

MASTER'S THESIS

Incremental maintenance of minimal and minimum bisimulation of cyclic graphs

Deng, Jintian

Date of Award:
2011

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Incremental Maintenance of Minimal and Minimum Bisimulation of Cyclic Graphs

DENG Jintian

A thesis submitted in partial fulfillment of the requirements
for the degree of
Master of Philosophy

Principal Supervisor: Assistant Professor CHOI Koon Kau

Hong Kong Baptist University

September 2011

Abstract

Graph databases have numerous recent applications, such as the Semantic Web, ontology representation, social networks and XML *etc.* A fundamental structural index for data graphs, namely *bisimulation*, has been reported useful in efficient path query processing and optimizations including selectivity estimation and graph compression, among many others. It is desirable to adapt a *minimal* or *minimum* bisimulation for query efficiency of their applications. In this thesis, we devise two incremental maintenance algorithms of *bisimulation* for a possibly *cyclic* data graph. Cyclic graphs are ubiquitous among the data on the Web, however, previous work on the maintenance of minimum bisimulation mostly focuses on *acyclic* graphs. In practice, graphs are subject to update. Therefore *bisimulation* has to be updated accordingly.

Firstly, in the context of database applications, it is natural to compute *minimal* bisimulation with merging algorithms. We propose an incremental maintenance algorithm for a *minimal* bisimulation of *cyclic* graphs, in the style of merging. In order to prune the computation on non-bisimilar SCCs, we propose a feature-based optimization. The features are designed to be constructed and used more efficiently than bisimulation minimization. We also conduct an experimental study that verifies the effectiveness and efficiency of our algorithm. For illustration purpose our features-based optimization can prune 50% (on average) unnecessary bisimulation computation on our synthetic data. Our experiment verifies that we extend the current state-of-the-art with a capability on handling *cyclic* graphs in maintenance of *minimal* bisimulation.

Secondly, while previous algorithm focuses on optimization issues and does not guarantee *minimum* bisimulation, we propose a hybrid incremental algorithm which guarantees *minimum* bisimulation even for a *cyclic* graph. To study the problem in the context of cyclic graphs, we first show that the two classes of existing minimization algorithms – merging algorithm and partition refinement – have their strengths and weaknesses. Second, we propose a novel hybrid algorithm and its analytical model. To the best of our knowledge, this is the first maintenance algorithm that guarantees *minimum* bisimulation of *cyclic* graphs. Third, while small updates on a graph may affect the *minimum* bisimulation arbitrarily, our preliminary experiments show that in practice, this is often rare. We propose to partially reuse the *minimum* bisimulation before an update to optimize maintenance performance. Finally, we present an experimental study on both synthetic and real-life datasets that verifies the efficiency and effectiveness of our algorithms.

Thirdly, while for path query answering, many indexes are *bisimulation*-based, there have been a large number of ad-hoc indexes for reachability query in graph databases. The structures and performances of these indexes on different graphs have known to be very different. In this thesis, we propose a hierarchical prediction framework to select the optimal index for a graph database.

Keywords: graph analysis, graph algorithms, evolving graphs, cyclic graphs, minimal bisimulation, minimum bisimulation, reachability query, path query.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Figures	viii
Chapter 1 Introduction	1
1.1 Graph Database	1
1.1.1 Graph Data	1
1.1.2 Graph Query	3
1.2 Bisimulation and Path Query Processing	4
1.3 Thesis Organization	6
Chapter 2 Preliminary Concepts of Graphs	8
Chapter 3 Background	12
3.1 Bisimulation related Indexes	12
3.1.1 Preliminary	12
3.1.2 DataGuide	14
3.1.3 1-Index	15
3.1.4 $A(k)$ -Index	16

3.1.5	D(k)-Index	18
3.2	Bisimulation related Algorithms	20
3.2.1	Paige-Tarjan Algorithm	20
3.2.2	Propagate Algorithm	21
3.2.3	Split/Merge Algorithm	23
3.2.4	Summary	25
Chapter 4 Optimizing Incremental Maintenance of Minimal Bisimulation		27
4.1	Bisimulation of Cyclic Graphs	27
4.2	Maintenance of Bisimulation	30
4.3	Feature-Based Optimization	33
4.3.1	Properties of Bisimulation of Cyclic Graphs	33
4.3.2	Features of SCCs	34
4.3.3	Offline versus Online Feature Construction	37
4.4	Experimental Evaluation	37
4.4.1	Hardware and software	37
4.4.2	Dataset	38
4.4.3	Performance analysis	40
4.5	Summary	46
Chapter 5 Incremental Maintenance of Minimum Bisimulation		47
5.1	Preliminaries	47
5.1.1	Bisimulation Minimization Algorithms	48
5.2	Bisimulation of Cyclic Graphs	49
5.2.1	Hybrid Minimization Algorithm	51
5.2.2	The Analytical Model	56
5.2.3	Generalization of Reduced Graphs	58
5.3	Incremental Maintenance	59
5.3.1	Deletions of edges	63

5.4	Optimization on the Maintenance Algorithm	67
5.4.1	Feature-based Optimization	67
5.4.2	Optimizing Partition Refinement	68
5.4.3	Optimizing Refinement Splitter	68
5.4.4	Optimization on Cross Nodes	69
5.5	Experimental Evaluation	69
5.5.1	Hardware and software	69
5.5.2	Dataset	70
5.5.3	Maintenance performance	71
5.5.4	Optimization effectiveness	74
5.5.5	Accuracy of the cost model	77
 Chapter 6 Predicting the Optimal Ad-Hoc Index for Reachability		
	Queries	83
6.1	Introduction	84
6.2	Related Work	85
6.3	Definitions and Preliminaries	86
6.4	Basic Prediction Model	88
6.4.1	The Basic Predictor	88
6.4.2	The Verifier and the Adjusted Predictor	89
6.5	Hierarchical Prediction Model	92
6.6	Experiment Evaluation	94
 Chapter 7 Conclusions and Future Work		97
7.1	Summary and Contributions	97
7.2	Future Work	98
 Bibliography		101
 Publications		107
 Curriculum Vitae		108