

Index

Note: Page numbers followed by f and t indicate figures and tables respectively

A

- AD network, 88–89
- Affinity propagation (AP) method, 41, 130
- Algorithm FUSE, 71–74
 - phases of, 72–73
- Alzheimer's disease (AD) network, 2, 59, 60, 140
- Annotation-driven clustering, 44–46
 - recursive function, 44–45
 - VI-Cut algorithm, 46
- Annotation loss, 136

B

- Betweenness-based clustering, 42–43
- Bimolecular fluorescence complementation (BIFC), 13–14
- Biological General Repository for Interaction Datasets (BioGRID), 16
- Biological networks, 1
 - analysis, 2
 - challenges, 4
 - components, 1
 - PPI network, 2, 5
 - scale-free distribution of, 2

C

- CFinder method, 39
- Clique percolation method (CPM), 39
- Clustering-based on Maximal Cliques (CMC) method, 40
- Clustering properties, PPI network, 139
 - cluster validation measures, 52
 - functional homogeneity-based validation, 50–51

- MIPS-based validation, 51
- functional modules, 23
- problem, *see* Clustering properties, PPI network clustering problem
- techniques, *see* Clustering properties, PPI network clustering techniques
- topological modules, 23
- Clustering properties, PPI network clustering problem
 - attributed, 26
 - categories, 29
 - clusters, overlapping nature of, 26
 - coefficient, 27
 - cohesiveness, 27–28
 - conductance, 28
 - Czeknowski-Dice distance, 28
 - dense and sparse clusters, 26
 - density, 27
 - edge betweenness, 28–29
 - full coverage of, 26
 - heuristic-based algorithms, 32
 - interactions, edge weights of, 26
 - log odd score, 28
 - modularity, 29
 - multiple clustering-based algorithms, 32
 - problem definition, 25
 - scalability, 26
 - techniques, 26–27, 30t
 - variational information distance, 29
- Clustering properties, PPI network clustering techniques
 - complete enumeration algorithms, 38–40
 - flow-based algorithms, 36–38

- genomic data-driven clustering
 - algorithms, 48–49
 - graph-cut and hierarchical clustering algorithms, 42–46
 - heuristic-based algorithms, 32–36
 - multiple clustering-based algorithms, 46–47
 - random walks and message passing algorithms, 40–42
- ClusterONE method, 35
- Complete enumeration algorithms
 - CFinder method, 39
 - CMC, 40
 - CPM, 39
 - k -clique modules, 39
 - MC-based approach, 38–39
 - module density, 40
 - near-cliques, types of, 39, 39f
 - SPC, 38
- Comprehensive summarization techniques, 141
- D**
- Database of Interacting Proteins (DIP), 16
- Deoxyribonucleic acid (DNA), 10, 11, 12
- Differential functional summarization
 - background, 117–119
 - DiffNet algorithm, 127–128
 - experimental study
 - annotation loss, effect of, 136
 - graph clustering algorithms, comparison with, 130–132
 - interaction noise, effect of, 135–136
 - MMS-treated/untreated dE-MAP network, functional analysis of, 129–130
 - running times, 133–135
 - various parameters, effect of, 132–133
 - functional subgraphs
 - constructing differential networks, 120–123
 - skewness and coherence measures, 123–124
 - motivation and overview, 119–120
 - problem
 - distinctiveness, 125–127
 - edge coverage, 125
 - subgraph coherence and skewness, 125
- Directed acyclic graph (DAG), 17
- Disease module, 141
- DOMINE database, 16
- DPPlus, 33, 34
- E**
- Edge betweenness, 28–29
- Edge coverage, 125
- Epistatic miniarray profile (E-MAP)
 - approach, 117
- Evaluation metrics
 - coverage metric, 75
 - precision and recall, 75–76
 - redundancy, 75
- Expansion operator, 36
- F**
- FACETS algorithm
 - initialization phase, 101
 - iteration phase, 102–105
- FACETS clusters
 - p -value significance of, 111
 - statistical significance of, 109–110
- Flow-based algorithms
 - expansion operator, 36
 - functional flow pattern, 38
 - inflation operator, 36
 - informative proteins, 37
 - MCL approach, 37
 - weight transition matrix, 36
- Functional flow pattern, 38
- Functional subgraphs
 - constructing differential networks, 120–123
 - negative differential, 122
 - positive differential, 122
 - S -score, 121
 - skewness measure, 123
 - coherence measure, 123–124
- Functional summarization
 - AD network, 88–89
 - algorithm FUSE, 71–74
 - phases of, 72–73
 - different parameters, effects of
 - annotation loss, effect of, 85–86
 - parameter k , effect of, 82–83
 - parameters b and d , effect of, 83–84
 - runtime and scalability, 87–88
 - statistical significance, 84–85
 - evaluation metrics
 - coverage metric, 75
 - precision and recall, 75–76
 - redundancy, 75
 - functional cluster, 63
 - FUSE, goal of, 63
 - FUSE vs. graph clustering methods

- cluster quality comparison, 77–79
 - dataset, 76
 - function representativeness comparison, 79–80
 - qualitative evaluation, 80–82
 - inferring functional cluster hubs, 90–92
 - motivation, 59–60
 - PPI clustering techniques, limitations of, problem
 - non-negative structural information value, 69
 - PPI, 65–68
 - structural information value, 70
 - related work, 63–65
 - Functional summary generator (FUSE), 139
 - goal of, 63
 - vs. graph clustering methods
 - cluster quality comparison, 77–79
 - dataset, 76
 - function representativeness comparison, 79–80
 - qualitative evaluation, 80–82
 - Functional summary graph (FSG), 65
- G**
- Gene Ontology (GO), database, 6
 - directed acyclic graph (GO DAG), 17, 18, 103, 112, 113
 - role of, 16
 - structure of, 17–19
 - Gene Ontology Annotation (GOA), 16
 - Genomic data-driven clustering algorithms, 48–49
 - Graph clustering algorithms, 130–132
 - Graph-cut and hierarchical clustering algorithms
 - annotation-driven clustering, 44–46
 - betweenness-based clustering, 42–43
 - shared neighbor-based clustering, 43–44
- H**
- Heuristic-based algorithms
 - molecular complex prediction, 33
 - post-processing
 - ClusterONE method, 35
 - DPCLUS, 33, 34
 - naive score, 35
 - partial coverage method, 36
 - RNSC, 34
 - scaled score, 35
 - seed-refine algorithm, 35
 - SPICi method, 35
 - vertex weighting, 32
 - High-throughput screening methods, noise in, 14–15
 - Human Protein Reference Database (HPRD), 16
- I**
- Imbalanced clusters, 37
 - Inflation operator, 36
 - Informative proteins, 37
 - IntAct database, 15–16
 - Interaction noise, 135–136
 - Inter-facet orthogonality, 96, 99–100
 - Intra-facet cluster modularity, 96, 99
- J**
- Jaccard index (JI), 85, 86, 105, 107, 136
 - Jaccard similarity, 135
 - Joint probability, 45
- K**
- k -clique modules, 39
 - Knowledge bases, 15–16, *see also* individual databases
 - Kyoto Encyclopedia of Genes and Genomes (KEGG) database, 15
- L**
- Laplacian graph, 26
 - Log odd score, 28
- M**
- Markov Clustering (MCL), 37, *see also* Flow-based algorithms
 - Methyl methanesulfonate (MMS), 118
 - Monte Carlo (MC) optimization-based technique, 38–39
 - Multi-faceted functional decomposition
 - experimental study
 - convergence, 109
 - experiment settings, 105
 - FACETS clusters, statistical significance of, 109–110
 - GO DAG, comparison with, 112
 - graph clustering approaches, 110–112
 - initial starting point, effect of, 109
 - quantitative assessment, 106
 - real data, validation on, 106–108
 - robustness, 108–109
 - running time, 110
 - FACETS algorithm
 - initialization phase, 101

- iteration phase, 102–105
- human autophagy system, 113–115
- motivation, 95–97
- problem statement
 - goal of, 98
 - inter-facet orthogonality, 99–100
 - intra-facet cluster modularity, 99
 - problem definition, 101
 - terminology, 97–98
- related work, 97
- Multiple clustering (MC)-based algorithms, 46–47
- Munich Information Center for Protein Sequences (MIPS) database, 16

- N**
- Naive score, 35
- Near-cliques, types of, 39, 39f
- Non-negative structural information value, 69

- O**
- Online Mendelian Inheritance in Man (OMIM) database, 17

- P**
- Phospho-ELM knowledge base, 16
- PHOSPIDA knowledge base, 16
- Protein-protein interaction (PPI) network, 2
 - biological processes, cell cycle, 9
 - clustering problem, *see* Clustering properties, PPI network clustering problem
 - clustering techniques, *see* Clustering properties, PPI network clustering techniques
 - cluster validation measures, 52
 - functional homogeneity-based validation, 50–51
 - MIPS-based validation, 51
 - databases, 15–16
 - GO, role of, 16
 - methods
 - BIFC, 13–14
 - high-throughput screening methods, noise in, 14–15
 - TAP, 13
 - Y2H, 12
 - molecule types, 11
 - stable, 11
 - transient, 11
- Proteins, 9

- 3D structured, 10

- Q**
- Quantitative assessment, of multi-faceted functional decomposition, 106
- Qualitative evaluation, in FUSE vs. graph clustering methods, 80–82
- Quantitative network summarization, 141, *see also* Differential functional summarization

- R**
- Random walks and message passing algorithms, 40–42
- REACTOME database, 15
- Repeated Random Walk (RRW) method, 41
- Restricted Neighborhood Search Clustering (RNSC) algorithm, 34

- S**
- Scaled score, 35
- Seed-refine algorithm, 35
- Shared neighbor-based clustering, 43–44
- SPICi method, 35
- STRING database, 15
- Superior clustering techniques, 140–141
- Superparamagnetic clustering (SPC), 38

- T**
- Tandem affinity purification (TAP), 13
- Tree-sniping approach, 45

- U**
- Unified probabilistic model, 48
- UniprotKB, knowledge base, 59

- V**
- Variational information distance, 29
- Vertex weighting, 32
- VI-Cut algorithm, 46

- W**
- Weight transition matrix, 36

- Y**
- Yeast-two-hybrid (Y2H) method, 12

- Z**
- Zero coherence, 124