

Contents

1	Networks and Fundamental Concepts	1
1.1	Network Adjacency Matrix	1
1.1.1	Connectivity and Related Concepts	2
1.1.2	Social Network Analogy: Affection Network	2
1.2	Analysis Tasks Amenable to Network Methods	3
1.3	Fundamental Network Concepts	4
1.3.1	Matrix and Vector Notation	5
1.3.2	Scaled Connectivity	5
1.3.3	Scale-Free Topology Fitting Index	6
1.3.4	Network Heterogeneity	8
1.3.5	Maximum Adjacency Ratio	8
1.3.6	Network Density	9
1.3.7	Quantiles of the Adjacency Matrix	10
1.3.8	Network Centralization	10
1.3.9	Clustering Coefficient	11
1.3.10	Hub Node Significance	11
1.3.11	Network Significance Measure	12
1.3.12	Centroid Significance and Centroid Conformity	12
1.3.13	Topological Overlap Measure	13
1.3.14	Generalized Topological Overlap for Unweighted Networks	14
1.3.15	Multinode Topological Overlap Measure	16
1.4	Neighborhood Analysis in PPI Networks	18
1.4.1	GTOM Analysis of Fly Protein–Protein Interaction Data	18
1.4.2	MTOM Analysis of Yeast Protein–Protein Interaction Data	20
1.5	Adjacency Function Based on Topological Overlap	21
1.6	R Functions for the Topological Overlap Matrix	21
1.7	Network Modules	22
1.8	Intramodular Network Concepts	24
1.9	Networks Whose Nodes Are Modules	25
1.10	Intermodular Network Concepts	26

1.11	Network Concepts for Comparing Two Networks	27
1.12	R Code for Computing Network Concepts	29
1.13	Exercises	30
	References	32
2	Approximately Factorizable Networks	35
2.1	Exactly Factorizable Networks	35
2.2	Conformity for a Non-Factorizable Network	36
2.2.1	Algorithm for Computing the Node Conformity	37
2.3	Module-Based and Conformity-Based Approximation of a Network	39
2.4	Exercises	42
	References	43
3	Different Types of Network Concepts	45
3.1	Network Concept Functions	46
3.2	CF-Based Network Concepts	48
3.3	Approximate CF-Based Network Concepts	49
3.4	Fundamental Network Concepts Versus CF-Based Analogs	50
3.5	CF-Based Concepts Versus Approximate CF-Based Analog	51
3.6	Higher Order Approximations of Fundamental Concepts	52
3.7	Fundamental Concepts Versus Approx. CF-Based Analogs	53
3.8	Relationships Among Fundamental Network Concepts	54
3.8.1	Relationships for the Topological Overlap Matrix	55
3.9	Alternative Expression of the Factorizability $F(A)$	56
3.10	Approximately Factorizable PPI Modules	56
3.11	Studying Block Diagonal Adjacency Matrices	61
3.12	Approximate CF-Based Intermodular Network Concepts	63
3.13	CF-Based Network Concepts for Comparing Two Networks	64
3.14	Discussion	65
3.15	R Code	67
3.16	Exercises	69
	References	74
4	Adjacency Functions and Their Topological Effects	77
4.1	Definition of Important Adjacency Functions	77
4.2	Topological Effects of the Power Transformation AF^{power}	79
4.2.1	Studying the Power AF Using Approx. CF-Based Concepts	80
4.2.2	MAR Is a Nonincreasing Function of β	80
4.3	Topological Criteria for Choosing AF Parameters	82
4.4	Differential Network Concepts for Choosing AF Parameters	83
4.5	Power AF for Calibrating Weighted Networks	84
4.6	Definition of Threshold-Preserving Adjacency Functions	84

4.7	Equivalence of Network Construction Methods	86
4.8	Exercises	87
	References	89
5	Correlation and Gene Co-Expression Networks	91
5.1	Relating Two Numeric Vectors	91
5.1.1	Pearson Correlation	93
5.1.2	Robust Alternatives to the Pearson Correlation	94
5.1.3	Biweight Midcorrelation	95
5.1.4	C-Index	96
5.2	Weighted and Unweighted Correlation Networks	97
5.2.1	Social Network Analogy: Affection Network	98
5.3	General Correlation Networks	99
5.4	Gene Co-Expression Networks	101
5.5	Mouse Tissue Gene Expression Data from of an F2 Intercross	103
5.6	Overview of Weighted Gene Co-Expression Network Analysis	108
5.7	Brain Cancer Network Application	110
5.8	R Code for Studying the Effect of Thresholding	112
5.9	Gene Network (Re-)Construction Methods	114
5.10	R Code	115
5.11	Exercises	117
	References	118
6	Geometric Interpretation of Correlation Networks	
	Using the Singular Value Decomposition	123
6.1	Singular Value Decomposition of a Matrix <i>datX</i>	123
6.1.1	Signal Balancing Based on Right Singular Vectors	124
6.1.2	Eigenvectors, Eigengenes, and Left Singular Vectors	125
6.2	Characterizing Approx. Factorizable Correlation Networks	126
6.3	Eigenvector-Based Network Concepts	129
6.3.1	Relationships Among Density Concepts in Correlation Networks	131
6.4	Eigenvector-Based Approximations of Intermodular Concepts	132
6.5	Networks Whose Nodes are Correlation Modules	134
6.6	Dictionary for Fundamental-Based and Eigenvector- Based Concepts	135
6.7	Geometric Interpretation	136
6.7.1	Interpretation of Eigenvector-Based Concepts	136
6.7.2	Interpretation of a Correlation Network	137
6.7.3	Interpretation of the Factorizability	138
6.8	Network Implications of the Geometric Interpretation	139
6.8.1	Statistical Significance of Network Concepts	140
6.8.2	Intramodular Hubs Cannot be Intermediate Nodes	140
6.8.3	Characterizing Networks Where Hub Nodes Are Significant	140

6.9	Data Analysis Implications of the Geometric Interpretation.....	141
6.10	Brain Cancer Network Application	143
6.11	Module and Hub Significance in Men, Mice, and Yeast	147
6.12	Summary	150
6.13	R Code for Simulating Gene Expression Data	153
6.14	Exercises	157
	References.....	159
7	Constructing Networks from Matrices.....	161
7.1	Turning a Similarity Matrix into a Network	161
7.2	Turning a Symmetric Matrix into a Network	162
7.3	Turning a General Square Matrix into a Network	163
7.4	Turning a Dissimilarity or Distance into a Network	164
7.5	Networks Based on Distances Between Vectors	165
7.6	Correlation Networks as Distance-Based Networks	166
7.7	Sample Networks for Outlier Detection	167
7.8	KL Dissimilarity Between Positive Definite Matrices	169
7.9	KL Pre-Dissimilarity for Parameter Estimation	170
7.10	Adjacency Function Based on Distance Properties	171
7.11	Constructing Networks from Multiple Similarity Matrices.....	172
7.11.1	Consensus and Preservation Networks	173
7.12	Exercises	175
	References.....	178
8	Clustering Procedures and Module Detection.....	179
8.1	Cluster Object Scatters Versus Network Densities	179
8.2	Partitioning-Around-Medoids Clustering.....	181
8.3	<i>k</i> -Means Clustering	182
8.4	Hierarchical Clustering	184
8.5	Cophenetic Distance Based on a Hierarchical Cluster Tree	186
8.6	Defining Clusters from a Hierarchical Cluster Tree: The Dynamictreecut Library for R.....	188
8.7	Cluster Quality Statistics Based on Network Concepts	192
8.8	Cross-Tabulation-Based Cluster (Module) Preservation Statistics	193
8.9	Rand Index and Similarity Measures Between Two Clusterings ..	195
8.9.1	Co-Clustering Formulation of the Rand Index	196
8.9.2	R Code for Cross-Tabulation and Co-Clustering	197
8.10	Discussion of Clustering Methods	198
8.11	Exercises	200
	References.....	205
9	Evaluating Whether a Module is Preserved in Another Network.....	207
9.1	Introduction	207
9.2	Module Preservation Statistics	209

9.2.1	Summarizing Preservation Statistics and Threshold Values	212
9.2.2	Module Preservation Statistics for General Networks.....	213
9.2.3	Module Preservation Statistics for Correlation Networks	214
9.2.4	Assessing Significance of Observed Module Preservation Statistics by Permutation Tests	218
9.2.5	Composite Preservation Statistic $Z_{summary}$	218
9.2.6	Composite Preservation Statistic <i>medianRank</i>	220
9.3	Cholesterol Biosynthesis Module Between Mouse Tissues.....	221
9.4	Human Brain Module Preservation in Chimpanzees	224
9.5	KEGG Pathways Between Human and Chimpanzee Brains.....	231
9.6	Simulation Studies of Module Preservation	233
9.7	Relationships Among Module Preservation Statistics	239
9.8	Discussion of Module Preservation Statistics	242
9.9	R Code for Studying the Preservation of Modules	244
9.10	Exercises	245
	References.....	245
10	Association Measures and Statistical Significance Measures	249
10.1	Different Types of Random Variables.....	249
10.2	Permutation Tests for Calculating p Values	250
10.3	Computing p Values for Correlations	252
10.4	R Code for Calculating Correlation Test p Values	254
10.5	Multiple Comparison Correction Procedures for p Values	255
10.6	False Discovery Rates and q -values.....	258
10.7	R Code for Calculating q -values	260
10.8	Multiple Comparison Correction as p Value Transformation.....	262
10.9	Alternative Approaches for Dealing with Many p Values	265
10.10	R Code for Standard Screening	266
10.11	When Are Two Variable Screening Methods Equivalent?	267
10.12	Threshold-Equivalence of Linear Significance Measures	269
10.13	Network Screening.....	271
10.14	General Definition of an Association Network	272
10.15	Rank-Equivalence and Threshold-Equivalence.....	272
10.16	Threshold-Equivalence of Linear Association Networks	273
10.17	Statistical Criteria for Choosing the Threshold τ	274
10.18	Exercises	274
	References.....	277
11	Structural Equation Models and Directed Networks	279
11.1	Testing Causal Models Using Likelihood Ratio Tests	279
11.1.1	Depicting Causal Relationships in a Path Diagram.....	280
11.1.2	Path Diagram as Set of Structural Equations	282
11.1.3	Deriving Model-Based Predictions of Covariances	283

11.1.4	Maximum Likelihood Estimates of Model Parameters . . .	285
11.1.5	Model Fitting p Value and Likelihood Ratio Tests	287
11.1.6	Model Fitting Chi-Square Statistics and LRT	287
11.2	R Code for Evaluating an SEM Model	289
11.3	Using Causal Anchors for Edge Orienting	294
11.3.1	Single Anchor Local Edge Orienting Score	295
11.3.2	Multi-Anchor LEO Score	297
11.3.3	Thresholds for Local Edge Orienting Scores	299
11.4	Weighted Directed Networks Based on LEO Scores	299
11.5	Systems Genetic Applications	300
11.6	The Network Edge Orienting Method	301
11.6.1	Step 1: Combine Quantitative Traits and SNPs	301
11.6.2	Step 2: Genetic Marker Selection and Assignment to Traits	303
11.6.3	Step 3: Compute Local Edge Orienting Scores for Aggregating the Genetic Evidence in Favor of a Causal Orientation	305
11.6.4	Step 4: For Each Edge, Evaluate the Fit of the Underlying Local SEM Models	305
11.6.5	Step 5: Robustness Analysis with Respect to SNP Selection Parameters	305
11.6.6	Step 6: Repeat the Analysis for the Next A–B Trait–Trait Edge and Apply Edge Score Thresholds to Orient the Network	307
11.6.7	NEO Software and Output	307
11.6.8	Screening for Genes that Are Reactive to <i>Insig1</i>	308
11.6.9	Discussion of NEO	308
11.7	Correlation Tests of Causal Models	310
11.8	R Code for LEO Scores	311
11.8.1	R Code for the <i>LEO.SingleAnchor</i> Score	311
11.8.2	R Code for the <i>LEO.CPA</i>	313
11.8.3	R Code for the <i>LEO.OCA</i> Score	315
11.9	Exercises	317
	References	318

12 Integrated Weighted Correlation Network Analysis

	of Mouse Liver Gene Expression Data	321
12.1	Constructing a Sample Network for Outlier Detection	321
12.2	Co-Expression Modules in Female Mouse Livers	324
12.2.1	Choosing the Soft Threshold β Via Scale-Free Topology	324
12.2.2	Automatic Module Detection Via Dynamic Tree Cutting	326
12.2.3	Blockwise Module Detection for Large Networks	327

12.2.4	Manual, Stepwise Module Detection	328
12.2.5	Relating Modules to Physiological Traits	330
12.2.6	Output File for Gene Ontology Analysis	333
12.3	Systems Genetic Analysis with NEO	334
12.4	Visualizing the Network	337
12.4.1	Connectivity, TOM, and MDS Plots	337
12.4.2	VisANT Plot and Software	339
12.4.3	Cytoscape and Pajek Software	339
12.5	Module Preservation Between Female and Male Mice	340
12.6	Consensus modules Between Female and Male Liver Tissues	344
12.6.1	Relating Consensus Modules to the Traits	345
12.6.2	Manual Consensus Module Analysis	348
12.7	Exercises	350
	References	351
13	Networks Based on Regression Models and Prediction Methods	353
13.1	Least Squares Regression and MLE	353
13.2	R Commands for Simple Linear Regression	355
13.3	Likelihood Ratio Test for Linear Model Fit	356
13.4	Polynomial and Spline Regression Models	358
13.5	R Commands for Polynomial Regression and Spline Regression ..	360
13.6	Conditioning on Additional Covariates	363
13.7	Generalized Linear Models	364
13.8	Model Fitting Indices and Accuracy Measures	365
13.9	Networks Based on Predictors and Linear Models	365
13.10	Partial Correlations and Related Networks	366
13.11	R Code for Partial Correlations	368
13.12	Exercises	368
	References	372
14	Networks Between Categorical or Discretized Numeric Variables	373
14.1	Categorical Variables and Statistical Independence	373
14.2	Entropy	375
14.2.1	Estimating the Density of a Random Variable	376
14.2.2	Entropy of a Discretized Continuous Variable	378
14.3	Association Measures Between Categorical Vectors	379
14.3.1	Association Measures Expressed in Terms of Counts	381
14.3.2	R Code for Relating Categorical Variables	381
14.3.3	Chi-Square Statistic Versus Cor in Case of Binary Variables	382
14.3.4	Conditional Mutual Information	383
14.4	Relationships Between Networks of Categorical Vectors	384
14.5	Networks Based on Mutual Information	385

14.6	Relationship Between Mutual Information and Correlation	387
14.6.1	Applications for Relating MI with Cor.....	390
14.7	ARACNE Algorithm	391
14.7.1	Generalizing the ARACNE Algorithm.....	393
14.7.2	Discussion of Mutual Information Networks	394
14.7.3	R Packages for Computing Mutual Information	395
14.8	Exercises	396
	References.....	399
15	Network Based on the Joint Probability Distribution	
	of Random Variables	401
15.1	Association Measures Based on Probability Densities	401
15.1.1	Entropy(X) Versus Entropy(Discretize(X))	403
15.1.2	Kullback–Leibler Divergence for Assessing Model Fit	405
15.1.3	KL Divergence of Multivariate Normal Distributions	406
15.1.4	KL Divergence for Estimating Network Parameters	407
15.2	Partitioning Function for the Joint Probability	408
15.3	Discussion	409
	References.....	410
	Index	413