Contents

1 Introduction and Overview 1
1.1 Why Networks? 1
1.2 Examples of Networks 3
1.2.1 Technological Networks 3
1.2.2 Social Networks 5
1.2.3 Biological Networks 7
1.2.4 Information Networks 9
1.3 About this Book 11

2 Preliminaries 15
2.1 Background on Graphs 15
2.1.1 Basic Definitions and Concepts 16
2.1.2 Families of Graphs 18
2.1.3 Graphs and Matrix Algebra 20
2.1.4 Graph Data Structures and Algorithms 21
2.2 Background in Probability and Statistics 24
2.2.1 Probability 25
2.2.2 Principles of Statistical Inference 31
2.2.3 Methods of Statistical Inference: Tutorials 32
2.3 Statistical Analysis of Network Data: *Prelude* 42
2.4 Additional Related Topics and Reading 45
Exercises 45

3 Mapping Networks 49
3.1 Introduction 49
3.2 Collecting Relational Network Data 50
3.2.1 Measurement of System Elements and Interactions 51
3.2.2 Enumerated, Partial, and Sampled Data 54
3.3 Constructing Network Graph Representations 56
3.4 Visualizing Network Graphs 58
3.4.1 Elements of Graph Visualization 58

3.4.2 Methods of Graph Visualization 60
3.5 Case Studies 63
3.5.1 Mapping ‘Science’ 65
3.5.2 Mapping the Internet 68
3.6 Mapping Dynamic Networks 74
3.7 Additional Related Topics and Reading 76
Exercises 77

4 Descriptive Analysis of Network Graph Characteristics 79
4.1 Introduction 79
4.2 Vertex and Edge Characteristics 80
4.2.1 Degree 80
4.2.2 Centrality 88
4.3 Characterizing Network Cohesion 94
4.3.1 Local Density 94
4.3.2 Connectivity 97
4.3.3 Graph Partitioning 102
4.3.4 Assortativity and Mixing 111
4.4 Case Study: Analysis of an Epileptic Seizure 114
4.5 Characterizing Dynamic Network Graphs 116
4.6 Additional Related Topics and Reading 119
Exercises 120

5 Sampling and Estimation in Network Graphs 123
5.1 Introduction 123
5.2 Background on Statistical Sampling Theory 126
5.2.1 Horvitz-Thompson Estimation for Totals 126
5.2.2 Estimation of Group Size 129
5.3 Common Network Graph Sampling Designs 131
5.3.1 Induced and Incident Subgraph Sampling 131
5.3.2 Star and Snowball Sampling 133
5.3.3 Link Tracing 136
5.4 Estimation of Totals in Network Graphs 137
5.4.1 Vertex Totals 137
5.4.2 Totals on Vertex Pairs 138
5.4.3 Totals of Higher Order 141
5.4.4 Effects of Design, Measurement, and Total 143
5.5 Estimation of Network Group Size 145
5.6 Other Network Graph Estimation Problems 149
5.7 Additional Related Topics and Reading 151
Exercises 151

6 Models for Network Graphs 153
6.1 Introduction 153
6.2 Random Graph Models 154
6.2.1 Classical Random Graph Models 156
6.2.2 Generalized Random Graph Models 158
6.2.3 Simulating Random Graph Models 159
6.2.4 Statistical Application of Random Graph Models 162
6.3 Small-World Models 169
6.3.1 The Watts-Strogatz Model 169
6.3.2 Other Small-World Network Models 171
6.4 Network Growth Models 172
6.4.1 Preferential Attachment Models 173
6.4.2 Copying Models 176
6.4.3 Fitting Network Growth Models 178
6.5 Exponential Random Graph Models 180
6.5.1 Model Specification 180
6.5.2 Fitting Exponential Random Graph Models 185
6.5.3 Goodness-of-Fit and Model Degeneracy 187
6.5.4 Case Study: Modeling Collaboration Among Lawyers 188
6.6 Challenges in Modeling Network Graphs 191
6.7 Additional Related Topics and Reading 193
Exercises 195

7 Network Topology Inference 197
7.1 Introduction 197
7.2 Link Prediction 199
7.2.1 Informal Scoring Methods 201
7.2.2 Probabilistic Classification Methods 202
7.2.3 Case Study: Predicting Lawyer Collaboration 205
7.3 Inference of Association Networks 207
7.3.1 Correlation Networks 209
7.3.2 Partial Correlation Networks 212
7.3.3 Gaussian Graphical Model Networks 216
7.3.4 Case Study: Inferring Genetic Regulatory Interactions 220
7.4 Tomographic Network Topology Inference 223
7.4.1 Tomographic Inference of Tree Topologies 225
7.4.2 Methods Based on Hierarchical Clustering 228
7.4.3 Likelihood-based Methods 231
7.4.4 Summarizing Collections of Trees 234
7.4.5 Case Study: Computer Network Topology Identification 236
7.5 Additional Related Topics and Reading 241
Exercises 242

8 Modeling and Prediction for Processes on Network Graphs 245
8.1 Introduction 245
8.2 Nearest Neighbor Prediction 246
8.3 Markov Random Fields 249
8.3.1 Markov Random Field Models 249
8.3.2 Inference and Prediction for Markov Random Fields 252
8.3.3 Related Probabilistic Models 256
8.4 Kernel-based Regression 257
8.4.1 Kernel Regression on Graphs 258
8.4.2 Designing Kernels on Graphs 262
8.5 Case Study: Predicting Protein Function 266
8.6 Modeling and Prediction for Dynamic Processes 271
8.6.1 Epidemic Processes: An Illustration 272
8.6.2 Other Dynamic Processes 280
8.7 Additional Related Topics and Reading 281
Exercises 282

9 Analysis of Network Flow Data 285
9.1 Introduction 285
9.2 Gravity Models 287
9.2.1 Model Specification 288
9.2.2 Inference for Gravity Models 292
9.3 Traffic Matrix Estimation 297
9.3.1 Static Methods 298
9.3.2 Dynamic Methods 306
9.3.3 Case Study: Internet Traffic Matrix Estimation 310
9.4 Estimation of Network Flow Costs 316
9.4.1 Link Costs from End-to-end Measurements 317
9.4.2 Path Costs from End-to-end Measurements 321
9.5 Additional Related Topics and Reading 328
Exercises 330

10 Graphical Models 333
10.1 Introduction 333
10.2 Defining Graphical Models 334
10.2.1 Directed Graphical Models 335
10.2.2 Undirected Graphical Models 339
10.3 Inference for Graphical Models 342
10.4 Additional Related Topics and Reading 344

Glossary of Notation 345

References 347

Author Index 373

Subject Index 381