

Index

Locators in **bold** refer to tables; those in *italic* to figures

- adjacency
 - lists, 130, 130
 - matrices, 129, 129–130, 291, 294, 295, 491–492, 499
 - matrix visualizations, 541, 544
- Affymetrix SNP microarrays, 10, 12–14, **15**
- aging, 220–223
 - PPI network analysis, 194, 199, 212, 220, 221–223
- algorithms; *see also* network alignment
 - alignment, 23, 24
 - clustering, 242, 256–270, 549, 585
 - force-directed, 549–550, 574, 575
 - FUSE, 394–397, 395, 397
 - gene prioritization, 437
 - genotypes, 10, 16, 14–18, 21–27
 - graph search, 130–131
 - hierarchical clustering, 257, 264
 - Hungarian, 128, 375
 - Isorank, 200, 217, 382–384, 383
 - layouts, 549
 - machine learning, 33–39, 601
 - mapping, 372
 - optimization, 374
 - orbit counting, **227**, 228
 - pattern mining, 329, 330, 340, 361
 - permutation testing, 347
- AlignMCL, pairwise network alignment, 382
- alignment
 - algorithms, 23, 24
 - graphs, 374
 - network *see* network alignment
 - strategy, 216–217
- alignment scoring schemes, 375
 - agreements and trade-offs, 378, 378–379
 - F-score, 376
 - global network alignments, 376–378
 - local network alignments, 375–376
 - multiple network alignments, 392
 - symmetric sub-structure score, 377
- AlignNemo, pairwise network alignment, 381–382
- Alzheimer's disease, 517–518
- animations, node-link diagrams, 551
- annotations
 - databases, **166**, 165–167, 170
 - PPI networks, 171, 370, 371
- apriori property, pattern mining, 331
- articulation points, 168, 177
- assortative networks, 135
- asthma, 441, 445, 447, 474
- asymmetric interactions, 113, 118
- automorphism orbits, 203, 384, 384, 385
 - edge, 204
 - node, 205
- average closeness centrality, 511
- average clustering coefficient, 135, 511
- average edge betweenness centrality, 512
- average efficiency, 511
- average node betweenness centrality, 511–512
- average node degree, 510
- average radiality, 512
- Bayesian networks, 39, 41–42, 292, 291–293
- betweenness centrality, 136, 168, 178
- biases
 - microarrays, 75, 97
 - PPI datasets, 158–159, 438
- BioFabric, 542, 545
- bioinformatics
 - Hi-C analysis, 82
 - lncRNAs analysis, 88–89
 - protein-DNA interactions, 600
- biological heterogeneity, 159–160, 288–290
- biological interpretation, disease modules, 445–448
- biological networks (BN), 111, 167, 193–194, 256, 271–272; *see also* molecular networks; visualizing networks
 - data integration, 229
 - metabolic networks, 113, 418
- biomarkers
 - cancer precision medicine, 287
 - cancer stem cells, **605**, **606**, 603–612
 - covariate factors, 349–359
 - discovery, 313–314, 359–361
 - exercises, 362–364
 - ovarian cancer, 608, 609, 608–610
 - pattern mining, 315–328

622 INDEX

- biomarkers (cont.)
 - prostate cancer, 606–608
 - statistical redundancy, 341–349
 - Tarone's method for discovery, 329–341
- bipartite graphs, 113, 123, 123, 375, 374–375
- bisulfite based arrays, DNA methylation, 72–73
- bisulfite conversion, 72–73
- BLUEPRINT epigenome, 79, 96
- Bonferroni correction, 326–327
- Boolean variables, 254–255
- brain
 - anatomy, 515, 516
 - connectomes, 490–491, 499, 500, 514
 - geometry, 492
 - MRI scanning, 492–493
 - topology, 492
- brain networks
 - disorders, 517–519
 - functional, 499–503
 - structural, 492–499
 - tools for analysis, 505–506
- BRCA1* gene, 476, 480
- breadth first search (BFS), 130, 131
- breast cancer, 79, 96, 473–474, 560, 572
- cancer; *see also* tumors
 - BRCA1* gene, 476, 480
 - breast, 79, 96, 473–474, 560, 572
 - gene mutations, 175–179, 177, 300, 462, 548–549
 - genome atlas, 22, 95, 462, 479–480, 577
 - precision medicine, 287
 - prostate, 606–608
 - proteins, 223–224
 - stem cells, 595–598, 603–612, 605, 606
- CART (classification and regression tree classifiers), 37
- categorical values, proximity calculation, 256
- causal variants, 466–470, 473
- cell signaling networks, 112, 113
- centrality
 - average closeness, 511
 - average edge betweenness, 512
 - average node betweenness, 511–512
 - betweenness, 136, 168, 178
 - closeness, 135, 136, 139
 - eccentricity, 135, 137
 - eigenvector, 135, 137
 - GDV, 207, 221, 225
 - subgraphs, 137
- characteristic path length, 511
- chemokines, 597, 609
- ChIA-PET technology, chromatin conformation, 81, 83
- ChIP *see* chromatin immunoprecipitation
- chromatin, 77
 - conformation, 67, 81–83
 - higher order organization, 80–87
 - modifications, 70, 315
 - topological associated domains, 86, 86–87
- chromatin immunoprecipitation (ChIP), 69, 77–78, 418–419
 - data analysis, 79, 78–79
 - differential binding, 80
- chromosome territories, 70
- cis-acting, 67
- classification and regression tree classifiers (CARTs), 37
- cliques (complete subgraphs), 120, 170, 210, 265, 269
- clonal theory, 598
- closeness centrality, 135, 136, 139
- cluster analysis, 241–243, 277
 - definitions, 243–245
 - exercises, 277–280
 - preprocessing, 246–251
 - proximity calculation, 252–256
 - workflow, 245, 245–246
- cluster evaluation
 - external, 271–272
 - internal, 272–274
 - optimization strategies, 275–277
 - validity indices, 270, 271
- clustering
 - algorithms, 242, 256–270, 549, 585
 - coefficients, 170
 - data formats, 244–245
 - networks, 209, 210, 209–210
 - partitional, 243
 - types, 243
- clusters, number of, 275, 276
- Cochran-Mantel-Haenszel (CMH) test, 351–354
 - minimum attainable P-value, 354–355, 356, 358
 - pruning condition, 355–359
- co-expression networks, 419–420
- colored graphlets, 195, 200–202
- columns, cortex, 491
- comparative genomics, 156
- complex diseases, network approaches, 473–474
- computational biology workflows, 175–179
- computational complexity, 117–118, 128, 197, 200, 226, 297, 338
- computational methods, PPI, 222–223, 601

- conditional probabilities, 41–42, 291–293, 320
- confounding effect, pattern mining, 350, 351
- connectedness
 - graph theory, 122, 122
 - subgraphs, 119
- connectomes, brain, 490–491, 499, 500, 514
- contagious diseases spread, 423–424, 426
 - transportation networks, 424–425
- context-sensitive interactomes, 479–480
- contingency table analysis, 31–32
- continues variables, proximity calculation, 252–254
- correlation networks, 539
- correlation, continues variables, 253, 254
- covariate factors, pattern mining, 349–359
- CpG islands, 67, 70
- curated databases, **161**, 162, 161–162, 179
- cystic fibrosis (CF), 473, 480
- Cytoscape, 177, 179, 533–536, 553
 - apps, 574–577
 - command language, 587, 586–589
 - control panel, 555, 556
 - example workflow, 577–585
 - exercises, 589
 - hierarchical clustering, 541, 579, 580
 - importing data files, 561, 563
 - importing from public databases, 560
 - importing networks, 560
 - integration of data, 559–562
 - k*-means clustering, 580, 580
 - network analysis, 574
 - network panel, 557, 558
 - results panel, 557
 - scripting, 586
 - STRING network, 540, 560, 572, 576, 579, 581
 - table panel, 557
 - user interface, 554, 555–559
 - view menu, 559
 - visualizing data, 565, 562–573, 574, 575, 586
 - visualizing networks, 534, 540–552
- DAG1* gene, 476
- data integration, 159–160, 288–290
 - Bayesian approaches, 292, 291–293
 - biological networks, 229
 - early, 289
 - heterogeneous, 289, 300–306
 - homogeneous, 289, 294–300
 - intermediate, 289
 - kernel-based methods, 293, 293–294
 - late, 289
 - network-based, 291, 290–291
 - precision medicine, 287–288, 290–294
 - protein-protein interactions, 159–160
 - training methods, 290, 289–290
- databases
 - annotation, **166**, 165–167, 170
 - curated, **161**, 162, 161–162, 179
 - epigenetic, **75**, 93–96, **97**
 - integrated, 160, **164**, **165**, 163–179
 - interaction, 160–167, 372, **463**, 602
 - interactome, 439
 - lncRNA, **91**
 - molecular interaction, **463**
 - prediction, 163, **164**, 162–165, 179
 - protein-protein interactions, **112**, 160–167, 602, **603**
 - public, 2, 22, 72, 74, 427, 559–562
- Davies-Bouldin index, 273
- DBSCAN, 267, 266–268
- degree distribution, 133, 133, 178, 428
- degree of vertices, 119, 119, 168
- de-noising networks, 212, 213
- density, 133, 134
- density based clustering algorithms, 257, 266–268
- depth first search (DFS), 130–131
- deterministic measures, network topology, 507, 510–513
- differential binding, 79, **80**
- differential methylation CpGs (DMC), **76**, 76–77
- differential methylation regions (DMR), **76**, 76–77
- differential network analysis, 481
- diffusion tensor imaging (DTI), 495, 498, 500
- diffusion-based methods, disease identification, 173
- diffusion-weighted MRI scanning, 495, 496
- direct to consumer services, 3, 9, **13**, 45
 - predictive genetic risk models, 39–44
- directed acyclic graphs (DAG), 39
- directed graphlets, 197–198
- directed graphs, 113, 118, 118
- directed networks, 402, 402
- disassortative networks, 135
- discriminative pattern mining *see* pattern mining
- disease gene prioritization, 440, 442
 - connectivity-based methods, 440–443
 - diffusion-based methods, 443–444
 - path-based methods, 443
- diseases, 437, 474
 - analysis, 438
 - biological interpretation, 445–448
 - contagious diseases spread, 423–426

624 INDEX

- diseases (cont.)
 enrichment, 445, 444–445
 epigenetic mechanisms, 65–66
 gene prioritization, 440–444
 identification, PPI, 152, 173–174
 interactome analysis, 430
 interactome construction, 427, 438
 molecular basis, 474, 470–475
 network approaches, 422–423, 437, 438, 440, 442, 473–474
 resources, 439
 seed clusters, 438–440
 treatment, PPI, 166
 validation, 444
- DNA methylation, 66–68, 479; *see also* epigenetics
 bisulfite based arrays, 72–73
 experimental strategies, 69–71
 microarrays, 73–77
 role in genomic profiles, 69
- DNA modifications, 68, 70
- DNA sequencing, next generation, 18–27
- domains, proteins, 156
- dominant models, 32
- dominating set (DS), 211, 210–211
- drug repurposing, 174, 290, 294, 423
 precision medicine, 287
- drug targeting, 174
- DTC services *see* direct-to-consumer services
- Dunn index, 273
- dynamic graphlets, 199, 198–200
- dynamic networks, 199, 198–200, 221, 225
- eccentricity centrality, 135, 137
- edge conservation, 217, 375
- edge correctness, alignment scoring, 377
- edge-colored graphlets, 200, 201
- edges, 463, 464
- edgetic perturbations, 466, 470
- edgotype prediction tools, 468
- edgotype scenarios, 467, 467
- effect sizes, SNPs, 5
- eigenvector centrality, 135, 137
- electroencephalography (EEG), 501–503, 504
- Encyclopedia of DNA elements (ENCODE), 93
- enrichment, disease modules, 445, 444–445
- epigenetic databases, 75, 93–96, 97
 BLUEPRINT Epigenome, 79, 96
 Encyclopedia of DNA elements, 93
 Functional Annotation of the Mammalian Genome, 95
 International Human Epigenome Consortium, 96
 Roadmap Epigenomics Project, 95
- epigenetics, 65–66, 66, 98; *see also* DNA
 methylation
 changes in tumors, 595
 disease mechanisms, 65–66
 exercises, 98–100
 higher order chromatin organization, 80–87
 histone modifications, 77–79
 long non-coding RNAs, 87–93
 mapping mechanisms, 72, 74
- epigenomics, 67
- Erdos-Rényi (ER) random graphs, 138–139
- error rates, PPI datasets, 156–158
- euchromatin, 67
- Euclidean distances, 252
- Eulerian circuits, 126–127
- events, dynamic networks, 198
- exact cluster ratio, 392
- expectation-maximization (EM) algorithm, 16
- false discovery rate (FDR) control, 361
- family-wise error rate (FWER), 326, 361
 Bonferroni correction, 326–327
 empirical approximations, 344–346
 Tarone's improved Bonferroni correction, 327–328
- FANTOM (Functional Annotation of the Mammalian Genome), 95, 477
- feature selection, cluster analysis, 247–248
- Fischer's exact test, 323–325, 336, 337
- 5C technology, chromatin conformation, 81, 83
- force-directed algorithms, 549–550, 574, 575
- 4C technology, chromatin conformation, 81, 83
- frequent pattern mining *see* apriori property, pattern mining
- Fruchterman-Reingold algorithm, 549
- F-score, alignment scoring, 376
- Functional Annotation of the Mammalian Genome (FANTOM), 95, 477
- functional annotations, PPI networks, 174–175
- functional brain networks, 499–503
- functional consistency, alignment scoring, 376
- functional MRI (fMRI), 499, 502, 518
- FUSE, multiple network alignment method, 395, 394–397
- fuzzy clustering, 243–244
- fuzzy C-means (FCM) clustering, 262
- FWER *see* family-wise error rate
- gain-of-interactions, 466
- gap statistic, 276–277
- GATK (Genome Analysis Toolkit), 22, 27
- GCD (graphlet correlation distance), 215

- GCM (graphlet correlation matrix), 207, 208–209
- GDD (graphlet degree distributions), 207, 208
- GDDA (graphlet degree distribution agreement), 214
- GDV *see* graphlet degree vectors
- GDV-centrality, 207, 221, 225
- GDV-matrices, 207, 208–209
- GDV-similarity, 205–207, 210, 218, 219
- gene co-expression networks, 113
- gene duplication and divergence (SF-GD), 139
- gene expression analysis, 241
- gene mutations, 152, 459, 461
 - BRCA1* gene, 476, 480
 - cancer, 177, 175–179, 300, 462, 548–549
 - cystic fibrosis, 480
 - DAG1* gene, 476
 - edgetic perturbations, 466
 - KRAS* gene, 610
 - loss-of-function, 465, 466
 - monogenic diseases, 470–473
 - oncogenes, 595
 - Parkinson's disease, 473, 480
 - personalized genetic tests, 3, 7, 13, 45
 - RAS* genes, 462
 - sickle cell disease, 461–462, 479–480
 - tumor suppressor genes, 595
- gene ontology (GO) annotation set, 370, 371
- gene prioritization
 - algorithms, 437
 - disease analysis, 440–444
- gene regulatory networks, 418–419
- gene signature improvements, PPI networks, 174
- generalized random graph models (ER-DD), 138
- genetic data, risk prediction, 1–6, 45–47
 - exercises, 47–50
 - glossary, 4–6
 - SNP-disease association, 31–44
 - SNPs identification, 9, 30
 - tests in healthcare, 6–9
- genetic interactions, 113–114, 420–422, 463, 481, 536
- genetic tests, healthcare, 6–9
- Genome Analysis Toolkit (GATK), 22, 27
- genome atlas, cancer, 22, 95, 462, 479–480, 577
- genome-wide association studies (GWAS), 2, 473, 479
- genotype-phenotype relationships, 460–461, 482
 - definitions, 460–461
 - exercises, 482–483
 - molecular networks, 459, 461–464
 - network approaches, diseases, 464–475
 - network-based tools, 471, 472
- tissue interactomes, 480–482
- tissue-sensitive molecular interaction networks, 475–480
- genotypes
 - algorithms, 14–18
 - calling algorithms, 10, 16, 21–27
 - definition, 460, 461
- geometric graph with gene duplications and mutations (GEO-GD), 136, 139
- geometric networks, 139–140
- Gini importance, 40
- global pairwise network alignment methods
 - GRAAL, 384–387
 - IsoRank, 382–384
 - other, 387–390
- GRAAL, global network aligner, 384–387
- GRAFENE, 215–216, 220
- graph(s), 167
 - alignment, 374
 - bipartite, 113, 123, 375, 374–375
 - density, 169, 169
 - kernels, 216
 - regularization, 300, 299–300
 - types, 122–126
 - weighted, 113, 123, 123
- graph based clustering algorithms, 257, 268–270
- graph search algorithms, 130–131
- graph theory, 111–114, 140
 - classic problems, 126–128
 - computational complexity, 117–118
 - data structures, 128–130
 - definitions, 118–119
 - degree and neighborhood, 119–120
 - exercises, 140–142
 - mathematical basis, 114–116
 - network measures, 132–140
 - search algorithms, 130–131
 - spectral graph theory, 131–132
 - subgraphs and connectedness, 120–122
 - trees, 124, 124
- GraphCrunch, 226, 227, 227
- graphlet correlation distance (GCD), 215
- graphlet correlation matrix (GCM), 207, 208–209
- graphlet counting, 196, 197, 227, 226–229
 - orbit-aware, 206, 226
 - orbit-unaware, 199, 226
- graphlet degree distribution agreement (GDDA), 214
- graphlet degree distributions (GDD), 207, 208
- graphlet degree vectors, 203–205, 384, 385, 384–387; *see also* GDV-centrality; GDV-matrices; GDV-similarity

626 INDEX

- graphlet degree vectors (cont.)
 - edge, 204
 - node-pair, 205
 - non-edge, 205
- graphlet frequency vector (GFV), 208
- graphlet kernel, 293
- graphlet-based alignment-free network
 - approach (GRAFENE), 215–216, 220
- graphlets, 195, 193–196, 205
 - biological applications, 218–226
 - colored, 195, 200–202
 - computational approaches, 209–218
 - directed, 197–198
 - dynamic, 199, 198–200
 - edge-colored, 200, 201
 - exercises, 230–234
 - heterogeneous, 195, 200–202
 - homogeneous, 195, 196, 201–202
 - network topology, 196–209
 - node-colored, 200–201
 - orbits, 203, 203, 204, 205
 - ordered, 202, 202
 - software tools, 230, 226–230
 - static, 195, 196, 202
 - undirected, 196, 202
 - unordered, 196, 202
- graph-structured samples, 318
- GWAS *see* genome-wide association studies
- Hall's theorem, 128
- Hamiltonian paths, 127–128
- hedghog signaling pathway, 182, 597
- heterogeneity
 - biological, 288–290
 - cancers, 287
 - condition specific, 160
 - data integration, 289, 300–306
 - experimental, 159
 - graphlets, 195, 200–202
 - molecular, 159
 - nomenclature, 160
- Hi-C analysis, bioinformatics, 82
- Hi-C technology, chromatin conformation, 67, 77, 82
 - mapping and filtering, 82
 - normalization, 82, 84
 - statistical analysis, 84
 - tools, 85
 - topological associated domains, 86, 86–87
 - visualization, 84, 86
- hierarchical clustering, 244, 262, 263, 265–266; *see also* linkage functions
 - algorithms, 257, 264
 - Cytoscape, 541, 579, 580
- high angular resolution diffusion imaging (HARDI), 497, 498
- high-throughput methods (HT), 154
- histone modifications, 77–79
- homogeneity
 - data integration, 289, 294–300
 - graphlets, 195, 196, 201–202
- Hotelling's T2 statistic, 32–33
- human tissues, mapping, 463, 477
- Hungarian algorithm, 128, 375
- Huntington's disease, 470–473
- hypergraphs, 125, 124–125, 543, 546
- hyper-networks, 403, 403
- IHEC (International Human Epigenome Consortium), 96
- Illumina NGS Platform, 19–21
- Illumina SNP BeadChips, 14, 72, 74
- inborn error of metabolism (IEM), 465, 466
- induced cancer stem cells, 596
- induced conserved sub-structure score (ICS),
 - alignment scoring, 377
- integrated databases, 160, 164, 165, 163–179
- interaction databases, 160–167, 372, 463, 602
- interaction networks, 536–537, 540
- interactome analysis, 151, 163, 162–165, 427–437, 464
 - basic properties, 428, 427–429
 - biological function, 429–430
 - construction, 427, 438
 - context-sensitive, 479–480
 - databases, 439
 - diseases, 430
 - network localization, 430–432
 - randomization, 431–437
- interactomics, 601
- intermediate data integration, 289
- International Human Epigenome Consortium (IHEC), 96
- inter-organismal networks, 474–475
- isomorphic graphs, 120, 121, 373
- isomorphism, 194, 199, 203, 212, 220
- IsoRank, global network aligner, 200, 217, 383, 382–384
- Jacard index, 256, 271–272
- joint probabilities, 320
- k*-correctness, alignment scoring, 375
- k*-coverage, 392

- k*-means clustering, 258–262, 259, 260, 267, 580, 580
- k*-partite matching, 397
- Kamada-Kawai algorithm, 549
- kernel functions, 38
- kernel-based methods, data integration, 293, 293–294
- KRAS gene, 606
- Lance-Williams recurrence formula, 265, **266**
- large scale clustering algorithms, 258
- largest connected component, alignment scoring, 378
- late data integration, 289
- layouts, 550, 551, 570–573
 - algorithms, 549
 - force directed, 549–550
 - node-link diagrams, 549, 550, 550–551
- leukemia, cancer stem cells, 598
- linear algebra, 114
- link prediction, 213, 212–213, 511–513
- linkage disequilibrium, 5, 31, 34
- linkage functions, 257, 263, 264–265, 267
 - average linkage, 265
 - complete linkage, 265
 - Lance-Williams recurrence formula, 265, **266**
 - single linkage, 265
- linkage methods, disease identification, 173
- lncRNA *see* long non-coding RNAs
- local community paradigm (LCP) theory, 512–513, 517
- local network alignments, 375–376
- localization
 - network, 430–432, 435, 440, 505
 - subcellular, 154, 166, 600
- logistic regression models, multi-SNP, 3, 34, 36–37
- long non-coding RNAs (lncRNA), 87–88
 - algorithms, 92, 94
 - bioinformatic tools, 88–89
 - databases, **91**
 - epigenetics, 87–93
 - precision medicine, 88
 - previously annotated, 89–93
 - unannotated, 93
- long-read sequencing, 71
- low-throughput methods (LT), 154
- machine learning
 - algorithms, 33–39, 601
 - data integration, 290–294
 - non-negative matrix factorization, 294–295
 - pattern mining, 313–314, 359–361
 - precision medicine, 287, 306
- macroscale, neuronal connectivity, 491
- magnetic resonance imaging *see* MRI
- mapping algorithms, 372
- mapping mechanisms, epigenetics, 72, 74
- marginal probabilities, 320
- matching
 - bipartite, 375, 374–375
 - graph theory, 128
 - index, 137
 - k*-partite, 397
- matrices
 - adjacency, 129, 129–130, 291, 294, 295, 491–492, 499
 - GDV, 207, 208–209
 - operations, 115
 - special, 115–116
 - spectral decomposition, 116
- MaWish, pairwise network alignment, 381
- mean normalized entropy (MNE), 392
- medicine, 414, 448
 - disease module analysis, 437–448
 - disease networks, 422–423
 - exercises, 449
 - interactome analysis, 427–437
 - molecular networks, 415–422
 - social networks, 423–426
 - types of network, 415
- mesoscale, neuronal connectivity, 491
- metabolic networks, 113, 418, 419
- metastases, 595–598
- microarrays, 1, 11, 9–18, 70
 - biases, 75, 97
 - DNA methylation, 73–77
 - genotyping algorithms, **15**, 16
 - limitations, 71
 - vs. next generation sequencing, 26–28
 - normalization, 75–76
- microscale, neuronal connectivity, 491
- minicolumns, cortex, 491
- minimum attainable *P*-value, 328, 332–337, 336, 337, 346, 354–355, 356, 358
- Minkowski distances, 252–253
- mixed graphs, 122–123
- mixture models, 16
- model based clustering algorithms, 257–258
- modularity, 509
- module-based methods, disease identification, 173
- molecular networks, 415, 464
 - causal variants, 466–470

628 INDEX

- molecular networks (cont.)
 - co-expression networks, 419–420
 - databases, **463**
 - diseases, 470–475
 - genetic interactions, 420–422
 - genotype-phenotype relationships, 459, 461–464
 - metabolic networks, 113, 418, 419
 - protein-protein interactions, 415–417
 - regulatory networks, 418–419
 - tissue-sensitive, 475–480
- monogenic diseases, 470–473
- motifs, network, 138, 197
- MRI (magnetic resonance imaging)
 - brain structure, 492–493
 - diffusion tensor imaging, 495, 498, 500
 - diffusion-weighted, 495, 496
 - high angular resolution diffusion imaging, 497, 498
 - T1-weighted, 68, 494, 496
- multi-graphs, 122, 123, 543, 546
- multilayer networks, 125, 125–126, 401, 402, 401–402
- multiple network alignment, 391
 - definitions, 390–391
 - FUSE example method, 394–397
 - other methods, 397–399
 - scoring alignments, 392
 - SMETANA example method, 392–394
- multiple testing correction, pattern mining, 325–326
- multi-SNP association studies, **35**, 33–39
- multi-threshold permutation correction (MTPC), 506
- mutations *see* gene mutations
- myoglobin, 370
- neighborhood of vertices, 119, 119
- NetAligner, pairwise network alignment, 381
- netdis, 215
- network(s)
 - alignment-based comparison, 216–218
 - alignment-free comparison, 214–216
 - brains *see* brain networks
 - clustering, 209, 210, 209–210
 - comparison, 213–214
 - construction, 167–168
 - definition, 167
 - de-noising, 212, 213
 - edges, 463, 464
 - genetic interactions, 113–114, 463, 481, 536
 - genotype-phenotype relationships, 464–475
 - geometry, 492, 515, 516, 513–516
 - inter-organismal, 474–475
 - localization, 430–432, 435, 440, 505
 - measures, 132–140
 - molecular *see* molecular networks
 - molecular basis of diseases, 474, 470–475
 - motifs, 138, 197
 - neuroscience, 491–492
 - nodes, 194, 199, 212, 220, 435–436, 463, 464
 - pathways, 537–538, 539, 541
 - properties, 135, 132–138
 - protein-protein interactions, 111, **112**, 153, 167–170
 - similarity, 538–539, 542, 543
 - taxonomy, 536–539
 - theory *see* graph theory
 - topology, 196–209, 492, 506–508
- network alignment (NA), 223, 369–373, 374, 403
 - alternative formalisms, 399
 - directed networks, 402
 - exercises, 404–407
 - hyper-networks, 403, 403, 403
 - methods, 400
 - multilayer networks, 125–126, 401, 402, 401–402
 - multiple *see* multiple network alignment
 - pairwise *see* pairwise network alignment
 - probabilistic networks, 400, 399–401
 - protein-protein interactions, 372, 371–373
 - search-based method, 217–218
 - two-stage method, 216–218
- network analysis
 - Cytoscape, 574
 - differential, 481
 - protein-protein interactions, 173–175, 176, 194, 199, 212, 220, 221–223
- network-based data integration, 291, 290–291
- network-based disease modules, 437–438, 440, 442
- network-based statistics (NBS), 506
- network-based tools, **471**, **472**
- network models, 138, **462**, 463
 - Erdos–Renyi Random Graphs, 138–139
 - geometric networks, 139
 - scale-free networks, 139
- NetworkBLAST, pairwise network alignment, 381
- neuroscience, 490–492
 - brain network analysis tools, 505–506
 - brain network disorders, 517–519
 - exercises, 519–520
 - functional brain networks, 499–503

- network geometry, 513–516
- network topology, 506–513
- nodes, 503–505
- structural brain networks, 492–499
- next generation sequencing, 2, 18–27, 70, 477
- vs. microarrays, 26–28
- node(s)
 - brain networks, 503–505
 - conservation, 216–217, 389
 - correctness/coverage, alignment scoring, 377
 - network, 194, 199, 212, 220, 435–436, 463, 464
 - removals, 466
- node-colored graphlets, 200–201
- node-link diagrams, 542–547
 - animations, 551
 - combining layouts, 550, 551
 - force directed layouts, 549–550
 - layering visualizations, 552, 552
 - layouts, 549, 550, 550–551
 - network embedding, 550
 - simple layout algorithms, 549
 - visual mappings, 548–549
 - visualizing networks, 547, 547–548
- non-negative matrix factorization (NMF), 295, 294–295
 - homogeneous data integration, 298–300
 - precision medicine, 294–300
 - solutions, 295–298
- non-negative matrix tri-factorization (NMTF), 301, 300–301
 - FUSE, 394–397
 - heterogeneous data integration, 305
 - precision medicine, 300–306
 - solutions, 301–305
- normalization, cluster analysis, 246–247
- Notch signaling pathway, 597
- odds ratio, 5, 43
- oncogenes, 595
- one-mode data format, 244
- optimization algorithms, 374
- orbit aware graphlet counting, 206, 226
- orbit aware quad census (Oaqc), 227, 228
- orbit counting algorithm (Orca), 227, 228
- orbit unaware graphlet counting, 199, 226
- orbit weights, 386
- orbits, graphlets, 203, 203, 204, 205
- ordered graphlets, 202, 202
- ovarian cancer, 305, 476
 - biomarkers, 608, 609, 608–610
 - cancer stem cells, 598
- overlapping clustering, 243
- pairwise network alignment, 391
 - definitions, 373–375
 - example method, PathBlast, 379–381
 - global methods, 382–390
 - other methods, 381–382
 - scoring alignments, 375–379
- parallel parameterized graphlet decomposition (PGD) library, 227, 229
- Parkinson's disease, 34, 473–475, 480, 518–519
- partitional clustering, 243
- partitioning around medoids (PAM), 261–262
- path lengths, 169–170
- PathBlast, pairwise network alignment, 380, 379–381
- pathogenicity, graphlet-based approach, 224
- pathway enrichment analysis, 179, 180
- pathways, networks, 537–538, 539, 541
- patient subtyping, cancer precision medicine, 287
- pattern enumeration tree, 330, 331
- pattern mining, 315–328
 - algorithms, 329, 330, 340, 361
 - confounding effect, 350, 351
 - covariate factors, 349–359
 - machine learning, 313–314, 359–361
 - permutation testing, 346–349
 - software tools, 360
 - statistical redundancy, 341–349
 - Tarone's method, 329–341
- pattern occurrence indicator, 315
- peak calling, 79, 79
- Pearson's chi-squared test, 323–325, 336
- permutation importance, 40
- permutation testing
 - algorithms, 347
 - pattern mining, 346–349
- personalized genetic tests (PGT), 3, 7, 13, 45
- personalized medicine *see* precision medicine
- personalized oncology, 287–288, 594
- phenotypes, definition, 460, 461; *see also*
 - genotype-phenotype relationships
- power-lawness, 509
- PPI *see* protein-protein interactions
- precision medicine, 286, 288, 306, 596
 - cancer, 287
 - data integration, 287–288
 - data integration methods, 290–294
 - data integration types, 288–290
 - drug repurposing, 287
 - exercises, 306–308
 - long non-coding RNAs, 88
 - machine learning, 287, 306

630 INDEX

- precision medicine (cont.)
 - non-negative matrix factorization, 294–300
 - non-negative matrix tri-factorization, 300–306
- prediction databases, 163, **164**, 162–165, 179
- predictive genetic risk models, 39–44
- preferential attachment, 194, 199, 212, 213, 220
- preprocessing, cluster analysis, 242, 246–251
- principal component analysis (PCA), 206, 215–216, 226, 251, 248–251
- probabilistic networks, 400, 399–401
- prostate cancer, biomarkers, 606–608
- protein complex detection, 241
- protein function prediction, 218–220
- protein homology detection, 241–242
- protein structure network (PSN), 193
- protein-DNA interactions, 600
- protein-protein interactions (PPI), 151–154, 179–181, 193, 536, 537
 - annotations, 171, 370, 371
 - biases, 158–159, 438
 - computational biology workflows, 175–179
 - computational method types, 156
 - computational prediction, 156–159
 - data integration, 159–160
 - databases, **112**, 160–167, 602, **603**
 - dominating set, 211
 - exercises, 181
 - experimental detection, **155**, 154–156
 - high-throughput methods (HT), 154
 - human aging, 220–223
 - interaction types, 155
 - limitations, detection methods, 154, 155
 - low-throughput methods (LT), 154
 - molecular networks, 415–417
 - network alignment, 372, 371–373
 - network analysis, 173–175, 176, 194, 199, 212, 220, 221–223
 - network visualizations, 170–172
 - networks, 111, **112**, 153, 167–170
 - stem cell therapy, 598–603, 603
 - tissue interactomes, 479
- proteins
 - cancer, 223–224, 598–603
 - domains, 156
 - functions, 370, 369–371
 - sequencing, 156
- proteomics, 599–601
- prototype based clustering algorithms, 257
- proximity calculation, cluster analysis, 252–256
- pruning condition, 338, 339
 - Cochran-Mantel-Haenszel test, 355–359
- public databases, 2, 22, 72, 74, 427, 559–562
- qualitative annotations, 172
- quantitative annotations, 172
- Rand index, 271
- random forest methods, 37–40
- random walks, 132
- randomization, network properties, 431–433, 435, 436–437
 - nodes, 435–436
 - topology, 433–435
- rapid graphlet enumerator (RAGE), **227**, 228
- RAS genes, mutations, 462
- recessive models, 32
- redundancy, pattern mining, 342, 341–349
- regulatory networks, 418–419
- relative graphlet frequency distance (RGFD), 214
- relative risk, 43
- repression, transcriptional, 491
- resting state networks (RSN), 499
- rich-clubness, 509–510, 517
- risk indicators, 43
- Roadmap Epigenomics Project, 95
- SAND/SAND-3D subgraph tools, **227**, 229
- scale-free networks, 139
- schema, 160
- scoring schemes *see* alignment scoring schemes
- search algorithms, graph theory, 130–131
- Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), 606–608
- search-based network alignment, 217–218
- seed clusters, disease module analysis, 438–440
- semantic similarity, alignment scoring, 376
- sequencing
 - alignment, 369
 - by synthesis, 71
 - proteins, 156
- SF-GD (scale-free gene duplication and divergence), 139
- short-read sequencing, 71
- sickle cell disease, 461–462, 479–480
- significant itemset mining, 317, 315–317
- significant pattern mining *see* pattern mining
- significant subgraph mining, 317–318, 319
- silhouette values, 273, 274
- similarity networks, 538–539, 542, 543
- simultaneous decomposition, 298, 299
- single nucleotide polymorphisms (SNP), 1
 - calling and genotyping, 21–27
 - definition, 6
 - disease association, 10, 41, 31–47
 - effect sizes, 5

- identification, 9, 30
- significant itemset mining, 315
- single-SNP association studies, 31–33
- small-worldness, 508
- SMETANA, multiple network alignment
 - method, 392–394
- social networks, 225, 224–226, 423–426, 538
 - contagious diseases spread, 423–424
 - social contagion, 425
 - transportation, 424–425
- software tools, pattern mining, **360**
- spectral clustering, 132
- spectral graph theory, 131–132
- spinocerebellar ataxia type 1 (SCA1), 473
- spreadsheets, visualizing networks, 541, 544
- standardization, cluster analysis, 246–247
- static graphlets, 195, 196, 202
- statistical association testing, pattern mining, 318
- statistical redundancy, pattern mining, 341–349
- stem cell therapy, 594–595
 - cancer stem cells, 595–598, **605**, **606**, 603–612
 - exercises, 612
 - protein interactions, 598–603
 - tumor stemness biomarkers, 596–597, **605**, **606**, 610, 603–612
- stochastic measures, network topology, 507–510
- STRING (Search Tool for the Retrieval of Interacting Genes/Proteins), 606–608
- structural brain networks, 492–499
- structural consistency, 509
- subcellular localization, 154, 166, 600
- subgraphs, 120, 120
 - centrality, 137
 - connectedness, 119
 - isomorphism problem, 373
- subset/superset relationships, pattern mining, 341, 343
- support vector machines, 34–38
- suppressor genes, tumors, 595
- symmetric sub-structure scores, alignment scoring, 377
- T1-weighted MRI scanning, 68, 494, 496
- targeted therapy, 594
- targeting, drugs, 174
- Tarone's improved Bonferroni correction, 327–328
- Tarone's method, pattern mining, 329–341
- ten-eleven-translocation (TET) proteins, 69
- tertiary structure, 156–158
- testability, pattern mining, 329, 337, 359
- 1000 genomes project, 1, 18, 28–29
- 3C technology, chromatin conformation, 81, 83
- time-respecting path, 199, 226
- tissue annotation, 166
- tissue interactomes, 478–482
 - differential network analysis, 481
 - genome-wide association studies, 479
 - meta-analysis, 481–482
 - PPI networks, 479
 - tools, **478**
- tissue profiles, **477**
- tissue-sensitive molecular networks, 475–480
- tissue-specific interactions, 476
- topological associated domains (TAD),
 - chromatin conformation, **86**, 86–87
- topology, networks, 196–209, 492, 506–508
- training methods, data integration, 290, 289–290
- transcriptional regulation networks, 113
- transcriptional, repression, 491
- transitivity clustering, 268–270
- transportation networks, contagious diseases, 424–425
- trees, graph theory, 124, 124
- tumors; *see also* cancer
 - metastases, 595–598
 - stemness biomarkers, 596–597, 603–612, **605**, **606**, 610
 - suppressor genes, 595
- Turing machines, 117
- two-mode data format, 244
- two-stage network alignment, 216–218
- undirected graphlets, 196, 202
- undirected graphs, 118
- unordered graphlets, 196, 202
- validity indices, cluster evaluation, 270–271
- variable importance measures (VIMs), 37
- variant calling algorithms, 25–26
- variety, databases, 161
- variety/velocity/veracity, databases, 160
- vector spaces, 116
- velocity, databases, 161
- veracity, databases, 161
- VIM (variable importance measures), 37
- visual mappings
 - continuous, 548
 - discrete, 548

632 INDEX

- visual mappings (cont.)
 - node-link diagrams, 548–549
 - passthrough, 548
- visualizing networks, 170–172, 533, 534, 535, 540–552
 - adjacency matrices, 541, 544
 - BioFabric, 542, 545
 - Cytoscape, 534
 - node-link diagrams, 547, 542–548
 - spreadsheets, 541, 544
 - walks, graph theory, 121, 121
 - weighted graphs, 113, 123, 123
 - weighted transitive graph projection
 - problem (WTGPP), 269, 268–269
 - Wnt signaling pathway, 597
 - writer enzymes, 68
- yeast two-hybrid (Y2H) assays, 155, 417, 427