Contents

List of Figures xvii

List of Tables xxi

List of Algorithms xxiii

Preface xxv

Acknowledgments xxvii

I Algorithms 1

1 Graphs and Matrices 3
J. Kepner
1.1 Motivation . . . 3
1.2 Algorithms . . . 4
1.2.1 Graph adjacency matrix duality . . . 4
1.2.2 Graph algorithms as semirings . . . 5
1.2.3 Tensors . . . 6
1.3 Data . . . 6
1.3.1 Simulating power law graphs . . . 6
1.3.2 Kronecker theory . . . 7
1.4 Computation . . . 7
1.4.1 Graph analysis metrics . . . 7
1.4.2 Sparse matrix storage . . . 8
1.4.3 Sparse matrix multiply . . . 9
1.4.4 Parallel programming . . . 9
1.4.5 Parallel matrix multiply performance . . . 10
1.5 Summary . . . 12
References . . . 12

2 Linear Algebraic Notation and Definitions 13
E. Robinson, J. Kepner, and J. Gilbert
2.1 Graph notation . . . 13

2.2 Array notation . . . 14
2.3 Algebraic notation . . . 14
2.3.1 Semirings and related structures . . . 14
2.3.2 Scalar operations . . . 15
2.3.3 Vector operations . . . 15
2.3.4 Matrix operations . . . 16
2.4 Array storage and decomposition . . . 16
2.4.1 Sparse . . . 16
2.4.2 Parallel . . . 17

3 Connected Components and Minimum Paths 19
C. M. Rader
3.1 Introduction . . . 19
3.2 Strongly connected components . . . 20
3.2.1 Nondirected links . . . 21
3.2.2 Computing **C** quickly . . . 22
3.3 Dynamic programming, minimum paths, and matrix exponentiation . . . 23
3.3.1 Matrix powers . . . 25
3.4 Summary . . . 26
References . . . 27

4 Some Graph Algorithms in an Array-Based Language 29
V. B. Shah, J. Gilbert, and S. Reinhardt
4.1 Motivation . . . 29
4.2 Sparse matrices and graphs . . . 30
4.2.1 Sparse matrix multiplication . . . 31
4.3 Graph algorithms . . . 32
4.3.1 Breadth-first search . . . 32
4.3.2 Strongly connected components . . . 33
4.3.3 Connected components . . . 34
4.3.4 Maximal independent set . . . 35
4.3.5 Graph contraction . . . 35
4.3.6 Graph partitioning . . . 37
4.4 Graph generators . . . 39
4.4.1 Uniform random graphs . . . 39
4.4.2 Power law graphs . . . 39
4.4.3 Regular geometric grids . . . 39
References . . . 41

5 Fundamental Graph Algorithms 45
J. T. Fineman and E. Robinson
5.1 Shortest paths . . . 45
5.1.1 Bellman–Ford . . . 46
5.1.2 Computing the shortest path tree (for Bellman–Ford) . . . 48
5.1.3 Floyd–Warshall . . . 53

5.2 Minimum spanning tree . . . 55
5.2.1 Prim's . . . 55
References . . . 58

6 Complex Graph Algorithms 59
E. Robinson
6.1 Graph clustering . . . 59
6.1.1 Peer pressure clustering . . . 59
6.1.2 Matrix formulation . . . 66
6.1.3 Other approaches . . . 67
6.2 Vertex betweenness centrality . . . 68
6.2.1 History . . . 68
6.2.2 Brandes' algorithm . . . 69
6.2.3 Batch algorithm . . . 75
6.2.4 Algorithm for weighted graphs . . . 78
6.3 Edge betweenness centrality . . . 78
6.3.1 Brandes' algorithm . . . 78
6.3.2 Block algorithm . . . 83
6.3.3 Algorithm for weighted graphs . . . 84
References . . . 84

7 Multilinear Algebra for Analyzing Data with Multiple Linkages 85
D. Dunlavy, T. Kolda, and W. P. Kegelmeyer
7.1 Introduction . . . 86
7.2 Tensors and the CANDECOMP/PARAFAC decomposition . . . 87
7.2.1 Notation . . . 87
7.2.2 Vector and matrix preliminaries . . . 88
7.2.3 Tensor preliminaries . . . 88
7.2.4 The CP tensor decomposition . . . 89
7.2.5 CP-ALS algorithm . . . 89
7.3 Data . . . 91
7.3.1 Data as a tensor . . . 91
7.3.2 Quantitative measurements on the data . . . 93
7.4 Numerical results . . . 93
7.4.1 Community identification . . . 94
7.4.2 Latent document similarity . . . 95
7.4.3 Analyzing a body of work via centroids . . . 97
7.4.4 Author disambiguation . . . 98
7.4.5 Journal prediction via ensembles of tree classifiers . . . 103
7.5 Related work . . . 106
7.5.1 Analysis of publication data . . . 106
7.5.2 Higher order analysis in data mining . . . 107
7.5.3 Other related work . . . 108
7.6 Conclusions and future work . . . 108
7.7 Acknowledgments . . . 110
References . . . 110

8 Subgraph Detection **115**
J. Kepner
8.1 Graph model . . . 115
8.1.1 Vertex/edge schema . . . 116
8.2 Foreground: Hidden Markov model . . . 118
8.2.1 Path moments . . . 118
8.3 Background model: Kronecker graphs . . . 120
8.4 Example: Tree finding . . . 120
8.4.1 Background: Power law . . . 120
8.4.2 Foreground: Tree . . . 121
8.4.3 Detection problem . . . 121
8.4.4 Degree distribution . . . 123
8.5 SNR, PD, and PFA . . . 124
8.5.1 First and second neighbors . . . 125
8.5.2 Second neighbors . . . 125
8.5.3 First neighbors . . . 126
8.5.4 First neighbor leaves . . . 126
8.5.5 First neighbor branches . . . 127
8.5.6 SNR hierarchy . . . 128
8.6 Linear filter . . . 129
8.6.1 Find nearest neighbors . . . 129
8.6.2 Eliminate high degree nodes . . . 129
8.6.3 Eliminate occupied nodes . . . 130
8.6.4 Find high probability nodes . . . 130
8.6.5 Find high degree nodes . . . 131
8.7 Results and conclusions . . . 132
References . . . 133

II Data **135**

9 Kronecker Graphs **137**
J. Leskovec
9.1 Introduction . . . 138
9.2 Relation to previous work on network modeling . . . 140
9.2.1 Graph patterns . . . 140
9.2.2 Generative models of network structure . . . 142
9.2.3 Parameter estimation of network models . . . 142
9.3 Kronecker graph model . . . 143
9.3.1 Main idea . . . 143
9.3.2 Analysis of Kronecker graphs . . . 147
9.3.3 Stochastic Kronecker graphs . . . 152
9.3.4 Additional properties of Kronecker graphs . . . 154
9.3.5 Two interpretations of Kronecker graphs . . . 155
9.3.6 Fast generation of stochastic Kronecker graphs . . . 157
9.3.7 Observations and connections . . . 158

9.4 Simulations of Kronecker graphs . . . 159
9.4.1 Comparison to real graphs . . . 159
9.4.2 Parameter space of Kronecker graphs . . . 161
9.5 Kronecker graph model estimation . . . 163
9.5.1 Preliminaries . . . 165
9.5.2 Problem formulation . . . 166
9.5.3 Summing over the node labelings . . . 169
9.5.4 Efficiently approximating likelihood and gradient . . . 172
9.5.5 Calculating the gradient . . . 173
9.5.6 Determining the size of an initiator matrix . . . 173
9.6 Experiments on real and synthetic data . . . 174
9.6.1 Permutation sampling . . . 174
9.6.2 Properties of the optimization space . . . 180
9.6.3 Convergence of the graph properties . . . 181
9.6.4 Fitting to real-world networks . . . 181
9.6.5 Fitting to other large real-world networks . . . 187
9.6.6 Scalability . . . 190
9.7 Discussion . . . 193
9.8 Conclusion . . . 195
Appendix: Table of Networks . . . 196
References . . . 198

10 The Kronecker Theory of Power Law Graphs 205
J. Kepner
10.1 Introduction . . . 205
10.2 Overview of results . . . 206
10.3 Kronecker graph generation algorithm . . . 208
10.3.1 Explicit adjacency matrix . . . 208
10.3.2 Stochastic adjacency matrix . . . 209
10.3.3 Instance adjacency matrix . . . 211
10.4 A simple bipartite model of Kronecker graphs . . . 211
10.4.1 Bipartite product . . . 212
10.4.2 Bipartite Kronecker exponents . . . 213
10.4.3 Degree distribution . . . 215
10.4.4 Betweenness centrality . . . 216
10.4.5 Graph diameter and eigenvalues . . . 218
10.4.6 Iso-parametric ratio . . . 219
10.5 Kronecker products and useful permutations . . . 220
10.5.1 Sparsity . . . 220
10.5.2 Permutations . . . 220
10.5.3 Pop permutation . . . 221
10.5.4 Bipartite permutation . . . 221
10.5.5 Recursive bipartite permutation . . . 221
10.5.6 Bipartite index tree . . . 224
10.6 A more general model of Kronecker graphs . . . 225
10.6.1 Sparsity analysis . . . 226

10.6.2 Second order terms . . . 227
10.6.3 Higher order terms . . . 230
10.6.4 Degree distribution . . . 231
10.6.5 Graph diameter and eigenvalues . . . 231
10.6.6 Iso-parametric ratio . . . 233
10.7 Implications of bipartite substructure . . . 234
10.7.1 Relation between explicit and instance graphs . . . 234
10.7.2 Clustering power law graphs . . . 237
10.7.3 Dendragram and power law graphs . . . 238
10.8 Conclusions and future work . . . 238
10.9 Acknowledgments . . . 239
References . . . 239

11 Visualizing Large Kronecker Graphs 241
H. Nguyen, J. Kepner, and A. Edelman
11.1 Introduction . . . 241
11.2 Kronecker graph model . . . 242
11.3 Kronecker graph generator . . . 243
11.4 Analyzing Kronecker graphs . . . 243
11.4.1 Graph metrics . . . 243
11.4.2 Graph view . . . 245
11.4.3 Organic growth simulation . . . 245
11.5 Visualizing Kronecker graphs in 3D . . . 246
11.5.1 Embedding Kronecker graphs onto a sphere surface . . . 247
11.5.2 Visualizing Kronecker graphs on parallel system . . . 247
References . . . 250

III Computation 251

12 Large-Scale Network Analysis 253
D. A. Bader, C. Heitsch, and K. Madduri
12.1 Introduction . . . 254
12.2 Centrality metrics . . . 255
12.3 Parallel centrality algorithms . . . 258
12.3.1 Optimizations for real-world graphs . . . 262
12.4 Performance results and analysis . . . 264
12.4.1 Experimental setup . . . 264
12.4.2 Performance results . . . 266
12.5 Case study: Betweenness applied to protein-interaction networks 268
12.6 Integer torus: Betweenness conjecture . . . 272
12.6.1 Proof of conjecture when n is odd . . . 274
12.6.2 Proof of conjecture when n is even . . . 276
References . . . 280

13 Implementing Sparse Matrices for Graph Algorithms 287
A. Buluç, J. Gilbert, and V. B. Shah
13.1 Introduction . . . 287
13.2 Key primitives . . . 291
13.3 Triples . . . 293
13.3.1 Unordered triples . . . 294
13.3.2 Row ordered triples . . . 298
13.3.3 Row-major ordered triples . . . 302
13.4 Compressed sparse row/column . . . 305
13.4.1 CSR and adjacency lists . . . 305
13.4.2 CSR on key primitives . . . 306
13.5 Case study: Star-P . . . 308
13.5.1 Sparse matrices in Star-P . . . 308
13.6 Conclusions . . . 310
References . . . 310

14 New Ideas in Sparse Matrix Matrix Multiplication 315
A. Buluç and J. Gilbert
14.1 Introduction . . . 316
14.2 Sequential sparse matrix multiply . . . 317
14.2.1 Layered graphs for different formulations of SpGEMM . . . 318
14.2.2 Hypersparse matrices . . . 320
14.2.3 DCSC data structure . . . 321
14.2.4 A sequential algorithm to multiply hypersparse matrices . . . 322
14.3 Parallel algorithms for sparse GEMM . . . 326
14.3.1 1D decomposition . . . 326
14.3.2 2D decomposition . . . 326
14.3.3 Sparse 1D algorithm . . . 327
14.3.4 Sparse Cannon . . . 327
14.3.5 Sparse SUMMA . . . 328
14.4 Analysis of parallel algorithms . . . 328
14.4.1 Scalability of the 1D algorithm . . . 329
14.4.2 Scalability of the 2D algorithms . . . 330
14.5 Performance modeling of parallel algorithms . . . 331
References . . . 334

15 Parallel Mapping of Sparse Computations 339
E. Robinson, N. Bliss, and S. Mohindra
15.1 Introduction . . . 339
15.2 Lincoln Laboratory mapping and optimization environment . . . 340
15.2.1 LLMOE overview . . . 341
15.2.2 Mapping in LLMOE . . . 343
15.2.3 Mapping performance results . . . 347
References . . . 352

16 Fundamental Questions in the Analysis of Large Graphs 353
J. Kepner, D. A. Bader, B. Bond, N. Bliss, C. Faloutsos, B. Hendrickson, J. Gilbert, and E. Robinson
16.1 Ontology, schema, data model 354
16.2 Time evolution 354
16.3 Detection theory 355
16.4 Algorithm scaling 355
16.5 Computer architecture 356

Index 359