# Contents

Preface xiii

1 An Introduction to Social Network Data Analytics 1
   *Charu C. Aggarwal*
   1. Introduction 1
   2. Online Social Networks: Research Issues 5
   3. Research Topics in Social Networks 8
   4. Conclusions and Future Directions 13
   References 14

2 Statistical Properties of Social Networks 17
   *Mary McGlohon, Leman Akoglu and Christos Faloutsos*
   1. Preliminaries 19
      1.1 Definitions 19
      1.2 Data description 24
   2. Static Properties 26
      2.1 Static Unweighted Graphs 26
      2.2 Static Weighted Graphs 27
   3. Dynamic Properties 32
      3.1 Dynamic Unweighted Graphs 32
      3.2 Dynamic Weighted Graphs 36
   4. Conclusion 39
   References 40

3 Random Walks in Social Networks and their Applications: A Survey 43
   *Purnamrita Sarkar and Andrew W. Moore*
   1. Introduction 43
   2. Random Walks on Graphs: Background 45
      2.1 Random Walk based Proximity Measures 46
      2.2 Other Graph-based Proximity Measures 52
      2.3 Graph-theoretic Measures for Semi-supervised Learning 53
      2.4 Clustering with random walk based measures 56
   3. Related Work: Algorithms 57
      3.1 Algorithms for Hitting and Commute Times 58
      3.2 Algorithms for Computing Personalized Pagerank and Simrank 60
<table>
<thead>
<tr>
<th>Contents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Related approaches</td>
<td></td>
</tr>
<tr>
<td>6.1 Inference using Graphical Models</td>
<td>139</td>
</tr>
<tr>
<td>6.2 Metric labeling</td>
<td>140</td>
</tr>
<tr>
<td>6.3 Spectral Partitioning</td>
<td>141</td>
</tr>
<tr>
<td>6.4 Graph Clustering</td>
<td>142</td>
</tr>
<tr>
<td>7. Variations on Node Classification</td>
<td></td>
</tr>
<tr>
<td>7.1 Dissimilarity in Labels</td>
<td>142</td>
</tr>
<tr>
<td>7.2 Edge Labeling</td>
<td>143</td>
</tr>
<tr>
<td>7.3 Label Summarization</td>
<td>144</td>
</tr>
<tr>
<td>8. Concluding Remarks</td>
<td></td>
</tr>
<tr>
<td>8.1 Future Directions and Challenges</td>
<td>145</td>
</tr>
<tr>
<td>8.2 Further Reading</td>
<td>146</td>
</tr>
<tr>
<td>References</td>
<td>146</td>
</tr>
</tbody>
</table>

6
Evolution in Social Networks: A Survey
Myra Spiliopoulou
1. Introduction                                                        149
2. Framework                                                           151
   2.1 Modeling a Network across the Time Axis                          151
   2.2 Evolution across Four Dimensions                                 152
3. Challenges of Social Network Streams                                154
4. Incremental Mining for Community Tracing                            156
5. Tracing Smoothly Evolving Communities                               160
   5.1 Temporal Smoothness for Clusters                                 160
   5.2 Dynamic Probabilistic Models                                    162
6. Laws of Evolution in Social Networks                                167
7. Conclusion                                                          169
References                                                             170

7
A Survey of Models and Algorithms for Social Influence Analysis
Jimeng Sun and Jie Tang
1. Introduction                                                        177
2. Influence Related Statistics                                         178
   2.1 Edge Measures                                                     178
   2.2 Node Measures                                                     180
3. Social Similarity and Influence                                      183
   3.1 Homophily                                                         183
   3.2 Existential Test for Social Influence                             188
   3.3 Influence and Actions                                             189
   3.4 Influence and Interaction                                         195
4. Influence Maximization in Viral Marketing                           200
   4.1 Influence Maximization                                            200
   4.2 Other Applications                                                206
5. Conclusion                                                          208
References                                                             209

8
A Survey of Algorithms and Systems for Expert Location in Social Networks
Theodoros Lappas, Kun Liu and Evimaria Terzi
1. Introduction                                                        216
2. Definitions and Notation  
3. Expert Location without Graph Constraints  
   3.1 Language Models for Document Information Retrieval  
   3.2 Language Models for Expert Location  
   3.3 Further Reading  
4. Expert Location with Score Propagation  
   4.1 The PageRank Algorithm  
   4.2 HITS Algorithm  
   4.3 Expert Score Propagation  
   4.4 Further Reading  
5. Expert Team Formation  
   5.1 Metrics  
   5.2 Forming Teams of Experts  
   5.3 Further Reading  
6. Other Related Approaches  
   6.1 Agent-based Approach  
   6.2 Influence Maximization  
7. Expert Location Systems  
8. Conclusions  
References  

9  
A Survey of Link Prediction in Social Networks  
Mohammad Al Hasan and Mohammed J. Zaki  
1. Introduction  
2. Background  
3. Feature based Link Prediction  
   3.1 Feature Set Construction  
   3.2 Classification Models  
4. Bayesian Probabilistic Models  
   4.1 Link Prediction by Local Probabilistic Models  
   4.2 Network Evolution based Probabilistic Model  
   4.3 Hierarchical Probabilistic Model  
5. Probabilistic Relational Models  
   5.1 Relational Bayesian Network  
   5.2 Relational Markov Network  
6. Linear Algebraic Methods  
7. Recent development and Future Works  
References  

10  
Privacy in Social Networks: A Survey  
Elena Zheleva and Lise Getoor  
1. Introduction  
2. Privacy breaches in social networks  
   2.1 Identity disclosure  
   2.2 Attribute disclosure  
   2.3 Social link disclosure  
   2.4 Affiliation link disclosure  
3. Privacy definitions for publishing data  
   3.1 $k$-anonymity  
References
Contents

3.2 \(l\)-diversity and \(t\)-closeness 290
3.3 Differential privacy 291
4. Privacy-preserving mechanisms 292
4.1 Privacy mechanisms for social networks 292
4.2 Privacy mechanisms for affiliation networks 297
4.3 Privacy mechanisms for social and affiliation networks 300
5. Related literature 302
6. Conclusion 302
References 303

Visualizing Social Networks 307
Carlos D. Correa and Kwan-Liu Ma
1. Introduction 307
2. A Taxonomy of Visualizations 309
2.1 Structural Visualization 309
2.2 Semantic and Temporal Visualization 313
2.3 Statistical Visualization 315
3. The Convergence of Visualization, Interaction and Analytics 316
3.1 Structural and Semantic Filtering with Ontologies 319
3.2 Centrality-based Visual Discovery and Exploration 319
4. Summary 322
References 323

Data Mining in Social Media 327
Geoffrey Barbier and Huan Liu
1. Introduction 327
2. Data Mining in a Nutshell 328
3. Social Media 330
4. Motivations for Data Mining in Social Media 332
5. Data Mining Methods for Social Media 333
5.1 Data Representation 334
5.2 Data Mining - A Process 335
5.3 Social Networking Sites: Illustrative Examples 336
5.4 The Blogosphere: Illustrative Examples 340
6. Related Efforts 344
6.1 Ethnography and Netnography 344
6.2 Event Maps 345
7. Conclusions 345
References 347

Text Mining in Social Networks 353
Charu C. Aggarwal and Haixun Wang
1. Introduction 354
2. Keyword Search 356
2.1 Query Semantics and Answer Ranking 357
2.2 Keyword search over XML and relational data 358
2.3 Keyword search over graph data 360
3. Classification Algorithms 366
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.</td>
<td>Clustering Algorithms</td>
<td>369</td>
</tr>
<tr>
<td>5.</td>
<td>Transfer Learning in Heterogeneous Networks</td>
<td>371</td>
</tr>
<tr>
<td>6.</td>
<td>Conclusions and Summary</td>
<td>373</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>374</td>
</tr>
<tr>
<td>14</td>
<td>Integrating Sensors and Social Networks</td>
<td>379</td>
</tr>
<tr>
<td></td>
<td>Charu C. Aggarwal and Tarek Abdelzaher</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Introduction</td>
<td>379</td>
</tr>
<tr>
<td>2.</td>
<td>Sensors and Social Networks: Technological Enablers</td>
<td>383</td>
</tr>
<tr>
<td>3.</td>
<td>Dynamic Modeling of Social Networks</td>
<td>385</td>
</tr>
<tr>
<td>4.</td>
<td>System Design and Architectural Challenges</td>
<td>387</td>
</tr>
<tr>
<td>4.1</td>
<td>Privacy-preserving data collection</td>
<td>388</td>
</tr>
<tr>
<td>4.2</td>
<td>Generalized Model Construction</td>
<td>389</td>
</tr>
<tr>
<td>4.3</td>
<td>Real-time Decision Services</td>
<td>389</td>
</tr>
<tr>
<td>4.4</td>
<td>Recruitment Issues</td>
<td>390</td>
</tr>
<tr>
<td>4.5</td>
<td>Other Architectural Challenges</td>
<td>390</td>
</tr>
<tr>
<td>5.</td>
<td>Database Representation: Issues and Challenges</td>
<td>391</td>
</tr>
<tr>
<td>6.</td>
<td>Privacy Issues</td>
<td>399</td>
</tr>
<tr>
<td>7.</td>
<td>Sensors and Social Networks: Applications</td>
<td>402</td>
</tr>
<tr>
<td>7.1</td>
<td>The Google Latitude Application</td>
<td>402</td>
</tr>
<tr>
<td>7.2</td>
<td>The Citysense and Macrosense Applications</td>
<td>403</td>
</tr>
<tr>
<td>7.3</td>
<td>Green GPS</td>
<td>404</td>
</tr>
<tr>
<td>7.4</td>
<td>Microsoft SensorMap</td>
<td>405</td>
</tr>
<tr>
<td>7.5</td>
<td>Animal and Object Tracking Applications</td>
<td>405</td>
</tr>
<tr>
<td>7.6</td>
<td>Participatory Sensing for Real-Time Services</td>
<td>406</td>
</tr>
<tr>
<td>8.</td>
<td>Future Challenges and Research Directions</td>
<td>407</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>407</td>
</tr>
<tr>
<td>15</td>
<td>Multimedia Information Networks in Social Media</td>
<td>413</td>
</tr>
<tr>
<td></td>
<td>Liangliang Cao, GuoJun Qi, Shen-Fu Tsai, Min-Hsuan Tsai, Andrey Del Pozo, Thomas S. Huang, Xuemei Zhang and Suk Hwan Lim</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Introduction</td>
<td>414</td>
</tr>
<tr>
<td>2.</td>
<td>Links from Semantics: Ontology-based Learning</td>
<td>415</td>
</tr>
<tr>
<td>3.</td>
<td>Links from Community Media</td>
<td>416</td>
</tr>
<tr>
<td>3.1</td>
<td>Retrieval Systems for Community Media</td>
<td>417</td>
</tr>
<tr>
<td>3.2</td>
<td>Recommendation Systems for Community Media</td>
<td>418</td>
</tr>
<tr>
<td>4.</td>
<td>Network of Personal Photo Albums</td>
<td>420</td>
</tr>
<tr>
<td>4.1</td>
<td>Actor-Centric Nature of Personal Collections</td>
<td>420</td>
</tr>
<tr>
<td>4.2</td>
<td>Quality Issues in Personal Collections</td>
<td>421</td>
</tr>
<tr>
<td>4.3</td>
<td>Time and Location Themes in Personal Collections</td>
<td>422</td>
</tr>
<tr>
<td>4.4</td>
<td>Content Overlap in Personal Collections</td>
<td>422</td>
</tr>
<tr>
<td>5.</td>
<td>Network of Geographical Information</td>
<td>423</td>
</tr>
<tr>
<td>5.1</td>
<td>Semantic Annotation</td>
<td>425</td>
</tr>
<tr>
<td>5.2</td>
<td>Geographical Estimation</td>
<td>425</td>
</tr>
<tr>
<td>5.3</td>
<td>Other Applications</td>
<td>426</td>
</tr>
<tr>
<td>6.</td>
<td>Inference Methods</td>
<td>427</td>
</tr>
<tr>
<td>6.1</td>
<td>Discriminative vs. Generative Models</td>
<td>427</td>
</tr>
<tr>
<td>6.2</td>
<td>Graph-based Inference: Ranking, Clustering and Semi-supervised Learning</td>
<td>428</td>
</tr>
</tbody>
</table>
An Overview of Social Tagging and Applications

Manish Gupta, Rui Li, Zhijun Yin and Jiawei Han

1. Introduction
   1.1 Problems with Metadata Generation and Fixed Taxonomies
   1.2 Folksonomies as a Solution
   1.3 Outline

2. Tags: Why and What?
   2.1 Different User Tagging Motivations
   2.2 Kinds of Tags
   2.3 Categorizers Versus Describers
   2.4 Linguistic Classification of Tags
   2.5 Game-based Tagging

3. Tag Generation Models
   3.1 Polya Urn Generation Model
   3.2 Language Model
   3.3 Other Influence Factors

4. Tagging System Design

5. Tag analysis
   5.1 Tagging Distributions
   5.2 Identifying Tag Semantics
   5.3 Tags Versus Keywords

6. Visualization of Tags
   6.1 Tag Clouds for Browsing/Search
   6.2 Tag Selection for Tag Clouds
   6.3 Tag Hierarchy Generation
   6.4 Tag Clouds Display Format
   6.5 Tag Evolution Visualization
   6.6 Popular Tag Cloud Demos

7. Tag Recommendations
   7.1 Using Tag Quality
   7.2 Using Tag Co-occurrences
   7.3 Using Mutual Information between Words, Documents and Tags
   7.4 Using Object Features

8. Applications of Tags
   8.1 Indexing
   8.2 Search
   8.3 Taxonomy Generation
   8.4 Public Library Cataloging
   8.5 Clustering and Classification
   8.6 Social Interesting Discovery
   8.7 Enhanced Browsing

9. Integration
9.1 Integration using Tag Co-occurrence Analysis and Clustering 485
9.2 TAGMAS: Federated Tagging System 486
9.3 Correlating User Profiles from Different Folksonomies 487

10. Tagging problems 488
10.1 Spamming 488
10.2 Canonicalization and Ambiguities 489
10.3 Other Problems 490

11. Conclusion and Future Directions 490
11.1 Analysis 491
11.2 Improving System Design 491
11.3 Personalized Tag Recommendations 491
11.4 More Applications 492
11.5 Dealing With Problems 492

References 492

Index 499
Social Network Data Analytics
Aggarwal, C.C. (Ed.)
2011, XIV, 502 p., Hardcover
ISBN: 978-1-4419-8461-6