Index

A
adjacency
  correlation, 28
  intermodular, 25
adjacency function, 77
  based on topological overlap, 21
  correlation-mutual information, 388
  examples, 77
  Moebius, 78, 386
  power, 78
  rank-preserving, 85
  threshold, 78
  threshold-preserving, 84
adjacency matrix, 1
  based on categorical vectors, 384
  based on mutual information, 385
  block diagonal, 61
  eigenvector-based, 127
  intermodular, 63
agglomerative hierarchical clustering, 184
AIC criterion
  NEO, 304
algorithm
  for computing CF, 36
  MM, 39
anchor
  common pleiotropic, 298
  orthogonal causal, 298
angle between vectors, 91
apoptosis genes
  humans versus chimpanzees, 231
approximate conformity based adjacency, 39
approximately factorizable
  protein interaction network, 56
ARACNE approach, 161, 373
area hyperbolic tangent, 253
area under ROC curve, *see* C-index
arrow
  causal direction, 280
association measure, 249, 379
  in terms of counts, 381
  between numeric random variables, 402
association network
  categorical vectors, 373
  joint probability distribution, 401
assumption
  bounded moment, 52
  maximum conformity, 50
Aten
  Jason, xi, 279
average linkage hierarchical clustering, 184

B
Barabasi
  Laszlo, vii
Bayesian network, 408, 409
bell curve, 401
biclustering procedure, 199
binary variables
  cor and chi-square statistic, 382
biweight midcorrelation, 95
biweight midcorrelation, 243, 392
  R code, 116
block diagonal, 61
Bonferroni corrected
  p-value, 257
Bonferroni correction, 249
bounded moment assumption, 52
  correlation network, 130
branch cutting
  cluster tree, 188

C
C-index, 96
  between p-value and q-value, 256
  for network construction, 366

S. Horvath, *Weighted Network Analysis: Applications in Genomics and Systems Biology*,
DOI 10.1007/978-1-4419-8819-5, © Springer Science+Business Media, LLC 2011

C-index (*cont.*)
for network definition, 365
threshold-implication, 269
calibrating networks, 84
R code, 348
Canberra distance, 165
sample network, 169
canonical Euclidean distance based network, 166
categorical variable, 249
causal anchor, 295
centralization, 10
centroid, 12
CF-based approximation
module-based, 39
Chebyshev distance, 165
cholesterol biosynthesis module
preservation in mouse tissues, 221
cliquishness, 11
cluster
cohesiveness measure, 213
tree, 185
validation statistics, 208
cluster quality statistics, 192, 244
mouse network, 342
cluster validation statistics, *see* module preservation statistics
based on prediction error, 208
clustering coefficient, 11
mean, 54
preservation statistic, 214
clustering procedure, 179
biclustering, 199
co-clustering, 199
two-way, 199
co-clustering, 199
R code, 197
Rand index, 196
co-expression information, 102
coefficients
SEM model, 282
coexpression network, 91
common pleiotropic anchors, 298
complete linkage hierarchical clustering, 184
composite module preservation statistics, 212
for correlation networks, 218
concordance index, *see* C-index
conditional mutual information, 383
conformity, 13, 35
algorithm, 37, 41
in a general network, 36
intramodular centroid, 25
uniqueness proof, 42
connectivity, 2
correlation, 28
differential, 27
frequency, 6
scaled, 6
connectivity plot, 337
connectivity preservation statistics, 214
consensus
module, 173
module R code, 344
network, 173
network R code, 344
consensus module, 3
R code, 344
cophenetic distance, 164
based on a dendrogram, 186
Coppola
Giovanni, xi
correction
Bonferroni, 257
Sidak, 258
correlation
biweight, 95
Goodman-Kruskal gamma, 96
partial, 367
Pearson, 93
R code, 116
related to mutual information, 387
robust alternatives, 94
Spearman, 94
test and p-value, 252, 311
correlation network, 91
general definition, 99
signed, 99
social network interpretation, 98
correlation test
R code, 254
cosine correlation, 93
covariance, 94
model based prediction, 283
covariate, 249
CPA, 298
criterion for choosing parameters
hard threshold with p-value, 274
mean connectivity, 82
scale free topology, 83
SFT in brain cancer, 110
Cytoscape software, 339

D

data reduction methods, 110, 141
decomposition
module- and CF-based, 39

degree, 2
dendogram, 185
dendrogram, 185
density, 9
  relationship among concepts, 131
density preservation statistics, 213
dependence
  statistical, 410
dictionary
  for relating fundamental and eigenvector-based concepts, 135
differential network analysis, 27
  CF-based, 64
differential network concepts, 27
  for choosing parameters, 83
directed network
  based on LEO scores, 299
  constructed from a matrix, 300
discretize
  equal-width, 377, 381
  function, 6
discriminatory power, 96
dissimilarity measure, 164
dist R function, 166
distance
  Canberra, 165
  Chebyshev, 165
  cophenetic, 164
  Euclidean, 165
  Euclidean and correlation, 166
  general properties, 164
  infinity norm, 169
  Manhattan, 165
  p-norm, 165
  R code, 166
  taxicab, 165
  ultra, 185
distance based network, 166
disturbance, 281
  structural error, 282
Dong
  Jun, 35, 45, 123
dot product, 92
Drake
  Tom, xi
dynamicTreeCut, 179
  R package, 188

E
edge orienting score, 295
eigengene, 125
  network, 135
eigennode, 125
eigenvector, 125
  factorizability, 126
  network, 134
endogenous variable, 281
energy, 408
entropy
  joint between categorical vectors, 376
  continuous variable, 402
  defining properties, 375
  estimation methods, 378
  interpretation with question-asking strategy, 403
  normal distribution, 402, 404
  of numeric variable and its discretized version, 403
  related to mutual information, 385
  relative, 405
equivalence
  network construction methods, 86
error rate
  family-wise, 257
error variable, 281
estimation method
  for probability density, 377
Euclidean distance, 165
  relation to correlation, 166
exogenous variable, 281
exponentially truncated power law, 7, 113

F
factor analysis, 37
  oblique, 135
factorizability
  eigennode, 240
  eigenvector, 126
factorizability measure, 36
  alternative expressions, 56
  module based, 40, 41
factorizable
  protein interaction network, 56
factorizable network, 35
  characterization of correlation, 126
false discovery rate, 258
  R code, 260
false positive rate, 258
family-wise error rate, 257
FDR, *see* false discovery rate
feature
  types, 249
Fisher
  exact test, 26, 380
  exact test for overlap, 198
  exact test R code, 383

Fisher (*cont.*)
Z statistic of cor, 253
Z transformation, 253
fitting criterion
SEM model, 288
flashClust R function, 185
Fox
John, 279, 289, 307
frequency distribution, 377
categorical variable, 374
joint, categorical variables, 375
Frobenius norm, 5

G
Gaussian distribution, 401
multivariate, 406
gene expression data
mouse module preservation, 340
mouse tissues, 103
gene screening
in brain cancer, 112
genetic polymorphism, 300
geometric interpretation
of correlation network, 137
of eigenvector-based concepts, 136
of factorizability, 138
Geschwind
Dan, xi, 98
glioblastoma multiforme, 110
GO enrichment analysis, 333
GTOM
neighborhood analysis, 18

H
Hamiltonian, 408
hard-thresholding, 78
hclust R function, 185
heatmap, 155, 342
heterogeneity, 8
hierarchical clustering, 184
in brain cancer, 110
R functions, 185
Hotelling statistic, 253
hub gene significance, 11
hub node significance, 11
characterizing networks, 140
hyperbolic function, 253
hypergeometric distribution, 26

I
IC algorithm
causal inference, 309
identity matrix, 5
IGP statistic, 208, 243
cholesterol biosynthesis, 224
correlation with Zsummary, 228
dependence on module size, 228
in-group proportion, *see* IGP statistic
independence
categorical vectors, 374
inequality
between vectors, 91
Cauchy-Schwarz, 92
Hoelder, 92
triangle, 164
infinity norm distance, 169
inner product, 92
inter-cluster dissimilarity, 184
inter-group distance, 184
intermediate nodes, 4, 140
intermodular, 26
eigenvector-based approximation, 132
intramodular connectivity
preservation statistic, 214

J
joint entropy, 376
numeric variables, 402
just-identified, 287

K
k-means clustering, 183
k-medoid clustering, 181
Kullback-Leibler, 405
dissimilarity as objective function, 170
dissimilarity between positive definite matrices, 169
for model fit, 405
multivariate normal distributions, 406
pre-dissimilarity, 169
SEM model fitting criterion, 288

L
Lange
Kenneth, xi, 41
Langfelder
Peter, 31, 95, 116, 188, 207, 321
Laplace principle, 406
latent variable, 281
LEO score, 295
multi-anchor, 297
R code, 311

single anchor, 295
thresholds, 299
LEO.OCA score, 298
Li
Ai, 16
likelihood ratio test
categorical vectors, 379
causal models, 279
for causal fit, 280
linear model
fitting index, 365
linear transformation, 93
linkage method
in hierarchical clustering, 184
list notation, 172
local edge orienting score, 295, *see* LEO score
local structural equation models, 295
Lusis
Jake, xi, 279, 321

M
Manhattan distance, 165
R code, 169
MAR
maximum adjacency ratio, preservation statistic, 214
Markov network, 408
matrix
multiplication, 5
positive definite, 169, 406
maximum
of a matrix, 5
maximum adjacency ratio, 8
non-increasing function of power, 80
preservation statistic, 214
maximum conformity assumption, 50
block diagonal example, 63
maximum likelihood estimate
SEM model parameters, 285
MCC transformation, 263
MDI
principle, 406
MDS, *see* multi-dimensional scaling
medianRank
human chimp module preservation, 228
module preservation statistic, 220
module preservation statistic, 212
sex differences in mouse livers, 343
medoid, 181
partitioning around, 181
meta analysis, 267
metric, *see* distance
microarray
brain cancer network, 110
microarray data
description, 101
Miller
Jeremy, xi
Miller-Madow estimator, 378
minimum
of a matrix, 5
Mischel
Paul, xi, 98, 110
MM algorithm, 39
model fitting criterion
SEM model, 288
model fitting index, 365
model fitting p-value, 287, 288
model p-value, 288
module, 22
general definition, 208
based on hierarchical clustering, 23
module membership
module preservation statistic, 216
module membership measure
fuzzy, 4
intramodular connectivity, 141
module preservation
strong versus weak, 210
module preservation statistics, 29
chapter, 207
comparisons in simulations, 243
composite, 212
cross-tabulation based, 193
density based, 215
dependence on module size, 212
discussion and limitations, 243
for correlation networks, 214
for general networks, 213
human versus chimpanzee brains, 224
KEGG pathways human versus chimpanzees, 231
R code, 214, 244
relationships among them, 239
significance test and p-value, 218
simulations, 233
thresholds, 219
module quality statistics, 243
module reproducibility, *see* module preservation statistics
module separability statistics, 217
module significance
in mouse, yeast, and humans, 147
module- and conformity-based approximation of A, 39
Moebius transformation, 78, 386

mQTL, 335
MTOM
  neighborhood analysis, 18
multi-dimensional scaling
  plot, 337
multiple comparison correction, 249, 255
  abstract discussion, 262
  transformation, 263
multiplication
  of matrices, 5
multivariate normal distribution, 406
mutual information
  for bivariate normal distribution, 403
  categorical vectors, 380
  conditioning variables, 383
  R packages, 395
  related to entropy, 385
  relation to LRT, 380
  relation with correlation, 387
mutual information network
  discretized numeric variables, 385
  discussion, 394
  R package, 395

N
nats
  unit of entropy, 380
neighborhood analysis, 18
Nelson
  Stan, xi, 98, 110
NEO method, 301
network
  among categorical vectors, 373
  among correlation modules, 134
  based on a dissimilarity or distance, 164
  based on a similarity, 161
  based on a symmetric matrix, 162
  based on C-index, 366
  based on distances between vectors, 165
  based on general square matrix, 163
  based on partial correlations, 366
  based on predictive accuracy, 365
  based on several similarity matrices, 172
  based on the C-index, 365
  based on the Euclidean distance, 166
  between discretized numeric variables, 385
  between modules, 25
  between random variables, 407
  calibration, 84
  calibration R code, 348
  constructed from a matrix, 161
  constructed using Kullback-Leibler, 170
  eigenvector-based, 134
  exactly factorizable, 35
  module, *see* module
  protein-protein interaction, 56
  screening, 271
  unweighted correlation, 97
  weighted correlation, 97
network concept, 4
  for measuring similarity, 28
  approximate CF-based, 49
  conformity based, 45, 48
  differential, 27
  differential CF-based, 64
  eigenvector-based, 123, 129
  for comparing networks, 27
  for module preservation, 209
  function, 46
  fundamental, 5, 47
  higher order approximation, 52
  intermodular, 26
  intermodular CF-based, 63, 242
  intermodular eigenvector based, 132
  intramodular, 24
  motivational example, 104
  overview, 45
  R code, 29, 67
  relationships in approx. factorizable networks, 54
  scaled, 6
  second order eigenvector based, 159
  special relationships, 54
  types, 46
  used for module preservation, 240
network concept function
  higher order approximation, 52
  Taylor expansion, 53
network edge orienting
  method, 301
network significance, 12
node conformity, 35
node significance measure, 2
  eigenvector-based, 127
  trait-based, 97
norm
  Frobenius, 5
normal distribution, 401
  multivariate, 406
null hypothesis, 250
  truly null, 255

O
object scatter, 180
object significance measure, 268
oblique factor analysis, 135

Oldham
Mike, xi, 167, 199, 225
Ophoff
Roel, xi
orthogonal causal anchor, 298
overview
WGCNA, 108

P
p-norm, 92
distance, 165
p-value
for model fit, 287
histogram, 256
SEM model fitting, 288
Pajek software, 340
pam, 181
parallel
maximum and minimum, 173
mean, 172
quantile transformation, 173
parsimonious parametrization
of network, 39
partial correlation, 367
R code, 368
partition function, 408
partitioning-around-medoids, *see* pam
paste function, 374
path analysis, 279
path diagram, 280
PC algorithm
causal inference, 309
Pearl
Judea, 310
Pearson
chi-square statistic, 379
chi-square vs. cor, 382
correlation, 93
polynomial regression
fitting index, 365
positive definite matrix, 169, 406
potential, 408
power adjacency function, 78
effects on approx. CF-based concepts, 80
topological effects, 79
preservation networks, 174
preservation of connectivity patterns, 229
preservation of intramodular connectivity
human chimp comparison, 229
Presson
Angela, xi
principle
insufficient reason, 406
Laplace, 406
maximum entropy, 406
of minimum discrimination, 406
probability density
estimation, 376
Kullback-Leibler divergence, 405
probability theory, 401
proportion of variance explained, 125
module preservation statistic, 216
protein-protein interaction, 56

Q
q-value, 249, 258
R code, 260
quantile, 10
quantile transformation, 173

R
R-squared
model fitting index, 365
RAM formulation of causal model, 289
Rand index, 195
rank-equivalence
mutual information networks, 386
object significance measures, 269
of adjacency matrices, 85
rankPvalue, 267
Ranola
John, xi, 41
rcorr.cens R function, 97
reticular action model, 289
ROC curve, 96

S
sample mean
random vector, 407
sample network
correlation based, 168
Euclidean distance based, 167, 321
sample networks, 167
for enhancing cluster analysis, 199
R code, 321
sample trait, 97
scale free topology, 7
fitting index, 7
scale free topology criterion, 83, 324
in brain cancer, 110
scale function, 92
scatter
of objects or nodes, 180

Schadt
  Eric, xi, 321
screening
  network based, 271
  standard marginal, 266
screening method, 267
sector plot, 27
SEM, *see* structural equation model,sem
sem
  R code, 289
  R package, 279
separability score
  of clusters, 217
separability statistic
  eigenvector-based, 133
separability statistics, 26
  CF-based, 64
  of modules, 217
Shannon
  entropy, 375
signal balancing, 124
signed correlation network, 99
significance
  chapter, 249
  object, 268
  of network concepts, 140
  p-value, 250
  set of variables, 260, 267
similarity measure, 161
simulation
  R code for expression data, 153
single linkage hierarchical clustering, 184
single nucleotide polymorphism, 300
singular value decomposition, 123
singular values, 124
singular vector, 124
SNP
  causal anchor, 300
soft thresholding, 78
Song
  Lin, 390
Spearman correlation, 94
  p-value, 254
standard screening, 266
static tree cutting, 188
statistical criteria: for choosing a hard threshold, 274
statistical dependence
  universal definition, 410
statistical significance, *see* significance
statistics
  cluster quality, 192, 244
straight line
  fitting index, 365
structural equation model, 279, 408
Student t statistic
  for comparing 2 means, 253
  for correlations, 252
systems genetics, 300

T
Taylor expansion, 53
threshold-equivalence
  of linear significance measures, 269
  unweighted linear networks, 273
threshold-implication
  of matrices, 384
  correlation and mutual information, 391
thresholding
  of correlation, 98
TOM plot, 337
topological criteria
  differential network concepts for choosing parameters, 83
  for choosing parameters/thresholds, 82, 324
topological effects
  of adjacency functions, 79
  of thresholding, 112
  R code, 112
topological overlap matrix, 13
  generalized, 14
  multinode, 16
  neighborhood analysis, 18
  plot, 337
  relationship to other concepts, 55
trace of matrix product
  convenient calculation, 287
transpose, 5
triangle-inequality, 164
two-way clustering, 199

U
ultra
  distance, 164, 171, 185
  metric, *see* ultra distance, 171
universal distance
  based on mutual information, 386

V
variable
  types, 249
  categorical, 249
  discrete numeric, 250

variable selection
  marginal, 266
  network based, 271
variance-covariance matrix, 407
vectorize a matrix, 9
VisANT software, 339
visualization
  of network, 337

W
Wald test, 291
weighted network, 1
Wright
  Sewall, 279

Y
Yip
  Andy, xi, 14

Z
Zhang
  Bin, xi, 98, 188
Zsummary
  human chimp module preservation, 228
  cholesterole biosynthesis module preservation, 223
  module preservation statistic, 212
  sex differences in mouse livers, 343
ZsummaryADJ
  module preservation statistic, 220