## CONTENTS

Preface xiii  
Acknowledgments xv  
Contributors xvii  

1 INTRODUCTION 1  
_Lawrence B. Holder and Diane J. Cook_  
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  
_Horst Bunke and Michel Neuhaus_  
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  
_Walter Didimo and Giuseppe Liotta_  
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 SGRAM 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
CONTENTS

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
Istvan Jonyer
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
Kouzou Ohara, Phu Chien Nguyen, Akira Mogi, Hiroshi Motoda,
and Takashi Washio
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 (Cl-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
Michel Liquière
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
*Thomas Gärtner, Tamás Horváth, Quoc V. Le, Alex J. Smola, and Stefan Wrobel*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.1 Introduction</td>
<td>253</td>
</tr>
<tr>
<td>11.2 Graph Classification</td>
<td>254</td>
</tr>
<tr>
<td>11.3 Vertex Classification</td>
<td>266</td>
</tr>
<tr>
<td>11.4 Conclusions and Future Work</td>
<td>279</td>
</tr>
<tr>
<td>References</td>
<td>280</td>
</tr>
</tbody>
</table>

## 12 KERNELS AS LINK ANALYSIS MEASURES
*Masashi Shimbo and Takahiko Ito*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.1 Introduction</td>
<td>283</td>
</tr>
<tr>
<td>12.2 Preliminaries</td>
<td>284</td>
</tr>
<tr>
<td>12.3 Kernel-based Unified Framework for Importance and Relatedness</td>
<td>286</td>
</tr>
<tr>
<td>12.4 Laplacian Kernels as a Relatedness Measure</td>
<td>290</td>
</tr>
<tr>
<td>12.5 Practical Issues</td>
<td>297</td>
</tr>
<tr>
<td>12.6 Related Work</td>
<td>299</td>
</tr>
<tr>
<td>12.7 Evaluation with Bibliographic Citation Data</td>
<td>300</td>
</tr>
<tr>
<td>12.8 Summary</td>
<td>308</td>
</tr>
<tr>
<td>References</td>
<td>308</td>
</tr>
</tbody>
</table>

## 13 ENTITY RESOLUTION IN GRAPHS
*Indrajit Bhattacharya and Lise Getoor*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.1 Introduction</td>
<td>311</td>
</tr>
<tr>
<td>13.2 Related Work</td>
<td>314</td>
</tr>
<tr>
<td>13.3 Motivating Example for Graph-Based Entity Resolution</td>
<td>318</td>
</tr>
<tr>
<td>13.4 Graph-Based Entity Resolution: Problem Formulation</td>
<td>322</td>
</tr>
<tr>
<td>13.5 Similarity Measures for Entity Resolution</td>
<td>325</td>
</tr>
<tr>
<td>13.6 Graph-Based Clustering for Entity Resolution</td>
<td>330</td>
</tr>
<tr>
<td>13.7 Experimental Evaluation</td>
<td>333</td>
</tr>
<tr>
<td>13.8 Conclusion</td>
<td>341</td>
</tr>
<tr>
<td>References</td>
<td>342</td>
</tr>
</tbody>
</table>

## Part III APPLICATIONS

## 14 MINING FROM CHEMICAL GRAPHS
*Takashi Okada*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.1 Introduction and Representation of Molecules</td>
<td>347</td>
</tr>
<tr>
<td>14.2 Issues for Mining</td>
<td>355</td>
</tr>
<tr>
<td>14.3 CASE: A Prototype Mining System in Chemistry</td>
<td>356</td>
</tr>
<tr>
<td>14.4 Quantitative Estimation Using Graph Mining</td>
<td>358</td>
</tr>
<tr>
<td>14.5 Extension of Linear Fragments to Graphs</td>
<td>362</td>
</tr>
</tbody>
</table>
14.6 Combination of Conditions 366
14.7 Concluding Remarks 375
References 377

15 UNIFIED APPROACH TO ROOTED TREE MINING:
ALGORITHMS AND APPLICATIONS 381
Mohammed Zaki
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
Andrew Tomkins and Ravi Kumar
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
Sherry E. Marcus, Melanie Moy, and Thayne Coffman
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