Contents

Introduction *page* 1
- 0.1 Background 3
- 0.2 Book Organization 4
- 0.3 Acknowledgments 7

Part I. Introduction to Graph Theory

1 Notations, Properties, and Representations 11
- 1.1 Graph Terminology and Notations 11
- 1.2 Graph Properties 13
- 1.3 Graph Types 14
- 1.4 Representing Graphs as Matrices 15
- 1.5 Using Matrices to Compute Graph Properties 16
- 1.6 Representing Graphs as Linked Lists 17
- 1.7 Eigenvalues and Eigenvectors 18

2 Graph-Based Algorithms 20
- 2.1 Depth-First Graph Traversal 20
- 2.2 Breadth-First Graph Traversal 22
- 2.3 Minimum Spanning Trees 23
- 2.4 Shortest-Path Algorithms 26
- 2.5 Cuts and Flows 29
- 2.6 Graph Matching 31
- 2.7 Dimensionality Reduction 32
- 2.8 Stochastic Processes on Graphs 34
- 2.9 Harmonic Functions 38
- 2.10 Random Walks 40
- 2.11 Spreading Activation 41
- 2.12 Electrical Interpretation of Random Walks 42
- 2.13 Power Method 44

2.14 Linear Algebra Methods for Computing Harmonic Functions 45
2.15 Method of Relaxations 46
2.16 Monte Carlo Methods 47

Part II. Networks
3 Random Networks 53
3.1 Networks and Graphs 53
3.2 Random Graphs 54
3.3 Degree Distributions 54
3.4 Power Laws 57
3.5 Zipf's Law 58
3.6 Preferential Attachment 61
3.7 Giant Component 62
3.8 Clustering Coefficient 62
3.9 Small Worlds 63
3.10 Assortativity 65
3.11 Centrality 67
3.12 Degree Centrality 67
3.13 Closeness Centrality 68
3.14 Betweenness Centrality 69
3.15 Network Example 70
3.16 Dynamic Processes: Percolation 72
3.17 Strong and Weak Ties 74
3.18 Assortative Mixing 76
3.19 Structural Holes 76

4 Language Networks 78
4.1 Co-Occurrence Networks 78
4.2 Syntactic Dependency Networks 80
4.3 Semantic Networks 81
4.4 Similarity Networks 85

Part III. Graph-Based Information Retrieval
5 Link Analysis for the World Wide Web 91
5.1 The Web as a Graph 91
5.2 PageRank 92
5.3 Undirected Graphs 95
5.4 Weighted Graphs 95

5.5 Combining PageRank with Content Analysis 97
5.6 Topic-Sensitive Link Analysis 97
5.7 Query-Dependent Link Analysis 100
5.8 Hyperlinked-Induced Topic Search 101
5.9 Document Reranking with Induced Links 103

6 Text Clustering 106
6.1 Graph-Based Clustering 108
6.2 Spectral Methods 111
6.3 The Fiedler Method 113
6.4 The Kernighan–Lin Method 114
6.5 Betweenness-Based Clustering 115
6.6 Min-Cut Clustering 117
6.7 Text Clustering Using Random Walks 119

Part IV. Graph-Based Natural Language Processing

7 Semantics 123
7.1 Semantic Classes 123
7.2 Synonym Detection 125
7.3 Semantic Distance 126
7.4 Textual Entailment 129
7.5 Word-Sense Disambiguation 131
7.6 Name Disambiguation 134
7.7 Sentiment and Subjectivity 135

8 Syntax 140
8.1 Part-of-Speech Tagging 140
8.2 Dependency Parsing 141
8.3 Prepositional-Phrase Attachment 144
8.4 Co-Reference Resolution 146

9 Applications 149
9.1 Summarization 149
9.2 Semi-supervised Passage Retrieval 150
9.3 Keyword Extraction 154
9.4 Topic Identification 156
9.5 Topic Segmentation 161
9.6 Discourse 162
9.7 Machine Translation 165

9.8 Cross-Language Information Retrieval 166
9.9 Information Extraction 169
9.10 Question Answering 171
9.11 Term Weighting 174

Bibliography 179
Index 191