Contents

Preface xi

Acknowledgements xiii

List of abbreviations xv

Introduction xix

References xxii

1 Multivariate analysis techniques for representing graphs and contingency tables 1

1.1 Quadratic placement problems for weighted graphs and hypergraphs 1

1.1.1 Representation of edge-weighted graphs 2

1.1.2 Representation of hypergraphs 5

1.1.3 Examples for spectra and representation of simple graphs 8

1.2 SVD of contingency tables and correspondence matrices 12

1.3 Normalized Laplacian and modularity spectra 16

1.4 Representation of joint distributions 21

1.4.1 General setup 21

1.4.2 Integral operators between L^2 spaces 22

1.4.3 When the kernel is the joint distribution itself 23

1.4.4 Maximal correlation and optimal representations 25

1.5 Treating nonlinearities via reproducing kernel Hilbert spaces 28

1.5.1 Notion of the reproducing kernel 29

1.5.2 RKHS corresponding to a kernel 32

1.5.3 Two examples of an RKHS 33

1.5.4 Kernel – based on a sample – and the empirical feature map 37

References 40

2 Multiway cuts and spectra 44

2.1 Estimating multiway cuts via spectral relaxation 44

2.1.1 Maximum, minimum, and ratio cuts of edge-weighted graphs 45

2.1.2 Multiway cuts of hypergraphs 54

2.2 Normalized cuts 57

2.3 The isoperimetric number and sparse cuts 64

2.4 The Newman–Girvan modularity 76

2.4.1 Maximizing the balanced Newman–Girvan modularity 78

2.4.2 Maximizing the normalized Newman–Girvan modularity 81
2.4.3 Anti-community structure and some examples 84
2.5 Normalized bicuts of contingency tables 88
References 91

3 Large networks, perturbation of block structures 96
3.1 Symmetric block structures burdened with random noise 96
3.1.1 General blown-up structures 99
3.1.2 Blown-up multipartite structures 109
3.1.3 Weak links between disjoint components 112
3.1.4 Recognizing the structure 114
3.1.5 Random power law graphs and the extended planted partition model 121
3.2 Noisy contingency tables 124
3.2.1 Singular values of a noisy contingency table 127
3.2.2 Clustering the rows and columns via singular vector pairs 129
3.2.3 Perturbation results for correspondence matrices 132
3.2.4 Finding the blown-up skeleton 138
3.3 Regular cluster pairs 142
3.3.1 Normalized modularity and volume regularity of edge-weighted graphs 142
3.3.2 Correspondence matrices and volume regularity of contingency tables 150
3.3.3 Directed graphs 156
References 157

4 Testable graph and contingency table parameters 161
4.1 Convergent graph sequences 161
4.2 Testability of weighted graph parameters 164
4.3 Testability of minimum balanced multiway cuts 166
4.4 Balanced cuts and fuzzy clustering 172
4.5 Noisy graph sequences 175
4.6 Convergence of the spectra and spectral subspaces 177
4.7 Convergence of contingency tables 182
References 187

5 Statistical learning of networks 189
5.1 Parameter estimation in random graph models 189
5.1.1 EM algorithm for estimating the parameters of the block-model 189
5.1.2 Parameter estimation in the α and β models 192
5.2 Nonparametric methods for clustering networks 197
5.2.1 Spectral clustering of graphs and biclustering of contingency tables 199
5.2.2 Clustering of hypergraphs 201

5.3 Supervised learning 203
References 205

Appendix A Linear algebra and some functional analysis 207
A.1 Metric, normed vector, and Euclidean spaces 207
A.2 Hilbert spaces 209
A.3 Matrices 217
References 233

Appendix B Random vectors and matrices 235
B.1 Random vectors 235
B.2 Random matrices 239
References 245

Appendix C Multivariate statistical methods 246
C.1 Principal component analysis 246
C.2 Canonical correlation analysis 248
C.3 Correspondence analysis 250
C.4 Multivariate regression and analysis of variance 252
C.5 The *k*-means clustering 255
C.6 Multidimensional scaling 257
C.7 Discriminant analysis 258
References 261

Index 263