and a single-holed region (Egenhofer and Vasardani 2007). Each single-holed region $\mathcal{B}i$ is equal to what remains after extracting region Hi's interior from the generalized region B* (Eqn. 5.1b). A depiction of a four-element topological relation between region A and a four-holed region 4-B, is given in Figure 5.3.

$$t(A, n-B) = \bigcup_{i=1}^{n} t(A, \mathcal{B}i)$$
, where $i=1...n$ and $t(A, \mathcal{B}i) \in 23-t_{RRh}$ (5.1a)

$$\forall i: i=1...n, \mathcal{B}i=B^*-Hi^\circ$$
 (5.1b)

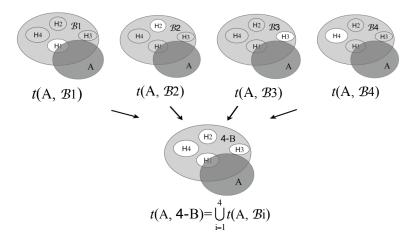


Figure 5.3 Depiction of the different $t(A, \mathcal{B}i)$ that together make up the four-element relation t(A, 4-B) between hole-free region A and multi-holed region 4-B.

5.2 Frequency Distribution Method

It follows from the multi-element relation model that each such relation is characterized by certain element frequencies, thus topologically different relations exhibit different frequencies. To compare multi-element relations for their topological similarity, the *Frequency Distribution Method* (FDM) first summarizes the relations in the frequencies of their participating elements. Subsequently, transformation of one multi-element relation into another is achieved using the concept of redistributing their summary frequencies. By assigning a cost to this transformation and translating the cost into a similarity value, FDM enables the assessment of relation similarity.

Two multi-element relations are used as reference for calculating the cost of relation transformation: Relation t(A, n-B) between hole-free region A and multi-holed region n-B and relation t(C, n-D) between hole-free region C and multi-holed region n-D, n being the number of holes. For short, the relations are denoted t_1 and t_2 , respectively.

5.2.1 Topological Dissimilarity and Relation Difference

Relations t_1 and t_2 may consist of different t_{RRh} s as their elements. The normalized frequencies of t_1 and t_2 's elements are summarized in vectors with m elements each, m being the collective number of different t_{RRh} s present in t_1 and t_2 (Eqn. 5.2). Vector V_1 records the normalized frequencies $Fi_{i=1}^{m}$ of t_1 , where Fi is the count (fi) of t_1 's elements of type $t_{RRh}i$, divided by the total number of holes n for normalization. Accordingly, vector V_2 records the normalized frequencies of t_2 . Normalization of the frequencies enables the comparison of the different frequency vectors, since each value becomes a percentage of the total number of different $t_{RRh}s$ in t_1 and t_2 . Elements with a zero frequency value in a vector correspond to $t_{RRh}s$ not present in the respective multi-element relation.

$$V = \begin{bmatrix} F1 \\ \vdots \\ Fm \end{bmatrix} \text{ where } \forall i = 1...m : Fi = \frac{fi}{n}$$
 (5.2)

The sum of a frequency vector V is defined as the sum of the frequency values F of its m elements. Since the frequency values in vectors V_1 and V_2 are normalized by the total number of holes, they are counted in frequency units and lie in the interval [0, 1]. A *frequency unit* is equal to the division of the unit by n (i.e., 1/n). Each frequency vector records all of the n elements of either t_1 or t_2 ; therefore, the sum of each vector is equal to the unit (Eqn. 5.3).

$$sum(V) = \sum_{i=1}^{m} F_i = 1$$
 (5.3)

If the vectors differ in frequency values for their corresponding elements, relations t_1 and t_2 are different. A quantification of this difference is their *topological dissimilarity*

 $\delta(t_1,t_2)$ (Def. 1). In order to determine this dissimilarity, one relation is transformed into the other, and the cost of this transformation is evaluated.

Definition 1: The *topological dissimilarity* $\delta(t_1,t_2)$ of two multi-element relations, t_1 and t_2 , is determined by the minimum cost of transforming frequency vector V_1 into vector V_2 by redistributing the normalized frequencies of V_1 so that they are identical to those of V_2 .

Matching vector frequencies for corresponding elements imply that the topology of the two relations is very similar. Nevertheless, t_1 and t_2 should not necessarily be regarded as exactly the same, unless their graphic depictions on the plane verify so.

The total cost of the frequency vector transformation is the weighted sum of the distances along the 23- t_{RRh} CNG between the elements, among which the redistribution of frequency units occurs. The number of frequency units moved between two t_{RRh} s represents the weight assigned to their distance on the 23- t_{RRh} CNG (Table 3.2). The maximum cost is incurred when the maximum possible amount of frequency units is redistributed over the longest shortest path on the CNG. The maximum amount of frequency units of vector V is the sum(V), which is equal to the unit (Eqn. 5.3), and the longest shortest path of the 23- t_{RRh} CNG, d_{max23} , is eight (Section 3.5.2). Subsequently, the *maximum cost of transformation* $\delta_{max}^{t_1-t_2}$ between t_1 and t_2 , is eight (Eqn. 5.4).

$$\delta_{\max}^{t_1 - t_2} = \text{sum}(V) * d_{\max 23} = 8$$
 (5.4)

In order to identify the amount of frequency units and the elements between which they are redistributed, the *relation difference* is calculated.

Definition 2: The *relation difference* ($\Delta_{t_1t_2}$) between two relations, t_1 and t_2 , is a frequency vector, defined as the difference of the relations' normalized frequency vectors (Eqn. 5.5).

$$\Delta_{t_1,t_2} = V_1 - V_2 \tag{5.5}$$

The values of the elements in $\Delta_{t_1t_2}$ are the frequency units that need to be redistributed. These values are both positive and negative, with the positive corresponding to excess frequency units that will be transferred to balance the negative ones. Consequently, $\Delta_{t_1t_2}$ carries the weight information necessary for computing the minimum cost of transforming V_1 into V_2 , in the form of the exchanged frequency units. The elements of $\Delta_{t_1t_2}$ represent the difference units between the two relations' frequency vectors. The vectors are normalized by the same number of holes n, therefore, the sum of the positive values in $\Delta_{t_1t_2}$ is equal to the sum of the negative values, canceling each other out (Eqn. 5.6).

$$\operatorname{sum}(\Delta_{t_1 t_2}) = 0 \tag{5.6}$$

5.2.2 Computing the Minimum Cost of Relation Transformation with the Transportation Algorithm

The determination of the minimum cost for transforming frequency vector V_1 into V_2 can be translated into a balanced transportation problem (Murty 1976; Strayer 1989). This section discusses how this translation is done, and the use of the transportation algorithm for acquiring a solution. The transportation algorithm, as well as the Hungarian (Kuhn 1955) and other algorithms are used to solve the assignment and the balanced transportation problems (Munkres 1957), which are cases of the minimum cost flow

problem, which in turn is a special case of the linear programming problem (Dantzig 1963; Dantzig and Thapa 1997).

5.2.2.1 Balanced Transportation Problem

An analogy of *warehouses* and *markets* is used for discussing the transportation problem. This problem, which considers the *supplies* of all the warehouses and the *demands* of all the markets, along with the *unit costs* for transportation between all pairs of warehouses and markets, is about finding the minimum cost for transporting all supplies from the warehouses to the markets so that all demands can be met. This analogy's components have their counterparts in the case of transforming one frequency vector into another. In particular, the *p* different elements with positive frequency differences in the relation-difference $\Delta_{t_1t_2}$ correspond to warehouses and the *n* different elements with negative frequency differences correspond to markets. The *i*th warehouse's (W_i) supply, denoted by s_i , is equal to the value of its corresponding frequency difference in $\Delta_{t_1t_2}$. Similarly, the *j*th market's (M_j) demand, denoted by d_j , equals the value of the analogous frequency difference in $\Delta_{t_1t_2}$.

The unit cost c_{ij} for moving a supply unit (i.e., a frequency unit) from W_i to M_j is the distance $dist(W_i, M_j)$ between the two t_{RRh} that correspond to W_i and M_j on the 23- t_{RRh} CNG (Fig. 3.9). The sum of the relation-difference is zero (Eqn. 5.6); therefore, the sum of all supplies equals the sum of all demands (Table 5.3) and, due to this equality, this transportation problem is a *balanced* transportation problem.

Table 5.3 The balanced transportation problem for $\Delta_{t_1t_2}$, with p warehouses ($W_{i=1}^p$) and p supplies ($S_{i=1}^p$), n markets ($M_{j=1}^n$) and n demands ($d_{j=1}^n$), and $p \times n$ unit costs ($C_{i,j=1}^{p,n}$).

If x_{ij} frequency units are to be transferred from warehouse W_i to market M_j , then the transportation problem is expressed as determining those values of x for which the total cost z is a minimum (Eqn. 5.7) and the *supply* and *demand constraints* are satisfied. The supply constraint states that the sum of all frequency units distributed out from a certain warehouse must equal the total supply capacity of that warehouse (Eqn. 5.8a), and similarly, the demand constraint states that the sum of all frequency units received by that market must equal the total demand of that market (Eqn. 5.8b). Cost z is then the minimum cost for transforming frequency vector V_1 into vector V_2 .

$$Z = \min(\sum_{i=1}^{p} \sum_{j=1}^{n} c_{ij} X_{ij})$$
 (5.7)

$$\forall i: \quad \sum_{j=1}^{n} \mathbf{x}_{ij} = \mathbf{s}_{i} \tag{5.8a}$$

$$\forall j \colon \sum_{i=1}^{p} \mathbf{x}_{ij} = \mathbf{d}_{j}$$
 (5.8b)

5.2.2.2 Similarity Values Obtained from the Transportation Algorithm

Solutions to the balanced transportation problem of identifying the topological dissimilarity between two multi-element relations or the minimum cost for transforming one relation into the other can be obtained by employing the *transportation algorithm* (Murty 1976; Strayer 1989). This algorithm operates in two steps: First a basic feasible

solution is found by randomly redistributing the frequency units in $\Delta_{t_1t_2}$ so that the supply and demand constraints are met (Eqn. 5.8a-b). Since the basic solution might not be optimal as far as the minimum cost z of the transportation is concerned, the algorithm, in a second step, iteratively improves the redistribution, until an optimal solution is found (Strayer 1989). There may be more than one way to distribute the frequency units, so that the minimum value of z is obtained. The higher the final solution, the more costly is the redistribution of frequency units.

The computed value of z actually stands for the topological dissimilarity $\delta(t_1,t_2)$ between two multi-element relations. The similarity $\sin(t_1,t_2)$ and dissimilarity $\delta(t_1,t_2)$ of two relations are complement values; therefore, they add up to a constant (i.e., g) (Eqn. 5.9). Dissimilarity $\delta(t_1,t_2)$ lies in the interval $[0, \delta_{\max}^{t_1-t_2}]$, and can be normalized in the closed interval [0, 1] if divided by $\delta_{\max}^{t_1-t_2}$ (Eqn. 5.10a). In this case, similarity is the dissimilarity's complement from the unit, and it also lies in the closed interval [0, 1] (Eqn. 5.10b).

$$sim(t_1, t_2) + \delta(t_1, t_2) = g \tag{5.9}$$

$$\delta(t_1, t_2) = \frac{Z(t_1, t_2)}{d_{\text{max}}^{t_1 - t_2}}$$
 (5.10a)

$$sim(t_1, t_2) = 1 - \delta(t_1, t_2)$$
(5.10b)

5.3 Example of Relation Similarity Evaluation

The similarity between relations t(A, 8-C) and t(B, 8-D) (Fig. 5.4), or t_1 and t_2 for short, is evaluated using the FDM. Since, in this case, *overlap* is the principal relation for both cases, five different t_{RRh} s are possible (Fig. 3.9). Table 5.4 displays these five t_{RRh} along

with the single-holed regions that participate in each t_{RRh} with A or B ($C_{i=1}^8$ and $\mathcal{D}_{i=1}^8$ respectively), according to the placements of the holes.

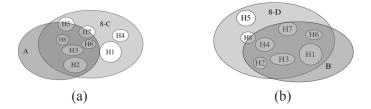


Figure 5.4 Example topological multi-element relations: (a) t(A, 8-C) and (b) t(B, 8-D).

Table 5.4 The t_{RRh} elements of the multi-holed relations t(A, 8-C) and t(B, 8-D).

	[<i>o d</i>]	[o m]	$[o\ o]$	[o cv]	[<i>o ct</i>]
A	C4	<i>C</i> 1	C7 + C5	-	C2 + C3 + C6 + C8
В	\mathcal{D} 5	-	$\mathcal{D}8$	\mathcal{D} 7	$\mathcal{D}1 + \mathcal{D}2 + \mathcal{D}3 + \mathcal{D}4 + \mathcal{D}6$

Table 5.4 reveals also the elements' frequencies for each of the multi-element relations. It is, therefore, possible to construct the normalized frequency vectors V_1 and V_2 for each multi-element relation, respectively (Eqn. 5.11). Since there are eight holes, normalization is achieved by dividing each frequency by eight (i.e., f/8, where f is the frequency of the element) and the frequency unit is 1/8.

$$V_{1} = \begin{bmatrix} [o \ d] & 0.125 \\ [o \ m] & 0.125 \\ [o \ o] & 0.25 \\ [o \ cv] & 0 \\ [o \ ct] & 0.5 \end{bmatrix} \quad \text{and} \quad V_{1} = \begin{bmatrix} [o \ d] & 0.125 \\ [o \ m] & 0 \\ [o \ o] & 0.125 \\ [o \ cv] & 0.125 \\ [o \ cv] & 0.125 \\ [o \ ct] & 0.625 \end{bmatrix}$$
 (5.11)

Vectors V_1 and V_2 do not share the same frequencies for corresponding types of elements; therefore, the transportation algorithm is employed in order to produce a dissimilarity value between them. Equation 5.12 gives their relation difference $\Delta_{t_1t_2}$.

$$\Delta_{t_1 t_2} = \begin{bmatrix}
[o \ d] & 0 \\
[o \ m] & -0.125 \\
[o \ o] & -0.125 \\
[o \ cv] & 0.125 \\
[o \ ct] & 0.125
\end{bmatrix}$$
(5.12)

By enforcing the sums constraint (Table 5.3), frequency units need to be redistributed so that balance is acquired. The relation difference shows which elements with excess (positive) frequency differences will offer units— $[o\ cv]$ and $[o\ ct]$ in this case—to the receiving elements with the negative frequency differences— $[o\ m]$ and $[o\ d]$. The distances between the elements on the 23- t_{RRh} CNG (Fig. 3.9) provide the costs of the transportations (Table 3.2). A basic feasible solution—which is also optimal for this simple case—is obtained if 0.125 units are transferred from $[o\ cv]$ to $[o\ m]$ and 0.125 units from $[o\ ct]$ to $[o\ o]$. Table 3.2 records the costs for these transfers as 2 and 2, respectively. Therefore, the value z for the cost of redistribution in $\Delta_{t_1t_2}$ is 0.5 (Eqn. 5.13).

$$Z = 0.125*dist([o\ cv] \rightarrow [o\ m]) + 0.125*dist([o\ ct] \rightarrow [o\ o]) = 0.5$$
 (5.13)

This value z is the cost of transforming V_1 into V_2 (or vise versa). Using Equation 5.10a, the dissimilarity $\delta(t_1,t_2)$ between the two multi-element relations is evaluated (Eqn. 5.14); subsequently, using Equation 5.10b, the similarity $\sin(t_1,t_2)$ between the two relations is estimated (Eqn. 5.15).

$$\delta(t_1, t_2) = \frac{z}{d_{\text{max}}^{t_1 - t_2}} = \frac{0.5}{8} = 0.0625$$
 (5.14)

$$sim(t_1, t_2) = 1 - \delta(t_1, t_2) = 1 - 0.0625 = 0.9375$$
(5.15)

Converted to a percentage, this similarity value is 93.75%. However, FDM is for comparing relations for their similarity to another relation. Similarity evaluation between two relations implies that one of them is the reference relation. To enable similarity comparisons, the similarity between the reference and at least one more relation has to be calculated (Janowicz 2006). The similarity computed with FDM does not imply that two relations are as similar, percentage wise, as the calculated number indicates. Occasionally, multiple different distributions yield the same minimal result. For more

complex frequency units distributions, the second phase of the transportation algorithm performs iteratively, starting from a basic feasible solution until an optimal distribution is found.

5.4 Properties of the Frequency Distribution Method

The FDM has certain characteristics:

- The *number of holes* n is the same between query and random relations that are ranked for their similarity against the query.
- There are three possible ways in which two multi-element relations—between a hole-free and a multi-holed region—may differ: (1) the principal relation is the same, while the refining relations—those associated with the placement of the holes—are different, (2) the principal relation differs but the placement of the holes remains invariable, and (3) the principal relation and some or all of the refining relations differ.
- All edges in the 23- t_{RRh} CNG have the *same weight* (i.e., the unit), so that the distance between any two relations on the graph solely depends on the length of the shortest path between them.

This section examines in detail the three ways that multi-element relations can differ, as well as how the weight constant affects the similarity evaluation procedure.

5.4.1 Effect of the Principal Relation on the Similarity Evaluation

There are five relations—disjoint, meet, equal, covers, and contains—whose strong influence on the topology forces the hole-free region to be in the same relation with all holes—disjoint for principal relations disjoint and meet, and contains for the remaining

three principal relations, in the given order (Fig. 5.5). Other multi-element relations may differ only in their principal relations, or both their principal and refining relations. They cannot, nonetheless, differ only in their refining relations. As this information is imprinted in the 23- t_{RRh} CNG, in case two multi-element relations share any of these five principal relations, their possible topological distinguishability is not recognized in the FDM and no further calculations are performed.

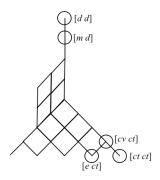


Figure 5.5 The five multi-elements indistinguishable relations.

For the remaining three principal relations—overlap, coveredBy, and inside—there are no restrictions; therefore, multi-element relations may differ in any of their constituent relations, or in both simultaneously. There are, however, a few exceptions for the principal relation inside. While having eight possibilities for different hole placements, inside shares only five common refining relations with principal relations coveredBy and overlap—namely disjoint, meet, overlap, covers, and contains. When the principal relation is inside and the refining relations are equal, coveredBy, or inside, multi-element relations can vary only in their principal relation or in both their constituent relations, but not solely in their refining relations. Figure 5.6 depicts various paths on the 23-t_{RRh} CNG between two different elements with overlap, coveredBy, or inside for their principal relation. These paths connect relations that have the same principal relation and different hole placements (Fig. 5.6a), or the same hole placements

and different principal relations (Fig. 5.6b), or both principal and refining relations different (Fig. 5.6c).

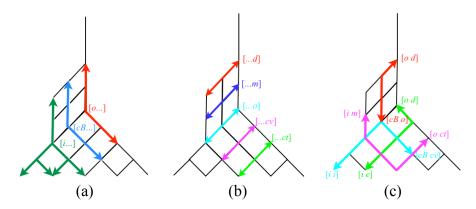


Figure 5.6 Paths connecting different elements. The elements have (a) the same principal, but different refining relations, (b) different principal, but the same refining relations, or (c) different principal and refining relations.

5.4.2 Ramifications of the Weight Constant

The weight constant—the fact that all edges in the 23- t_{RRh} CNG have the same weight—is related to the dependency of the method on the 23- t_{RRh} topological relations. In order to evaluate the ramifications of this constant on the FDM, the ranking of a few especially chosen example relations against a specific query is examined. The choice of the unit as the weight for all the edges of the 23- t_{RRh} CNG equalizes the amount of change from one principal relation to a neighboring one while keeping the refining relation the same, with the amount of change of keeping the same principal relation and moving to a neighboring refining relation (Fig. 5.7). It implies that a topological change in the relation between the hole-free region and the generalized region has the same topological importance as a topological change in the relation between the hole-free region and any one of the holes.

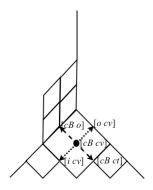


Figure 5.7 Moving from relation [*cB cv*] to any of its four neighboring relations causes the same amount of topological change.

The importance of the invariant distance unit weight is highlighted in the following examples. Figure 5.8a shows the query topological scene against which four other scenes will be evaluated for their degree of similarity. They are chosen so that the first one has the same principal but different refining relations (Fig. 5.8b), the second and third have different principal but the same refining relations (Fig. 5.8c-d), and the fourth one has different principal and some but not all different refining relations (Fig. 5.8e).

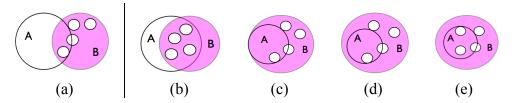


Figure 5.8 Similarity evaluation of multi-element relations: (a) the query relation and (b)–(e) the relations to be assessed according to their similarity with the query.

For the first relation (Fig. 5.8a), using the FDM to calculate the cost associated with transforming it to the query, four frequency units need to be transferred from the $[o\ ct]$ elements, to each of the query's elements. The normalized frequency unit is 1/4—there are four holes in total—and the distances over which the units need to be transferred are $dist([o\ ct] \rightarrow [o\ d])=4$, $dist([o\ ct] \rightarrow [o\ m])=3$, $dist([o\ ct] \rightarrow [o\ o])=2$ and $dist([o\ ct] \rightarrow [o\ cv)=1$, as indicated from Table 3.2. The similarity sim(b, query) is

calculated to be 68.75%. For the second relation (Fig. 5.8c), four frequency units need to be transferred again, but this time the principal relation changes, while the holes' placement remains the same. Applying FDM gives sim(c, query) = 87.5%.

The third relation (Fig. 5.8d) has different principal but the same refining relations again, only this time the elements are farther apart from the query's respective relations on the CNG. The analogous calculations give sim(d, query)=75%. For the fourth relation (Fig. 5.8e), the principal and some, but not all, refining relations are different. The redistribution of frequency units gives sim(e, query) = 62.5%.

While relation 5.8b exposes no change from the query with respect to the principal relation, it does not rank the highest in the similarity assessment. Instead relations 5.8c and 5.8d, with different principal relations but with the same topological placement of their holes, rank higher. Last in ranking comes 5.8e, with both principal and refining relations different from the query's corresponding relations. Figure 5.9 orders the relations according to their ranking.

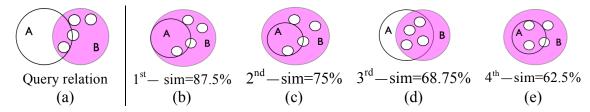


Figure 5.9 The ranking of multi-element relations according to their similarity with the query relation. (a) the query relation and (b)–(e) the ranked relations.

The weight constant is responsible for the specific ranking of the results. The 23- t_{RRh} CNG captures all topological changes among the 23 possible elements of a multi-element relation. The path between two relations on the graph represents a path of least topological difference, implying that the closer the relations are on the graph, the smaller is the amount of change required in order to gradually transform one relation to the other

(Bruns and Egenhofer 1996). Topological changes of the 23- t_{RRh} , however, have the characteristic of comprising two, not totally independent, variables: (1) change in the principal and (2) change in the refining relation. Neighboring relations on the graph may differ by a step of change in either of their constituent relations. Therefore, the placement of the elements on the 23- t_{RRh} CNG dictates the similarity ranking. Closeness between t_{RRh} s on the graph has a strong bearing on a multi-element relation's similarity to the query relation, even if their principal relations are different. In a different similarity model, where the two variables are treated separately by using a different principle than least topological change, the lengths between neighbors on the CNG would vary. In such a case, differences in their principal relation and differences in their refining relations would play distinct roles in the similarity assessment between relations over holed regions. In the model presented here, however, topological changes in both constituent relations are of the same importance.

5.4.3 Analysis of Random Query Experiments

With the help of a software prototype for randomly generating multi-element topological relations, experiments were conducted for ranking a set of ten such random relations (Fig. 5.10b-k) according to their similarity with a query multi-element relation (Fig. 5.10a), using FDM. The results of the similarity evaluation are given in Table 5.5.

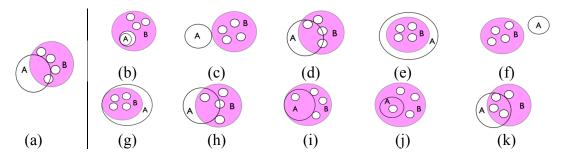


Figure 5.10 Randomly generated archived topological scenarios: (a) the query relations and (b)–(k) the relations to be ranked according to their similarity to the query.

Table 5.5 Similarity evaluation for the ten randomly generated relations (Fig. 5.10b-k) with respect to the query relation (Fig. 5.10a).

The similarity assessment determines that the relation topologically closest to the query is relation 5.10d with sim(d) = 90.6%, followed by relation 5.10h with sim(h) = 87.5%. In this case, the two most highly ranked scenarios not only share the same principal relation, but also display hole placements that resemble the topology of the query. Therefore, the elements of each relation appear very close to the query's corresponding elements on the 23- t_{RRh} CNG, and are responsible for the high similarity ranking. The scenarios that follow in the ranking present a combination of different principal and also quite different refining relations from the analogous relations of the query. The last rank is occupied by case 5.10e that has elements located the farthest from the corresponding query elements, on the 23- t_{RRh} CNG. Specifically, [ct ct] is the farthest from elements [o m] and [o o] of the query, with distances four and five, respectively. Such distances are the longest distances that both elements can have from any other relation on the graph. Therefore, their placement on the 23- t_{RRh} CNG is responsible for placing 5.10e to the lowest ranking.

5.5 Summary

The Frequency Distribution Method (FDM) enables the similarity comparison among relations between a hole-free and a multi-holed region, when the number of holes is the same. The relations under comparison are summarized in vectors that display the frequency of the relations' elements—the t_{RRh} s identified in each relation, when each hole is considered as unique in the host region. Modeling relations in frequency vectors allows

the similarity evaluation to be described as calculating the cost for transforming one relation into the other in the form of a balanced transportation problem, where the differences in the frequencies of their corresponding elements are considered. Positive frequency elements transfer frequency units to elements with frequency deficiency, to acquire equilibrium.

By calculating the overall cost of these transfers for the relation transformation, using the transportation algorithm, a dissimilarity between relations is evaluated. Subsequently, dissimilarity is expressed in its complement similarity value. The costs are identified as distances between two elements on the 23- t_{RRh} CNG. The FDM addresses this way, questions of similarity ranking of a group of relations against a specific query. Example similarity rankings reveal that the positioning of the holes relative to the hole-free region is as important for the similarity evaluation as is the overall relation between the hole-free region and the host of the holes.

Chapter 6 Similarity Assessment with Varying Numbers of Holes

In Chapter 5 the comparisons among relations between a hole-free and a multi-holed region were restricted to the cases when the multi-holed regions have the same number of holes, as this requirement is one of the basic properties of the Frequency Distribution Method (FDM). In many real-world scenarios, however, the need to compare such relations extends to comparing relations with regions that have different numbers of holes (Fig. 6.1).

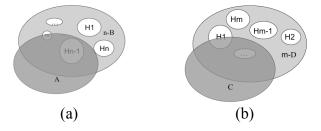


Figure 6.1 Relations between a hole-free and a multi-holed region with different numbers of holes: (a) relation t(A, n-B) and (b) relation t(C, m-D).

Apart from the different number of holes, such relations may also vary in the way that the hosts and the hole-free regions are related, in the way the holes are placed with respect to the hole-free region, or in all three ways. Assuming that the holes are all of the same importance and that their placements, as well as the host region's placement with

respect to the hole-free region are the same for the different relations, the difference in the number of holes can be interpreted as comparing scenarios where holes have been dropped from or added to the multi-holed region. If a number of parameters are different, however, the question is which of the topological scenarios under examination are more similar to the query relation and how do the number of holes, the principal relation, and the hole positioning affect the similarity ranking. This chapter examines whether FDM, unaltered, can handle these questions to compare relations with different numbers of holes and, if not, what adjustments are needed to allow FDM to deal with such cases.

Figure 6.2 displays a sketched geological scenario where a comparison among relations between a hole-free and a multi-holed region, with regions of different numbers of holes is needed. Even though geologic formations exist in the three dimensions, geologists often study and make decisions over the two-dimensional representations of the formations on geological maps. Such representations are used in this example. Region B represents an underground sedimentary rock formation with concentrations of oil residing in the holes H1 through H8. Regions D and F are similar rock formations with holes H1 through H6 and H1 through H7, respectively, filled with oil as well. Regions A, C, and E are ground hole-free regions that have been deemed appropriate for excavation in order to reach the oil deposits from the reservoirs in D, B, and F, correspondingly. All holes are of equal profit importance and the topological relation between excavation region A and reservoir B, for which the profit from the oil exploitation is known, is considered the reference relation. Which one of the remaining two topological relations is then more similar topologically to the reference relation and, therefore, has a similar economic value? Is it the relation between C and D, where D has the same hole placements as B but is missing two holes, or is it the relation between E and reservoir F,

which presents different hole placements than B with respect to the excavation region, but is only missing one hole?

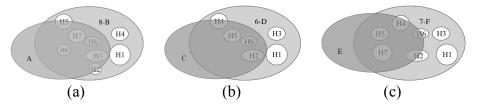


Figure 6.2 Sketched geological scenarios of multi-holed oil baring reservoirs. The reservoirs are represented by regions 8-B, 6-D, and 7-F, while the hole-free regions A, C, and E have been identified as excavation areas.

When the number of holes between the regions in the relations under comparison is different, the supply and demand constraints (Section 5.2.2.1) are no longer in effect, therefore the transportation problem gets out of balance and the transportation algorithm cannot be applied. As a result, FDM is inappropriate for addressing such questions in a reasoned matter. To allow for comparisons of relations where the regions have different number of holes, this chapter modifies FDM, and in particular, the 23- t_{RRh} conceptual neighborhood graph to account for the costs related with the change in the number of holes between the multi-holed regions in the relations. The extension of the FDM implies that future alterations can also accommodate comparisons among relations between regions that are both multi-holed.

Section 6.1 points out the limitations of the FDM for dealing with relations of multi-holed regions with varying numbers of holes. The required changes result in an extended version of the 23- t_{RRh} CNG (Section 6.2). Section 6.3 examines different cost assigning methods for producing the table of costs for the extended CNG and Section 6.4 presents the alternative FDM and demonstrates its use with an example. The chapter is summarized in Section 6.5.

6.1 Limitations of the FDM

The Frequency Distribution Method relies on satisfying the supply and demand constraints in order to maintain the transportation problem in balance, ensuring that the transportation algorithm can be applied. When the number of holes is the same in the regions of the relations under comparison, it suffices to redistribute the element frequencies in order to transform one relation into the other and achieve balance. Therefore, when the number of holes is different, the surplus or deficit in element frequencies cannot be met by mere frequency units' redistribution, and the system gets out of balance (Section 6.1.1). A solution for rebalancing the system comes from studying the connection between the 8- t_{RR} and the 23- t_{RRh} CNGs (Section 6.1.2).

6.1.1 Out-of-Balance System

An example featuring two multi-element topological relations t(A, 5-B) and t(C, 4-D) (Fig. 6.3) is used as reference for discussing how the difference in the number of holes brings the system out of balance, rendering the unaltered FDM insufficient for evaluating the similarity in such cases. For simplicity, the principal relation and the placement of the holes is the same for the two relations; however, this is not a general precondition for the alternative FDM studied in this chapter.

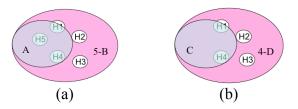


Figure 6.3 Example multi-element relations with different numbers of holes: (a) relation between hole-free region A and region B with five holes and (b) relation between hole-free region C and region D with four hole. Both multi-element relations share the principal relation *coveredBy* and the same refining relations.

When applying the FDM for evaluating the similarity, the first step is to record the frequencies of the t_{RRh} s that compose the relations, in the frequency vectors V_1 and V_2 for t(A, 5-B) and t(C, 4-D), respectively (Eqn. 6.1a). The maximum number of holes in one region is 5, thus one frequency unit equals 1/5 (Section 5.3.1). After recording the normalized frequencies, it is also possible to calculate the relation difference Δ , which carries the weights information necessary for computing the minimum cost of transforming V_1 into V_2 in the form of the exchanged frequency units (Eqn. 6.1b).

$$V_{1} = \begin{bmatrix} [cB \ d] & 0.2 \\ [cB \ m] & 0.2 \\ [cB \ o] & 0.2 \\ [cB \ cv] & 0.2 \\ [cB \ ct] & 0.2 \end{bmatrix} \text{ and } V_{2} = \begin{bmatrix} [cB \ d] & 0.2 \\ [cB \ m] & 0.2 \\ [cB \ o] & 0.2 \\ [cB \ cv] & 0.2 \\ [cB \ cv] & 0.2 \\ [cB \ ct] & 0 \end{bmatrix}$$

$$(6.1a)$$

$$\Delta = V_1 - V_2 = \begin{bmatrix} [cB \ d] & 0 \\ [cB \ m] & 0 \\ [cB \ o] & 0 \\ [cB \ cv] & 0 \\ [cB \ ct] & 0.2 \end{bmatrix}$$
(6.1b)

The sum of the positive frequency differences is no longer equal to the sum of the negative frequency differences, however, as required by the nil sum of the relation difference (Eqn. 5.6). The inequality is explained by the extra hole in 5-B (i.e., hole H5), which has been dropped in 4-D. This extra hole's presence creates an excess of frequency units and is responsible for throwing the system out of balance. The transportation algorithm cannot be applied, because there is no indication as to where the excess frequency unit should be transferred, as there is no demand in the form of a negative element of equal absolute value in the relation difference (Eqn. 6.1b).

6.1.2 Rebalancing the System

In order to bring the system back into balance, a way to create the appropriate demand to counterbalance the excess positive frequency units in the relation difference is necessary. The frequency vectors record the frequencies of the t_{RRh} s and summarize the relations, but the vectors' numerical difference (Δ) does not allow for the change in the number of holes to affect the similarity evaluation. This change, therefore, needs to be represented in Δ as well. The current section discusses the concept of adding a new quantity in Δ , of absolute value equal to the excess created by reducing the number of holes. This new quantity allows for the difference in the number of holes to affect the similarity evaluation, rebalances the system by bringing the sum of Δ back to zero, and allows for the transportation algorithm to perform.

The approach described in this chapter is based on the following basic assumption. Each value in Δ represents the weight (i.e., the frequency units) assigned to the cost of transforming a t_{RRh} into another. The cost is the distance on the 23- t_{RRh} CNG between the two t_{RRh} s that transform into each other—the shortest path between them. Therefore, the additional weight quantity that rebalances Δ is related to a distance between two relations that transform into each other as well. In particular, it is related to the distance that connects the relations over the region, which drops or gains the hole. From now on, since adding and dropping a hole are inverse operations of equal cost, only the dropping of a hole will be considered, without loss of generality.

It is necessary then that new distances, which represent costs associated with dropping holes from the t_{RRh} s and to which the new weight elements in Δ will be assigned, are appended to the neighborhood graph. Each time a hole is dropped, it is dropped from one of the single-holed regions that collectively make up for the multi-

holed region, as dictated from the HFM (Chapter 5). The single-holed region becomes then hole-free, and the t_{RRh} to which it was involved (i.e., the relation between the hole-free and the single-holed region that drops the hole) becomes one of the 8- t_{RR} . Specifically, the former t_{RRh} is replaced by its t_{RR} principal relation (Fig. 6.4).

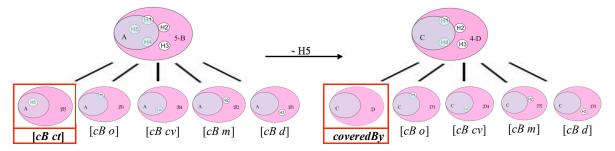


Figure 6.4 Example of transforming a multi-element relation into another, when the number of holes between them is different. Single-holed region $\mathcal{D}5$ becomes hole free region D—marked with a dashed border—as it loses hole H5. The t_{RRh} [cB ct] between C and $\mathcal{D}5$ becomes the t_{RR} coveredBy between hole-free regions C and D.

6.2 Bridging Different Topological Changes

The new distances added to the CNG connect relations of different types—a t_{RRh} with a t_{RR} —implying that the 8- t_{RR} CNG is appended to the 23- t_{RRh} CNG, with additional edges that connect each t_{RRh} with its principal t_{RR} . Each of these two graphs is a depiction of steps of topological change between relations of two-dimensional regions—both hole-free for the 8- t_{RR} and a hole-free and a single-holed for the 23- t_{RRh} . Before studying the changes along the connections of the two graphs, section 6.2.1 examines the topological changes portrayed on each separate graph.

6.2.1 Changes in the Relation between Regions

The reasoning behind constructing the 8- t_{RR} CNG is that the edges correspond to gradual changes between the relations on the graph. Particularly, each edge is a step of

topological change between the two relations it connects. It is the necessary topological change that transforms one topological relation into a different one, with no intermediate cases (Egenhofer and Al Taha 1992). The 8- t_{RR} CNG considered here is the one that depicts the A-neighbors, where none of the three topological changes of scaling, translation, or rotation may transform relation *equal* to relations *overlap*, *inside*, or *contains*, directly. The transformation needs to go through relations *coveredBy* and *covers*, respectively. A discussion of the B-neighbors or the C-neighbors graph where such direct transformations are possible is included in the Future Work section of the thesis (Section 8.3). The change occurs on one of the two objects of the relation and it alters the relation to one of its immediate neighbors on the graph. The topological structure of the object (i.e., region) itself, however, remains intact. The typical topological changes considered for the 8- t_{RR} are *translation* of the object in any direction in the plane, *scaling* (expansion or reduction), or *rotation*. Figure 6.5a depicts the 8- t_{RR} CNG with labeled edges according to the possible change of one region in relation to the other.

When a hole is added to one of the two regions, each of the 8- t_{RR} on the graph is replaced by its set of refining t_{RRh} s, resulting in the 23- t_{RRh} CNG. The refinements are due to the additional possibilities for relations created by the different placements of the hole relative to the hole-free region. The reasoning for connecting the relations on the 23- t_{RRh} CNG remains the same—each edge represents a step of topological change between the relations it connects (Section 5.2). The change occurs on one of the two regions of the relation, altering either the principal or the refining relation. Consequently, the edges on the 23- t_{RRh} CNG connect t_{RRh} s that either share the same principal relation and their refining relations are immediate neighbors on the 8- t_{RR} CNG, or they share the same

refining relations and their principal relations are immediate neighbors on the 8- t_{RR} CNG. The changes considered are, similarly, translation, scaling and rotation. When the principal relation remains invariant, the change happens to either the hole-free region or the hole, while when the refining relation remains the same the change occurs to either the hole-free region or the host of the hole. Figure 6.5b depicts the 23- t_{RRh} CNG with labeled edges according to the possible change to the hole, the generalized, or the hole-free region.

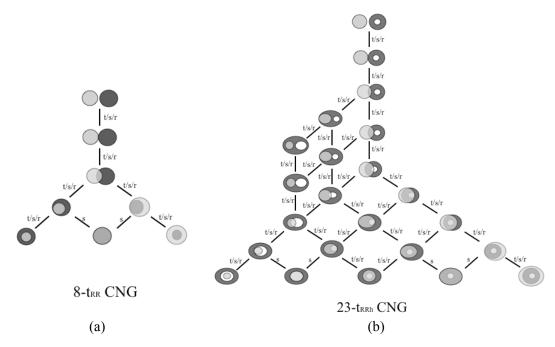


Figure 6.5 The Conceptual Neighborhood Graphs for (a) 8- t_{RR} and (b) 23- t_{RRh} with edge labels for possible topological changes from one relation to the other: t=translation, s=scaling, r=rotation

The three topological changes—translation, scaling and rotation—that may occur to a region are of equal importance and the precondition in this work is that only one topological change occurs between two neighboring relations. Examination of the relations in both graphs shows that in most cases one relation can be transformed to its neighbor by translation, scaling, or rotation, equally. There are exceptions, however,

when either the relation itself (for the t_{RR} s) or one of its constituent relations (for the t_{RR} hs) is the relation *equal*. For the 8- t_{RR} CNG, the only topological change for the pairs of relations *coveredBy-equal* and *covers-equal* is scaling. In either case, one of the two regions has to increase or decrease from *coveredBy* or *covers* to *equal*, respectively. The reverse is necessary in the opposite direction (Fig. 6.5a). The change, nonetheless, cannot be achieved by translation of one region.

Due to the same reasoning, for the 23- t_{RRh} CNG the only topological change for the pairs of relations [i cB]-[i e] and [i e]-[i cv] is scaling. The hole-free region, for example, has to scale up in order to change from coveredBy to being equal to the hole, or scale down in the reverse direction. Similarly, the hole-free region scales up when changing from equal to covers the hole and scales down in the reverse direction (Fig. 6.5b). Scaling can only hold for the pairs [cB e]-[e ct] and [e ct]-[ev e] as well. The difference this time is that the hole-free region has to scale up or down to turn into or turn out of equal with the generalized region, instead of the hole (Fig. 6.5b).

6.2.2 Changes in the Topological Structure of a Region Between Relations

Starting with two hole-free regions, the formation of a hole in one of the regions transforms the initial t_{RR} into one of its refining t_{RRh} s. It is, therefore, possible to align the 8- t_{RR} CNG with the 23- t_{RRh} CNG by adding edges between them (Fig. 6.6). During transformations between relations on the two separate levels—the 8- t_{RR} and the 23- t_{RRh} CNG—of the complete graph, the topology of the regions themselves remains unaltered; however, the same is not true for the changes occurring along the edges that align the two levels. Such edges connect relations between regions whose topological structure changes from a modification—addition or elimination—of a hole. These edges, therefore, carry

different semantics from the ones occurring on the separate graphs. They do not indicate topological transformations (translation, scaling or rotation) anymore. Rather, they represent a change in the topological structure of the region that drops or gains a hole, while the principal topological relation with the hole-free region remains the same.

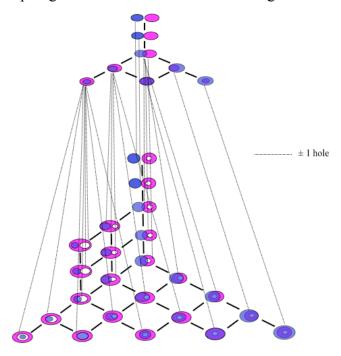


Figure 6.6 The alignment of the 8- t_{RR} with the 23- t_{RRh} CNG. Solid lines indicate a topological change in the relation of the regions, while the dotted lines indicate a change in the topological structure of one of the regions.

6.2.3 The Hybrid Conceptual Neighborhood Graph

The resulting graph is an extended, hybrid 8- t_{RRh} and 23- t_{RRh} CNG built by connecting the two separate graphs (Fig. 6.6). The additional edges link the 8- t_{RR} with the sets of their refining relations, yielding 23 new edges: one between *disjoint* and { $[d\ d]$ }, one between *meet* and { $[m\ d]$ }, five between *overlap* and each of { $[o\ d]$, $[o\ m]$, $[o\ o]$, $[o\ cv]$, $[o\ ct]$ }, one between *covers* and { $[cv\ ct]$ }, one between *equal* and { $[e\ ct]$ }, five between *coveredBy* and each of { $[cB\ d]$, $[cB\ m]$, $[cB\ o]$, $[cB\ cv]$, $[cB\ ct]$ }, one between *contains* and { $[ct\ ct]$ } and eight between *inside* and each of { $[i\ d]$, $[i\ m]$, $[i\ o]$, $[i\ cv]$, $[i\ cB]$, $[i\ e]$,

[$i\ ct$], [$i\ i$]}. The two separate graphs are each occupying their own *level* where relations of the same type reside—one level for t_{RR} s and another level for t_{RRh} s—resulting in a two-level graph. The edges that connect relations of the same type are called *intra-level* edges—8 for the t_{RR} level and 23 for the t_{RRh} level, whereas the 23 additional edges that connect relations of different type are called *cross-level edges*.

6.2.4 Cost of the Change in the Topological Structure of the Region

In the two separate graphs (Fig. 6.6), the topological changes of translation, scaling, and rotation are considered of equal importance. Since each edge on each separate graph represents one possible topological change, all edges are assigned the unit length, which is one of the basic characteristics of the FDM. Therefore, the cost of transforming one t_{RRh} into another solely depends on the length of the shortest path between the two relations, along the 23- t_{RRh} CNG (Table 5.2). Given the unit length of each edge, this length corresponds to the least amount of edges between the two t_{RRh} S.

Dropping a hole, however, is a topological change that differs from translation, scaling, and rotation. The change in the topological structure of a single-holed region that occurs when its hole is dropped corresponds to a certain cost as well, assigned to the cross-level edge that connects the t_{RRh} —between the hole-free region and the single-holed region that loses the hole—with its principal t_{RR} , the relation that remains between the hole-free region and the newly formed hole-free region after dropping the hole. Once the costs for all the cross-level edges are identified, the hybrid CNG provides the complete costs information necessary for enabling the FDM to evaluate the similarity between relations of regions featuring different numbers of holes.

6.2.4.1 Cost of Hole Modification: Cost Model A

There are several reasoning procedures for evaluating the cost of hole elimination and specifying how this cost relates to the translation, scaling, or rotation costs. For the performance of the FDM, however, a consistent cost assigning technique is needed for the creation of the table of costs that enables the exchange of frequency units between relations with regions of different numbers of holes.

One approach is to consider the hole elimination as an atomic change in the topology of the region and to assign to it the unit cost, depicted as course B in Figure 6.7. This reasoning equalizes the topological change in the relation (i.e., translation, scaling, or rotation) with the change in the topological structure of a region (i.e., hole elimination) and results in a hybrid CNG with edges of equal length along and across the two levels. A different line of thinking considers the possible sub-atomic changes. The hole collapsing to a cut in the region, the cut collapsing to a piercing in the region, and then the piercing disappearing are three possible consecutive steps that change the topology of the region, leaving it hole-free (Fig. 6.7 course C). If each of these changes is assigned the unit cost, then the total cost for dropping a hole from a region totals three units. This cost model, called Cost Model A (CMA) is used in the construction of the table of costs in Section 6.2.4.2. It differentiates the cost of hole elimination from the topological changes of translation and scaling. However, it is not considered the only or the most reasonable method for evaluating the cost of transforming every one of the 23- t_{RRh} to its principal t_{RR} , and different cost assigning methods are considered in Section 6.3.

When more than one holes exist in a region, one possible approach to hole reduction is to consider the moving and merging of neighboring holes (Fig. 6.7). If the moving of the neighboring holes is assigned the unit cost, and the merging of the holes is

assigned the unit cost as well, then the cost of hole number reduction by hole merging would be two. However, in the intermediate step the holes *meet*, therefore, this approach is not considered.

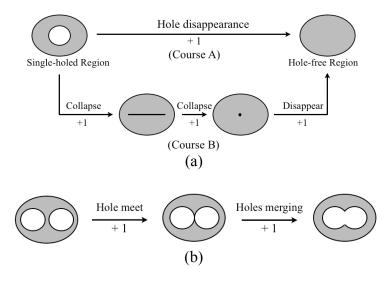


Figure 6.7 Different hole elimination procedures: (a) the atomic change in the region's topology when the hole is eliminated as course A, and the gradual change in the region's topology in three steps as course B, and (b) hole merging.

6.2.4.2 The Weighted Hybrid CNG and the Table of Costs from CMA

The assignment of costs to the cross-level edges that connect the 23- t_{RRh} CNG and the 8- t_{RR} CNG using CMA completes the costs information for the hybrid CNG. The graph has been modified in Figure 6.8 for clarity, as the costs assigned to each edge are added. The additional costs are summarized in Table 6.1 in the form of distances between relations of different type—23- t_{RRh} and 8- t_{RR} —of the hybrid CNG. Distances are calculated as the sum of the costs assigned to the edges belonging to the shortest path between two relations.

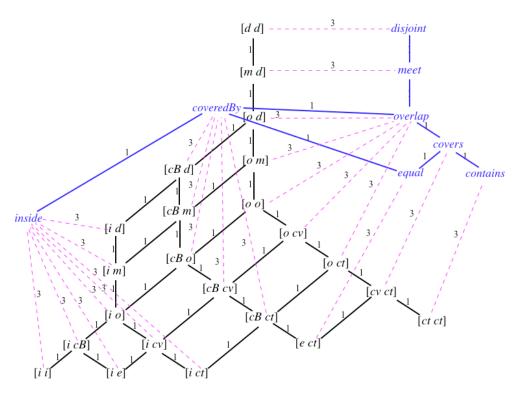


Figure 6.8 Hybrid CNG with costs of relation transformations assigned to the edges for CMA. One cost unit is assigned to the topological change of a relation, three cost units to the change in the topological structure of a region.

The table of costs for CMA breaks down into four sub-tables: (1) the table of costs for the 23- t_{RRh} ×23- t_{RRh} (Table 3.2), (2) the table of costs for the 23- t_{RRh} ×8- t_{RR} (Table 6.1), (3) its inverse table 8- t_{RR} ×23- t_{RRh} , and (4) the table of costs for the 8- t_{RR} ×8- t_{RR} that records the distances between all possible pairs of the 8- t_{RR} . The extended table (Table 6.2) corresponds to the complete record of the costs necessary for employing the transportation algorithm to the assessment of similarity between relations with different numbers of holes. The absolute value of the additional quantity in Δ represents the frequency units that are equal to the number of holes that have been dropped and acts as the weight assigned to the distance between the t_{RRh} and the t_{RR} formed after dropping the hole.

Table 6.1 Distances between relations of different type (t_{RRh} and t_{RR}) of the hybrid CNG, assigned using CMA.

	disjoint	meet	overlap	covers	contains	equals	coveredBy	inside
$\overline{[d d]}$	3	4	5	6	7	7	6	7
[m d]	4	3	4	5	6	6	5	6
[o d]	5	4	3	4	5	5	4	5
[o m]	5	4	3	4	5	5	4	5
$[o\ o]$	5	4	3	4	5	5	4	5
$[o \ cv]$	5	4	3	4	5	5	4	5
[<i>o ct</i>]	5	4	3	4	5	5	4	5
$[cv \ ct]$	6	5	4	3	4	4	5	6
$[ct \ ct]$	7	6	5	4	3	5	6	7
[<i>e ct</i>]	7	6	5	4	5	3	4	5
[cBd]	6	5	4	5	6	4	3	4
[cB m]	6	5	4	5	6	4	3	4
$[cB\ o]$	6	5	4	5	6	4	3	4
$[cB\ cv]$	6	5	4	5	6	4	3	4
$[cB\ ct]$	6	5	4	5	6	4	3	4
[i d]	7	6	5	6	7	5	4	3
[i m]	7	6	5	6	7	5	4	3
$[i \ o]$	7	6	5	6	7	5	4	3
[i cv]	7	6	5	6	7	5	4	3
[i ct]	7	6	5	6	7	5	4	3
[i e]	7	6	5	6	7	5	4	3
[i cB]	7	6	5	6	7	5	4	3
$[i \ i]$	7	6	5	6	7	5	4	3

Table 6.2 Schematic depiction of the complete table of costs for Cost Model A.

$$t_{RRh}$$
i, i=1..23 t_{RR} j, j=1..8 t_{RRh} i, i=1..23 Table 3.2 Table 6.1 t_{RR} j, j=1..8 $\left(\text{Table 6.1}\right)^{\text{T}}$ 8- t_{RR} × 8- t_{RR}

6.3 Alternative Cost Assigning Methods

The reasoning behind quantifying the change in the topological structure of a region when a hole is eliminated, and expressing it as a cost assigned to transforming a t_{RRh} into a t_{RR} , may vary. Common factors in the logic are the need to quantify the change in the topology and to apply a consistent reasoning while evaluating the hole elimination from each of the $23-t_{RRh}$. The hole elimination costs are recorded in tables, which are

subsequently used in the FDM for assessing similarity among relations with varying numbers of holes. This section examines three more reasoning techniques for evaluating the cost of dropping a hole. These techniques produce distinct tables of costs.

6.3.1 Balanced Hybrid CNG: Cost Model B

For the similarity evaluation between relations with regions of different numbers of holes, the costs of transforming relations to other relations of the same or different type are necessary. The costs for transformations of the same type are the weighted distances between relations along one of the levels of the hybrid CNG. The costs of transformation of different type, however, are weighted distances between relations across the two levels. The path of edges that leads from a t_{RRh} to a t_{RR} has to traverse one of the cross-level edges that connect the two separate levels (i.e., neighborhood graphs) on the hybrid CNG.

When the relations under comparison share the same principal relation, the path between the t_{RRh} that loses its hole and the newly formed t_{RR} comprises only the cross-level edge between the t_{RRh} and its principal relation. If the two relations that are compared have different principal relations, the newly formed t_{RR} after the hole elimination has to be transformed into a different t_{RR} , which means that a movement along the 8- t_{RR} level has to be performed as well (Fig. 6.9 path A). Alternatively, the t_{RRh} can first be transformed to the t_{RRh} with the different principal relation and then the single-holed region can drop its hole, which means that a traverse along the 23- t_{RRh} CNG is followed by a cross-traverse between the two graphs (Fig. 6.9 path B).

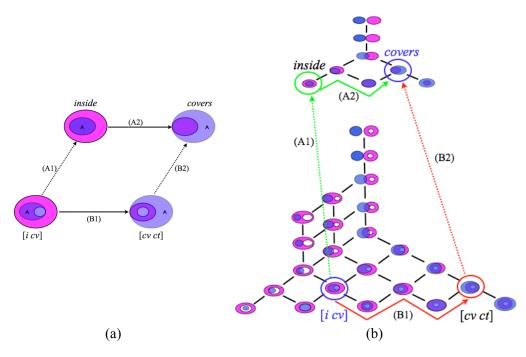


Figure 6.9 Different relation transformation paths. (a) Transformation paths A: $[i \ cv]$ $\xrightarrow{A1}$ $[cv \ ct]$ $\xrightarrow{A2}$ covers, and B: $[i \ cv]$ $\xrightarrow{B1}$ inside $\xrightarrow{B2}$ covers, and (b) display of the paths A and B on the hybrid CNG.

The movements along the two separate graphs may be considered of known cost, equal to the number of steps of topological deformations of translation, scaling, or rotation (e.g., movements A2 and B1 in Figure 6.9b). It is the movements along the cross-level edges on the hybrid graph that are of unknown value (e.g., movements A1 and B1 in Figure 6.9b). The basic premise for computing Cost Model B (CMB) is that if a specific cross-level edge is considered as reference and assigned a cost, then values for the rest of the cross-level edges on the hybrid CNG may be estimated. The assumption is that the two paths A and B, that lead from any random t_{RRh} to the t_{RR} at the end of the reference cross-level edge, are of equal overall cost.

As the cross-level reference edge, the edge between $[d\ d]$ and disjoint, denoted by $E([d\ d]-disjoint)$ is chosen. The choice is based on the fact that since the region that loses the hole is in disjoint relation with the hole-free region, the change in its topological

structure will not alter in any way the relation between the two regions. The reference edge is assigned the unit cost by considering the disappearance of the hole as an atomic step in the change of the region's topology (Fig. 6.10), in contrast to Cost Model A, in which the elimination of the hole is a three-step procedure (Fig. 6.7 B).

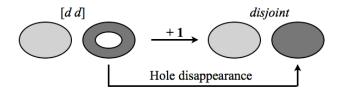


Figure 6.10 The reference edge $E([d\ d]-disjoint)$ for Cost Model B, assigned the unit cost after hole disappearance.

The cost assigned to the cross-level edge connecting a random $t_{RRh}x$ with its principal relation $\pi(t_{RRh}x)$, denoted by $E(t_{RRh}x-\pi(t_{RRh}x))$, may be evaluated using $E([d\ d]-disjoint)$ and the known costs of the traverses along the separate CNGs of the hybrid graph. These traverses are either the distance between $t_{RRh}x$ and $[d\ d]$ on the 23- t_{RRh} CNG, denoted by $dist(t_{RRh}x\rightarrow [d\ d])$, or the distance between $\pi(t_{RRh}x)$ and disjoint on the 8- t_{RR} CNG, denoted by $dist(\pi(t_{RRh}x)\rightarrow disjoint)$. According to CMB's basic assumption, the two paths that lead from $t_{RRh}x$ to disjoint are considered of equal cost (Eqn. 6.2a). Since the reference edge is assigned the unit cost, each of the 23 edges that connect the two separate CNGs on the hybrid graph may then be calculated using Equation (6.2b).

$$E(t_{RRh}x - \pi(t_{RRh}x)) + dist(\pi(t_{RRh}x) \rightarrow disjoint) = dist(t_{RRh}x \rightarrow [d\ d]) + E([d\ d] - disjoint)(6.2a)$$

$$E(t_{RRh}x - \pi(t_{RRh}x)) = dist(t_{RRh}x \rightarrow [d\ d]) - dist(\pi(t_{RRh}x) \rightarrow disjoint) + 1, \ x = 1..23$$
 (6.2b)

As an example, the cost assigned to $E([i\ i]-inside)$ is calculated (Fig. 6.11). The distance between $[i\ i]$ and $[d\ d]$ on the 23- t_{RRh} CNG is eight and the distance between

inside and *disjoint* on the 8- t_{RR} CNG is four, therefore, using Equation 6.2b, E([i i]—inside) equals five (Eqn. 6.3).

$$E([i\ i]-inside) = dist([i\ i] \rightarrow [d\ d]) - dist(inside \rightarrow disjoint) + 1 = 8-4+1 = 5$$
 (6.3)

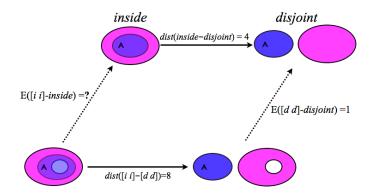


Figure 6.11 Schematic example of calculating the cost of an intra-level link.

The assignment of costs following this reasoning also implies that the cost of hole elimination increases in parallel with the increasing interaction between the hole-free and the single-holed region. As the relation between the hole-free region and the host region or the hole changes (from completely *disjoint*, to sharing boundary parts when *meet*, to sharing both interior and boundary parts when *overlap*, etc.), so do the distances $dist(t_{RRh}x \rightarrow [d\ d])$ and $dist(\pi(t_{RRh}x) \rightarrow disjoint)$. For principal relations with more than one refinement, while $dist(\pi(t_{RRh}x) \rightarrow disjoint))$ remains constant, $dist(t_{RRh}x \rightarrow [d\ d])$ increases as the hole-free region gets more involved with the hole (Eqn. 6.2b). For principal relations with one refinement, both distances increase as the interaction between the hole-free and the single-holed region changes. So there is an outward expansion from the reference edge that implies an increasing involvement between the hole-free and the host region initially, and subsequently with its hole. This expansion on the hybrid graph is responsible for the differentiation in the costs assigned to the edges connecting the principal t_{RRS} and their refining t_{RRh} s.

Using Table 3.2 and the 8- t_{RR} ×8- t_{RR} table of costs, the costs assigned to all 23 cross-level edges that connect the two separate graphs of the hybrid CNG are calculated, using Equation 6.2b. Once the costs are known, the distances from any t_{RRh} to any t_{RR} on the hybrid graph are evaluated (Table 6.3). With these costs, the distance between any two relations on the hybrid graph is calculated as the sum of the costs on the path between the two relations. The result is the Table of Costs (Table 6.4) and the hybrid CNG for CMB (Fig. 6.12).

Table 6.3 Distances between relations of different type (t_{RRh} and t_{RR}) on the hybrid CNG, assigned using CMB.

	disjoint	meet	overlap	covers	contains	equals	coveredBy	inside
$\overline{[d d]}$	1	2	3	4	5	5	4	5
[m d]	2	1	2	3	4	4	3	4
[o d]	3	2	1	2	3	3	2	3
[o m]	4	3	2	3	4	4	3	4
$[o\ o]$	5	4	3	4	5	5	4	5
$[o \ cv]$	6	5	4	5	6	6	5	6
[<i>o ct</i>]	7	6	5	6	7	7	6	7
$[cv \ ct]$	8	7	6	5	6	6	7	8
$[ct \ ct]$	9	8	7	6	5	7	8	9
[<i>e ct</i>]	9	9	7	6	7	5	6	7
[cBd]	4	3	2	3	4	2	1	2
[cB m]	5	4	3	4	5	3	2	3
$[cB\ o]$	6	5	4	5	6	4	3	4
$[cB\ cv]$	7	6	5	6	7	5	4	5
$[cB\ ct]$	8	7	6	7	8	6	5	6
[i d]	5	4	3	4	5	3	2	1
[i m]	6	5	4	5	6	4	3	2
$[i \ o]$	7	6	5	6	7	5	4	3
[i cv]	8	7	6	7	8	6	5	4
[i ct]	9	8	7	8	9	7	6	5
[i e]	9	8	7	8	9	7	6	5
[i cB]	8	7	6	7	8	6	5	4
$[i \ i]$	9	8	7	8	9	7	6	5

Table 6.4 Schematic depiction of the complete table of costs for Cost Model B.

	$t_{RRh}i$, i=123	t_{RR} j, j=18
$t_{RRh}i$, $i=123$	Table 3.2	Table 6.3
t_{RR} j , j =18	$(Table 6.3)^T$	$8-t_{RR} \times 8-t_{RR}$

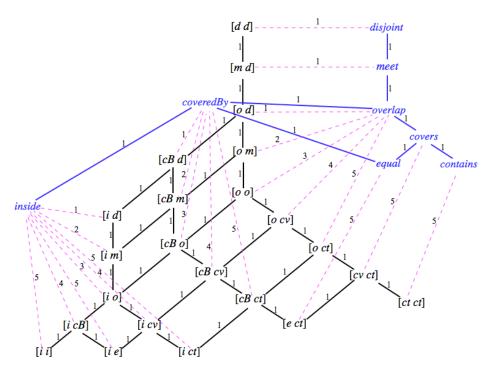


Figure 6.12 Hybrid CNG with costs of relation transformations assigned to the edges from Cost Model B.

6.3.2 Grouped Hybrid CNG: Cost Model C

The conjecture for model C is that the cost for eliminating a hole from t_{RRh} s that share the same principal relation should be the same. In contrast to CMB, the different degrees of interaction between the hole-free region and the hole are not important. Rather, the cost is only affected by the degree of interaction between the hole-free region and the host of the holes, which is reflected in the principal relation. Therefore, the cross-level edges extending from a specific principal t_{RR} to its refinement(s) are all assigned the same cost. For the multi-refinement relations, the multiple edges may be schematically grouped in a single edge. This grouping can be depicted as morphing the 23- t_{RRh} CNG into a copy of the 8- t_{RR} CNG (Fig. 6.13).

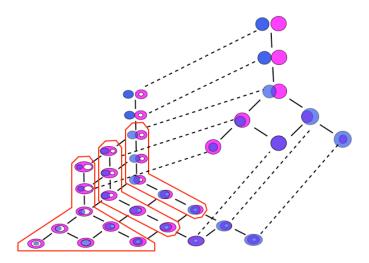


Figure 6.13 The grouped hybrid CNG.

The edges or the groups of edges extending from principal relations that indicate less interaction between the hole-free and the host region are attributed lower costs than edges or groups extending from principal relations that indicate more interaction between the two regions. The increase in interaction is determined by the sequence of the principal relations on the 8- t_{RR} CNG (Fig. 6.14a). Starting from two completely *disjoint* regions, the interaction between them increases gradually up to complete containment or conversely inclusion, as the relations change one step at a time on the CNG. Each step of change produces the relation(s) of the next level of interaction. The linking edges extending from the principal relations of each level are assigned one additional cost unit from the edges of the previous level, starting with one unit for $E([d\ d]-disjoint)$, up to a total of 5 units (Fig. 6.14b). Edges of principal relations belonging to the same level are assigned equal costs.

The result of the cost assigning reasoning is the grouped hybrid graph according to CMC, depicted in Figure 6.15. Once the costs are known, the distances from any t_{RRh} to any t_{RR} on the hybrid graph are evaluated (Table 6.5). With these costs, the distance between any two relations on the hybrid graph is calculated as the path between the two

relations with the smaller overall cost. The result is the Table of Costs for CMC (Table 6.6).

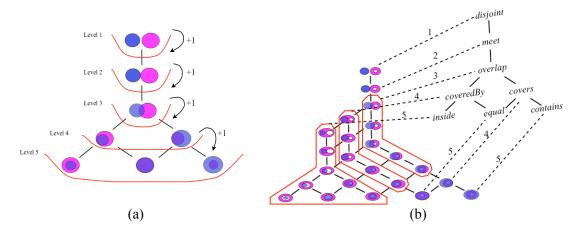


Figure 6.14 Grouped hybrid CNG reasoning: (a) levels of region interaction on the 8- t_{RR} CNG, and (b) costs assigned to plain and grouped linking edges on the hybrid CNG, according to CMC.

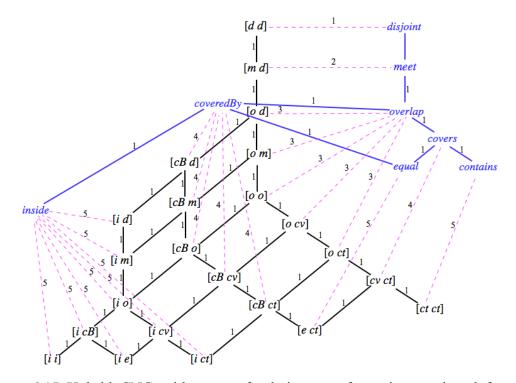


Figure 6.15 Hybrid CNG with costs of relation transformations assigned from Cost Model C.

Table 6.5 Distances between relations of different type (t_{RRh} and t_{RR}) on the hybrid CNG, assigned using CMC.

	disjoint	meet	overlap	covers	contains	equals	coveredBy	inside
[d d]	1	2	3	4	5	5	4	5
[m d]	2	2	3	4	5	5	4	5
[o d]	3	3	3	4	5	5	4	5
[o m]	4	4	3	4	5	5	4	5
$[o\ o]$	5	4	3	4	5	5	4	5
[o cv]	5	4	3	4	5	5	4	5
[<i>o ct</i>]	5	4	3	4	5	5	4	5
$[cv \ ct]$	6	5	4	4	5	5	6	6
$[ct \ ct]$	7	6	5	5	5	7	6	7
[<i>e ct</i>]	7	6	5	5	6	5	5	7
[cBd]	4	5	4	5	6	5	4	5
[cB m]	5	5	4	5	6	5	4	5
$[cB\ o]$	6	5	4	5	6	5	4	5
$[cB\ cv]$	6	5	4	5	6	5	4	5
$[cB \ ct]$	6	5	4	5	6	5	4	5
[i d]	5	6	5	6	7	6	5	5
[i m]	6	6	5	6	7	6	5	5
$[i \ o]$	7	6	5	6	7	6	5	5
[i cv]	7	6	5	6	7	6	5	5
[i ct]	7	6	5	6	7	6	5	5
[i e]	8	7	6	7	8	7	6	5
[i cB]	8	7	6	7	8	7	6	5
$[i \ i]$	9	8	7	8	9	7	6	5

Table 6.6 Schematic depiction of the complete table of costs for Cost Model C.

	$t_{RRh}i$, i=123	t_{RR} j, j=18
t_{RRh} i, i=123	Table 5.2	Table 6.5
t_{RR} j , j =18	$(Table 6.5)^T$	$8-t_{RR} \times 8-t_{RR}$

6.3.3 Modified Grouped Hybrid CNG - Longest Paths: Cost Model D

The last alternative cost model examined is a modified version of CMC. As described in Section 6.3.1 for CMB, there are two paths connecting two relations of different type on the hybrid CNG: (1) the traverse along the 23- t_{RRh} CNG followed by the cross over to the 8- t_{RR} CNG and (2) the cross over to the 8- t_{RR} CNG followed by the traverse along that graph. For all previous models, except CMB, these two paths are of different overall cost,

and the one with the lower cost is the one always recorded in the tables of costs. In CMB, these paths are of equal cost. In order to examine how the similarity evaluation results change from those of CMC, model D's table of costs is built by recording the more costly path between relations, produced from the same cost assigning method. For this model, the most costly path actually coincides with the path that first crosses over the 8- t_{RR} CNG and then traverses the graph to the desired principal relation. Consequently, models CMC and CMD share the same hybrid CNG (Fig. 6.15), but have different tables of costs by recording different paths between relations. Table 6.7 shows the distances between relations of different type for CMD, while the depiction of the overall cost table is given in Table 6.8.

Table 6.7 Distances between relations of different type (t_{RRh} and t_{RR}) on the hybrid CNG, assigned using CMD.

	disjoint	meet	overlap	covers	contains	equals	coveredBy	inside
$\overline{[d d]}$	1	2	3	4	5	5	4	5
[m d]	3	2	3	4	5	5	4	5
[o d]	5	4	3	4	5	5	4	5
[o m]	5	4	3	4	5	5	4	5
$[o\ o]$	5	4	3	4	5	5	4	5
$[o \ cv]$	5	4	3	4	5	5	4	5
$[o\ ct]$	5	4	3	4	5	5	4	5
$[cv \ ct]$	7	6	5	4	5	5	6	7
$[ct \ ct]$	9	8	7	6	5	7	8	9
[<i>e ct</i>]	9	8	7	6	7	5	6	7
[cBd]	7	6	5	6	7	5	4	5
[cB m]	7	6	5	6	7	5	4	5
$[cB\ o]$	7	6	5	6	7	5	4	5
$[cB\ cv]$	7	6	5	6	7	5	4	5
$[cB\ ct]$	7	6	5	6	7	5	4	5
[i d]	9	8	7	8	9	7	6	5
[i m]	9	8	7	8	9	7	6	5
$[i \ o]$	9	8	7	8	9	7	6	5
[i cv]	9	8	7	8	9	7	6	5
[i ct]	9	8	7	8	9	7	6	5
[i e]	9	8	7	8	9	7	6	5
[i cB]	9	8	7	8	9	7	6	5
$[i \ i]$	9	8	7	8	9	7	6	5

Table 6.8 Schematic depiction of the complete table of costs for Cost Model D.

$$t_{RRh}$$
i, i=1..23 t_{RR} j, j=1..8
 t_{RRh} i, i=1..23 Table 5.2 Table 6.7
 t_{RR} j, j=1..8 $\left(\text{Table 6.7}\right)^{\text{T}}$ 8- t_{RR} × 8- t_{RR} Table of Costs

6.4 Adapted FDM for Relations of Regions with Different Numbers of Holes

With the hybrid CNG acting as the source of information for the costs associated with dropping a hole, the FDM is altered to accommodate similarity evaluation between relations of regions with varying numbers of holes. This section discusses the alternative FDM (Section 6.4.1) and illustrates its application with an example, using Cost Model A (Section 6.4.2).

6.4.1 Frequency Units that Rebalance the System

The new element necessary for bringing the sum of relation difference $\Delta(t_1,t_2)$ back to zero is actually the difference in the values of a new frequency element \hbar added to vectors V. This new element denotes the normalized number of holes that have been dropped, creating the inequality. If n_1 and n_2 are the different numbers of holes in the two multi-holed regions, \hbar equals the normalized difference $|n_1-n_2|$ in the vector that corresponds to the relation with number of holes equal to $\min(n_1, n_2)$ and it is zero for the relation with number of holes $\max(n_1, n_2)$ (Eqn. 6.3). Normalization is achieved by division with n, where $n=\max(n_1, n_2)$. The remaining m elements in V are the same as described in Equation 5.3.

$$V = \begin{bmatrix} F1 \\ \vdots \\ Fm \\ \vec{h} \end{bmatrix}, \text{ where } \vec{h} = \begin{cases} 0, \text{ for the relation with } \# \text{ holes} = \max(n_1, n_2) \\ \frac{|n_1 - n_2|}{n}, \text{ where } n = \max(n_1, n_2), \text{ for the relation with } \# \text{ holes} = \min(n_1, n_2) \end{cases}$$
 (6.3)

Since h is zero in one of the vectors, its value is the normalized difference $|\mathbf{n}_1 - \mathbf{n}_2|$ in the relation difference Δ as well, and its addition to Δ brings sum(Δ) back to zero. Then the sum of all supplies equals the sum of all demands, yielding a balanced transportation problem, for which the transportation algorithm may be applied. The methodology for relation similarity evaluation, in the form of calculating the minimum cost of transforming one relation's frequency vector into the other relation's vector, is the same as the one described in Chapter 5, now that the additional element in the vectors and the relation difference (Δ) has rebalanced the system.

However, the maximum cost of transformation—the maximum dissimilarity $\delta_{\max Hybrid}^{t_1-t_2}$ between relations t_1 and t_2 —needs to be reevaluated. The maximum possible amount of frequency units that can be redistributed is sum(V). The addition of \hbar matches the deficit in frequency units in the vector of the relation with the fewer holes and makes sum(V) equal to the unit again (Section 5.2.1). The frequency units introduced through \hbar , nevertheless, are distributed on a different path on the hybrid CNG than the rest of the units, because they make up for the dropped holes. The value of \hbar is distributed between the two levels of the graph, from a t_{RRh} to a t_{RR} . The redistribution of the rest of the frequency units, which are responsible for the transformation of a t_{RRh} to a different t_{RRh} , is limited at the 23- t_{RRh} CNG level.

Accordingly, there are two maximum costs of transformation: (1) $\delta_{\text{max}}^{\text{RRh-RRh}}$, equal to the product of sum(V) with the maximum shortest path on the 23- t_{RRh} , which is eight (Eqn. 6.4a), and (2) $\delta_{\text{max}}^{\text{RRh-RR}}$, equal to the product of sum(V) with the maximum distance across the two graphs and along the 8- t_{RR} CNG, denoted by $d_{\text{max}}(t_{RRh}-t_{RR})$, which is equal to seven (Eqn. 6.4b). The value of $d_{\text{max}}(t_{RRh}-t_{RR})$ is seven, because it is equal to the sum of the

three units for any cross-level edge and the maximum distance along the 8- t_{RR} graph, which is four.

$$d_{max}^{RRh-RRh} = sum(V) * d_{max23} = 1 * 8 = 8$$
 (6.4a)

$$d_{\max}^{RRh-RR} = sum(V) * d_{\max}(t_{RRh} - t_{RR}) = 1 * 7 = 7$$
(6.4b)

The calculation of z, the minimum cost for transforming vector V_1 into vector V_2 , is the same as the calculation in the initial FDM (Eqn. 5.2.2). It is essential, however, to differentiate the part of z that represents the distribution between t_{RRh} s, from the part that represents the cost of distribution between a t_{RRh} and a t_{RR} . To achieve this, first the redistribution scheme is decided (or is performed by the transportation algorithm if the algorithm is applied) and two different costs are calculated: (1) z_1 , equal to the sum of the products of the frequency units that are distributed between t_{RRh} s with the appropriate distances on the 23- t_{RRh} CNG, and (2) z_2 , equal to the sum of the products of the frequency units that are distributed between a t_{RRh} and a t_{RR} with the appropriate distances from the hybrid graph. The overall cost z then is the sum of z_1 and z_2 (Eqn. 6.5).

$$Z = Z_1 + Z_2 \tag{6.5}$$

The computation of the dissimilarity $\delta(t_1,t_2)$ is also modified to reflect the different parts of the overall cost of transforming t_1 into t_2 . To normalize the dissimilarity δ it is now necessary to normalize each part of the cost with the appropriate maximum cost of redistribution, $\delta_{\max}^{RRh-RRh}$ or δ_{\max}^{RRh-RR} , respectively (Eqn. 6.6). Similarity and dissimilarity are still complement values; therefore, since $\delta(t_1,t_2)$ is normalized to the interval [0 1] by division with the maximum costs for redistribution, similarity is still the complement of dissimilarity from the unit (Eqn. 6.7).

$$\delta(t_1, t_2) = \frac{z_1}{\delta_{\max}^{RRh-RRh}} + \frac{z_2}{\delta_{\max}^{RRh-RR}}$$
 (6.6)

$$\sin(t_1, t_2) = 1 - \delta(t_1, t_2) \tag{6.7}$$

6.4.2 Example of Similarity Evaluation with the Modified FDM

The multi-element relations t(A, 5-B) and t(C, 4-D) used as an example in Section 6.1.1 are used again here, only this time, t(A, 5-B) is the reference relation t_1 against which t(C, 4-D) and two more relations, t(E, 4-F) and t(G, 4-H), or t_2 , t_3 and t_4 for short, are ranked according to their similarity. All three relations have a multi-holed region with one less hole than the reference relation, two of them (t(C, 4-D)) and t(E, 4-F) share the same principal relation with the reference relation—that is, coveredBy—while the third (t(G, 4-H)) has a different one—inside.

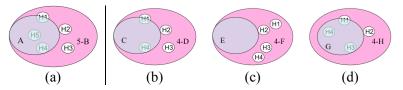


Figure 6.16 Multi-element relation comparison. (b) Relation t(C, 4-D), (c) relation t(E, 4-E), and (d) relation t(G, 4-H) compared against reference relation (a) t(A, 5-B).

For calculating the similarity of relation t(C, 4-D), the frequency vectors V_1 and V_2 (Section 6.1.1) are modified accordingly. The reference relation has one extra hole, which is dropped in t(C, 4-D), therefore, 1/5=0.2 frequency units are excess for V_1 . In V_2 these 0.2 frequency units are assigned to the additional element \hat{h} . The relation name of element \hat{h} is *coveredBy*, since the $t_{RRh}=[cB\ ct]$ between hole-free region C and single-holed region $\mathcal{D}5$ is transformed to $t_{RR}=coveredBy$ between hole-free regions C and D, after $\mathcal{D}5$ drops its hole (Fig. 6.4). The value of \hat{h} is zero for V_1 that presents no change in the number of holes (Eqn. 6.8a). The relation difference is calculated accordingly (Eqn. 6.8b).

$$V_{1} = \begin{bmatrix} [cB \ d] & 0.2 \\ [cB \ m] & 0.2 \\ [cB \ o] & 0.2 \\ [cB \ cv] & 0.2 \\ [cB \ ct] & 0.2 \\ [coveredBy \ 0] \end{bmatrix} \text{ and } V_{2} = \begin{bmatrix} [cB \ d] & 0.2 \\ [cB \ m] & 0.2 \\ [cB \ o] & 0.2 \\ [cB \ cv] & 0.2 \\ [cB \ ct] & 0 \\ [cb \ ct] & 0 \\ [coveredBy \ 0.2] \end{bmatrix}$$
(6.8a)

$$\Delta(t2,t1) = V_2 - V_1 = \begin{bmatrix} [cB \ d] & 0 \\ [cB \ m] & 0 \\ [cB \ o] & 0 \\ [cB \ cv] & 0 \\ [cB \ ct] & -0.2 \\ coveredBy & 0.2 \end{bmatrix}$$
(6.8b)

The relation difference now reflects the necessary redistribution of frequency units for transforming t_2 into t_1 : one frequency unit will be transferred from *coveredBy* to $[cB\ ct]$, since one hole that was contained in D is dropped. There is no other redistribution of frequency units that would take place between $t_{RRh}s$ only, therefore, the minimum cost of redistribution is of type z_2 only. The distance $dist(coveredBy \rightarrow [cB\ ct])$ is three cost units (Table 6.1) and so the cost of redistribution, z_2 , using Equation 5.7 is 0.6 (Eqn. 6.9).

$$Z_2 = 0.2*dist(coveredBy \rightarrow [cB\ ct]) = 0.2*3 = 0.6$$
 (6.9)

With z determined, the dissimilarity $\delta(t_2,t_1)$ and similarity $\sin(t_2,t_1)$ values can be evaluated using Equations 6.6 and 6.7, respectively (Eqn. 6.10a-b).

$$\delta(t_2, t_1) = 0 + \frac{Z_2}{\delta_{\text{maxHybrid}}^{t_1 - t_2}} = \frac{0.6}{7} = 0.086$$
 (6.10a)

$$\sin(t_2, t_1) = 1 - \delta(t_2, t_1) = 1 - 0.086 = 0.914 \text{ or } 91.4\%$$
 (6.10b)

Following the same methodology, $sim(t_3,t_1)$ is evaluated. Relation t_3 has one less hole and the placements of the remaining holes are different than t_1 . The relation's frequency vector V_3 and the relation difference $\Delta(t_3,t_1)$ are given in Equation 6.11.

$$V_{3} = \begin{bmatrix} [cB \ d] & 0.6 \\ [cB \ m] & 0.2 \\ [cB \ o] & 0 \\ [cB \ cv] & 0 \\ [cB \ ct] & 0 \\ [coveredBy \ 0.2 \end{bmatrix} \text{ and } \Delta(t_{3},t_{1}) = V_{3} - V_{1} = \begin{bmatrix} [cB \ d] & 0.4 \\ [cB \ m] & 0 \\ [cB \ o] & -0.2 \\ [cB \ cv] & -0.2 \\ [cB \ ct] & -0.2 \\ [coveredBy \ 0.2 \end{bmatrix}$$

$$(6.11)$$

Since the cost of transforming *coveredBy* to any of the $t_{RRh} = [cB \ x]$ is three, the transportation algorithm distributes first the frequency units from $[cB \ d]$ to the demands that are the closest. In this case, the shorter distances from $[cB \ d]$ are

 $dist([cB\ d] \rightarrow [cB\ o])=2$ and $dist([cB\ d] \rightarrow [cB\ cv])=3$. The t_{RRh} that will receive the additional 0.2 frequency units that are assigned to coveredBy after one hole has been dropped is $[cB\ ct]$. Now, both types of $costs-z_1$ and z_2 -exist, since frequency units are distributed both between t_{RRh} s and between a t_{RR} and a t_{RRh} . The results after applying Equations 6.12a-c and 6.13 are: $z_1=1$, $z_2=0.6$, $\delta(t_3,t_1)=0.21$ and $sim(t_3,t_1)=79\%$

$$Z_{t3t1} = Z_1 + Z_2 \tag{6.12a}$$

$$Z_1 = 0.2*dist([cB\ d] \rightarrow [cB\ o]) + 0.2*dist([cB\ d] \rightarrow [cB\ cv])$$

$$Z_2 = 0.2*dist(coveredBy \rightarrow [cB\ ct])$$
(6.12b)

$$\delta(t_3, t_1) = \frac{Z_1}{\delta_{\max}^{RRh-RRh}} + \frac{Z_2}{\delta_{\max}^{RRh-RR}}$$
 (6.12c)

$$sim(t_3,t_1)=1-\delta(t_3,t_1)=0.79 \text{ or } 79\%$$
 (6.13)

Finally, the similarity between relations t_4 and t_1 is evaluated. In this case, while the refining relations that denote the placement of the holes relative to the hole-free region are the same, the principal relation has changed to *inside*. The modified V_1 and also V_4 are given in Equation 6.14a and the relation difference in Equation 6.14b.

$$V_{1} = \begin{bmatrix} [cB \ d] & 0.2 \\ [cB \ m] & 0.2 \\ [cB \ o] & 0.2 \\ [cB \ cv] & 0.2 \\ [cB \ cv] & 0.2 \\ [i \ m] & 0 \\ [i \ o] & 0 \\ [i \ cv] & 0 \\ [i \ ct] & 0 \\ inside & 0 \end{bmatrix} \quad \text{and} \quad V_{4} = \begin{bmatrix} [cB \ d] & 0 \\ [cB \ m] & 0 \\ [cB \ cv] & 0 \\ [cB \ cv] & 0 \\ [cB \ cv] & 0 \\ [i \ cB \ cv] & 0 \\ [i \ m] & 0.2 \\ [i \ o] & 0.2 \\ [i \ cv] & 0.2 \\ [i \ ct] & 0.2 \\ [i \ ct]$$

$$\Delta(t_{4},t_{1}) = V_{4} - V_{1} = \begin{bmatrix} [cB\ d] & -0.2 \\ [cB\ m] & -0.2 \\ [cB\ co] & -0.2 \\ [cB\ cv] & -0.2 \\ [cB\ ct] & -0.2 \\ [i\ m] & 0.2 \\ [i\ co] & 0.2 \\ [i\ cv] & 0.2 \\ [i\ ct] & 0.2 \\ [i\ ct] & 0.2 \\ [i\ sinside & 0.2 \end{bmatrix}$$

$$(6.14b)$$

The difference from t_2 and t_3 is that a hole is dropped from relation *inside*, therefore, the t_{RRh} between the hole-free region G and single-holed region $\mathcal{H}5$ is changed to t_{RR} *inside* for hole-free regions G and H after the hole is dropped (Fig. 6.16d). All distances between the neighboring t_{RRh} s that share the same refining relation (e.g., $dist([i\ m] \rightarrow [cB\ m]), dist([i\ o] \rightarrow [cB\ o]))$ are equal to one cost unit. This is the distribution with the minimum cost, since the distance from *inside* to any of the t_{RRh} s with *coverdBy* as their principal relation is four units. The results after applying Equations 6.15a-c, and 6.16. are: z_1 =0.8, z_2 =0.8, $\delta(t_4,t_1)$ =0.21 and $\sin(t_4,t_1)$ =79%.

$$Z_{t4t1} = Z_1 + Z_2 \tag{6.15a}$$

$$z_2 = 0.2 * [dist([i\ m] \rightarrow [cB\ m]) + dist([i\ o] \rightarrow [cB\ o]) + dist([i\ cv] \rightarrow [cB\ cv]) + dist([i\ ct] \rightarrow [cB\ ct])]$$

$$(6.15b)$$

$$z_2 = 0.2 * dist(inside \rightarrow [cB\ d])$$

$$\delta(t_4, t_1) = \frac{z_1}{\delta_{\text{max}}^{\text{RRh-RRh}}} + \frac{z_2}{\delta_{\text{max}}^{\text{RRh-RR}}}$$
(6.15c)

$$\sin(t_4, t_1) = 1 - \delta(t_4, t_1) = 0.79 \text{ or } 79\%$$
 (6.16)

The cost of redistribution is the same as for t_3 , therefore, t_3 and t_4 rank in second place together (Fig. 6. 16). The tie can be attributed to the fact that even though t_4 has a different principal relation than the reference relation t_1 , the placement of the holes is the same. Relations *inside* and *coveredBy* are immediate neighbors on the CNG. The relative placement of the elements of each relation on the graph in a close local neighborhood is

responsible for the ranking. Relations whose majority of elements are placed closer to the elements of the reference relation on the hybrid CNG will rank higher, even if the principal relation is different.

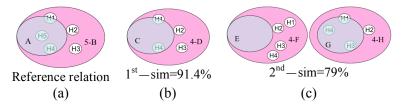


Figure 6.17 Ranking results against reference relation t(A, 5-B): (b) Relation t(C, 4-D), which ranks the closest to the reference, and (c) relations t(E, 4-F) and t(G, 4-H), which both rank second.

6.5 Summary

This chapter discussed the modifications in the reasoning process of the FDM so that the method can address cases of similarity ranking among multi-element relations, when the multi-holed regions have different numbers of holes. The key differentiation point from FDM for relations of regions with the same number of holes is the assignment of a cost to the modification of the topology of the region due to the elimination or addition of a hole. It is a rational addition to reestablish an equilibrium to the system that was thrown out of balance due to the difference in the topological structure of the participating regions. Therefore, the main characteristic of the altered FDM is the coexistence of topological changes that alter the relation between two regions with changes that modify the topological structure of the regions themselves. Coexistence is achieved through hybrid CNGs that connect the 23- t_{RRh} CNG with the 8- t_{RR} CNG.

It has been displayed that there are different reasoning practices for selecting the cost that will be assigned to changes due to hole eliminations. Four possible models—CMA, CMB, CMC, and CMD—have been analyzed. The results of the different cost

models are: (1) hybrid CNGs with alternative tables of costs, (2) extended frequency vectors that depict the deficit in frequency units due to the elimination of holes and, (3) appropriately modified calculations for the minimum cost of transformation between two relations and subsequently, their similarity. The example similarity evaluation shows that the relative placement of the elements of each multi-element relation on the graph, in locally close neighborhoods, is responsible for higher similarity rankings, even if the overall principal relation differs from that of the reference relation.

Chapter 7 Evaluation of Similarity Results

The various cost-assigning models developed in Chapter 6—CMA, CMB, CMC, and CMD—are expected to result in—at least numerically—different similarity rankings. Each model produces a distinct table of costs, from which the similarity evaluation method acquires the distances between relations. Other than the disparity in the numerical percentages, however, the possible variations in the rankings within and across the models need to be studied. This chapter examines the results of using FDM and employing the various cost models to compare relations between a hole-free and a multi-holed region, when the regions have different numbers of holes. For text brevity, we shall refer to relations as having a number of holes, even though the number of holes is a property of the multi-holed regions. By testing how such parameters as the different number of holes from the query relation or the change in the constituent relations shape the results, the goal is to compare the models, assess whether they produce the same, comparable or even contradictory similarity rankings, and evaluate the method. The closeness or deviation of each model from what would be considered an expected behavior is also appraised.

Section 7.1 describes the experimental framework for testing FDM using the different cost models. The analysis of the similarity rankings against specific criteria is

discussed in Section 7.2 while the overall assessment of the FDM for comparing relations of regions with different numbers of holes is presented in Section 7.3. The chapter is summarized in Section 7.4.

7.1 Experimental Setup

The evaluation of the FDM and the different cost models is performed by examining the results of ranking specific groups of relations for their similarity against particular queries. Both the queries and the relation groups are chosen so that certain general characteristics of a cost model may be examined. Section 7.1.1 describes the criteria against which each model is tested, while section 7.1.2 discusses the different datasets and the reasoning behind their selection as test beds. Section 7.1.3 illustrates the general evaluation procedure.

7.1.1 Evaluation Criteria

There are certain expectations against which the different cost models are evaluated according to the similarity rankings they produce (Section 7.2). These expectations are based on the fact that the cost models are internally linked to the hybrid neighborhood graph, therefore, the positioning of the various relations on the graph (e.g., their local neighborhoods) and the distances among the relations are the main determining factors for the similarity ranking against specific queries. Furthermore, the FDM that uses these cost models evaluates relations whose number of holes differs from that in the query; therefore, the number of holes is another determining factor for the similarity rankings. The criteria for evaluating the behavior of the different cost models are: (1) the *number of*

holes criterion, (2) the *principal relation* criterion, and (3) the *refining relation* criterion, and they are summarized below.

- Relations with a smaller difference in the number of holes from the query relation, are expected to rank closer to the query than relations with a bigger difference in the number of holes.
- 2. Relations with the same principal relation as the query are expected to be more similar to the query relation than those with a different principal relation.
- 3a. Among relations with the same principal relation as the query, those with most or all refining relations equal to the query relation are expected to be more similar to the query than those with most or all refining relations different.
- 3b. Among relations with the same principal relation, but which differs from the query's principal relation, those with most or all refining relations equal to the query relation are expected to be more similar to the query than those with most or all refining relations different.

Deviations from this baseline are anticipated to show how the different cost-assigning methods influence the similarity rankings, and whether there is a preferable model for similarity ranking of relations with different numbers of holes. The behavior of the models against the evaluation criteria and the interpretation of the results will also verify whether the expected model behavior is justified or not.

7.1.2 Datasets and Queries

An experimental evaluation with synthetic datasets was performed. The use of synthetic, rather than actual data, allowed for an experimental set up that is tailored to cover all possibilities of relations presenting a higher or lower degree of resemblance to the query, according to the evaluation criteria. Synthetic data are also free of the need to be cleaned up for errors and inconsistencies. They provide a systematic and controlled approach for testing how a variety of relations over a multi-holed region rank against specific queries, as well as how their similarity changes with monitored changes in their number of holes and their constituent relations.

Two datasets, D_I and D_{II}, are examined for their similarities against query QO, with principal relation *overlap* and queries QI₂ and QI₅, with principal relation *inside* (Fig. 7.1). Each dataset comprises a group of eight relations and each such dataset relation has a different principal relation, so that all eight principal relations may be evaluated against the query. Relations with one of the five single-refinement principal relations—*disjoint, meet, covers, contains*, and *equal*—are common for all datasets. Relations with the remaining three multi-refinement principal relations—*overlap*, *coveredBy*, and *inside*—vary in the two datasets. The queries are also selected from the group of the multi-refinement relations so that query and dataset relations exhibit variable combinations of common and different constituent relations (Fig. 7.1). The subscript index of a query or multi-refinement dataset relation reveals the number of refining relations in each case. Table 7.1 summarizes the refining relations for each query and each multi-refinement relation in the datasets.

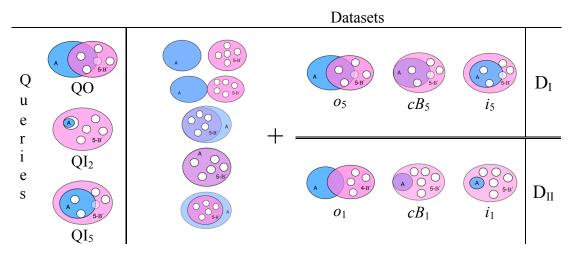


Figure 7.1 Depiction of instances of queries QO, QI₂, QI₅ and datasets D_I and D_{II}.

Table 7.1 Refining relations for queries and multi-refinement dataset relations.

	disjoint	meet	overlap	covers	contains	equal	coveredBy	inside
QO		$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	-	-	-
QI_2							$\sqrt{}$	
QI_5		$\sqrt{}$	$\sqrt{}$	V				
D_{I} - o_{5}	V	V	$\sqrt{}$		$\sqrt{}$	-	-	-
D_{II} - o_1						-	-	-
D_I - cB_5		$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	-	-	-
D_{II} - cB_1						-	-	-
D_{I} - i_{5}		$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$			
$\mathrm{D_{II}} ext{-}i_1$								

7.1.3 Evaluation Procedure

All datasets are evaluated for their similarity against all queries, yielding six combinations. The combination of query and dataset for pairs D_I -QO and D_I -QI₅ examines the cases where query and dataset relations with multi-refinement principal relations *overlap, coveredBy,* and *inside* share all of their common refining relations. Pairs D_{II} -QO and D_{II} -QI₅ examine the case where query and multi-refinement relations share some of their common refinements. Finally, pairs D_I -QI₂ and D_{II} -QI₂ examine the

cases where query and multi-refinement relations share some of their refinements and query *inside* introduces one uncommon refinement (Table 7.1).

Each query relation has five holes. The relations in the datasets also start with five holes and each dataset iterates over five rounds of evaluation. In each iteration, the dataset relations each drop one hole; thus, at the last round the relations comprise hole-free regions. This setup results in five rankings of eight relations against each query, for each dataset and each model. Not all possible sequences with which the holes are dropped were examined, as we deemed necessary to rather test one sequence that provided certain combinations of relations, the comparison of which would show the behavior of the altered FDM against the evaluation criteria. The individual similarity results would differ for different hole-dropping sequences. However, the final conclusions that may be drawn for the similarity evaluation method and the cost-assigning models are independent of the sequence with which the holes are dropped. The selection of five as the initial number of holes is equally significant as any other number of holes. The choice is based on the fact that having five consecutive single hole eliminations for all eight relations creates a large enough, but still manageable, body of data for observing the behavior of each model.

7.2 Analysis of Similarity Rankings

This section examines the similarity rankings of all datasets for the four cost models against the evaluation criteria. The common characteristics of the models are presented for each criterion separately, followed by additional observations and variations in the models' performance.

The similarity values are displayed along three continua, in which the dataset relations are depicted according to three local neighborhoods that their principal relations form on the 8- t_{RR} CNG (Fig. 7.2): (1) the left neighborhood *disjoint-meet-overlap-coveredBy-inside*, (2) the right neighborhood *disjoint-meet-overlap-covers-contains*, and (3) the bottom neighborhood *inside-coveredBy-equal-covers-contains*. The union of these three neighborhoods covers all principal relations and each of them allows for a continuous depiction of the similarity values for relations with neighboring principal relations, avoiding gaps and jumps to non-immediate neighbors. The dataset relations are depicted during the first round of evaluation, having four holes.

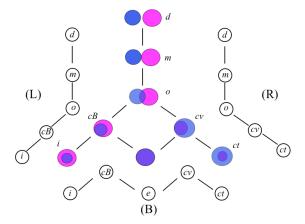


Figure 7.2 The left (L), right (R), and bottom (B) continua of the principal relations on the 8- t_{RR} .

7.2.1 First Criterion: Influence of the Number of Holes

Examination of the models' rankings for all datasets against the first criterion (Section 7.1.1) shows the following patterns:

 Relations with fewer holes may rank higher than relations with number of holes closer to the query's number. Depending on the combination of cost model and dataset, the decrease in the similarity of relations with the reduction in the number of holes may be monotonic or non-monotonic.

Figure 7.3 depicts dataset D_{II} 's similarity ranking against QI_5 with model C, along the three continua. There are relations with three or two holes that rank higher than relations with four holes, especially pronounced in Figure 7.3b. The similarity ranking for this dataset and model, therefore, is non-monotonic with a decreasing number of holes. However, the combination of model D and dataset D_I produces monotonic similarity rankings against the same query (QI_5) (Fig. 7.4), as relations with more holes always rank higher than relation with fewer holes.

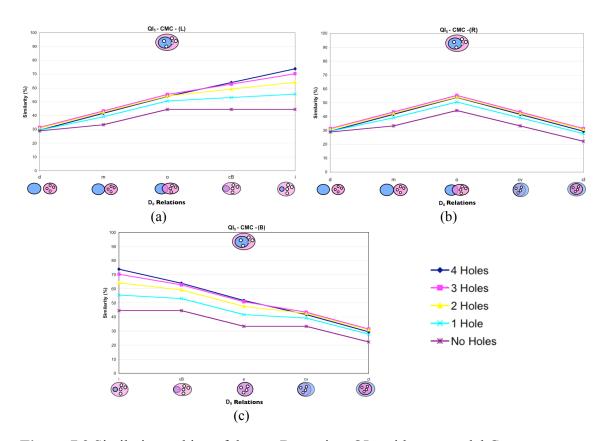


Figure 7.3 Similarity ranking of dataset D_{II} against QI₅, with cost model C.

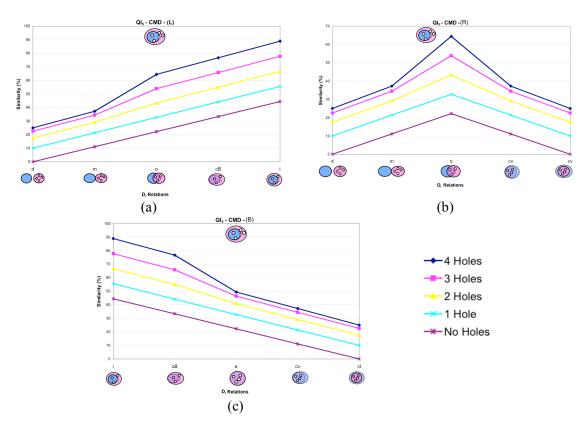


Figure 7.4 Similarity rankings of dataset D_I against QI₅, with cost model D.

7.2.2 Second Criterion: Influence of the Principal Relation

Examination of the models' rankings for all datasets against the second criterion (Section 7.1.1) displays the following patterns:

 Among relations with the same number of holes (i.e., relations of a certain evaluation round) and with the same refining relations that are very different than the query's, the relation that shares the same principal relation with the query ranks first.

Such is the case exemplified in Figure 7.5 for dataset D_1 's rankings against QI_2 , with cost model A. Relations *inside*, with four or even three holes rank above the remaining relations.

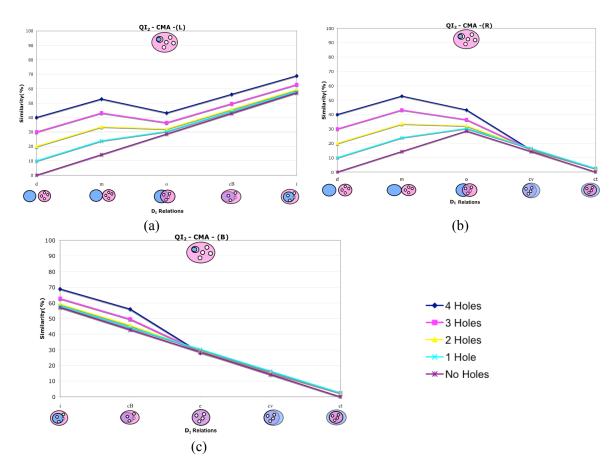


Figure 7.5 Similarity rankings of dataset D_I against QI₂, with cost model A.

7.2.3 Third Criterion: Influence of the Refining Relations

Examination of the models' rankings for all datasets against the third criterion (Section 7.1.1) reveals the following results:

All models produce rankings that validate the influence of the refining relations. Among dataset relations with the same principal relation, which may or may not be the same as the query's, those with most or all refining relations the same as the query's rank higher than those with most or all refining relations different than the query's. By implication, the combination of both principal and most or all refining relations the same as the query's rank higher than the rest of the relations, in the same evaluation round.

The differentiating role that the refining relations play for dataset relations that share the same principal relation as the query is important, since the difference in the similarity value may be substantial. Such a case is the different similarity values for relations with principal relation *inside* in datasets D_I and D_{II}, against QI₅, for cost model B (Fig. 7.6). The relation in D_I (Fig. 7.6b) has four common refining relations with the query QI₅ (Fig. 7.6a), while D_{II}'s relation has only one refining relation common with QI₅ (Fig. 7.6c). The approximate difference in the value is 24%.

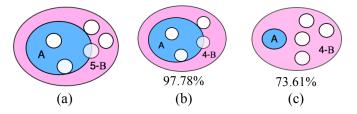


Figure 7.6 Similarity values after the first evaluation round of cost model B. The relations are *inside* in datasets $D_I(b)$ and $D_{II}(c)$, against query relation $QI_5(a)$.

Accordingly, there is a noticeable difference in the similarity among relations with the same principal relation that is different from the query's. For example, the similarity value for cost model D of the *overlap* relations in D_I (Fig. 7.7b) is higher than the value of the corresponding relation in D_{II} (Fig. 7.7c) because of the fewer refining relations it has in common with query QI_5 (Fig. 7.7a).

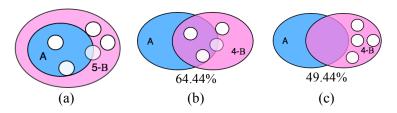


Figure 7.7 Similarity values after the first evaluation round of cost model C. The relations are *overlap* in datasets D_{II} (b) and D_{II} (c), against query relation QI_5 (a).

Figure 7.8 displays an example where relations *inside*, which have not only the same principal, but also very similar refining relations as query QI₂'s not only rank

higher than the rest in every evaluation round, but also posses the first few ranking positions, even with fewer holes than the rest of the relations.

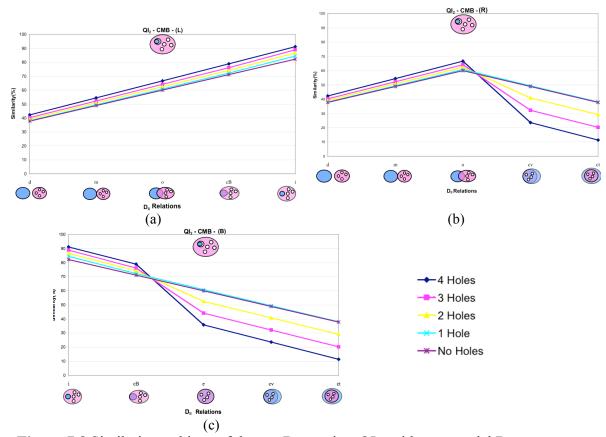


Figure 7.8 Similarity rankings of dataset D_{II}, against QI₂, with cost model B.

7.2.4 Additional Observations On the Similarity Rankings

Apart from the observations related with the evaluation criteria, the analysis of the similarity rankings further reveals characteristics that some or all models display through their results, as well as some differences among them.

• An additional consequence of the influence of the refining relations is verified in all models. Among relations with different principal relations, none of which is the same as the query's, the ones with most or all refining relations equal to the query's rank higher than those with different refining relations.

• The ranking of relations with both principal and most or all refining relations different than the query's, depends on their relative distances from the query on the hybrid CNG, regardless of the cost model.

Figure 7.5 is an example of the difference in the similarity value due to the refining relations between two relations from the same dataset. Relations with *meet* and *overlap* as their principal relations both have different principal relations than query QI₂. However, *meet* shares the same refining relations with the query and ranks higher for the first three evaluation rounds, while *overlap* has most refining relations different ranking lower for these rounds. Even if the principal relation *overlap* is closer to relation inside on the 8- t_{RR} CNG, the refinements' distance on the 23- t_{RRh} CNG is what eventually indicates which relation ranks higher in similarity. In the example illustrated in Figure 7.8, the remaining refinement [m d] is closer to the query's [i d] refinement when one, two, and three holes have been dropped. After the fourth iteration, *overlap* refinements sit closer to [i cB], which is the QI₂'s other refining relation, thus ranking higher than the *meet* relations.

7.3 Assessment of the Similarity Evaluation Method for Relations with Different Numbers of Holes

The overall assessment of the outcomes of applying the alternative FDM (Chapter 6) to various datasets, using different cost models, reveals some interesting insights about the method and the cost assigning techniques described. They are summarized and discussed below.

1. All models produce mostly unsurprising similarity rankings that differ in the absolute values.

The results mostly fulfill the expectations set by the evaluation criteria. This realization has implications for the method and the different cost models. The fact that none of the models produced especially surprising or controversial rankings implies the soundness of the method as an evaluation technique for relation similarity. It also implies that none of the described approaches for evaluating the cost of dropping a hole is irrational or interferes with the rankings in an uncontrolled manner.

2. The patterns of the similarity rankings for a specific dataset are more similar between pairs of models. In particular, the pairs with similar patterns are CMA-CMD, and CMB-CMC. Responsible for the formation of the pairs are the cost assigning techniques.

In general, CMA and CMD more distinctively differentiate the cost of dropping a hole from a region from the change in the topological relation between two regions. The part of the dissimilarity $\delta(t_1,t_2)$ that corresponds to the normalized cost of the dropped holes, $\frac{Z^2}{\delta_{\max}^{RRh-RR}}$, is higher for these two models, than it is for the pair CMB-CMC (Eqn. 7.1).

$$\delta(t_1, t_2) = \frac{z_1}{\delta_{\text{max}}^{\text{RRh-RRh}}} + \frac{z_2}{\delta_{\text{max}}^{\text{RRh-RR}}}$$
(7.1)

For CMA, the cost z_2 of the dropped holes is higher as a consequence of normalization by a smaller $\delta_{\text{max}}^{\text{RRh-RR}}$ than the rest of the models (Section 6.4.2). For the CMD, cost z_2 is higher, because the distances between relations of different type that are recorded in model D's table of costs are the longer of the two paths that may lead from one relation to the other. The part $\frac{z_1}{\delta_{\text{max}}^{\text{RRh-RRh}}}$, which corresponds to the transformation of one t_{RRh} relation to another, is common for all models (Chapter 6).

The common patterns appear according to the combination of query and dataset relations. While the pattern for query QO and dataset D_I is almost the same for all models, the patterns for the datasets compared against queries QI_2 and QI_5 are different between the two pairs of models. For the pair CMA-CMD, the pattern for QI_2 and any dataset resembles the one in Figure 7.8, and for QI_5 and any dataset the one in Figure 7.4. Accordingly, for the pair CMB-CMC, the pattern for QI_2 and any dataset resembles the one depicted in Figure 7.5, and finally for QI_5 and any dataset, the pattern follows the one illustrated in Figure 7.3.

3. The closer the elements of the dataset relation are to the elements of the query relation, the closer the evaluation results of the different models are. When the distances between the elements of the relation and the elements of the query are bigger, the two pairs of models behave differently.

This realization becomes apparent in the rankings of the different datasets. For dataset D_I, when compared against QO, the rankings of all models are very similar and look like the one depicted in Figure 7.9. The elements of query QO are placed centrally on the hybrid CNG. In addition, relations *coveredBy* and *inside* in dataset D_I share the same refining relations as the query, therefore, the distances of their elements on the graph are small (Fig. 7.10). The different cost-assigning methods that are responsible for the grouping of the models in two pairs do not vary greatly when the distances between the elements of query and relation remain small.

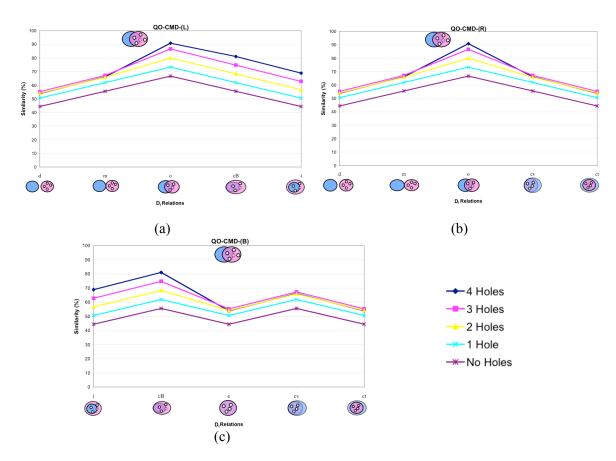


Figure 7.9 Similarity rankings of dataset D_I against QO, with cost model D.

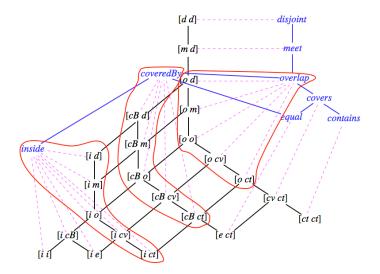


Figure 7.10 Hybrid CNG highlighting the *overlap*, *coveredBy*, and *inside* relations that participate in dataset D_I .

The differences in the similarity rankings between the pairs of models, however, are more accentuated when the elements of the query relation and the elements of the dataset

relations are placed farther away on the graph. Such an example is when dataset D_{II} 's relations are compared against QI_2 . The elements $[i\ d]$ and $[i\ cB]$ of query QI_2 are placed farther away from the elements of the rest of the relations, as they are situated close to the edges of the graph (Fig. 7.11). In such cases, the magnitude of the difference between the similarity values of relations in consecutive evaluation rounds is bigger for the pair CMA-CMD (Fig. 7.8 and Fig. 7.4) than the magnitude of the respective similarity differences for pair CMB-CMC (Fig. 7.5 and Fig. 7.3).

The magnitude of the similarity differences changes for the two pairs of models, when it comes to relations with elements that are in even greater distances from the elements of QI_2 on the graph, such as the elements of relations with principal relation *covers*, *contains*, and *equal* (i.e., elements [$cv\ ct$], [$ct\ ct$] and [$e\ ct$]) (Fig. 7.11). Models CMA and CMD have the same similarity value for any number of holes and get a zero value when no holes are left (Fig. 7.8b,c), whereas models CMB and CMC keep increasing the similarity values of these relations, as more holes are dropped (Fig. 7.5b,c).

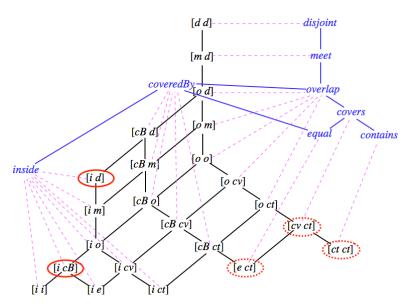


Figure 7.11 Highlighted elements. Elements of query QI_2 [id] and [icB] with solid ellipses and elements [cvct], [ctct] and [ect] from the relations in dataset D_{II} with dashed ellipses.

The rankings in this case depend on how the different costs in $\delta(t_1,t_2)$ contribute to the overall dissimilarity value (Eqn. 7.1). For both pairs of models, cost $\frac{Z_1}{\delta^{RRh-RRh}_{max}}$ steadily decreases in value, since the more holes are dropped, the less remaining $[cv\ ct]$, $[ct\ ct]$ and $[e\ ct]$ elements need to be transformed to the $[i\ d]$ and $[i\ cB]$ elements. For pair CMA-CMD, cost $\frac{Z_2}{\delta^{RRh-RR}_{max}}$ changes with big increments as more holes are dropped. The bigger value in each evaluation round balances out the decreases in $\frac{Z_1}{\delta^{RRh-RRh}_{max}}$ and keeps the similarity constant for almost all rounds (Fig. 7.8b,c). For pair CMB-CMC, however, the increments in $\frac{Z_2}{\delta^{RRh-RR}_{max}}$ are much smaller and do not balance out the decrease in cost $\frac{Z_1}{\delta^{RRh-RRh}_{max}}$, making the overall $\delta(t_1,t_2)$ value lower, as more holes are dropped. Consequently, the similarity value becomes higher in each evaluation round, imposing relations with more holes to rank lower than relations with fewer or no holes (Fig. 7.5b,c).

4. Overall, cost-assigning methods that more notably differentiate the cost of dropping a hole from the cost of change in the topological relation between regions produce more distinct and expected—according to the evaluation criteria—similarity rankings.

Such cost models also avoid increasing similarity values with number of holes that gets less and less equal to that of the query, when the relations under comparison are very different. So it may be suggested that such models, as for example CMA and CMD, are to be preferred for the similarity evaluation between relations with different number of holes. However, the attractiveness of the FDM for comparing such relations is its flexibility to adjust to any cost-assigning approach that is deemed appropriate according to the application at hand, without the fear of inexplicable or totally unexpected similarity results. This stability is provided by the basic infrastructure of the method, which is the

specific placement of the relations on the hybrid CNG. The infrastructure provides a benchmark against which the rankings of the various cost assigning techniques are criticized and with which the results may also be explained.

7.4 Summary

The assessment of the FDM as a similarity evaluation technique between relations with different numbers of holes and the interpretation of the similarity results were discussed. Three evaluation criteria set a benchmark against which four different cost models were examined for their rankings of various combinations of datasets and queries. None of the results of the cost models greatly deviates from the expected rankings set by the criteria. All models rank first those relations that have most of the three parameters—principal relation, refining relations, and number of holes—in common with the query. However, models with cost-assigning techniques that more solidly differentiate the costs of changing the topological structure of a region from the change that happens in the topological relation between two regions produce fewer controversial results. The controversy is generally generated with highly different relations whose elements are separated from the query's elements by longer distances on the hybrid CNG. Nevertheless, the results demonstrate that FDM is a reliable method for comparing relations with different numbers of holes, with some evidence for an ability to adapt to any cost model that may better accommodate the needs of an application.

Chapter 8 Conclusions

Versatile spatial data types are required for modeling complex spatial objects, often characterized by the presence of holes. To date, however, relations between hole-free regions have been the prevailing tools for spatial qualitative reasoning, limiting the ability to query and compare more complex spatial configurations. This thesis focused on the topological relations of single-holed regions and their composition inferences, examining how the presence of holes affects spatial reasoning and requires the development of new qualitative models and quantitative measurements for analysis. For advanced query answering, similarity evaluation among topological relations is a desirable asset of any geographic database. Using the topological relations for singleholed regions, a method of comparing relations featuring a multi-holed region complements the analytic framework developed in this thesis. Insights about the differences between reasoning over hole-free regions and reasoning over holed regions is expected to contribute to the design of future geographic information systems that more adequately process complex spatial phenomena and are better equipped to answer similarity database queries. This chapter summarizes the thesis (Section 8.1), describes the major findings and contributions (Section 8.2), and discusses possible future research avenues motivated by the results of this thesis (Section 8.3).

8.1 Summary of the Thesis

This thesis developed a formal framework for spatial reasoning with topological relations featuring single-holed regions using the eight possible relations between two hole-free regions defined in the 9-intersection (Egenhofer and Herring 1990). The Single-Holed Regions Model (S-HRM) comprises a set of 23 relations between a hole-free and a single-holed region, a set of 152 relations between two single-holed regions, composition tables of both sets, as well as their conceptual neighborhood graphs. Both sets involve relations that are jointly exhaustive and pairwise disjoint, differentiating between different topological structures imposed by the hole's presence. A custom-made checker verified that the spatial configuration of each relation and relation composition is valid according to Mackworth's (1977) network consistency constraints.

The compositions of the new sets of relations enabled qualitative inferences that were complemented with such quantitative measures as the *composition crispness* and the *cumulative frequencies*. The quantitative composition analysis supports the assumption that the presence of holes is responsible for more refined inference results.

Spatial phenomena of the real word often contain multiple holes. To examine relations featuring a multi-holed region, this thesis developed the Multi-Holed Region Model (M-HRM) and a method, called the Frequency Distribution Method (FDM), for comparing relations between a hole-free and a multi-holed region for their similarity. FDM builds on the representation of relations between a hole-free and a multi-holed region as multi-element relations, comprising as many elements—relations between the hole-free and a single-holed region—as the number of holes, considering each hole as if it were the only one.

Based on the conceptual neighborhood graph of the elements, FDM assesses the similarity between relations with a multi-holed region having the same number of holes as the complement of their dissimilarity. Dissimilarity is computed as the minimum cost of transforming one relation's elements to those of the other. The cost of the element transformation equals the distance of the two elements along the neighborhood graph. The transportation algorithm (Murty 1976; Strayer 1989) is employed to calculate the minimum cost in cases for which multiple transformations are possible. The similarity evaluation method demonstrated that among relations whose multi-holed regions have the same number of holes, the placement of the holes with respect to the hole-free region is as influencing for the similarity rankings as is the relation between the hole-free region and the host of the holes.

The FDM was, subsequently, modified to compare for their similarity relations featuring a multi-holed region when the regions have different numbers of holes. The difference in the number of holes was added to the representation of the relation with the fewer holes in the form of the relation between two hole-free regions created by the hole elimination. As a result, this thesis introduced the concept of simultaneously evaluating the change in the topological relation between regions with the change in a region's topology caused by eliminating a hole. To accommodate such changes, a hybrid graph is required so that the neighborhood graph of relations between a hole-free and a single-holed region is connected to the graph of relations between hole-free regions. The cost for hole elimination from each relation was added to the graph, enabling the method to compute the dissimilarity between relations with different numbers of holes.

There are different approaches for assigning a cost to hole elimination from a single-holed region. This thesis examined four different cost-assigning methods and their

similarity rankings for two synthetic datasets. The behavior of each model was studied against a baseline for the similarity results. The analysis of all rankings verified the expected behavior for the models and the dependability of the method, as none of the models produced completely extraordinary results. Differences between the rankings are intensified in cases where relations under comparison have the fewest characteristics in common and it is suggested that cost-assigning techniques that more distinctly differentiate the cost of hole elimination from that of change in the topological relation produce the most expected patterns in the similarity rankings.

8.2 Contributions and Major Findings

In this thesis, a comprehensive formal framework for qualitative reasoning with relations featuring single-holed regions was developed. This framework differs from previous approaches that either ignored the holes or treated them with techniques suited to relations for hole-free regions. This section discusses the major contributions of this work and the research findings that address the hypothesis of the thesis.

8.2.1 Contributions

The main contributions of this thesis are the following:

• A formal framework for relations featuring single-holed regions.

The Single-Holed Region Model is an exhaustive collection of binary relations involving one or two single-holed regions, explicit conceptual neighborhood graphs that define the sequence of relation change, and complete composition tables with valid inference results. New quantitative measures introduced with the model, help quantify the results of the composition analysis and reveal patterns related with the existence of the holes.

The set of 23 relations between a hole-free and a single-holed region is novel in its approach to cover relations between *multi-sorted* regions—regions originating from two different domains. Members of other explicit sets of relations are typically of the same type, namely either simple, hole-free regions, or complex regions with possible separations of the exterior as well as the interior. Through the converse property of the constituent relations, the concept of a converse relation still exists (i.e., the relations between a single-holed and a hole-free region, $t_{RhR} = \overline{t_{RRh}}$). However, there is neither an identity relation, nor is any relation reflexive, symmetric, or transitive. Therefore, the basic requirements are not met for a relation algebra. Lack of symmetric relations results in an asymmetric 23- t_{RRh} CNG. Many of these properties, such as an identity relation, symmetric relations, and transitivity are restored for the 152 relations between two singholed regions.

• A model for comparing relations between a hole-free and a multi-holed region for their similarity, which is independent of the number of holes.

The Frequency Distribution Method developed in this thesis is a model for comparing relations with multi-holed regions that is independent of the number of holes. Instead of explicitly accounting for each hole separately, multi-holed region relations are summarized with the Multi-Holed Region Model, which uses the frequencies of the relations' single-holed region elements. FDM departs from previous efforts involving the tedious enumeration of binary relations between the holes, the host, and the external regions. The utilization of the transportation algorithm for evaluating the minimum cost of relation transformation makes for an elegant approach to similarity evaluation that relies on the actual conceptual neighborhood graph of the relations for producing similarity values. Such an approach enables the quantification of the qualitative change of

relations traced on the neighborhood graph, providing a numerical representation of the differences and commonalities between relations with a multi-holed region.

• Concurrent evaluation of the change in the topological relation between regions with the change in the topological structure of a region.

This thesis introduced a novel way of thinking about relations similarity, with the concept of assessing in parallel the change in a binary with the change in a unary topological relation. Creating the hybrid graph that connects the neighborhood graph of single-holed region relations with that of the hole-free region relations, FDM makes up for a reference model of bringing the two changes together and enabling the similarity evaluation between relations with different numbers of holes. Such an approach takes into account the difference in the topological structure of the participating regions, in contrast to previous similarity evaluation methods that only assess either the numerical difference of regions as components of a relation, or the differences in the topological relations between the regions.

• A reliable similarity evaluation method for relations with different numbers of holes, adaptable to different cost-assigning methods.

The underlying structure of the method—the hybrid conceptual neighborhood graph—ensures that the FDM may be used with different cost-assigning methods for evaluating the cost for eliminating a hole, without resulting in utterly contentious similarity rankings. The cost values assigned to the edges that connect the single-holed region relations with their principal relations do not affect the neighboring of either the hole-free or the single-holed region relations. The reliance on the basic structure of the neighborhood graph allows flexibility on the cost assigning procedure for meeting the needs of an application.

8.2.2 Major Findings

The hypothesis of the thesis stated that taking into consideration the existence of holes in two-dimensional regions, imposes new constraints on the topological relations that can hold between such regions, and affects the inferences of the reasoning processes towards more refined results. This thesis addressed the hypothesis with the following findings:

• Reasoning with holed regions differs from reasoning with hole-free regions, producing more refined composition inferences.

The quantification of the composition analysis verifies that new constraints imposed on the topological relations from holes differentiate the reasoning with relations over holed regions from that with relations over hole-free regions. The numerical analysis indicates that acknowledging the holes, instead of favoring a hole-free region approach, leads to less ambiguous inferences for a little over 50% of the cases. Such results indicate that a model of topological relations of holed regions favors better decision making by providing more refined and accurate outcomes. In addition, the inference results demonstrate that the more holes that are initially involved in the composition of relations, the higher the percentage of unique relation results, which implies complete certainty in the outcome.

- Consideration of the placement of the holes during the similarity evaluation among relations with a multi-holed region enables more exact similarity rankings. In particular:
 - When the multi-holed regions contain equal numbers of holes, the holes' placement with respect to the hole-free region affects the similarity, as much as the relation between the hole-free region and the host does.

When comparing relations over regions with the same number of holes, FDM's basis is the structure provided by the neighborhood graph of the 23 single-holed region relations. Therefore, two relations that have different principal relations, but share all or most refining relations, may be regarded as more similar than relations that share the same principal relation, but differ in all or most refining relations. Responsible for this similarity evaluation result is the shorter distance between relations that share the same refining relations on the graph. Conversely, longer distances separate relations with different refining relations. The similarity rankings in this thesis demonstrate that taking into consideration the placement of the holes, instead of solely relying on the overall relation between the hole-free and the host region, returns more detailed similarity rankings.

- When the multi-holed regions have different numbers of holes, the placement of the holes matters even more for the similarity rankings, especially when the relations under comparison are topologically very different.

The numerical similarity differences among the various cost-assigning models demonstrated that the placement of the holes with respect to the hole-free region affects the similarity ranking more strongly when the relations under comparison have fewer or no elements in common. Longer distances on the hybrid graph between relations that do not share common refining relations more explicitly display the difference among various cost models, resulting in more prominent differences in the numerical rankings of such relations. The similarity rankings examined in this thesis suggest that the more distinct the costs assigned to cross-level edges are from the costs of the distances between intralevel relations, the more expected are the rankings.

These findings verify the hypothesis that spatial reasoning about relations over regions with holes differs from reasoning about relations over hole-free regions. The additional constraints imposed by the holes lead to sets of more fine-grained relations, more refined composition inferences, and more accurate similarity comparisons. The topological relation between a hole-free region and the host region of the holes is a key controlling parameter of the inference mechanisms. However, it is the relations between the hole-free region and the holes that refine the composition results and finalize the similarity rankings, especially for relations over multi-holed regions with very different hole placements.

8.3 Future Work

This thesis presented a formal framework for reasoning with single-holed region relations and used it as a basis for developing a similarity assessment method for multi-holed region relations. The results open new research avenues and anticipate further development of the reasoning tools presented in this thesis. This section first discusses some research alternatives about the S-HRM and then provides guidance about expanding the range of the FDM.

8.3.1 Alternatives to the S-HRM

The S-HRM developed in this thesis is based on two principal assumptions: (1) the relation between the generalized region B^* and its hole B_H is always $t(B^*, B_H) = contains$, and (2) the neighborhood graph of the eight relations between two hole-free regions captures only the A-neighbors, so that if a region expands from the *inside* relation it can be transformed to *coveredBy* in one step, but not to *equal*. Conversely, if a region is

reduced in size when it *contains* the other region, it can be converted to *covers*, but not to *equal*. It is possible, however, to allow for different or additional relations to hold between the generalized region and the hole, which would result in different relation sets than the ones derived in this thesis. The CNGs of the sets are also modified with the addition or elimination of certain edges, due to different topological constraints for the different host-to-hole relations.

8.3.1.1 Relaxing the Definition of the Single-Holed Region

Chapter 3 defined that a single-holed region B comprises the generalized region B^* and the hole B_H , which is completely *inside* B^* . A less restrictive model allows the hole to be *coveredBy* (Fig. 8.1a) or even *equal* to B^* (Fig. 8.1b), leading to different semantics of a region with a hole (Egenhofer *et al.* 1994).

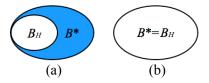


Figure 8.1 Single-holed region B in different relations with the hole B_H : (a) B_H coveredBy B^* or (b) B_H equal to B^* .

According to the constraint of the hole placement, the sets of relations between a hole-free and a single-holed region, or between two single-holed regions, would be different than the ones developed in S-HRM. Seven different sets of t_{RRh} s may be developed, allowing either one of the three different relations between the hole and the generalized region or combinations thereof: (B_H inside B^*), (B_H coveredBy B^*), (B_H equal B^*), (B_H inside or coveredBy B^*), (B_H inside or equal B^*) and (B_H inside or coveredBy or equal B^*).

The relaxation of the hole placement definition is interesting, because it implies a variation in the set of realizable relations. For instance, the set of t_{RRh} s for which the hole is always coveredBy the generalized region has the same number of relations as the set developed in this thesis, namely 23. However, five relations— $[i\ e]$, $[i\ cv]$, $[i\ ct]$, $[cB\ ct]$, and $[e\ ct]$ —have been replaced by five different ones (Fig. 8.2)— $[m\ m]$, $[cB\ cB]$, $[cB\ e]$, $[e\ cv]$, and $[cv\ cv]$ —due to the new limitations imposed by the coveredBy relation, on the relation between the hole-free region and the hole. Accordingly, the conceptual neighborhood graph for the 23- t_{RRh} based on the coveredBy relation is different (Fig. 8.3).

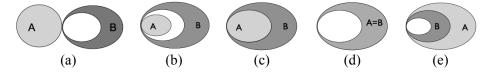


Figure 8.2 The five new t_{RRh} relations when the hole is *coveredBy* the generalized region: (a) $[m \ m]$, (b) $[cB \ cB]$, (c) $[cB \ e]$, (d) $[e \ cv]$, and (e) $[cv \ cv]$.

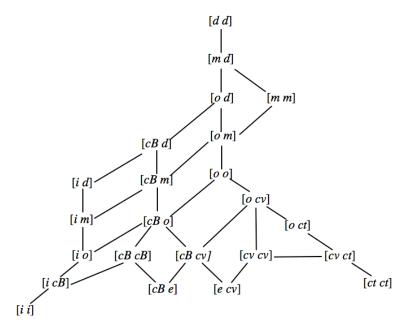


Figure 8.3 The 23- t_{RRh} CNG for which B_H coveredBy B^* .

It is expected that allowing combinations of two or all three relations between the hole and the generalized region will result in sets of relations larger than 23. For example,

if both *inside* and *coveredBy* are allowed, the set comprises 28 relations. The increased sets of relations produces larger composition tables, and inferences are expected to be less refined than the ones derived in this thesis, due to fewer topological constraints.

Accordingly, the CNGs will expand, and the FDM for evaluating similarity between relations featuring multi-holed regions will have to be based on larger tables of costs. Further work is then required for providing all possible topological configurations of relations with single-holed regions and making the S-HRM a more complete framework for extracting composition inferences. Such a complete framework would be used with the FDM for similarity rankings among a wider range of holed-region relations.

8.3.1.2 Relaxing the Scaling Transformation Requirements

The conceptual neighborhood graphs developed in this thesis support connectivity of the A-neighbors only. In A-neighbors, the scaling transformation is responsible for restraining connectivity of relation *equal* with relations *coveredBy* and *covers* only. Such a constraint forces the expansion or reduction of a region in relations *inside* and *contains*, respectively, to be a two-step procedure in order to change to *equal*. However, if that constraint is relaxed, the scaling transformation from *inside* or *contains* to *equal* can be a one step procedure. This is the scenario of two regions being concentric, and one of them expanding or collapsing to equality with the other region. Such a neighborhood captures the C-neighbors (Egenhofer and Al-Taha 1992; Freksa 1992).

Accordingly, if the regions have the same size and shape, the translation or rotation transformation between relations *overlap* and *equal* can also be reduced to one step, instead of two—through *coveredBy* or *covers*—and vise versa, as it is in the A-

neighborhood graph. The one step transformations result in the B-neighbors (Egenhofer and Al-Taha 1992; Freksa 1992).

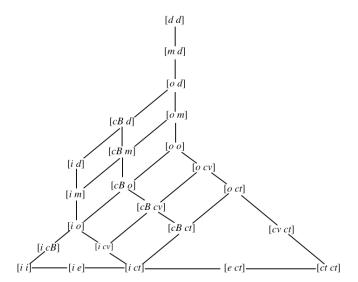


Figure 8.4 The 23- t_{RRh} CNG featuring the C-neighbors.

Further work is needed in order to evaluate how the change in the neighborhoods of the 23- t_{RRh} CNG and the table of costs affect the similarity results. It is expected that

relations with *equal* in any of their constituent relations would now rank closer to the *inside* and *contains* constituent relations, rather than *coveredBy* and *covers*, which was the case for the similarity results examined in this thesis. Analogous work would have to examine the effects on the similarity rankings of the B-neighbors, whereby translation and rotation may directly change the relation from *overlap* to *equal*, and vise versa.

8.3.2 Relations of Holed Regions on the Sphere

The regions participating in the relations developed in S-HRM are embedded in the two-dimensional plane. There are many GIS applications, however, that deal with phenomena that spread across the entire globe, such as a national minority's world-wide distribution or the spatio-temporal spread of a disease over two or more continents. Such applications require semantic models of spatial relations proper for the sphere, the two-dimensional surface embedded in the three-dimensional space, and its particular properties (Usery 2002).

While most models of topological relations apply to regions embedded either in the two- or three-dimensional space, a set of 11 relations based on the 9-intersection, has been developed for the sphere \mathbb{P}^2 , where $\mathbb{P} \subset \mathbb{R}$ such that \mathbb{P} is connected and $\min(\mathbb{P})=\max(\mathbb{P})$ (Egenhofer 2007). For applications such as monitoring the long-term change of the position of the ozone hole in the earth's atmosphere, a model of relations over holed regions on the sphere is of value. Further work is required then to develop the set of relations between a hole-free and a single-holed region and between two single-holed regions, using the set of 11 sphere relations, in the same way such sets were developed for the plane, in this thesis. It is expected that the corresponding sets will be larger, since there are three more binary relations—attach, entwined, and embrace—

realizable only on the sphere, in addition to the eight realizable both on the plane and the sphere.

In Chapter 3, a region with a hole on the plane was defined as a spatial region with a separated exterior. The separations are the *outer exterior*, an unbounded set separated from the interior of the region by the *outer boundary*, and the *inner interior*, bounded from the *inner boundary* of the region. The inner interior fills the region's hole. On the sphere, however, this separation is not so clear, since the outer exterior is now bounded as well, and homeomorphic to a half-sphere. The fact that both the inner and the outer exterior of the region on the sphere are bounded by the boundaries of the region and are homeomorphic to half-spheres makes is difficult to distinguish which one is the hole and which one is the exterior, especially in the case where the region's extend surrounds the sphere (Fig. 8.5). Furthermore, in such a depiction, both regions could either be in relation *meet* or *attach* with the region.

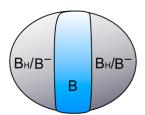


Figure 8.5 An ambiguous depiction of the hole in a region that sits on the sphere.

It is, therefore, crucial for the derivation of the sets of relations with a single-holed region on the sphere to enforce some constraint about the relation between the region and its hole, as for example the one used in this thesis for regions in the plane—the hole is always *inside* the region. Then it is only the region's outer exterior that is in relation *attach* with the region. This kind of constraint ensures that the relations between a hole-free region and the generalized region or the hole of the single-holed region can be

clearly specified on the sphere. Absence of such a constraint would create yet another visual confusion, as the mere demarcation of the holed region on the sphere using Jordan curves does not suffice for distinguishing the hole, from the outer exterior of the region.

8.3.3 Extending the FDM

The similarity evaluation method developed in this thesis enables the comparison of relations with one multi-holed region. There are certain aspects of the method that need to be further enhanced in order for the FDM to provide more complete answers to database similarity queries. Improvements include enabling the method to compare relations where both holes are multi-holed, to differentiate holes according to their distinctive roles, and to possibly group holes according to their properties. In addition, FDM as developed in this thesis, offers numerical similarity results. It is highly desirable to examine how these numbers are matched against natural language expressions that people use when asked to compare relations for their similarity.

8.3.3.1 M: N-Holed Regions Relations

In case both regions in a relation have holes, a technique analogous to the one developed in this thesis, the M-HRM, for summarizing the relations in hole placement frequencies needs to be adopted. With the sets of relations for single-holed regions, two possibilities exist for summarizing the m: n-holed relations: (1) using the set of $23-t_{RRh}$ s, one region at a time is considered as hole-free and the t_{RRh} s with the other region are recorded in frequency vectors (Fig. 8.6), or (2) using the set of $152-t_{RhRh}$ s, both regions are considered as single-holed and the t_{RhRh} s between them are recorded for all pairs of holes (Fig. 8.7).

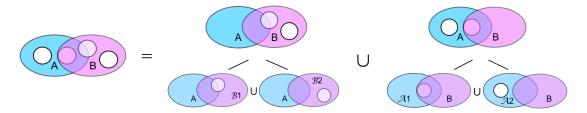


Figure 8.6 The 2-holed region to 2-holed region relation between A and B as the union of two multi-element relations, each of which may be broken down to two elements.

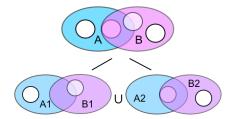


Figure 8.7 The two t_{RhRh} s that form the 2-holed to 2-holed region relation between A and B.

With the set of 23 t_{RRh} s, two frequency vectors V_A and V_B are defined for each of the two relations under comparison. As in the 2-holed region to 2-holed region relation in the example (Fig. 8.6a), this duplexity is necessary since each of the regions, A and B, are in turn considered as hole-free and their relations with the holes of the other region are recorded (Fig. 8.6b and 8.6c). The deviation from the method developed in this thesis is that rather than one cost, two costs—c1 and c2—of transformations of the two vectors summarizing one relation into the vectors of the other relation, are recorded. The final cost c of the relation transformation then is the sum of these two costs.

Using the set of 152- t_{RhRh} , the relation in the example would be the union of two single-holed to single-holed region relations (Fig. 8.7). The frequencies would then be recorded in an analogous way and the costs for relation transformations would have to be acquired by the 152- t_{RhRh} CNG. It is an open question whether the two methods based on the two different relation sets produce the same or different similarity results.

8.3.3.2 Holes with Different Roles

The relations compared in this thesis are over regions in which all holes have the same role. The role of a hole may have a two-fold interpretation—it may have to do with its relative importance in comparison with the rest of the holes, or it may be related to its semantics. In the first case, all holes may be of the same nature, for example, coverage holes in a network setting, various lakes in the same plain, patches of different land use in an agricultural division, or oil deposits in an underground geologic formation. In the second case, holes or their content may belong to ontologically different categories. For example, in an urban environment, holes may be open-air recreational parks, state protected land, a chemically contaminated and restricted area, or a natural discontinuity such as a lake or a hill.

In case the holes are all of the same nature, the importance of each hole is relative. It may be measured by the size of the hole, or any of the other properties of the hole or its content that is appropriate for the application. In the example of the oil-baring geologic formation, the bigger the size of a hole, the higher its economic value. In other cases the importance of a hole may be related to the hole's relative placement or neighboring with an accessibility point. The cost of relation transformations examined in this thesis, did not take into consideration different weights for holes. To enable the differentiation of holes according to their importance, a weight factor w_{ij} needs to be added to the calculation of the cost (Eqn. 8.1). The values of the weight factor and the property to which it is associated are depended on the application under consideration and are the subject of further research.

$$Z = \min(\sum_{i=1}^{p} \sum_{j=1}^{n} w_{ij} c_{ij} X_{ij})$$
(8.1)

However, if the holes vary semantically, defining a weight factor is a more complicated, multi-variable procedure. The mere addition of the weight factor is not enough, as the weights themselves would be attributed to properties of holes of different nature, which may not be comparable. It is left for future work to decipher the appropriate linking property that would allow semantically different holes to be attributed a weight from the same scale.

8.3.3.3 Hole Generalization

Examining holes for their properties may also allow for certain groupings of holes. For example, if holes are judged for their size, smaller holes that are in proximity (Fig. 8.8a) may be seen as one single hole (Fig. 8.8b). In such a case, when compared with other relations, it is interesting to examine whether relations over regions with grouped holes and their ungrouped versions, share the same similarity against other relations. Groups of holes may also have different implications for map generalization. When moving to smaller scale map representations, the size of map objects is compared against a certain threshold. For example, holes smaller than the threshold are eliminated. A different approach is to replace a number of neighboring small holes with a single bigger one hole, much like in Gestalt theory of perception (Wertheimer 1923), which would meet the size threshold. The single hole would need to cover the spatial extent of the group of holes. What is necessary then, is a method to determine the relation between the newly formed conglomeration and the rest of the regions involved in a topological relation.

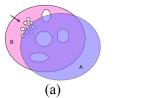




Figure 8.8 Hole generalization: (a) a number of small holes getting replaced by (b) a single hole.

8.3.4 Introducing Human Language Expressions

With the development of any new GIS application, the goal is to facilitate interaction between user and system in a way that feels intuitive and natural to the user. Therefore, for FDM to be a complete similarity evaluation method, it needs to be calibrated according human subjects' similarity assessments, and to incorporate natural-language expressions in the results' presentation.

To calibrate the method, a series of human subject tests need to be performed in order to examine how people rank the relations in the same datasets evaluated by FDM. Examination of the test results will verify whether or not people judge similarity between relations with multi-holed regions with the same criteria as FDM does. It is also desirable to enable FDM to express the similarity results with natural-language expressions much like people do. To realize this, various similarity percentage ranges need to be matched against certain assessment terms that people use, such as *almost identical*, *very similar*, *similar*, and *not similar*. This matching resembles the effort to extend the terminology of expressing topological relations involving regions with broad boundaries, since the firm terms that concern regions with crisp boundaries cannot cover vague spatial objects (Bejaoui *et al.* 2008). The difference in the case of FDM is that ranges of absolute numerical similarity values will have to be matched with vague natural language terms. Achieving so, will facilitate a human-computer system collaboration that resembles more

the person-to-person communication and will contribute to the development of more interactive, innovative systems.

REFERENCES

- Ahmed, N., Kanhere, S.S., and Jha, S. (2005). The Holes Problem in Wireless Sensor Networks: A Survey. *ACM SIGMOBILE Mobile Computing and Communications Review* **9**(2): 4.
- Alexandroff, P. (1961). *Elementary Concepts of Topology*. Dover Publications New York, NY.
- Bejaoui, L., Pinet, F., Schneider, M., and Bédard, Y. (2008). An Adverbial Approach for the Formal Specification of Topological Constraints Involving Regions with Broad Boundaries. In Li, Q., Spaccapietra, S., Yu, E., and Olivé, A. (eds) *Proceedings of the 27th International Conference on Conceptual Modeling*, Barcelona, Spain, *Lecture Notes in Computer Science*, **5231** pp. 383-396.
- Bennett, B. (1998). Determining Consistency of Topological Relations. *Constraints* **3**(2-3): 213-225.
- Bertamini, M. (2006). Who Owns the Contour of a Visual Hole? *Perception* **35**(7): 883-894.
- Bertamini, M. and Croucher, C.J. (2003). The Shape of Holes. *Cognition* 87(1): 33-54.
- Bittner, T. and Stell, J. (2000). Rough Sets in Approximate Spatial Reasoning. In Second International Conference on Rough Sets and Current Trends in Computing, Lecture Notes in Computer Science (LNCS), 2005 pp. 445-453 Berlin: Springer-Verlag.
- Bruns, H. and Egenhofer, M. (1996). Similarity of Spatial Scenes. In Kraak, M. J. and Molenaar, M. (eds) *Seventh International Symposium on Spatial Data Handling* (*SDH*'96), Delft, The Netherlands, pp. 4A 31-42.
- Burrough, P. and Frank, A. (1996). *Geographic Objects with Indeterminate Boundaries*. Gisdata 2, Taylor & Francis, London.
- Casati, R. and Varzi, A. (1994). *Holes and Other Superficialities*. The MIT Press Cambridge, MA.
- Chang, S., Jungert, E., and Li, Y. (1989). The Design of Pictorial Databases Based Upon the Theory of Symbolic Projections. In Buchmann, A., Günther, O., Smith, T., and Wang, Y. (eds) First Symposium on Design and Implementation of Large Spatial Databases, Lecture Notes in Computer Science, 409 pp. 303-323 Springer-Verlag, New York, NY.

- Clarke, B. (1981). A Calculus of Individuals Based on Connection. *Notre Dame Journal of Formal Logic* **22**(3): 204-218.
- Clarke, B. (1985). Individuals and Points. *Notre Dame Journal of Formal Logic* **26**(1): 61-75.
- Clementini, E. and Di Felice, P. (1996a). An Algebraic Model for Spatial Objects with Indeterminate Boundaries. In Burrough, P. and Frank, A. (eds) *Geographic Objects with Indeterminate Boundaries*. Taylor & Francis, Bristol, PA: 155-170.
- Clementini, E. and Di Felice, P. (1996b). A Model for Representing Topological Relationships between Complex Geometric Features in Spatial Databases. *Information Sciences* **90**(1-4): 121-136.
- Clementini, E., Di Felice, P., and Califano, G. (1995). Composite Regions in Topological Queries. *Information Systems* **20**(7): 579-594.
- Clementini, E., Di Felice, P., and Van Osterom, P. (1993). A Small Set of Formal Topological Relationships Suitable for End-User Interaction. In Abel, D. and Ooi, B. (eds) *Third International Symposium on Advances On Spatial Databases, SSD'93*, Singapore, *Lecture Notes in Computer Science*, **692** pp. 277-295 Springer, Berlin, Germany.
- Cohn, A. (1997). Qualitative Spatial Representation and Reasoning Techniques. In Brewka, G., Habel, C., and Nebel, B. (eds) *KI-97: Advances in Artificial Intelligence*. Springer-Verlag. Freiburg, Germany **1303**: 1-30.
- Cohn, A., Bennett, B., Gooday, J., and Gotts, N. (1997). Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. *GeoInformatica* 1: 275-316.
- Cohn, A., Bennett, B., Gooday, J., Gotts, N., and Stock, O. (1997). Representing and Reasoning with Qualitative Spatial Relations. In *Spatial and Temporal Reasoning*. Kluwer Academic Publishers, Dordrecht: 97-134.
- Cohn, A. and Gotts, N. (1996a). The 'Egg-Yolk' Representation of Regions with Indeterminate Boundaries. In Burrough, P. and Frank, A. (eds) *Geographic Objects with Undetermined Boundaries*, pp. 171-187, Taylor and Francis, Bristol, PA
- Cohn, A. and Gotts, N. (1996b). Representing Spatial Vagueness: A Mereological Approach. In *Principles of Knowledge Representation and Reasoning: Proc. of 5th International Conference (KR96)*, San Francisco, Morgan Kaufmann.
- Cohn, A. and Hazarika, S. (2001). Qualitative Spatial Representation and Reasoning: An Overview. *Fundamenta Informaticae* **46**(1-2): 2-32.

- Cohn, A. and Renz, J. (2007). Qualitative Spatial Representation and Reasoning. In Harmele, F., Lifschitz, V., and Porter, B. (eds) *Handbook of Knowledge Representation*, Elsevier.
- Cohn, A. and Varzi, A. (2003) Mereotopological Connection. *Journal of Philosophical Logic*, **32**: 357-390
- Dantzig, G. (1963) *Linear Programming and Extensions*. Princeton, NJ: Princeton University Press.
- Dantzig, G. and Thapa, M. (1997) *Linear Programming 1: Introduction*. Springer, New York
- Davis, E., Gotts, N., and Cohn, A. (1999). Constraint Networks of Topological Relations and Convexity. *Constraints* **4**(3): 241-280.
- Dilo, A., De By, R.A., and Stein, A. (2007). A System of Types and Operators for Handling Vague Spatial Objects. *International Journal of Geographical Information Science* **21**(4): 397-426.
- Du, S., Qin, Q., Wang, Q., and Ma, H. (2008). Reasoning About Topological Relations between Regions with Broad Boundaries. *International Journal of Approximate Reasoning* **47**: 219-232.
- Egenhofer, M. (1989). A Formal Definition of Binary Topological Relationships. In Litwin, W. and Schek, H. (eds) *Third International Conference on Foundations of Data Organization and Algorithms (FODO)*, Paris, France, *Lecture Notes in Computer Science*, **367** pp. 457-472.
- Egenhofer, M. (1991). Reasoning About Binary Topological Relations. In Gunther, O. and Scheck, H. (eds) *Second Symposium on Large Spatial Databases*, Zurich, Switzerland, *Lecture Notes in Computer Science*, **525** pp. 143-160 Springer-Verlag.
- Egenhofer, M. (1993). A Model for Detailed Binary Topological Relationships. *Geomatica* **47**(3&4): 261-273.
- Egenhofer, M. (1994). Deriving the Composition of Binary Topological Relations. *Journal of Visual Languages and Computing* **5**(2): 133-149.
- Egenhofer, M. (1997). Query Processing in Spatial-Query-by-Sketch. *Journal of Visual Languages and Computing* **5**(1): 133-149.
- Egenhofer, M. (2005). Spherical Topological Relations. *Journal on Data Semantics* III: 25-49.

- Egenhofer, M. and Al-Taha, K. (1992). Reasoning About Gradual Changes of Topological Relationships. In Frank, A., Campari, I., and Formentini, U. (eds) *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, Pisa, Italy, *Lecture Notes in Computer Science*, **639** pp. 196-219 Springer-Verlag.
- Egenhofer, M., Clementini, E., and Felice, P. (1994). Topological Relations between Regions with Holes. *International Journal of Geographical Information Systems* **8**(2): 129-142.
- Egenhofer, M. and Franzosa, R. (1991). Point-Set Topological Spatial Relations. *International Journal of Geographical Information Science* **5**(2): 161-174.
- Egenhofer, M. and Franzosa, R. (1995). On the Equivalence of Topological Relations. *International Journal of Geographical Information Systems* **9**(2): 133-152.
- Egenhofer, M. and Herring, J. (1990a). A Mathematical Framework for the Definition of Topological Relationships. In Brassel, K. and Kishimoto, H. (eds) *Fourth International Symposium on Spatial Data Handling*, Zurich, Switzerland, pp. 803-813.
- Egenhofer, M. and Herring, J. (1990b) Categorizing Binary Topological Relations between Regions, Lines, and Points in Geographic Databases Department of Surveying Engineering University of Maine Orono
- Egenhofer, M. and Mark, D. (1995). Naive Geography. In Frank, A. and Kuhn, W. (eds) *Third European Conference on Spatial Information Theory (COSIT* '95), Semmering, Austria, pp. 1-12.
- Egenhofer, M. and Sharma, J. (1993). Assessing the Consistency of Complete and Incomplete Topological Information. *Geographical Systems* **1**(1): 47.
- Egenhofer, M. and Vasardani, M. (2007). Spatial Reasoning with a Hole. In Winter, S., Duckham, M., Kulik, L., and Kuipers, B. (eds) *Spatial Information Theory–8th International Conference*, *COSIT 2007*, Melbourne, Australia, *Lecture Notes in Computer Science*, **4736** pp. 303-320 Springer, New York.
- Erwig, M. and Schneider, M. (1997). Vague Regions. In Scholl, M. and Voisard, A. (eds) 5th International Symposium on Advances in Spatial Databases (SSD'97), Lecture Notes in Computer Science, 1262 pp. 289-320 Springer, Heidelberg.
- ESRI (1995) ESRI Spatial Database Engine (SDE)
- Forbus, K. (1984). Qualitative Process Theory. Artificial Intelligence 24(1-3): 85.

- Frank, A. (1992). Qualitative Reasoning About Distances and Directions in Geographic Space. *Journal of Visual Languages and Computing* **3**(4): 343-371.
- Frank, A. (1995). Qualitative Spatial Reasoning: Cardinal Directions as an Example. *International Journal of Geographical Information Systems* **10**: 269-290.
- Frank, A. and Kuhn, W. (1986). Cell Graphs: A Provable Correct Method for the Storage of Geometry. *Proc. 2nd Intl. Symposium on Spatial Data Handling, Seattle*: 411.
- Freeman, J. (1975). The Modeling of Spatial Relations. *Computer Graphics and Image Processing* **4**(2): 156-171.
- Freksa, C. (1992). Temporal Reasoning Based on Semi-Intervals. *Artificial Intelligence* **54**: 199-227.
- Freksa, C. and Röhrig, R. (1993). Dimensions of Qualitative Spatial Reasoning. In Carreté, P. and Singh, M.G. (eds) *Qualitative Reasoning in Decision Technologies QUARDET* '93, CIMNE Barcelona, pp. 483-492.
- Gerevini, A. and Renz, J. (2002). Combining Topological and Size Information for Spatial Reasoning. *Artificial Intelligence* **137**(1-2): 1-42.
- Gotts, N. (1994). How Far Can We C? Defining a Doughnut Using Connection Alone. In Doyle, J., Sandewall, E., and Torasso, P. (eds) *4th Internation Conference on Principles of Knowledge Representation and Reasoning (KR94)*, San Francisco, CA, pp. 246-257 Morgan Kaufmann.
- Gotts, N. (1996). Topology from a Single Primitive Relation: Defining Topological Properties and Relations in Terms of Connection. Technical Report, 96-23, University of Leeds, School of Computer Studies.
- Goyal, R. (2000) Similarity Assessment for Cardinal Directions between Extended Spatial Objects. PhD Thesis, Spatial Information Science and Engineering, University of Maine, Orono, ME
- Hayes, P. (1979). The Naive Physics Manifesto. In Michie, D. (ed) *Expert Systems in the Micro-Electronic Age*. Edinburgh University Press, UK: 242-270.
- Hernández, D. (1991). Relative Representations of Spatial Knowledge. In Mark, D. and Frank, A. (eds) *Cognitive and Linguistic Aspects of Geographic Space*. Kluwer Academic, Dordrecht: 373-385.
- Hernández, D. (1994). *Qualitative Representation of Spatial Knowledge. Lecture Notes in Artificial Intelligence*, **804**, Springer-Verlag Berlin, Heidelberg, New York.

- Janowicz, K. (2006) Sim-DL: Towards a Semantic Similarity Measurement Theory for the Description Logic ALCNR in Geographic Information Retrieval. In Meersman, R., Tari, Z., and Herrero, R. (eds) *On the Move to Meaningful Internet Systems 2006*, SeBGIS 2006, OTM Workshops 2006, Lecture Notes in Computer Science, 4278, pp. 1681-1692, Springer Verlag, Berlin
- Jiang, J. and Worboys, M. (2006). Specifying Events by Changes in Topological Properties (Extended Abstract). In 4th Internation Conference on Geographic Information Science (GIScience 2006).
- Jiang, J. and Worboys, M. (2008). Detecting Basic Topological Changes in Sensor Networks by Local Aggregation. In *Proceedings of ACM-GIS 2008*, pp. 13-22.
- Jiang, J. and Worboys, M. (2009). Event-Based Topology for Dynamic Planar Areal Objects. *International Journal of Geographical Information Science* **23**(1): 33-60.
- Kainz, W., Egenhofer, M., and Greasley, I. (1993). Modeling Spatial Relations and Operations with Partially Ordered Sets. *International Journal of Geographical Information Systems* **7**(3): 215-229.
- Klippel, A., Worboys, M., and Duckham, M. (2008). Identifying Factors of Geographic Even Conceptualization. *International Journal of Geographical Information Science* **22**(2): 183-204.
- Kuhn, H. W. (1955) Variants of the Hungarian Method for Assignment Problems. *Naval Research Logistics Quarterly*, **3**: 253-258.
- Kuipers, B. (1994). *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. MIT Press.
- Kurata, Y. (2008). The 9⁺-Intersection: A Universal Framework for Modeling Topological Relations. In Cova, T., Beard, K., Frank, A., and Goodchild, M. (eds) *Fifth International Conference on Geographic Information Science-GIScience* 2008, Park City, UT, *Lecture Notes In Computer Science*, **5266** pp. 181-198 Springer-Verlag Berlin, Heidelberg.
- Kurata, Y. and Egenhofer, M. (2007). The 9⁺-Intersection for Topological Relations between a Directed Line Segment and a Region. In Gottfried, B. (ed) *1st International Symposium for Behavioral Monitoring and Interpretation*, pp. 62-76.
- Ladkin, P.B. and Maddux, R. (1994). On Binary Constraint Problems. *Journal of the Association of Computing Machinery* **41**(3): 435-469.

- Lehmann, F. and Cohn, A.G. (1994). The Egg/Yolk Reliability Hierarchy: Semantic Data Integration Using Sorts with Prototypes. In *Proceedings of the 3rd. International Conference on Information and Knowledge Management*, Gaithersburg, MD: ACM, pp. 272-279.
- Lewis, D. and Lewis, S. (1970). Holes. Australasian Journal of Philosophy 48(2): 206-212.
- Lewis, D. and Lewis, S. (1996). Casati and Varzi on Holes. *Philosophical Review* **105**: 77-79.
- Li, S. (2006). A Complete Classification of Topological Relations Using the 9-Intersection Method. *International Journal of Geographical Information Science* **20**(6): 589-610.
- Li, S. (2007). Combining Topological and Directional Information for Spatial Reasoning. In Veloso, M. (ed) 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, pp. 435-440.
- McKenney, M., Pauly, A., Praing, R., and Schneider, M. (2007). Local Topological Relationships for Complex Regions. In *10th International Symposium on Spatial and Temporal Databases*, pp. 203-220.
- McKenney, M., Praing, R., and Schneider, M. (2008). Deriving Topological Relationships between Simple Regions with Holes. In Ruas, A. and Gold, C. (eds) *Headway in Spatial Data Handling*. Springer Berlin Heidelberg: 521-531.
- Mackworth, A. (1977). Consistency in Networks of Relations. *Artificial Intelligence* **8**: 99-118.
- Mark, D. (1993). Human Spatial Cognition. In Medyckyj-Scott, D. and Hearnshaw, M.H. (eds) *Human Factors in Geographical Information Systems*. Belhaven Press: 51.
- Mark, D. and Freundschuh, S. (1995). Spatial Concepts and Cognitive Models for Geographic Information Use. In Nyerges, T., Mark, D., Laurini, R., and Egenhofer, M. (eds) Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems. Kluwer Academic Publishers, Dordrecht: 21.
- Montello, D. (1993). Scale and Multiple Psychologies of Space. In Frank, A. and Campari, I. (eds) *Spatial Information Theory: A Theoretical Basis for GIS*. Springer-Verlag Berlin **716**: 312-321.
- Munkres, J (1957). Algorithms for the Assignment and Transportation Problems. *Journal* of the Society and Applied Mathematics, **5**(1): 32-38.

- Munkres, J. (1975). Topology: A First Course. Prentice-Hall.
- Murty, K. (1976). *Linear and Combinatorial Programming*. John Wiley & Sons, Inc. New York.
- Nedas, K. (2006) *Semantic Similarity of Spatial Scenes*. Ph.D. Thesis, University of Maine, Orono, ME (http://www.library.umaine.edu/theses/pdf/NedasKA2006.pdf)
- Nelson, R. and Palmer, S.E. (2001). Of Holes and Wholes: The Perception of Surrounded Regions. *Perception* **30**(10): 1213-1226.
- OGC (1999) OGC Abstract Specifications OpenGIS Consortium (OGC) http://www.opengis.org/techno/specs.htm (accessed on 12/25/2007)
- OGC (2005) OpenGIS Geography Markup Language (GML) Encoding Standard http://www.opengeospatial.org/standards/gml (accessed on 12/25/2007)
- Paiva, J. (1998) Topological Equivalence and Similarity in Multi-Representation Geographic Databases Spatial Information Science and Engineering University of Maine Orono, ME
- Pequet, D. and Ci-Xiang, Z. (1987). An Algorithm to Determine the Directional Relationship between Arbitrary-Shape Polygons in the Plane. *Pattern Recognition* **20**(1): 65-76.
- Pawlak, Z. (1994). Rough Sets: Present State and Further Prospects. In Lin, T. and Wildberger, A. (eds) *Third International Workshop on Rough Set and Soft Computing (RSSC'94)*, San Jose, CA, pp. 3-5.
- Randell, D., Cohn, A., and Cui, Z. (1992a). Computing Transitivity Tables: A Challenge for Automated Theorem Provers. In Kapur, D. (ed) *Proceedings of the 11th International Conference on Automated Deduction (CADE-11)*, *Lecture Notes in Computer Science*, **607** pp. 786 Springer-Verlag.
- Randell, D., Cui, Z., and Cohn, A. (1992b). A Spatial Logic Based on Regions and Connection. In Nebel, B., Rich, C., and Swartout, W. (eds) *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, San Mateo, pp. 165-176 Morgan Kaufmann.
- Renz, J. (2002). Qualitative Spatial Reasoning with Topological Information, Lecture Notes in Artificial Intelligence, 2293 Springer-Verlag Berlin, Heidelberg.

- Rodríguez, A. and Egenhofer, M. (2000). A Comparison of Inferences About Containers and Surfaces in Small-Scale and Large-Scale Spaces. *Journal of Visual Languages and Computing* **11**(6): 639-662.
- Roy, A. and Stell, J. (2001). Spatial Relations between Indeterminate Regions. *International Journal of Approximate Reasoning* **27**: 205-234.
- Schneider, M. and Behr, T. (2006). Topological Relationships between Complex Spatial Objects. *ACM Transactions on Database Systems (TODS)* **31**(1): 39-81.
- Smith, T. and Park, K. (1992). Algebraic Approach to Spatial Reasoning. *International Journal of Geographical Information Science* **6**(3): 177-192.
- Spanier, E. (1966). Algebraic Topology. McGraw-hill Book Company, NY.
- Stefanidis, A. and Nittel, S. (2004). Geosensor Networks. CRC Press.
- Strayer, J. (1989). *Linear Programming and Its Applications*. Springer-Verlag New York.
- Tarski, A. (1941). On the Calculus of Relations. *Journal of Symbolic Logic* **6**(3): 73-89.
- Tversky, A. (1977). Features of Similarity. *Psychological Review* **84**: 327-352.
- Usery, E. (2002) University Consortium for Geographic Information Science Research Priorities: Global Representation and Modeling (http://www.ucgis.org/priorities/research/2002researchPDF/shotterm/0_global_representation.pdf)
- Varzi, A. (1993). Spatial Reasoning in a Holey World. In Torasso, P. (ed) *Advances in Artificial Intelligence (AI*IA '93)*, Torino, Italy, pp. 326-336.
- Varzi, A. (1996). Reasoning About Space: The Hole Story. *Logic and Logical Philosophy* **4**: 3-39.
- Vasardani, M. and Egenhofer, M. (2008). Single-Holed Regions: Their Relations and Inferences. In Cova, T., Beard, K., Frank, A., and Goodchild, M. (eds) Fifth International Conference on Geographic Information Science-GIScience 2008, Park City, UT, Lecture Notes in Computer Science, **5266** pp. 337-353 Springer-Verlag Berlin Heidelberg.

- Vasardani, M. and Egenhofer, M. (2009). Comparing Relations with a Multi-Holed Region. In Hornsby, K., Claramunt, C., Denis, M., and Ligozat, G. (eds) *Conference on Spatial Information Theory (COSIT'09)*, Aber Wrac'h, France, *Lecture Notes in Computer Science*, **5756** pp. 159-176 Springer-Verlag Berlin Heidelberg.
- Weisstein, E. (2007) Hole *MathWorld A Wolfram Web Resource* http://mathworld.wolfram.com/Hole.html (accessed on 12/17/2007)
- Wertheimer, M. (1923). Laws of Organization in Perceptual Form. In Ellis, W. (ed) *A Source Book of Gestalt Psychology*, pp.71-88. Routledge & Kegan Paul, London
- Worboys, M. (1996). Metrics and Topologies for Geographic Space. In Kraak, M. J. and Molenaar, M. (eds) *Advances in Geographic Information Systems Research II*, pp. 365-376, Taylor & Francis.
- Worboys, M. and Bofakos, P. (1993). A Canonical Model for a Class of Areal Spatial Objects. In Abel, D. and Ooi, B. (eds) *Proceedings of the Third International Symposium on Advances in Spatial Databases*, *Lecture Notes in Computer Science*, **692** pp. 36-52 Springer-Verlag.
- Worboys, M. and Duckham, M. (2004). *GIS: A Computing Perspective*. CRC Press, Boca Raton, FL.
- Zubin, K. (1989). Untitled. In Mark, D., Frank, A., Egenhofer, M., Freundschuh, S., McGranaghan, M., and White, R.M. (eds) *Languages of Spatial Relations: Initiative Two Specialist Meeting Report*. Technical Paper 89-2, National Center for Geographic Information and Analysis, Santa Barbara, CA: 13-17.

BIOGRAPHY OF THE AUTHOR

Maria Vasardani was born in Athens, Greece on September 5, 1978. She studied in the University of Athens and obtained her Bachelor of Science in Geology in 2000, graduating first in her class, with high distinction. During her studies at the Department of Geology, she developed an interest in Geographic Information Systems (GISs) and became a research assistant at the GIS and Remote Sensing Lab of the department. After working there for a year after her graduation, she spent six months in the University of Ottawa, working on remote sensing related projects in the Department of Geography. In 2002 she began her graduate studies at the Department of Geomatic Engineering at University College London, UK from where she obtained her Master of Science in Geographic Information Science in September of 2003, with distinction. Her master's thesis entitled "Cascading Web Map Servers and GIS Web Services" involved creating customized online GIS applications and integrating data from several web map servers. She resumed similar responsibilities as an employ of NAMA Geoinformatics, a GIS software developing company in Athens, Greece. Having decided to go back to academia, she spent two semesters at the University at Buffalo, NY, and then moved to the University of Maine where she continued her doctoral studies in Spatial Information Science and Engineering in 2005. She has since served as a Graduate Research Assistant and was awarded the UMaine Graduate Research Award in 2008.

Maria is a candidate for the Doctor of Philosophy degree in Spatial Information Science and Engineering from the University of Maine in December, 2009.