Exploratory Social Network Analysis with Pajek

*Revised and Expanded Second Edition*

This is the first textbook on social network analysis integrating theory, applications, and professional software for performing network analysis (Pajek). Step by step, the book introduces the main structural concepts and their applications in social research with exercises to test understanding. In each chapter, each theoretical section is followed by an application section explaining how to perform the network analyses with Pajek software. Pajek software and datasets for all examples are freely available, so the reader can learn network analysis by doing it. In addition, each chapter offers case studies for practicing network analysis. In the end, the reader will have the knowledge, skills, and tools to apply social network analysis in all social sciences, ranging from anthropology and sociology to business administration and history.

Wouter de Nooy is Associate Professor in the Department of Communication Science at the University of Amsterdam, The Netherlands, and a member of the Amsterdam School of Communication Research (ASCoR) and the Netherlands School of Communication Research (NESCOr).

Andrej Mrvar is Associate Professor of Social Science Informatics on the Faculty of Social Sciences, University of Ljubljana, Slovenia. He won several awards for graph drawings at competitions between 1995 and 2005. He has edited *Metodoloski zvezki – Advances in Methodology and Statistics* since 2000.

Vladimir Batagelj is Professor of Discrete and Computational Mathematics at the University of Ljubljana, Slovenia, and a member of the editorial boards of *Informatica* and *Journal of Social Structure*. He has authored several articles in *Communications of ACM, Psychometrika, Journal of Classification, Social Networks, Discrete Mathematics, Algorithmica, Journal of Mathematical Sociology, Quality and Quantity, Informatica, Lecture Notes in Computer Science, Studies in Classification, Data Analysis, and Knowledge Organization.*
Structural Analysis in the Social Sciences

Mark Granovetter, editor

The series *Structural Analysis in the Social Sciences* presents studies that analyze social behavior and institutions by reference to relations among such concrete social entities as persons, organizations, and nations. Relational analysis contrasts with reductionist methodological individualism on the one hand and with macro-level determinism on the other, whether based on technology, material conditions, economic conflict, adaptive evolution, or functional imperatives. In this more intellectually flexible structural middle ground, analysts situate actors and their relations in a variety of contexts. Since the series began in 1987, its authors have variously focused on small groups, history, culture, politics, kinship, aesthetics, economics, and complex organizations, creatively theorizing how these shape and in turn are shaped by social relations. Their style and methods have ranged widely, from intense, long-term ethnographic observation to highly abstract mathematical models. Their disciplinary affiliations have included history, anthropology, sociology, political science, business, economics, mathematics, and computer science. Some have made explicit use of “social network analysis,” including many of the cutting-edge and standard works of that approach, whereas others have eschewed formal analysis and used “networks” as a fruitful orienting metaphor. All have in common a sophisticated and subtle approach that forcefully illuminates our complex social world.

Other Books in the Series

1. Mark S. Mizruchi and Michael Schwartz, eds., *Intercorporate Relations: The Structural Analysis of Business*
2. Barry Wellman and S. D. Berkowitz, eds., *Social Structures: A Network Approach*
3. Ronald L. Breiger, ed., *Social Mobility and Social Structure*
4. David Knopke, *Political Networks: The Structural Perspective*
6. Kyriakos Kontopoulos, *The Logics of Social Structure*
7. Philippa Pattison, *Algebraic Models for Social Structure*
8. Stanley Wasserman and Katherine Faust, *Social Network Analysis: Methods and Applications*
9. Gary Herrigel, *Industrial Constructions: The Sources of German Industrial Power*
11. Per Hage and Frank Harary, *Island Networks: Communication, Kinship, and Classification Structures in Oceana*
12. Thomas Schweizer and Douglas R. White, eds., *Kinship, Networks and Exchange*
15. Rebecca Adams and Graham Allan, *Placing Friendship in Context*

(continued after the index)
Exploratory Social Network Analysis with Pajek

*Revised and Expanded Second Edition*

WOUTER DE NOOY
University of Amsterdam

ANDREJ MRVAR
University of Ljubljana

VLADIMIR BATAGELJ
University of Ljubljana
To Anuška,

who makes things happen
Contents

Figures

Table

Preface to the Second Edition

Preface to the First Edition

PART I – FUNDAMENTALS

1 Looking for Social Structure

1.1 Introduction

1.2 Sociometry and Sociogram

1.3 Exploratory Social Network Analysis

1.3.1 Network Definition

1.3.2 Manipulation

1.3.3 Calculation

1.3.4 Visualization

1.4 Assembling a Social Network

1.5 Summary

1.6 Questions

1.7 Assignment

1.8 Further Reading

1.9 Answers

2 Attributes and Relations

2.1 Introduction

2.2 Example: The World System

2.3 Partitions

2.4 Reduction of a Network

2.4.1 Local View

2.4.2 Global View

2.4.3 Contextual View

2.5 Vectors and Coordinates
2.6 Network Analysis and Statistics 57
2.7 Summary 60
2.8 Questions 61
2.9 Assignment 62
2.10 Further Reading 62
2.11 Answers 63

PART II – COHESION
3 Cohesive Subgroups 71
3.1 Introduction 71
3.2 Example 71
3.3 Density and Degree 73
3.4 Components 77
3.5 Cores 81
3.6 Cliques and Complete Subnetworks 84
3.7 Summary 90
3.8 Questions 92
3.9 Assignment 94
3.10 Further Reading 94
3.11 Answers 94

4 Sentiments and Friendship 97
4.1 Introduction 97
4.2 Balance Theory 97
4.3 Example 100
4.4 Detecting Structural Balance and Clusterability 101
4.5 Development in Time 107
4.6 Summary 110
4.7 Questions 111
4.8 Assignment 112
4.9 Further Reading 113
4.10 Answers 113

5 Affiliations 116
5.1 Introduction 116
5.2 Example 117
5.3 Two-Mode and One-Mode Networks 118
5.4 Islands 124
5.5 The Third Dimension 129
5.6 Summary 133
5.7 Questions 133
5.8 Assignment 134
5.9 Further Reading 135
5.10 Answers 136
### Contents

**PART III – BROKERAGE**

6 Center and Periphery

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Introduction</td>
<td>141</td>
</tr>
<tr>
<td>6.2 Example</td>
<td>141</td>
</tr>
<tr>
<td>6.3 Distance</td>
<td>143</td>
</tr>
<tr>
<td>6.4 Betweenness</td>
<td>150</td>
</tr>
<tr>
<td>6.5 Eigenvector Centrality</td>
<td>153</td>
</tr>
<tr>
<td>6.6 Summary</td>
<td>154</td>
</tr>
<tr>
<td>6.7 Questions</td>
<td>155</td>
</tr>
<tr>
<td>6.8 Assignment</td>
<td>155</td>
</tr>
<tr>
<td>6.9 Further Reading</td>
<td>156</td>
</tr>
<tr>
<td>6.10 Answers</td>
<td>157</td>
</tr>
</tbody>
</table>

7 Brokers and Bridges

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Introduction</td>
<td>159</td>
</tr>
<tr>
<td>7.2 Example</td>
<td>160</td>
</tr>
<tr>
<td>7.3 Bridges and Bi-Components</td>
<td>161</td>
</tr>
<tr>
<td>7.4 Ego-Networks and Constraint</td>
<td>166</td>
</tr>
<tr>
<td>7.5 Affiliations and Brokerage Roles</td>
<td>173</td>
</tr>
<tr>
<td>7.6 Summary</td>
<td>178</td>
</tr>
<tr>
<td>7.7 Questions</td>
<td>179</td>
</tr>
<tr>
<td>7.8 Assignment</td>
<td>180</td>
</tr>
<tr>
<td>7.9 Further Reading</td>
<td>181</td>
</tr>
<tr>
<td>7.10 Answers</td>
<td>182</td>
</tr>
</tbody>
</table>

8 Diffusion

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Example</td>
<td>186</td>
</tr>
<tr>
<td>8.2 Contagion</td>
<td>189</td>
</tr>
<tr>
<td>8.3 Exposure and Thresholds</td>
<td>193</td>
</tr>
<tr>
<td>8.4 Critical Mass</td>
<td>200</td>
</tr>
<tr>
<td>8.5 Summary</td>
<td>205</td>
</tr>
<tr>
<td>8.6 Questions</td>
<td>207</td>
</tr>
<tr>
<td>8.7 Assignment</td>
<td>208</td>
</tr>
<tr>
<td>8.8 Further Reading</td>
<td>208</td>
</tr>
<tr>
<td>8.9 Answers</td>
<td>209</td>
</tr>
</tbody>
</table>

**PART IV – RANKING**

9 Prestige

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1 Introduction</td>
<td>215</td>
</tr>
<tr>
<td>9.2 Example</td>
<td>216</td>
</tr>
<tr>
<td>9.3 Popularity and Indegree</td>
<td>217</td>
</tr>
<tr>
<td>9.4 Correlation</td>
<td>219</td>
</tr>
<tr>
<td>9.5 Domains</td>
<td>221</td>
</tr>
<tr>
<td>9.6 Proximity Prestige</td>
<td>225</td>
</tr>
</tbody>
</table>
Contents

9.7 Summary 228
9.8 Questions 229
9.9 Assignment 230
9.10 Further Reading 231
9.11 Answers 232

10 Ranking 234
10.1 Introduction 234
10.2 Example 235
10.3 Triadic Analysis 235
10.4 Acyclic Networks 243
10.5 Symmetric-Acyclic Decomposition 246
10.6 Summary 252
10.7 Questions 253
10.8 Assignment 255
10.9 Further Reading 255
10.10 Answers 256

11 Genealogies and Citations 259
11.1 Introduction 259
11.2 Example I: Genealogy of the Ragusan Nobility 259
11.3 Family Trees 260
11.4 Social Research on Genealogies 268
11.5 Example II: Citations Among Papers on Network Centrality 278
11.6 Citations 279
11.7 Summary 288
11.8 Questions 289
11.9 Assignment 1 290
11.10 Assignment 2 290
11.11 Further Reading 290
11.12 Answers 291

PART V – ROLES

12 Blockmodels 299
12.1 Introduction 299
12.2 Matrices and Permutation 300
12.3 Roles and Positions: Equivalence 306
12.4 Blockmodeling 315
  12.4.1 Blockmodel 315
  12.4.2 Blockmodeling 317
  12.4.3 Regular Equivalence 322
12.5 Summary 327
12.6 Questions 328
Contents

12.7 Assignment 330
12.8 Further Reading 331
12.9 Answers 332

13 Random Graph Models 336
13.1 Introduction 336
13.2 Example 338
13.3 Modeling Overall Network Structure 340
  13.3.1 Classic Uniform Models 341
  13.3.2 Small-World Models 345
  13.3.3 Preferential Attachment Models 349
13.4 Monte Carlo Simulation 356
13.5 Summary 360
13.6 Questions 362
13.7 Assignment 364
13.8 Further Reading 364
13.9 Answers 365

Appendix 1 Getting Started with Pajek 369
  A1.1 Installation 369
  A1.2 Network Data Formats 369
  A1.3 Creating Network Files for Pajek 371
    A1.3.1 Within Pajek 371
    A1.3.2 Helper Software 373
    A1.3.3 Word Processor 373
    A1.3.4 Relational Database 376
  A1.4 Limitations 381
  A1.5 Updates of Pajek 382

Appendix 2 Exporting Visualizations 383
  A2.1 Export Formats 383
    A2.1.1 Bitmap 383
    A2.1.2 Encapsulated PostScript 384
    A2.1.3 Scalable Vector Graphics 385
    A2.1.4 Virtual Reality Modeling Language and X3D 387
    A2.1.5 MDL MOL and Kinemages 388
  A2.2 Layout Options 389
    A2.2.1 Top Frame on the Left – EPS/SVG Vertex Default 390
    A2.2.2 Bottom Frame on the Left – EPS/SVG Line Default 392
    A2.2.3 Top Frame on the Right 394
    A2.2.4 Middle Frame on the Right 394
    A2.2.5 Bottom Frame on the Right – EPS Border 395
# Contents

| Appendix 3 | Shortcut Key Combinations | 396 |
| A3.1 | Main Screen | 396 |
| A3.2 | Hierarchy Edit Screen | 397 |
| A3.3 | Draw Screen | 397 |

*Glossary* 399

*Index of Pajek and R Commands* 409

*Subject Index* 414
## Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dependencies between the chapters (for the second edition).</td>
<td>xxvii</td>
</tr>
<tr>
<td>2</td>
<td>Sociogram of dining-table partners.</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Partial listing of a multiple relations network data file for Pajek.</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>Pajek Main screen.</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>Menu structure in Pajek.</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>An information box in Pajek.</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>Report screen in Pajek.</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>Dialog box of Info &gt; Network &gt; General command.</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>Draw screen in Pajek.</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>Continue dialog box.</td>
<td>20</td>
</tr>
<tr>
<td>11</td>
<td>A selected option in the Draw screen.</td>
<td>22</td>
</tr>
<tr>
<td>12</td>
<td>Options menu of the Draw screen.</td>
<td>23</td>
</tr>
<tr>
<td>13</td>
<td>Textual output from [Draw] Info &gt; All Properties.</td>
<td>23</td>
</tr>
<tr>
<td>14</td>
<td>A 3-D rendering of the dining-table partners network.</td>
<td>25</td>
</tr>
<tr>
<td>15</td>
<td>Random network without lines.</td>
<td>27</td>
</tr>
<tr>
<td>16</td>
<td>Edit Network screen.</td>
<td>28</td>
</tr>
<tr>
<td>17</td>
<td>World trade of manufactures of metal and world system position.</td>
<td>38</td>
</tr>
<tr>
<td>18</td>
<td>Edit screen with partition according to world system position.</td>
<td>39</td>
</tr>
<tr>
<td>19</td>
<td>Vertex colors according to a partition in Pajek.</td>
<td>41</td>
</tr>
<tr>
<td>20</td>
<td>Trade ties within South America.</td>
<td>44</td>
</tr>
<tr>
<td>21</td>
<td>The Partitions menu.</td>
<td>45</td>
</tr>
<tr>
<td>22</td>
<td>World system positions in South America: (2) semiperiphery and (3) periphery.</td>
<td>45</td>
</tr>
<tr>
<td>23</td>
<td>Trade in manufactures of metal among continents (imports in thousands of U.S. dollars).</td>
<td>46</td>
</tr>
<tr>
<td>24</td>
<td>Trade among continents in the Draw screen.</td>
<td>48</td>
</tr>
</tbody>
</table>
xvi  Figures

25  Contextual view of trade in South America.  49
26  Geographical view of world trade in manufactures of metal, ca. 1994.  52
27  Info > Vector dialog box.  52
28  Trade, position in the world system, and GDP per capita.  55
29  Aggregate trade in manufactures of metal among world system positions.  64
30  Contextual view of North American trade ties and (mean) GDP per capita.  65
31  Visiting ties in Attiro.  72
32  A simple unconnected directed network.  77
33  Strong components (contours) and family–friendship groupings (vertex colors and numbers) in the network of Attiro.  80
34  $k$-cores in the visiting network at Attiro.  82
35  $k$-cores.  83
36  Stacking or nesting of $k$-cores.  83
37  The complete triad and an example.  85
38  A hierarchy of cliques.  87
39  Viewing a hierarchy in an Edit screen.  88
40  Complete triads and family–friendship groupings (colors and numbers inside vertices).  89
41  Decision tree for the analysis of cohesive subgroups.  91
42  A Person–Other–Object (X) triple.  98
43  P-O-X triple as a signed digraph.  98
44  A balanced network.  100
45  First positive and negative choices between novices at T4.  102
46  Output listing of a Balance command.  105
47  Three solutions with one error.  106
48  Partial listing of Sampson.net.  108
49  Differences between two solutions with four classes.  114
50  A fragment of the Scottish directorates network.  119
51  One-mode network of firms created from the network in Figure 50.  120
52  One-mode network of directors derived from Figure 50.  121
53  Islands in the network of Scottish firms, 1904–1905 (contours added manually).  125
54  The islands network of Scottish firms (1904–1905) with industrial categories (class numbers) and capital (vertex size).  127
55  Islands in three dimensions.  130
56  Coordinate system of Pajek.  130
57  A landscape of islands in the Scottish firms network.  131
58  Communication ties within a sawmill.  142
### Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>59</td>
<td>Star-networks and line-networks.</td>
</tr>
<tr>
<td>60</td>
<td>Distances to or from Juan (vertex colors: Default GreyScale 1).</td>
</tr>
<tr>
<td>61</td>
<td>Geodesics between HP-1 and EM-4.</td>
</tr>
<tr>
<td>62</td>
<td>Betweenness centrality in the sawmill.</td>
</tr>
<tr>
<td>63</td>
<td>Communication network of striking employees.</td>
</tr>
<tr>
<td>64</td>
<td>Cut-vertices (gray) and bi-components (manually circled) in the strike network.</td>
</tr>
<tr>
<td>65</td>
<td>Hierarchy of bi-components and bridges in the strike network.</td>
</tr>
<tr>
<td>66</td>
<td>Three connected triads.</td>
</tr>
<tr>
<td>67</td>
<td>Alejandro’s ego-network.</td>
</tr>
<tr>
<td>68</td>
<td>Proportional strength of ties around Alejandro.</td>
</tr>
<tr>
<td>69</td>
<td>Constraints on Alejandro.</td>
</tr>
<tr>
<td>70</td>
<td>Energized constraint network.</td>
</tr>
<tr>
<td>71</td>
<td>Five brokerage roles of actor (v).</td>
</tr>
<tr>
<td>72</td>
<td>Bob’s ego-network.</td>
</tr>
<tr>
<td>73</td>
<td>Constraint inside groups.</td>
</tr>
<tr>
<td>74</td>
<td>Two overlapping cliques.</td>
</tr>
<tr>
<td>75</td>
<td>Friendship ties among superintendents and year of adoption.</td>
</tr>
<tr>
<td>76</td>
<td>Adoption of the modern math method: diffusion curve.</td>
</tr>
<tr>
<td>77</td>
<td>Diffusion by contacts in a random network ((N = 100, \text{vertex numbers indicate the distance from the source vertex})).</td>
</tr>
<tr>
<td>78</td>
<td>Diffusion from a central and a marginal vertex.</td>
</tr>
<tr>
<td>79</td>
<td>Adoption (vertex color) and exposure (in brackets) at the end of 1959.</td>
</tr>
<tr>
<td>80</td>
<td>Modern math network with arcs pointing toward later adopters.</td>
</tr>
<tr>
<td>81</td>
<td>Visiting ties and prestige leaders in San Juan Sur.</td>
</tr>
<tr>
<td>82</td>
<td><em>Partitions</em> menu in Pajek.</td>
</tr>
<tr>
<td>83</td>
<td>Distances to family 47 (represented by the numbers within the vertices).</td>
</tr>
<tr>
<td>84</td>
<td>Proximity prestige in a small network.</td>
</tr>
<tr>
<td>85</td>
<td>Student government discussion network.</td>
</tr>
<tr>
<td>86</td>
<td>An example of a network with ranks.</td>
</tr>
<tr>
<td>87</td>
<td>Triad types with their sequential numbers in Pajek.</td>
</tr>
<tr>
<td>88</td>
<td>Strong components in the student government discussion network.</td>
</tr>
<tr>
<td>89</td>
<td>Acyclic network with shrunk components.</td>
</tr>
<tr>
<td>90</td>
<td>Clusters of symmetric ties in the student government network.</td>
</tr>
</tbody>
</table>
Figures

91 Discussion network shrunk according to symmetric clusters. 247
92 Symmetric components in the (modified) student government discussion network. 248
93 The order of symmetric clusters according to the depth partition (acyclic). 250
94 Ranks in the student government discussion network. 251
95 Three generations of descendants to Petrus Gondola (years of birth). 261
96 Ore graph. 262
97 Descendants of Petrus Gondola and Ana Goce. 264
98 Shortest paths between Pauch and Margarita Gondola. 265
99 Structural relinking in an Ore graph. 270
100 P-graph. 271
101 Structural relinking in a P-graph. 272
102 Fragment of relinking grandchildren. 275
103 Centrality literature network in layers according to year of publication. 280
104 $k$-cores in the centrality literature network (without isolates). 282
105 Traversal weights in a citation network. 283
106 A main path in the centrality literature network. 286
107 Main path component of the centrality literature network (not all names are shown here). 287
108 Communication lines among striking employees. 300
109 The matrix of the strike network sorted by ethnic and age groups. 302
110 A network and a permutation. 303
111 Partial listing of the strike network as a binary matrix. 304
112 The strike network permuted according to ethnic and age groups. 305
113 Part of the permuted strike network displayed as a binary network. 306
114 Hypothetical ties among two instructors (i) and three students (s). 306
115 A dendrogram of similarities. 308
116 Imports of miscellaneous manufactures of metal and world system position in 1980. 309
117 Hierarchical clustering of the world trade network. 312
118 Hierarchical clustering of countries in the Hierarchy Edit screen. 313
119 An ideal core-periphery structure. 315
120 Image matrix and shrunk network. 316
121 Error in the imperfect core-periphery matrix. 318
Figures

122 *Optimize Partition* dialog box. 319
123 Output of the *Optimize Partition* procedure. 320
124 *Random Start* dialog box. 321
125 Matrix of the student government network. 323
126 Image matrix and error matrix for the student government network. 324
127 Assembling a blockmodel in Pajek. 326
128 Random versions of a small friendship network. 337
129 Political blogosphere, United States, February 8, 2005. 339
130 Small-world random graph generation: ring of local lines (left) and rewired lines (right). 346
131 Log-log degree distributions of the blogs network: absolute frequencies (left) and cumulative proportions (right). 352
132 *Read Network* dialog box. 370
133 A network in Pajek matrix format. 371
134 Editing vertex labels. 372
135 Edit Network screen. 372
136 An empty network in Pajek Arcs/Edges format. 374
137 A network in the Pajek Arcs/Edges format. 375
138 A network in the Pajek matrix format. 375
139 A two-mode network in the Pajek Arcs/Edges format. 376
140 Four tables in the world trade database (MS Access 97). 377
141 Contents of the *Countries* table (partial). 377
142 A lookup to the *Countries* table. 378
143 Export a report to plain text. 379
144 Tables and relations in the database of Scottish companies. 381
145 The *Options* screen. 389
146 Layout of a vertex and its label. 390
147 The x/y ratio of a vertex. 391
148 The position and orientation of a line label. 393
149 Gradients in SVG export: linear (left) and radial (right). 395
Tables

1. Tabular output of the command Info> Partition.  
2. Distribution of GNP per capita in classes.  
4. Cross-tabulation of world system positions (rows) and GDP per capita (columns).  
5. Frequency distribution of degree in the symmetrized network of visits.  
6. Error score with all choices at different moments ($\alpha = .5$).  
7. Error score with first choices only ($\alpha = .5$).  
8. Line multiplicity in the one-mode network of firms.  
9. Frequency tabulation of coordinator roles in the strike network.  
10. Adoption in the modern math network.  
11. Adoption rate and acceleration in the modern math diffusion curve.  
12. Fragment of Table 11.  
13. Indegree listing in Pajek.  
14. Input domain of $f_{47}$.  
15. Size of input domains in the visiting relations network.  
17. Triad census of the example network.  
18. Triad census of the student government network.  
19. Number of children of Petrus Gondola and his male descendants.  
22. Traversal weights in the centrality literature network.  
23. Dissimilarity scores in the example network.  
24. Cross-tabulation of initial (rows) and optimal partition (columns).
### Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>Final image matrix of the world trade network.</td>
<td>322</td>
</tr>
<tr>
<td>26</td>
<td>Monte Carlo simulation results: confidence intervals for the simple undirected blogs network.</td>
<td>357</td>
</tr>
<tr>
<td>27</td>
<td>Names of colors in Pajek.</td>
<td>391</td>
</tr>
</tbody>
</table>
Preface to the Second Edition

I go with him out in a shed in back and see he is selling a whole Harley machine in used parts, except for the frame, which the customer already has. He is selling them all for $125. Not a bad price at all.

Coming back I comment, “He’ll know something about motorcycles before he gets those together.”

Bill laughs. “And that’s the best way to learn, too.”

Robert M. Pirsig, *Zen and the Art of Motorcycle Maintenance*

To some of its readers, this book is an introduction to social network analysis; to other readers, it is a manual to Pajek software (http://pajek.imfm.si/doku.php). To us, it is both. As Patrick Doreian argued in his review of our book [In: *Social Networks* 28 (2006) 269–274], an understanding of social network analysis is required for proper use of Pajek, and, vice versa, understanding the concepts and logic of Pajek fosters comprehension of network concepts. In this second edition, we have aimed to strengthen both aspects, updating the discussion of the Pajek interface and commands to include several capabilities that have been implemented since we submitted the text of the first edition, such as multiplex networks (Section 1.3.1), eigenvector centrality (Section 6.5), matrix multiplication (Section 11.3), and using Pajek output in R (Chapters 5 and 13). The new capabilities cover some important advances in social network analysis, including random graph models to which we have dedicated a new chapter.

We expanded the Further Reading sections with references to seminal, often-cited texts. This should allow the reader to trace the literature on the selected topic in bibliographic and citation databases. For more comprehensive lists of literature, we refer to two other volumes in this series: S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge: Cambridge University Press, 1994) and P. J. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis* (Cambridge: Cambridge University Press, 2005).
Preface to the Second Edition


We hope that this second edition will continue to stimulate analysts to sharpen their understanding of social networks and expand their command of network analytic tools.
In the social sciences, social network analysis has become a powerful methodological tool alongside statistics. Network concepts have been defined, tested, and applied in research traditions throughout the social sciences, ranging from anthropology and sociology to business administration and history.

This book is the first textbook on social network analysis integrating theory, applications, and professional software for performing network analysis. It introduces structural concepts and their applications in social research with exercises to improve skills, questions to test understanding, and case studies to practice network analysis. In the end, the reader will have the knowledge, skills, and tools to apply social network analysis.

We stress learning by doing: Readers acquire a feel for network concepts by applying network analysis. To this end, we make ample use of professional computer software for network analysis and visualization: Pajek. This software, operating under Windows 95 and later, and all example datasets are provided on a Web site (http://vlado.fmf.uni-lj.si/pub/networks/book/) dedicated to this book. All the commands that are needed to produce the graphical and numerical results presented in this book are extensively discussed and illustrated. Step by step, the reader can perform the analyses presented in the book.

Note, however, that the graphical display on a computer screen will never exactly match the printed figures in this book. After all, a book is not a computer screen. Furthermore, newer versions of the software will appear, with features that may differ from the descriptions presented in this book. We strongly advise using the version of Pajek software supplied on the book’s Web site (http://vlado.fmf.uni-lj.si/pub/networks/book/) while studying this book and then updating to a newer version of Pajek afterward, which can be downloaded from http://vlado.fmf.uni-lj.si/pub/networks/pajek/default.htm.
Preface to the First Edition

Overview

This book contains five parts. The first part (Part I) presents the basic concepts of social network analysis. The next three parts present the three major research topics in social network analysis: cohesion (Part II), brokerage (Part III), and ranking (Part IV). We claim that all major applications of social network analysis in the social sciences relate to one or more of these three topics. The final part (Part V) discusses an advanced technique (viz., blockmodeling), which integrates the three research topics.

The first part, titled Fundamentals, introduces the concept of a network, which is obviously the basic object of network analysis, and the concepts of a partition and a vector, which contain additional information on the network or store the results of analyses. In addition, this part helps the reader get started with Pajek software.

Part II on cohesion consists of three chapters, each of which presents measures of cohesion in a particular type of network: ordinary networks (Chapter 3), signed networks (Chapter 4), and valued networks (Chapter 5). Networks may contain different types of relations. The ordinary network just shows whether there is a tie between people, organizations, or countries. In contrast, signed networks are primarily used for storing relations that are either positive or negative such as affective relations: liking and disliking. Valued networks take into account the strength of ties, for example, the total value of the trade from one country to another or the number of directors shared by two companies.

Part III on brokerage focuses on social relations as channels of exchange. Certain positions within the network are heavily involved in the exchange and flow of information, goods, or services; whereas others are not. This is connected to the concepts of centrality and centralization (Chapter 6) or brokers and bridges (Chapter 7). Chapter 8 discusses an important application of these ideas, namely, the analysis of diffusion processes.

The direction of ties (e.g., who initiates the tie) is not very important in the section on brokerage, but it is central to ranking, presented in Part IV. Social ranking, it is assumed, is connected to asymmetric relations. In the case of positive relations, such as friendship nominations or advice seeking, people who receive many choices and reciprocate few choices are deemed as enjoying more prestige (Chapter 9). Patterns of asymmetric choices may reveal the stratification of a group or society into a hierarchy of layers (Chapter 10). Chapter 11 presents a particular type of asymmetry, namely, the asymmetry in social relations caused by time: genealogical descent and citation.
The final section, Part V, on roles concentrates on rather dense and small networks. This type of network can be visualized and stored efficiently by means of matrices. Blockmodeling is a suitable technique for analyzing cohesion, brokerage, and ranking in dense, small networks. It focuses on positions and social roles (Chapter 12).

The book is intended for researchers and managers who want to apply social network analysis and for courses on social network analysis in all social sciences as well as other disciplines using social methodology (e.g., history and business administration). Regardless of the context in which the book is used, Chapters 1, 2, and 3 must be studied to understand the topics of subsequent chapters and the logic of Pajek. Chapters 4 and 5 may be skipped if the researcher or student is not interested in networks with signed or valued relations, but we strongly advise including them to be familiar with these types of networks. In Parts III (Brokerage) and IV (Ranking), the first two chapters present basic concepts and the third chapter focuses on particular applications.

Figure 1 shows the dependencies among the chapters of this book. To study a particular chapter, all preceding chapters in this flowchart must have been studied before. Chapter 10, for instance, requires understanding of Chapters 1 through 4 and 9. Within the chapters, there are no sections that can be skipped.

In an undergraduate course, Parts I and II should be included. A choice can be made between Part III and Part IV; or, alternatively, just the first chapter from each section may be selected. Part V on social roles and blockmodeling is quite advanced and more appropriate for a postgraduate course. For managerial purposes, Part III is probably more interesting than Part IV.
This book offers an introduction to social network analysis, which implies that it covers a limited set of topics and techniques, which we feel a beginner must master to be able to find his or her way in the field of social network analysis. We have made many decisions about what to include and what to exclude, and we want to justify our choices now.

As reflected in the title of this book, we restrict ourselves to exploratory social network analysis. The testing of hypotheses by means of statistical models falls outside the scope of this book. In social network analysis, hypothesis testing is important but complicated; it deserves a book on its own. Aiming our book at people who are new to social network analysis, our first priority is to have them explore the structure of social networks to give them a feel for the concepts and applications of network analysis. Exploration involves visualization and manipulation of concrete networks, whereas hypothesis testing boils down to numbers representing abstract parameters and probabilities. In our view, exploration yields the intuitive understanding of networks and basic network concepts that are a prerequisite for well-considered hypothesis testing.

From the vast array of network analytic techniques and indices we discuss only a few. We have no intention of presenting a survey of all structural techniques and indices because we fear that readers will not be able to see the forest for the trees. We focus on as few techniques and indices as are needed to present and measure the underlying concept. With respect to the concept of cohesion, for instance, many structural indices have been proposed for identifying cohesive groups: $n$-cliques, $n$-clans, $n$-clubs, $m$-cores, $k$-cores, $k$-plexes, lambda sets, and so on. We discuss only components, $k$-cores, 3-cliques, and $m$-slices ($m$-cores) because they suffice to explain the basic parameters involved: density, connectivity, and strength of relations within cohesive subgroups.

Our choice is influenced by the software that we use because we have decided to restrict our discussion to indices and techniques that are incorporated in this software. Pajek software is designed to handle very large networks (up to millions of vertices). Therefore, this software package concentrates on efficient routines, which are capable of dealing with large networks. Some analytical techniques and structural indices are known to be inefficient (e.g., the detection of $n$-cliques), and for others no efficient algorithm has yet been found or implemented. This limits our options; we present only the detection of small cliques (of size 3), and we cannot extensively discuss an important concept such as $k$-connectivity. In summary, this book is neither a complete catalogue of network analytic concepts and techniques nor an exhaustive manual to all commands of Pajek. It offers just enough concepts, techniques, and skills to understand and perform all major types of social network analysis.
Preface to the First Edition

In contrast to some other handbooks on social network analysis, we minimize mathematical notation and present all definitions verbatim. There are no mathematical formulae in the book. We assume that many students and researchers are interested in the application of social network analysis rather than in its mathematical properties. As a consequence, and this may be very surprising to seasoned network analysts, we do not introduce the matrix as a data format and display format for social networks until the end of the book.

Finally, there is a remark on the terminology used in the book. Social network analysis derives its basic concepts from mathematical graph theory. Unfortunately, different “vocabularies” exist within graph theory, using different concepts to refer to the same phenomena. Traditionally, social network analysts have used the terminology employed by Frank Harary, for example, in his book *Graph Theory* (Reading: Addison-Wesley, 1969). We choose, however, to follow the terminology that prevails in current textbooks on graph theory, for example, R. J. Wilson’s *Introduction to Graph Theory* (Edinburgh: Oliver and Boyd, 1972; published later by Wiley, New York). Thus, we hope to narrow the terminological gap between social network analysis and graph theory. As a result, we speak of a vertex instead of a node or a point, and some of our definitions and concepts differ from those proposed by Frank Harary.

Acknowledgments

The text of this book has benefited from the comments and suggestions from our students at the University of Ljubljana and the Erasmus University Rotterdam, who were the first to use it. In addition, Michael Frishkopf and his students of musicology at the University of Alberta gave us helpful comments. Mark Granovetter, who welcomed this book to his series, and his colleague Sean Farley Everton have carefully read and commented on the chapters. In many ways, they have helped us make the book more coherent and understandable to the reader. We are also very grateful to an anonymous reviewer, who carefully scrutinized the book and made many valuable suggestions for improvements. Ed Parsons (Cambridge University Press) and Nancy Hulan (TechBooks) helped us through the production process. Finally, we thank the participants of the workshops we conducted at the Sunbelt International Conferences on Social Network Analysis in New Orleans (XXII) and Cancun (XXIII) for their encouraging reactions to our manuscript.

Most datasets that are used in this book have been created from sociograms or listings printed in scientific articles and books. Notwithstanding our conviction that reported scientific results should be used
Preface to the First Edition

and distributed freely, we have tried to trace the authors of these articles and books and ask for their approval. We are grateful to have obtained explicit permission for using and distributing the datasets from them. Authors or their representatives whom we have not reached are invited to contact us.