List of Figures xv
List of Tables xxi
Preface xxiii

1
An Introduction to Graph Data 1
Charu C. Aggarwal and Haixun Wang
1. Introduction 1
2. Graph Management and Mining Applications 3
3. Summary 8
References 9

2
Graph Data Management and Mining: A Survey of Algorithms and Applications 13
Charu C. Aggarwal and Haixun Wang
1. Introduction 13
2. Graph Data Management Algorithms 16
   2.1 Indexing and Query Processing Techniques 16
   2.2 Reachability Queries 19
   2.3 Graph Matching 21
   2.4 Keyword Search 24
   2.5 Synopsis Construction of Massive Graphs 27
3. Graph Mining Algorithms 29
   3.1 Pattern Mining in Graphs 29
   3.2 Clustering Algorithms for Graph Data 32
   3.3 Classification Algorithms for Graph Data 37
   3.4 The Dynamics of Time-Evolving Graphs 40
4. Graph Applications 43
   4.1 Chemical and Biological Applications 43
   4.2 Web Applications 45
   4.3 Software Bug Localization 51
5. Conclusions and Future Research 55
References 55

3
Graph Mining: Laws and Generators 69
Deepayan Chakrabarti, Christos Faloutsos and Mary McGlohon
1. Introduction 70
2. Graph Patterns 71
2. Types of Dense Components
   2.1 Absolute vs. Relative Density
   2.2 Graph Terminology
   2.3 Definitions of Dense Components
   2.4 Dense Component Selection
   2.5 Relationship between Clusters and Dense Components
3. Algorithms for Detecting Dense Components in a Single Graph
   3.1 Exact Enumeration Approach
   3.2 Heuristic Approach
   3.3 Exact and Approximation Algorithms for Discovering Densest Components
4. Frequent Dense Components
   4.1 Frequent Patterns with Density Constraints
   4.2 Dense Components with Frequency Constraint
   4.3 Enumerating Cross-Graph Quasi-Cliques
5. Applications of Dense Component Analysis
6. Conclusions and Future Research

References

11 Graph Classification

Koji Tsuda and Hiroto Saigo

1. Introduction
2. Graph Kernels
   2.1 Random Walks on Graphs
   2.2 Label Sequence Kernel
   2.3 Efficient Computation of Label Sequence Kernels
   2.4 Extensions
3. Graph Boosting
   3.1 Formulation of Graph Boosting
   3.2 Optimal Pattern Search
   3.3 Computational Experiments
   3.4 Related Work
4. Applications of Graph Classification
5. Label Propagation
6. Concluding Remarks

References

12 Mining Graph Patterns

Hong Cheng, Xifeng Yan and Jiawei Han

1. Introduction
2. Frequent Subgraph Mining
   2.1 Problem Definition
   2.2 Apriori-based Approach
   2.3 Pattern-Growth Approach
   2.4 Closed and Maximal Subgraphs
   2.5 Mining Subgraphs in a Single Graph
   2.6 The Computational Bottleneck
3. Mining Significant Graph Patterns
   3.1 Problem Definition
   3.2 gboost: A Branch-and-Bound Approach
13 A Survey on Streaming Algorithms for Massive Graphs

Jian Zhang

1. Introduction 393
2. Streaming Model for Massive Graphs 395
3. Statistics and Counting Triangles 397
4. Graph Matching 400
   4.1 Unweighted Matching 400
   4.2 Weighted Matching 403
5. Graph Distance 405
   5.1 Distance Approximation using Multiple Passes 406
   5.2 Distance Approximation in One Pass 411
6. Random Walks on Graphs 412
7. Conclusions 416

References 417

14 A Survey of Privacy-Preservation of Graphs and Social Networks

Xintao Wu, Xiaowei Ying, Kun Liu and Lei Chen

1. Introduction 422
   1.1 Privacy in Publishing Social Networks 422
   1.2 Background Knowledge 423
   1.3 Utility Preservation 424
   1.4 Anonymization Approaches 424
   1.5 Notations 425
2. Privacy Attacks on Naive Anonymized Networks 426
   2.1 Active Attacks and Passive Attacks 426
   2.2 Structural Queries 427
   2.3 Other Attacks 428
3. $K$-Anonymity Privacy Preservation via Edge Modification 428
   3.1 $K$-Degree Generalization 429
   3.2 $K$-Neighborhood Anonymity 430
   3.3 $K$-Automorphism Anonymity 431
4. Privacy Preservation via Randomization 433
   4.1 Resilience to Structural Attacks 434
   4.2 Link Disclosure Analysis 435
   4.3 Reconstruction 437
   4.4 Feature Preserving Randomization 438
5. Privacy Preservation via Generalization 440
6. Anonymizing Rich Graphs 441

References 447
2.1 Dynamic Call Graphs 517
2.2 Bugs in Software 518
2.3 Bug Localization with Call Graphs 519
2.4 Graph and Tree Mining 520
3. Related Work 521
4. Call-Graph Reduction 525
4.1 Total Reduction 525
4.2 Iterations 526
4.3 Temporal Order 528
4.4 Recursion 529
4.5 Comparison 531
5. Call Graph Based Bug Localization 532
5.1 Structural Approaches 532
5.2 Frequency-based Approach 535
5.3 Combined Approaches 538
5.4 Comparison 539
6. Conclusions and Future Directions 542
Acknowledgments 543
References 543

18
A Survey of Graph Mining Techniques for Biological Datasets 547
S. Parthasarathy, S. Tatikonda and D. Ucar 548
1. Introduction 548
2. Mining Trees 549
2.1 Frequent Subtree Mining 550
2.2 Tree Alignment and Comparison 552
2.3 Statistical Models 554
3. Mining Graphs for the Discovery of Frequent Substructures 555
3.1 Frequent Subgraph Mining 555
3.2 Motif Discovery in Biological Networks 560
4. Mining Graphs for the Discovery of Modules 562
4.1 Extracting Communities 564
4.2 Clustering 566
5. Discussion 569
References 571

19
Trends in Chemical Graph Data Mining 581
Nikil Wale, Xia Ning and George Karypis 582
1. Introduction 582
2. Topological Descriptors for Chemical Compounds 583
2.1 Hashed Fingerprints (FP) 584
2.2 Maccs Keys (MK) 584
2.3 Extended Connectivity Fingerprints (ECFP) 584
2.4 Frequent Subgraphs (FS) 585
2.5 Bounded-Size Graph Fragments (GF) 585
2.6 Comparison of Descriptors 585
3. Classification Algorithms for Chemical Compounds 588
3.1 Approaches based on Descriptors 588
3.2 Approaches based on Graph Kernels 589
4. Searching Compound Libraries 590
## Contents

4.1 Methods Based on Direct Similarity 591  
4.2 Methods Based on Indirect Similarity 592  
4.3 Performance of Indirect Similarity Methods 594  
5. Identifying Potential Targets for Compounds 595  
5.1 Model-based Methods For Target Fishing 596  
5.2 Performance of Target Fishing Strategies 600  
6. Future Research Directions 600  
References 602  
Index 607
Managing and Mining Graph Data
Aggarwal, C.C.; Wang, H. (Eds.)
2010, XXII, 600 p. 20 illus., Hardcover