### MANAGING AND MINING GRAPH DATA

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Power Laws and Heavy-Tailed Distributions</td>
<td>72</td>
</tr>
<tr>
<td>2.2</td>
<td>Small Diameters</td>
<td>77</td>
</tr>
<tr>
<td>2.3</td>
<td>Other Static Graph Patterns</td>
<td>79</td>
</tr>
<tr>
<td>2.4</td>
<td>Patterns in Evolving Graphs</td>
<td>82</td>
</tr>
<tr>
<td>2.5</td>
<td>The Structure of Specific Graphs</td>
<td>84</td>
</tr>
<tr>
<td>3.</td>
<td>Graph Generators</td>
<td>86</td>
</tr>
<tr>
<td>3.1</td>
<td>Random Graph Models</td>
<td>88</td>
</tr>
<tr>
<td>3.2</td>
<td>Preferential Attachment and Variants</td>
<td>92</td>
</tr>
<tr>
<td>3.3</td>
<td>Optimization-based generators</td>
<td>101</td>
</tr>
<tr>
<td>3.4</td>
<td>Tensor-based</td>
<td>108</td>
</tr>
<tr>
<td>3.5</td>
<td>Generators for specific graphs</td>
<td>113</td>
</tr>
<tr>
<td>3.6</td>
<td>Graph Generators: A summary</td>
<td>115</td>
</tr>
<tr>
<td>4.</td>
<td>Conclusions</td>
<td>115</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>117</td>
</tr>
</tbody>
</table>

### Query Language and Access Methods for Graph Databases

**Huahai He and Ambuj K. Singh**

1. **Introduction**

2. **Operations on Graph Structures**

3. **Graph Query Language**

4. **Implementation of the Selection Operator**

5. **Experimental Study**

6. **Related Work**

7. **Future Research Directions**

8. **Conclusion**

Appendix: Query Syntax of GraphQL

References

### Graph Indexing

**Xifeng Yan and Jiawei Han**

1. **Introduction**
2. Feature-Based Graph Index 162
   2.1 Paths 163
   2.2 Frequent Structures 164
   2.3 Discriminative Structures 166
   2.4 Closed Frequent Structures 167
   2.5 Trees 167
   2.6 Hierarchical Indexing 168
3. Structure Similarity Search 169
   3.1 Feature-Based Structural Filtering 170
   3.2 Feature Miss Estimation 171
   3.3 Frequency Difference 172
   3.4 Feature Set Selection 173
   3.5 Structures with Gaps 174
4. Reverse Substructure Search 175
5. Conclusions 177
References 178

6
Graph Reachability Queries: A Survey 181
Jeffrey Xu Yu and Jiefeng Cheng
1. Introduction 181
2. Traversal Approaches 186
   2.1 Tree+SSPI 187
   2.2 GRIPP 187
3. Dual-Labeling 188
4. Tree Cover 190
5. Chain Cover 191
   5.1 Computing the Optimal Chain Cover 193
6. Path-Tree Cover 194
7. 2-HOP Cover 196
   7.1 A Heuristic Ranking 197
   7.2 A Geometrical-Based Approach 198
   7.3 Graph Partitioning Approaches 199
   7.4 2-Hop Cover Maintenance 202
8. 3-Hop Cover 204
9. Distance-Aware 2-Hop Cover 205
10. Graph Pattern Matching 207
   10.1 A Special Case: $A \rightarrow D$ 208
   10.2 The General Cases 211
11. Conclusions and Summary 212
References 212

7
Exact and Inexact Graph Matching: Methodology and Applications 217
Kaspar Riesen, Xiaoyi Jiang and Horst Bunke
1. Introduction 218
2. Basic Notations 219
3. Exact Graph Matching 221
4. Inexact Graph Matching 226
   4.1 Graph Edit Distance 227
   4.2 Other Inexact Graph Matching Techniques 229
5. Graph Matching for Data Mining and Information Retrieval 231
MANAGING AND MINING GRAPH DATA

6. Vector Space Embeddings of Graphs via Graph Matching 235
7. Conclusions 239
References 240

8
A Survey of Algorithms for Keyword Search on Graph Data 249
Haixun Wang and Charu C. Aggarwal
1. Introduction 250
2. Keyword Search on XML Data 252
   2.1 Query Semantics 253
   2.2 Answer Ranking 254
   2.3 Algorithms for LCA-based Keyword Search 258
3. Keyword Search on Relational Data 260
   3.1 Query Semantics 260
   3.2 DBXplorer and DISCOVER 261
4. Keyword Search on Schema-Free Graphs 263
   4.1 Query Semantics and Answer Ranking 263
   4.2 Graph Exploration by Backward Search 265
   4.3 Graph Exploration by Bidirectional Search 266
   4.4 Index-based Graph Exploration – the BLINKS Algorithm 267
   4.5 The ObjectRank Algorithm 269
5. Conclusions and Future Research 271
References 271

9
A Survey of Clustering Algorithms for Graph Data 275
Charu C. Aggarwal and Haixun Wang
1. Introduction 275
2. Node Clustering Algorithms 277
   2.1 The Minimum Cut Problem 277
   2.2 Multi-way Graph Partitioning 281
   2.3 Conventional Generalizations and Network Structure Indices 282
   2.4 The Girvan-Newman Algorithm 284
   2.5 The Spectral Clustering Method 285
   2.6 Determining Quasi-Cliques 288
   2.7 The Case of Massive Graphs 289
3. Clustering Graphs as Objects 291
   3.1 Extending Classical Algorithms to Structural Data 291
   3.2 The XProj Approach 293
4. Applications of Graph Clustering Algorithms 295
   4.1 Community Detection in Web Applications and Social Networks 296
   4.2 Telecommunication Networks 297
   4.3 Email Analysis 297
5. Conclusions and Future Research 297
References 299

10
A Survey of Algorithms for Dense Subgraph Discovery 303
Victor E. Lee, Ning Ruan, Ruoming Jin and Charu Aggarwal
1. Introduction 304
Contents

2. Types of Dense Components 305
   2.1 Absolute vs. Relative Density 305
   2.2 Graph Terminology 306
   2.3 Definitions of Dense Components 307
   2.4 Dense Component Selection 308
   2.5 Relationship between Clusters and Dense Components 309

3. Algorithms for Detecting Dense Components in a Single Graph 311
   3.1 Exact Enumeration Approach 311
   3.2 Heuristic Approach 314
   3.3 Exact and Approximation Algorithms for Discovering Densest Components 322

4. Frequent Dense Components 327
   4.1 Frequent Patterns with Density Constraints 327
   4.2 Dense Components with Frequency Constraint 328
   4.3 Enumerating Cross-Graph Quasi-Cliques 328

5. Applications of Dense Component Analysis 329

6. Conclusions and Future Research 331

References 333

11

Graph Classification 337
Koji Tsuda and Hirotaro Saigo

1. Introduction 337

2. Graph Kernels 340
   2.1 Random Walks on Graphs 341
   2.2 Label Sequence Kernel 342
   2.3 Efficient Computation of Label Sequence Kernels 343
   2.4 Extensions 349

3. Graph Boosting 349
   3.1 Formulation of Graph Boosting 351
   3.2 Optimal Pattern Search 353
   3.3 Computational Experiments 354
   3.4 Related Work 355

4. Applications of Graph Classification 358

5. Label Propagation 358

6. Concluding Remarks 359

References 359

12

Mining Graph Patterns 365
Hong Cheng, Xifeng Yan and Jiawei Han

1. Introduction 366

2. Frequent Subgraph Mining 366
   2.1 Problem Definition 366
   2.2 Apriori-based Approach 367
   2.3 Pattern-Growth Approach 368
   2.4 Closed and Maximal Subgraphs 369
   2.5 Mining Subgraphs in a Single Graph 370
   2.6 The Computational Bottleneck 371

3. Mining Significant Graph Patterns 372
   3.1 Problem Definition 372
   3.2 gboost: A Branch-and-Bound Approach 373
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