
MINING GRAPH DATA

EDITED BY

Diane J. Cook

*School of Electrical Engineering and Computer Science
Washington State University
Pullman, Washington*

Lawrence B. Holder

*School of Electrical Engineering and Computer Science
Washington State University
Pullman, Washington*



WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

CONTENTS

Preface	xiii
Acknowledgments	xv
Contributors	xvii
1 INTRODUCTION	1
<i>Lawrence B. Holder and Diane J. Cook</i>	
1.1 Terminology	2
1.2 Graph Databases	3
1.3 Book Overview	10
References	11
Part I GRAPHS	15
2 GRAPH MATCHING—EXACT AND ERROR-TOLERANT METHODS AND THE AUTOMATIC LEARNING OF EDIT COSTS	17
<i>Horst Bunke and Michel Neuhaus</i>	
2.1 Introduction	17
2.2 Definitions and Graph Matching Methods	18
2.3 Learning Edit Costs	24
2.4 Experimental Evaluation	28
2.5 Discussion and Conclusions	31
References	32
3 GRAPH VISUALIZATION AND DATA MINING	35
<i>Walter Didimo and Giuseppe Liotta</i>	
3.1 Introduction	35
3.2 Graph Drawing Techniques	38
3.3 Examples of Visualization Systems	48
	vii

3.4	Conclusions	55
	References	57

4 GRAPH PATTERNS AND THE R-MAT GENERATOR **65**

Deepayan Chakrabarti and Christos Faloutsos

4.1	Introduction	65
4.2	Background and Related Work	67
4.3	NetMine and R-MAT	79
4.4	Experiments	82
4.5	Conclusions	86
	References	92

Part II MINING TECHNIQUES **97**

5 DISCOVERY OF FREQUENT SUBSTRUCTURES **99**

Xifeng Yan and Jiawei Han

5.1	Introduction	99
5.2	Preliminary Concepts	100
5.3	Apriori-based Approach	101
5.4	Pattern Growth Approach	103
5.5	Variant Substructure Patterns	107
5.6	Experiments and Performance Study	109
5.7	Conclusions	112
	References	113

6 FINDING TOPOLOGICAL FREQUENT PATTERNS FROM GRAPH DATASETS **117**

Michihiro Kuramochi and George Karypis

6.1	Introduction	117
6.2	Background Definitions and Notation	118
6.3	Frequent Pattern Discovery from Graph Datasets—Problem Definitions	122
6.4	FSG for the Graph-Transaction Setting	127
6.5	SiGRAM for the Single-Graph Setting	131
6.6	GREW—Scalable Frequent Subgraph Discovery Algorithm	141
6.7	Related Research	149
6.8	Conclusions	151
	References	154

7 UNSUPERVISED AND SUPERVISED PATTERN LEARNING IN GRAPH DATA **159**

Diane J. Cook, Lawrence B. Holder, and Nikhil Ketkar

7.1	Introduction	159
-----	--------------	-----

7.2	Mining Graph Data Using Subdue	160
7.3	Comparison to Other Graph-Based Mining Algorithms	165
7.4	Comparison to Frequent Substructure Mining Approaches	165
7.5	Comparison to ILP Approaches	170
7.6	Conclusions	179
	References	179
8	GRAPH GRAMMAR LEARNING	183
	<i>Istvan Jonyer</i>	
8.1	Introduction	183
8.2	Related Work	184
8.3	Graph Grammar Learning	185
8.4	Empirical Evaluation	193
8.5	Conclusion	199
	References	199
9	CONSTRUCTING DECISION TREE BASED ON CHUNKINGLESS GRAPH-BASED INDUCTION	203
	<i>Kouzou Ohara, Phu Chien Nguyen, Akira Mogi, Hiroshi Motoda, and Takashi Washio</i>	
9.1	Introduction	203
9.2	Graph-Based Induction Revisited	205
9.3	Problem Caused by Chunking in B-GBI	207
9.4	Chunkingless Graph-Based Induction (CI-GBI)	208
9.5	Decision Tree Chunkingless Graph-Based Induction (DT-CIGBI)	214
9.6	Conclusions	224
	References	224
10	SOME LINKS BETWEEN FORMAL CONCEPT ANALYSIS AND GRAPH MINING	227
	<i>Michel Liquière</i>	
10.1	Presentation	227
10.2	Basic Concepts and Notation	228
10.3	Formal Concept Analysis	229
10.4	Extension Lattice and Description Lattice Give Concept Lattice	231
10.5	Graph Description and Galois Lattice	235
10.6	Graph Mining and Formal Propositionalization	240
10.7	Conclusion	249
	References	250

11	KERNEL METHODS FOR GRAPHS	253
	<i>Thomas Gärtner, Tamás Horváth, Quoc V. Le, Alex J. Smola, and Stefan Wrobel</i>	
11.1	Introduction	253
11.2	Graph Classification	254
11.3	Vertex Classification	266
11.4	Conclusions and Future Work	279
	References	280
12	KERNELS AS LINK ANALYSIS MEASURES	283
	<i>Masashi Shimbo and Takahiko Ito</i>	
12.1	Introduction	283
12.2	Preliminaries	284
12.3	Kernel-based Unified Framework for Importance and Relatedness	286
12.4	Laplacian Kernels as a Relatedness Measure	290
12.5	Practical Issues	297
12.6	Related Work	299
12.7	Evaluation with Bibliographic Citation Data	300
12.8	Summary	308
	References	308
13	ENTITY RESOLUTION IN GRAPHS	311
	<i>Indrajit Bhattacharya and Lise Getoor</i>	
13.1	Introduction	311
13.2	Related Work	314
13.3	Motivating Example for Graph-Based Entity Resolution	318
13.4	Graph-Based Entity Resolution: Problem Formulation	322
13.5	Similarity Measures for Entity Resolution	325
13.6	Graph-Based Clustering for Entity Resolution	330
13.7	Experimental Evaluation	333
13.8	Conclusion	341
	References	342
Part III	APPLICATIONS	345
14	MINING FROM CHEMICAL GRAPHS	347
	<i>Takashi Okada</i>	
14.1	Introduction and Representation of Molecules	347
14.2	Issues for Mining	355
14.3	CASE: A Prototype Mining System in Chemistry	356
14.4	Quantitative Estimation Using Graph Mining	358
14.5	Extension of Linear Fragments to Graphs	362

14.6	Combination of Conditions	366
14.7	Concluding Remarks	375
	References	377
15	UNIFIED APPROACH TO ROOTED TREE MINING: ALGORITHMS AND APPLICATIONS	381
	<i>Mohammed Zaki</i>	
15.1	Introduction	381
15.2	Preliminaries	382
15.3	Related Work	384
15.4	Generating Candidate Subtrees	385
15.5	Frequency Computation	392
15.6	Counting Distinct Occurrences	397
15.7	The SLEUTH Algorithm	399
15.8	Experimental Results	401
15.9	Tree Mining Applications in Bioinformatics	405
15.10	Conclusions	409
	References	409
16	DENSE SUBGRAPH EXTRACTION	411
	<i>Andrew Tomkins and Ravi Kumar</i>	
16.1	Introduction	411
16.2	Related Work	414
16.3	Finding the densest subgraph	416
16.4	Trawling	418
16.5	Graph Shingling	421
16.6	Connection Subgraphs	429
16.7	Conclusions	438
	References	438
17	SOCIAL NETWORK ANALYSIS	443
	<i>Sherry E. Marcus, Melanie Moy, and Thayne Coffman</i>	
17.1	Introduction	443
17.2	Social Network Analysis	443
17.3	Group Detection	452
17.4	Terrorist Modus Operandi Detection System	452
17.5	Computational Experiments	465
17.6	Conclusion	467
	References	468
	Index	469