

# Contents

Preface	xiii
1	
An Introduction to Social Network Data Analytics	1
<i>Charu C. Aggarwal</i>	
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
<i>Mary McGlohon, Leman Akoglu and Christos Faloutsos</i>	
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
<i>Purnamrita Sarkar and Andrew W. Moore</i>	
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 Sim-rank	60

3.3	Algorithms for Computing Harmonic Functions	63
4.	Related Work: Applications	63
4.1	Application in Computer Vision	64
4.2	Text Analysis	64
4.3	Collaborative Filtering	65
4.4	Combating Webspam	66
5.	Related Work: Evaluation and datasets	66
5.1	Evaluation: Link Prediction	66
5.2	Publicly Available Data Sources	68
6.	Conclusion and Future Work	69
	References	71
4		
	Community Discovery in Social Networks: Applications, Methods and Emerging Trends	79
	<i>S. Parthasarathy, Y. Ruan and V. Satuluri</i>	
1.	Introduction	80
2.	Communities in Context	82
3.	Core Methods	84
3.1	Quality Functions	85
3.2	The Kernighan-Lin(KL) algorithm	86
3.3	Agglomerative/Divisive Algorithms	87
3.4	Spectral Algorithms	89
3.5	Multi-level Graph Partitioning	90
3.6	Markov Clustering	91
3.7	Other Approaches	92
4.	Emerging Fields and Problems	95
4.1	Community Discovery in Dynamic Networks	95
4.2	Community Discovery in Heterogeneous Networks	97
4.3	Community Discovery in Directed Networks	98
4.4	Coupling Content and Relationship Information for Community Discovery	100
5.	Crosscutting Issues and Concluding Remarks	102
	References	104
5		
	Node Classification in Social Networks	115
	<i>Smriti Bhagat, Graham Cormode and S. Muthukrishnan</i>	
1.	Introduction	116
2.	Problem Formulation	119
2.1	Representing data as a graph	119
2.2	The Node Classification Problem	123
3.	Methods using Local Classifiers	124
3.1	Iterative Classification Method	125
4.	Random Walk based Methods	127
4.1	Label Propagation	129
4.2	Graph Regularization	132
4.3	Adsorption	134
5.	Applying Node Classification to Large Social Networks	136
5.1	Basic Approaches	137
5.2	Second-order Methods	137
5.3	Implementation within Map-Reduce	138

6.	Related approaches	139
6.1	Inference using Graphical Models	139
6.2	Metric labeling	140
6.3	Spectral Partitioning	141
6.4	Graph Clustering	142
7.	Variations on Node Classification	142
7.1	Dissimilarity in Labels	142
7.2	Edge Labeling	143
7.3	Label Summarization	144
8.	Concluding Remarks	144
8.1	Future Directions and Challenges	145
8.2	Further Reading	146
	References	146
6		
	Evolution in Social Networks: A Survey	149
	<i>Myra Spiliopoulou</i>	
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	177
	<i>Jimeng Sun and Jie Tang</i>	
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	215
	<i>Theodoros Lappas, Kun Liu and Evimaria Terzi</i>	
1.	Introduction	216

2.	Definitions and Notation	217
3.	Expert Location without Graph Constraints	219
3.1	Language Models for Document Information Retrieval	219
3.2	Language Models for Expert Location	220
3.3	Further Reading	221
4.	Expert Location with Score Propagation	221
4.1	The PageRank Algorithm	222
4.2	HITS Algorithm	223
4.3	Expert Score Propagation	224
4.4	Further Reading	226
5.	Expert Team Formation	227
5.1	Metrics	227
5.2	Forming Teams of Experts	228
5.3	Further Reading	232
6.	Other Related Approaches	232
6.1	Agent-based Approach	233
6.2	Influence Maximization	233
7.	Expert Location Systems	235
8.	Conclusions	235
	References	236
9		
A	Survey of Link Prediction in Social Networks	243
	<i>Mohammad Al Hasan and Mohammed J. Zaki</i>	
1.	Introduction	244
2.	Background	245
3.	Feature based Link Prediction	246
3.1	Feature Set Construction	247
3.2	Classification Models	253
4.	Bayesian Probabilistic Models	259
4.1	Link Prediction by Local Probabilistic Models	259
4.2	Network Evolution based Probabilistic Model	261
4.3	Hierarchical Probabilistic Model	263
5.	Probabilistic Relational Models	264
5.1	Relational Bayesian Network	266
5.2	Relational Markov Network	266
6.	Linear Algebraic Methods	267
7.	Recent development and Future Works	269
	References	270
10		
	Privacy in Social Networks: A Survey	277
	<i>Elena Zheleva and Lise Getoor</i>	
1.	Introduction	277
2.	Privacy breaches in social networks	280
2.1	Identity disclosure	281
2.2	Attribute disclosure	282
2.3	Social link disclosure	283
2.4	Affiliation link disclosure	284
3.	Privacy definitions for publishing data	286
3.1	$k$ -anonymity	288

3.2	<i>l</i> -diversity and <i>t</i> -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
11		
	Visualizing Social Networks	307
	<i>Carlos D. Correa and Kwan-Liu Ma</i>	
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
12		
	Data Mining in Social Media	327
	<i>Geoffrey Barbier and Huan Liu</i>	
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
13		
	Text Mining in Social Networks	353
	<i>Charu C. Aggarwal and Haixun Wang</i>	
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

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

6.3	Online Learning	429
7.	Discussion of Data Sets and Industrial Systems	432
8.	Discussion of Future Directions	434
8.1	Content-based Recommendation and Advertisements	434
8.2	Multimedia Information Networks via Cloud Computing	434
	References	436
16		
An Overview of Social Tagging and Applications		447
<i>Manish Gupta, Rui Li, Zhijun Yin and Jiawei Han</i>		
1.	Introduction	448
1.1	Problems with Metadata Generation and Fixed Taxonomies	449
1.2	Folksonomies as a Solution	449
1.3	Outline	450
2.	Tags: Why and What?	451
2.1	Different User Tagging Motivations	451
2.2	Kinds of Tags	452
2.3	Categorizers Versus Describers	453
2.4	Linguistic Classification of Tags	454
2.5	Game-based Tagging	455
3.	Tag Generation Models	455
3.1	Polya Urn Generation Model	456
3.2	Language Model	458
3.3	Other Influence Factors	459
4.	Tagging System Design	460
5.	Tag analysis	462
5.1	Tagging Distributions	463
5.2	Identifying Tag Semantics	464
5.3	Tags Versus Keywords	466
6.	Visualization of Tags	467
6.1	Tag Clouds for Browsing/Search	468
6.2	Tag Selection for Tag Clouds	468
6.3	Tag Hierarchy Generation	469
6.4	Tag Clouds Display Format	470
6.5	Tag Evolution Visualization	470
6.6	Popular Tag Cloud Demos	471
7.	Tag Recommendations	472
7.1	Using Tag Quality	472
7.2	Using Tag Co-occurrences	473
7.3	Using Mutual Information between Words, Documents and Tags	474
7.4	Using Object Features	474
8.	Applications of Tags	475
8.1	Indexing	475
8.2	Search	475
8.3	Taxonomy Generation	480
8.4	Public Library Cataloging	481
8.5	Clustering and Classification	482
8.6	Social Interesting Discovery	483
8.7	Enhanced Browsing	484
9.	Integration	485

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





<http://www.springer.com/978-1-4419-8461-6>

Social Network Data Analytics

Aggarwal, C.C. (Ed.)

2011, XIV, 502 p., Hardcover

ISBN: 978-1-4419-8461-6