

Contents

Preface	XI
1 Introduction	1
1.1 Background.....	1
1.2 Web Community	4
1.3 Outline of the Book	5
1.4 Audience of the Book.....	6
2 Preliminaries	7
2.1 Matrix Expression of Hyperlinks	7
2.2 Eigenvalue and Eigenvector of the Matrix	9
2.3 Matrix Norms and the Lipschitz Continuous Function	10
2.4 Singular Value Decomposition (SVD) of a Matrix	11
2.5 Similarity in Vector Space Models.....	14
2.6 Graph Theory Basics	14
2.7 Introduction to the Markov Model	15
3 HITS and Related Algorithms	17
3.1 Original HITS.....	17
3.2 The Stability Issues	20
3.3 Randomized HITS.....	22
3.4 Subspace HITS.....	23
3.5 Weighted HITS.....	24
3.6 The Vector Space Model (VSM).....	27
3.7 Cover Density Ranking (CDR)	29
3.8 In-depth Analysis of HITS.....	31
3.9 HITS Improvement	35
3.10 Noise Page Elimination Algorithm Based on SVD.....	38
3.11 SALSA (Stochastic algorithm)	43
4 PageRank Related Algorithms	49
4.1 The Original PageRank Algorithm.....	49
4.2 Probabilistic Combination of Link and Content Information	53
4.3 Topic-Sensitive PageRank	56

4.4	Quadratic Extrapolation.....	58
4.5	Exploring the Block Structure of the Web for Computing PageRank.....	60
4.6	Web Page Scoring Systems (WPSS).....	64
4.7	The Voting Model.....	71
4.8	Using Non-Affiliated Experts to Rank Popular Topics.....	75
4.9	A Latent Linkage Information (LLI) Algorithm.....	79
5	Affinity and Co-Citation Analysis Approaches.....	85
5.1	Web Page Similarity Measurement.....	85
5.1.1	Page Source Construction.....	85
5.1.2	Page Weight Definition.....	87
5.1.3	Page Correlation Matrix.....	89
5.1.4	Page Similarity.....	92
5.2	Hierarchical Web Page Clustering.....	95
5.3	Matrix-Based Clustering Algorithms.....	97
5.3.1	Similarity Matrix Permutation.....	97
5.3.2	Clustering Algorithm from a Matrix Partition.....	99
5.3.3	Cluster-Overlapping Algorithm.....	101
5.4	Co-Citation Algorithms.....	104
5.4.1	Citation and Co-Citation Analysis.....	104
5.4.2	Extended Co-Citation Algorithms.....	106
6	Building a Web Community.....	111
6.1	Web Community.....	111
6.2	Small World Phenomenon on the Web.....	113
6.3	Trawling the Web.....	115
6.3.1	Finding Web Communities Based on Complete Directed Bipartite Graphs.....	117
6.4	From Complete Bipartite Graph to Dense Directed Bipartite Graph.....	118
6.4.1	The Algorithm.....	119
6.5	Maximum Flow Approaches.....	123
6.5.1	Maximum Flow and Minimum Cut.....	124
6.5.2	FLG Approach.....	125
6.5.3	IK Approach.....	129
6.6	Web Community Charts.....	133
6.6.1	The Algorithm.....	135
6.7	From Web Community Chart to Web Community Evolution... ..	138
6.8	Uniqueness of a Web Community.....	141
7	Web Community Related Techniques.....	145

7.1	Web Community and Web Usage Mining.....	145
7.2	Discovering Web Communities Using Co-occurrence.....	147
7.3	Finding High-Level Web Communities	149
7.4	Web Community and Formal Concept Analysis.....	151
7.4.1	Formal Concept Analysis.....	152
7.4.2	From Concepts to Web Communities.....	152
7.5	Generating Web Graphs with Embedded Web Communities....	155
7.6	Modeling Web Communities Using Graph Grammars	157
7.7	Geographical Scopes of Web Resources	158
7.7.1	Two Conditions: Fraction and Uniformity.....	159
7.7.2	Geographical Scope Estimation	161
7.8	Discovering Unexpected Information from Competitors	161
7.9	Probabilistic Latent Semantic Analysis Approach	164
7.9.1	Usage Data and the PLSA Model	165
7.9.2	Discovering Usage-Based Web Page Categories	167
8	Conclusions.....	169
8.1	Summary.....	169
8.2	Future Directions.....	171
	References	173
	Index.....	181
	About the Authors.....	185