

Bibliography

1. Berkhin, P. Survey of Clustering Data Mining Techniques, in Technical Report, Accrue Software, San Jose, CA.
2. Han, J., & Kamber, M. *Data Mining: Concepts and Techniques*, Morgan Kaufmann, California, 2001.
3. Al-Ejeh, F. et al. Meta-analysis of the global gene expression profile of triple-negative breast cancer identifies genes for the prognostication and treatment of aggressive breast cancer, *Oncogenesis - Nature*, 2014.
4. Jiang, D., Pei, J., & Zhang, A. Interactive Exploration of Coherent Patterns in Time-Series Gene Expression Data, in Proceedings of Ninth ACM SIGKDD International Conference Knowledge Discovery and Data Mining (KDD '03), Washington, DC, 565-570.
5. Jiang, D., Pei, J., & Zhang, A. DHC: A Density-Based Hierarchical Clustering Method for Time-Series Gene Expression Data, in Proceedings of Third IEEE International Symposium on Bioinformatics and Bioengineering (BIBE2003), IEEE, 393 - 400.
6. Solomon, E. P., Berg, L. R., & Martin, D. W. *Biology*, 9th ed., Thomson Learning, California, 2011.
7. Domany, E. Cluster Analysis of Gene Expression Data, *Journal of Statistical Physics* **110** (3–6), 1117–1139, 2003.
8. Somasundaram, K., Mungamuri, S. K., & Wajapeyee, N. DNA Microarray Technology and its Applications in Cancer Biology, in Review, Department of Microbiology and Cell Biology , Indian Institute of Science, Bangalore.
9. Draghici, S. *Data Analysis Tools for DNA Microarrays*, First Edition ed., Chapman and Hall/CRC, London, 2003.
10. Mitra, S. et al. *Introduction to Machine learning and Bioinformatics*, First Edition ed., Chapman and Hall/CRC Computer Science & Data Analysis, London, 2008.
11. Hocine, S., Raymond, P., & Zenklusen, D. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast, *Nature Methods* **10**, 119–121, 2013.
12. Bhar, A. et al. Coexpression and coregulation analysis of time-series gene

- expression data in estrogen-induced breast cancer cell, *Algorithms for Molecular Biology* **8** (9), 1-11, 2013.
13. Sun, Y. et al. Cancer progression modeling using static sample data, *Genome Biology* **15** (440), 1-17, 2014.
 14. Gupta, A., & Bar-Joseph, Z. Extracting Dynamics from Static Cancer Expression Data, *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* **5** (2), 172-182, 2008.
 15. Auer, P. L., & Doerge, R. W. Investigations - Methods, technology, and resources: Statistical Design and Analysis of RNA Sequencing Data, *Genetics* **185**, 405-416, 2010.
 16. Lee, J. H. et al. Highly Multiplexed Subcellular RNA Sequencing in Situ, *Science* **343** (6177), 1360-1363, 2014.
 17. Babu, M. M. Introduction to Microarray Data Analysis, in *Computational Genomics*, edited by Grant, R., Horizon Press, UK, 2004.
 18. Quinlan, J. R. *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc, San Francisco, California, 1993.
 19. Breiman, L. et al. *Classification and Regression Trees*, Chapman and Hall, 1984.
 20. Aha, D. W., Kibler, D., & Albert, M. K. Instance-Based Learning Algorithms, *Machine Learning* **6**, 37-66, 1991.
 21. John, G. H., & Langley, P. Estimating Continuous Distributions in Bayesian Classifiers, in Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI '95), Morgan Kaufmann Publishers Inc. , San Francisco, CA, USA, 338-345.
 22. Şenbabaoğlu, Y., Michailidis, G., & Li, J. Z. Critical limitations of consensus clustering in class discovery, *Scientific Reports - Nature* **4** (6207), 2014.
 23. Kaufman, L., & Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, 1990.
 24. Raymond, T. N., & Han, J. Efficient and Effective Clustering Method for Spatial Data Mining, in Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94), Morgan Kaufmann Publishers Inc., San Francisco, CA,

- USA, 144-155.
25. Guha, S., Rastogi, R., & Shim, K. CURE: An Efficient Clustering Algorithm for Large Datasets, in Proceedings of ACM SIGMOD Conference (SIGMOD '98), ACM New York, 73-84.
 26. Guha, S., Rastogi, R., & Shim, K. ROCK: A Robust Clustering Algorithm for Categorical Attributes, *Information Systems* **25** (5), 345–366, 1999.
 27. Zhang, T., Ramakrishnan, R., & Linvy, M. BIRCH: An Efficient data Clustering Method for Very Large Datasets, *Data Mining and Knowledge Discovery* **1** (2), 141-182, 1997.
 28. Ester, M. et al. A density-based algorithm for discovering clusters in large spatial databases, in Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD'96), AAAI Press, Portland, OR, 226-231.
 29. Hinneburg, A., & Keim, D. An efficient approach to clustering in large multimedia datasets with noise, in Proceedings of 4th International Conference on Knowledge Discovery and Data Mining (KDD '98), AAAI Press, New York, 58-65.
 30. Fisher, D. Improving inference through conceptual clustering, in Proceedings of the National Conference Artificial Intelligence (AAAI-87), AAAI Press, Seattle, WA, 461-465.
 31. Kohonen, T. *Self-Organizing Maps*, 2nd Extended Edition, Springer Series in Information Sciences, 1997.
 32. Shamir, R., & Sharan, R. CLICK: A Clustering Algorithm for Gene Expression Analysis, in Proceedings of Eighth International Conference on Intelligent Systems for Molecular Biology (ISMB '00), AAAI Press, 307-316.
 33. Ben-Dor, A., Shamir, R., & Yakhini, Z. Clustering Gene Expression Patterns, *Journal of Computational Biology* **6** (3/4), 281-297, 1999.
 34. Dembele, D., & Kastner, P. Fuzzy C-Means method for clustering microarray data, *Bioinformatics* **19**, 973–980, 2003.
 35. Di-Gesu, V. et al. GenClust: A genetic algorithm for clustering gene expression data, *BMC Bioinformatics* **6**, 289, 2005.

36. Brun, M. et al. Model-based evaluation of clustering validation measures, *Pattern Recognition* **40** (3), 807-824, 2007.
37. Dumonceaux, F., Raschia, G., & Gelgon, M. An Algebraic Approach to Ensemble Clustering, in 22nd International Conference on Pattern Recognition (ICPR '14), IEEE, Stockholm, 1301 - 1306.
38. Zhang, H. et al. Improving accuracy for cancer classification with a new algorithm for genes selection, *BMC Bioinformatics* **13** (298), 2012.
39. Elder IV, J. F., & Pregibon, D. A Statistical Perspective on Knowledge Discovery in Databases, in *Advances in Knowledge Discovery and Data Mining*, edited by Fayyad, U. M. et al. , AAAI Press, Menlo Park. CA, 1996, 83-113.
40. Dietterich, T. G. Machine-Learning Research: Four Current Directions, *AI Magazine* **4** (18), 97-136, 1997.
41. Steinberg, D., & Colla, P. *CART- Classification and regression trees*, Salford Systems, San Diego, CA, 1997.
42. Breiman, L. Bagging predictors, *Machine Learning* **2** (24), 123-140, 1996b.
43. Kuncheva, L. Combining classifiers: Soft computing solutions, in *Pattern Recognition: From Classical to Modern Approaches*, edited by Pal, S. K., & Pal, A., World Scientific Publishing Co, Singapore, 2001, 427-452.
44. Hansen, L. K., & Salamon, P. Neural network ensembles, *IEEE Transactions Pattern Analysis & Machine Intelligence* **10** (12), 993-1001, 1990.
45. Tumer, K., & Ghosh, J. Error correlation and error reduction in ensemble classifiers, *Connection Science* **8** (3-4), 385-403, 1996.
46. Krogh, A., & Vedelsby, J. Neural network ensembles, cross validation and active learning, *Advances in Neural Information Processing Systems* **7**, 231-238, 1995.
47. Wolpert, D. H. Stacked generalization, *Neural Networks* **5**, 241-259, 1992.
48. Kuncheva, L. *Combining Pattern Classifiers: Methods and Algorithms*, Wiley, 2004.
49. Webb, G. I., & Zheng, Z. Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques, *IEEE Transactions on Knowledge and*

- Data Engineering* **8** (16), 980-991, 2004.
50. Breiman, L. Arcing classifiers (with discussion), *The Annals of Statistics* **3** (26), 801-849, 1998.
51. Schapire, R. E. et al. Boosting the margin: A new explanation for the effectiveness of voting methods, *The Annals of Statistics* **5** (26), 1651-1686, 1998.
52. Schapire, R. E. The strength of weak learnability, *Machine Learning* **5** (2), 197-227, 1990.
53. Freund, Y., & Schapire, R. E. Experiments with a new boosting algorithm, in Proceedings of the 13th International Conference in Machine Learning (ICML '96), Bari, Italy, 148-156.
54. Breiman, L. *Arcing Classifiers*, Statistics Department, University of California, California, 1996.
55. Merler, S., Caprile, B., & Furlanello, C. Parallelizing AdaBoost by weights dynamics, *Computational Statistics & Data Analysis* **51** (5), 2487-2498, 2007.
56. Zhang, C. X., & Zhang, J. S. A local boosting algorithm for solving classification problems, *Computational Statistics & Data Analysis* **52** (4), 1928-1941, 2008.
57. Bauer, E., & Kohavi, R. An empirical comparison of voting classification algorithms: Bagging, Boosting, and variants, *Machine Learning* **35**, 1-38, 1999.
58. Breiman, L. Random Forests, *Machine Learning* **1** (45), 5-32, 2001.
59. Kamath, C., & Cantu-Paz, E. Creating ensembles of decision trees through sampling, in Proceedings of 33rd symposium on the interface of computing science and statistics, Costa Mesa, CA.
60. Kamath, C., Cantu-Paz, E., & Littau, D. Approximate splitting for ensembles of trees using histograms, in Proceedings of 2nd SIAM international conference on data mining (SDM-2002), Arlington, VA, USA, 370-383.
61. Dougherty, E. R. et al. Performance of Error Estimators for Classification, *Current Bioinformatics* **1** (5), 53-67, 2010.
62. Ge, G., & Wong, G. Classification of premalignant pancreatic cancer mass-

- spectrometry data using decision tree ensembles, *BMC Bioinformatics* **9**, 275, 2008.
63. Wang, S. Q., Yang, J., & Chou, K. C. Using stacked generalization to predict membrane protein types based on pseudo-amino acid composition, *Journal of Theoretical Biology* **242** (4), 941–946, 2006.
64. Netzer, M. et al. A new ensemble-based algorithm for identifying breath gas marker candidates in liver disease using ion molecule reaction mass spectrometry, *Bioinformatics* **25** (7), 941–947, 2009.
65. Tsoumakas, G., Partalas, I., & Vlahavas, I. A Taxonomy and Short Review of Ensemble Selection, in Proceedings of Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications, SUEMA (ECAI 2008), Patras, Greece.
66. Banfield, R. E. et al. A comparison of decision tree ensemble creation techniques, *IEEE Transactions Pattern Analysis and Machine Intelligence* **1** (29), 173-180, 2007.
67. Rokach, L. Collective-agreement-based pruning of ensembles, *Computational Statistics & Data Analysis* **53** (4), 1015-1026, 2009.
68. Liu, H., Mandvikar, A., & Mody, J. An empirical study of building compact ensembles, *International Conference on Web-Age Information Management (WAIM), Lecture Notes in Computer Science (LNCS)* **3129**, 622-627, 2004.
69. Margineantu, D., & Dietterich, T. Pruning adaptive boosting, in Proceedings of 14th International Conference on Machine Learning (ICML-97), Morgan Kaufmann Publishers Inc. , San Francisco, CA, USA, 211-218.
70. Zhou, Z. H., Wu, J., & Tang, W. Ensembling neural networks: many could be better than all, *Artificial Intelligence* **137**, 239-263, 2002.
71. Rokach, L. Ensemble-based classifiers, *Artificial Intelligence Review* **1-2** (33), 1-3, 2010.
72. Ben-Dor, A. et al. Tissue Classification with Gene Expression Profiles, *Journal of Computational Biology* **3-4** (7), 559-583, 2000.
73. Dudoit, S., Fridlyand, J., & Speed, T. Comparison of discrimination methods for the classification of tumors using gene expression data, *Journal of the American*

- Statistical Association* **97**, 77-87, 2002.
74. Dettling, M., & Buhlmann, P. Boosting for tumor classification with gene expression data, *Bioinformatics* **9** (19), 1061-1069, 2003.
75. Long, P. Boosting and Microarray Data, *Machine Learning* **53**, 31-44, 2003.
76. Tan, A., & Gilbert, D. Ensemble machine learning on gene expression data for cancer classification, *Applied Bioinformatics* **3 Suppl** (2), S75-S83, 2003.
77. Lee, J. et al. An extensive comparison of recent classification tools applied to microarray data, *Computational Statistics and Data Analysis* **48**, 869-885, 2005.
78. Diaz-Uriarte, R., & Andres, S. d. Gene selection and classification of microarray data using random forest, *BMC Bioinformatics* **7**, 3, 2006.
79. Izmirlian, G. Application of the random forest classification algorithm to a SELDI-TOF proteomics study in the setting of a cancer prevention trial, *Annals of the New York Academy of Sciences* **1020**, 154-174, 2004.
80. Zhang, H., Yu, C., & Singer, B. Cell and tumor classification using gene expression data: Construction of forests, *Proceedings of National Academy Science (PNAS)* **7**, 4168-4172, 2003.
81. Geurts, P. et al. Proteomic mass spectra classification using decision tree based ensemble methods, *Bioinformatics* **15** (21), 3138-3145, 2005.
82. Topchy, A. et al. Adaptive clustering ensembles, in Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), Cambridge, UK, 272-275.
83. Xu, R., & Wunsch, D. I. Survey of clustering algorithms, *IEEE Transactions Neural Networks* (16), 645-678, 2005.
84. Dudoit, S., & Fridlyand, J. Bagging to improve the accuracy of a clustering procedure, *Bioinformatics* **9** (19), 1090-1099, 2003.
85. Gablentz, W., & Koppen, M. Robust clustering by evolutionary computation, in Proceedings Fifth Online World Conference Soft Computing in Industrial Applications (WSC5).
86. Hong, Y. et al. Unsupervised feature selection using clustering ensembles and population based incremental learning algorithm, *Pattern Recognition Society* **9**

- (41), 2742-2756, 2008.
87. Luo, H., Jing, F., & Xie, X. Combining multiple clusterings using information theory based genetic algorithm, in IEEE International Conference on Computational Intelligence and Security-2006, IEEE, Guangzhou, 84-89.
88. Fern, Z. X., & Brodley, E. C. Solving cluster ensemble problems by bipartite graph partitioning, in Proceedings Twenty-First International Conference on Machine Learning (ICML'04), ACM, New York, NY, USA, Chongqing, China, 36.
89. Kellam, P. et al. Comparing, contrasting and combining clusters in viral gene expression data, in 6th Workshop on Intelligent Data Analysis in Medicine and Pharmacology, 56-62.
90. Rand, M. W. Objective criteria for the evaluation of clustering methods, *Journal of American Statistical Association* **66** (336), 846-850, 1971.
91. Ben-Hur, A., Elisseeff, A., & Guyon, I. A stability based method for discovering structure in clustered data, in Proceedings of the Pacific Symposium on Biocomputing, Kauai, Hawaii (USA), 6-17.
92. Mirkin, G. B. *Mathematical Classification and Clustering*, Kluwer Academic Press, Dordrecht, 1996.
93. Zhao, Y., & Karypis, G. *Criterion functions for document clustering: Experiments and analysis*, 2002.
94. van Rijsbergen, C. J. Foundation of evaluation, *Journal of Documentation* **30** (4), 365-373, 1974.
95. van Dongen, S. Performance criteria for graph clustering and Markov cluster experiments, in Tech. Rep. INS-R0012, Centre for Mathematics and Computer Science.
96. Bakus, J., Hussain, M. F., & Kamel, M. A. SOM-based document clustering using phrases, in 9th International Conference on Neural Information Processing (ICONIP '02), IEEE, 2212-2216.
97. Strehl, A., & Ghosh, J. Cluster ensembles: A knowledge reuse framework for combining multiple partitions, *The Journal of Machine Learning Research* **3**, 583-617, 2002.

98. Meilă, M. Comparing clusterings - an information based distance, *Journal of Multivariate Analysis* **98** (5), 873-895, 2007.
99. Vega-Pons, S., Correa-Morris, J., & Ruiz-Shulcloper, J. Weighted cluster ensemble using a kernel consensus function, in Proceedings of the 13th Iberoamerican congress on Pattern Recognition: Progress in Pattern Recognition, Image Analysis and Applications (CIARP '08), Springer-Verlag, Berlin, Heidelberg, 195-202.
100. Vega-Pons, S., Correa-Morris, J., & Ruiz-Shulcloper, J. Weighted partition consensus via kernels, *Pattern Recognition* **8** (43), 2712-2724, 2010.
101. Fischer, B., & Buhmann, J. Bagging for path-based clustering, *IEEE Transactions Pattern Analysis and Machine Intelligence* **11** (25), 1411-1415, 2003.
102. Lam, L., & Suen, C. Y. Application of majority voting to pattern recognition: an analysis of its behavior and performance, *IEEE Transactions System, Man and Cybernetics* **5** (27), 553-568, 1997.
103. Kuhn, H. The Hungarian Method for the assignment problem, *Naval Research Logistics Quarterly* **2**, 83-97, 1995.
104. Dimitriadou, E., Weingessel, A., & Hornik, K. Voting-Merging: An ensemble method for clustering, in Proceedings of the International Conference on Artificial Neural Networks (ICANN '01), Lecture Notes in Computer Science, Vienna, Austria, 217-224.
105. Fred, A. L. N., & Jain, A. K. Combining multiple clustering using evidence accumulation, *IEEE Transactions Pattern Analysis and Machine Intelligence* **27** (6), 835-850, 2005.
106. Jain, A. K., Murty, M., & Flynn, P. Data clustering: A review, *ACM Comput. Surv. (CSUR)* **3** (31), 264-323, 1999.
107. Li, Y. et al. Clustering ensembles based on normalized edge. eds. Z. H. Zhou, H. Li and Q. Yang, *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Lecture Notes in Artificial Intelligence (LNAI)* **4426**, 664-671, 2007.
108. Karypis, G., & Kumar, V. A fast and high quality multilevel scheme for

- partitioning irregular graphs, *SIAM J. Scient. Comput.* **20**, 359-392, 1998.
109. Karypis, G. et al. Multilevel hypergraph partitioning: application in VLSI domain, in Proceedings 34th Annual Conference on Design Automation (DAC'97), ACM Press, New York, NY, USA, 526-529.
110. Abdala, D., Wattuya, P., & Jiang, X. Ensemble clustering via random walker consensus strategy, in 20th International Conference on Pattern Recognition, (ICPR 2010), IEEE, Istanbul, 1433 - 1436.
111. Wattuya, P. et al. A random walker based approach to combining multiple segmentations, in Proceedings of the 19th International Conference on Pattern Recognition (ICPR 2008), IEEE, Tampa, FL, 1-4.
112. Filkov, V., & Skiena, S. Integrating microarray data by consensus clustering, *International Journal of Artificial Intelligence Tools* **4** (13), 863-880, 2004.
113. Gionis, A., Mannila, H., & Tsaparas, P. Clustering aggregation, *ACM Transactions Knowledge Discovery from Data* **1** (1), 341-352, 2007.
114. Alion, N., Charikar, M., & Newman, A. Aggregating inconsistent information: ranking and clustering, *37th Symp. Theory of Computing (STOC)*, 684-693, 2005.
115. McLachlan, G., & Peel, D. *Finite Mixture Models*, John Wiley & Sons, New York, 2000.
116. Topchy, A., Jain, A. K., & Punch, W. A mixture model of clustering ensembles, in Proceedings of the SIAM International Conference on Data Mining, Lake Buena Vista, Florida, 379-390.
117. Yoon, H. S. et al. Heterogeneous clustering ensemble method for combining different cluster results, *Data Mining for Biomedical Applications, Lecture Notes in Computer Science* **3916**, 82-92, 2006.
118. Yoon, H.-S. et al. A novel framework for discovering robust cluster results, *Discovery Science, Lecture Notes in Computer Science* **4265**, 373-377, 2006.
119. Analoui, M., & Sadighian, N. Solving cluster ensemble problems by correlation's matrix & GA, *IFIP International Federation for Information Processing* **228** (1), 227-231, 2006.
120. Domeniconi, C. et al. Locally adaptive metrics for clustering high dimensional

- data, *Data Mining and Knowledge Discovery* **14**, 63-97, 2007.
121. Domeniconi, C., & Al-Razgan, M. Weighted cluster ensembles: methods and analysis, *ACM Transactions on Knowledge Discovery from Data (TKDD)* **2** (4), 1-17, 2009.
122. Scholkopf, B., & Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, 2002.
123. Vega-Pons, S., Correa-Morris, J., & Ruiz-Shulcloper, J. Clustering ensemble method for heterogeneous partitions, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Lecture Notes in Computer Science* **5856**, 481-488, 2009.
124. Gasch, P., & Eisen, M. Exploring the conditional regulation of yeast gene expression through fuzzy k-means clustering, *Genome Biology* **3** (11), 2002.
125. Avogadri, R., & Valentini, G. Fuzzy ensemble clustering based on random projections for DNA microarray data analysis, *Artificial Intelligence in Medicine* **45** (2-3), 173-183, 2009.
126. Dyrskjøt, L. et al. Identifying distinct classes of bladder carcinoma using microarrays, *Nature Genetics* **33**, 90-96, 2003.
127. Onken, M. et al. Gene Expression Profiling in Uveal Melanoma Reveals Two Molecular Classes and Predicts Metastatic Death, *Cancer Research* **64**, 7205-7209, 2004.
128. Dopazo, J. Functional Interpretation of Microarray Experiments, *OMICS: A Journal of Integrative Biology* **10** (3), 398-410, 2006.
129. Monti, S. et al. Consensus Clustering: A Resampling-based Method for Class Discovery and Visualization of Gene Expression Microarray Data, *Machine Learning* **52** (1-2), 91-118, 2003.
130. Grotkjaer, T. et al. Robust multi-scale clustering of large DNA microarray data sets with the consensus algorithm, *Bioinformatics* **22**, 58-67, 2006.
131. Bertoni, A., & Valentini, G. Randomized Embedding Cluster Ensembles for gene expression data analysis, in In SETIT 2007-IEEE International Conf. on Sciences of Electronic, Technologies of Information and Telecommunications, Hammamet, Tunisia.

132. Yu, Z., Wong, H. S., & Wang, H. Graph-based consensus clustering for class discovery from gene expression data, *Bioinformatics* **23** (21), 2888-2896, 2007.
133. Chapelle, O., Schölkopf, B., & Zien, A. *Semi-Supervised Learning*, MIT Press, Cambridge, Massachusetts, USA, 2006.
134. Zhu, X., & Goldberg, A. *Introduction to Semi-Supervised Learning*, Morgan & Claypool, 2009.
135. Yarowsky, D. Unsupervised word sense disambiguation rivaling supervised methods, in Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL '95), Association for Computational Linguistics Stroudsburg, PA, USA, 189-196.
136. Riloff, E., Weibe, J., & Wilson, T. Learning subjective nouns using extraction pattern bootstrapping, in Proceedings of the Seventh Conference on Natural Language Learning (CONLL '03), Association for Computational Linguistics Stroudsburg, PA, USA, 25-32.
137. Maeireizo, B., Litman, D., & Hwa, R. Co-training for predicting emotions with spoken dialogue data, in Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004), Association for Computational Linguistics, Stroudsburg, PA, USA.
138. Rosenberg, C., Hebert, M., & Schneiderman, H. Semi-supervised self-training of object detection models, in Proceedings of the 7th IEEE Workshop on Applications of ComputerVision (WACV/MOTIONS '05), IEEE, Breckenridge, CO, 29-36.
139. Castelli, V., & Cover, T. On the exponential value of labeled samples, *Pattern Recognition Letters* **16**, 105-111, 1995.
140. Castelli, V., & Cover, T. The relative value of labeled and unlabeled samples in pattern recognition with an unknown mixing parameter, *IEEE Transactions Information Theory* **42** (6), 2101-2117, 1996.
141. Ratsaby, J., & Venkatesh, S. Learning from a mixture of labeled and unlabeled examples with parametric side information, in Proceedings of the 8th Annual Conference on Computational Learning Theory (COLT '95), ACM Press, New York, NY, 412-417.

142. Nigam, K. et al. Text classification from labeled and unlabeled documents using EM, *Machine Learning - Special issue on information retrieval* **39** (2-3), 103-134, 2000.
143. Nigam, K. Using unlabeled data to improve text classification, in *Technical Report doctoral dissertation, CMU-CS-01-126*, Carnegie Mellon University, Pittsburgh, 2001.
144. Fujino, A., Ueda, N., & Saito, K. A hybrid generative/discriminative approach to semi-supervised classifier design, in The Twentieth National Conference on Artificial Intelligence (NCAI-05), AAAI Press, Pennsylvania, 764-769.
145. Dara, R., Kremer, S., & Stacey, D. Clustering unlabeled data with SOMs improves classification of labeled real-world data, in Proceedings of the World Congress on Computational Intelligence (WCCI '02), IEEE, Honolulu, HI, 2237 - 2242.
146. Goldberg, A. et al. Multi-manifold semi-supervised learning, *Journal of Machine Learning Research : Workshop and Conference Proceedings* **5**, 169-176, 2009.
147. Blum, A., & Mitchell, T. Combining labeled and unlabeled data with co-training, in COLT: Proceedings of the Workshop on Computational Learning Theory, Morgan Kaufmann Publishers, Morgan Kaufmann, San Francisco, 92-100.
148. Mitchell, T. The role of unlabeled data in supervised learning, in Proceedings of the Sixth International Colloquium on Cognitive Science, Kluwer Academic Press, San Sebastian, Spain.
149. Zhu, X. *Semi-supervised learning literature survey*, Madison, 2008.
150. Nigam, K., & Ghani, R. Analyzing the effectiveness and applicability of co-training, in Proceedings of 9th International Conference on Information and Knowledge Management, ACM, New York, 86-93.
151. Jones, R. Learning to extract entities from labeled and unlabeled text, in *PhD Thesis*, Carnegie Mellon University Doctoral Dissertation, Carnegie Mellon University, Pittsburgh, 2005.
152. Balcan, M.-F., & Blum, A. An augmented pac model for semi-supervised

- learning, in *Semi-Supervised Learning*, edited by Chapelle, O., Scholkopf, B., & Zien, A., MIT Press Scholarship Online, 2006.
153. Zhou, Z. H., & Xu, J. M. On the relation between multi-instance learning and semi-supervised learning, in Proceedings of the 24th International Conference on Machine Learning (ICML '07), ACM New York, NY, USA, 1167-1174.
154. Goldman, S., & Zhou, Y. Enhancing supervised learning with unlabeled data, in Proc. 17th International Conference on Machine Learning (ICML '00), Morgan Kauffmann, San Francisco, CA, 327-334.
155. Zhou, Y., & Goldman, S. Democratic co-learning, in Proceedings of the 16th International Conference on Tools with Artificial Intelligence (ICTAI 2004), IEEE, 594 - 602.
156. Zhou, Z. H., & Li, M. Tri-training: exploiting unlabeled data using three classifiers, *IEEE Transactions on Knowledge and Data Engineering* **17** (11), 1529-1541, 2005.
157. Johnson, R., & Zhang, T. Two-view feature generation model for semi-supervised learning, in The 24th International Conference on Machine Learning (ICML '07), ACM New York, NY, 25-32.
158. Li, T. et al. Gene functional classification by semi-supervised learning from heterogeneous data, in Proceedings of the 2003 ACM Symposium on Applied Computing (SAC'03), Melbourne, FL, USA, 78-82.
159. Qi, Y. et al. Semi-supervised analysis of gene expression profiles for lineage-specific development in the *Caenorhabditis elegans* embryo, *Bioinformatics* **22** (14), 417-423, 2006.
160. Sindhwani, V., Niyogi, P., & Belkin, M. A co-regularized approach to semi-supervised learning with multiple views, in Proceedings of the 22nd ICML Workshop on Learning with Multiple Views.
161. Brefeld, U. et al. Efficient co-regularized least squares regression, in 23rd International Conference on Machine Learning (ICML'06), ACM Press, Pittsburgh, USA, 137-144.
162. Brefeld, U., Biischer, C., & Scheffer, T. Multiview discriminative sequential learning, in Proceedings of the European Conference on Machine Learning

- (ECML '05), Springer, 60-71.
163. Brefeld, U., & Scheffer, T. Semi-supervised learning for structured output variables, in 23rd International Conference on Machine Learning (ICML'06), ACM New York, NY, Pittsburgh, USA, 145-152.
164. Vapnik, V. *Statistical Learning Theory*, Wiley-Interscience, 1998.
165. Shi, M., & Zhang, B. Semi-supervised learning improves gene expression-based prediction of cancer recurrence, *Bioinformatics* **27** (21), 3017-3023, 2011.
166. Wang, J., Shen, X., & Pan, W. On Efficient Large Margin Semisupervised Learning: Method and Theory, *Journal of Machine Learning Research* **10**, 719-742, 2009.
167. Kawano, S., Misumi, T., & Konishi, S. Semi-Supervised Logistic Discrimination Via Graph-Based Regularization, *Neural Process Letters* **36**, 203-216, 2012.
168. Blum, A., & Chawla, S. Learning from labeled and unlabeled data using graph mincuts, in Proceedings of the 18th International Conference on Machine Learning (ICML '01), Morgan Kaufmann Publishers Inc. , San Francisco, CA, USA, 19-26.
169. Blum, A. et al. Semi-supervised learning using randomized mincuts, in 21st International Conference on Machine Learning (ICML '04), ACM, New York, NY, USA , Banff, Canada, 13.
170. Pang, B., & Lee, L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL '04), Stroudsburg, PA, USA, 271-278.
171. Grady, L., & Funka-Lea, G. Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials, in Proceedings of the 8th ECVV Workshop on Computer Vision Approaches to Medical Image Analysis and Mathematical Methods in Biomedical Image Analysis 2004, Prague, Czech Republic, 230-245.
172. Belkin, M., Niyogi, P., & Sindhvani, V. *Manifold regularization: A geometric framework for learning from examples (Technical Report TR-2004-06)*,

- University of Chicago, 2004.
173. Belkin, M., Niyogi, P., & Sindhwani, V. On Manifold Regularization, in Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTAT 2005).
174. Wang, Y., & Pan, Y. Semi-supervised consensus clustering for gene expression data analysis, *Journal of BioData Mining* **7** (1), 2014.
175. Nguyen, T. P., & Ho, T. B. A Semi-supervised Learning Approach to Disease Gene Prediction, in Proceedings of the International Conference on Bioinformatics and Biomedicine (BIBM 2007), IEEE, Fremont, CA, 423-428.
176. Nguyen, T. P., & Ho, T. B. Detecting disease genes based on semi-supervised learning and protein-protein interaction networks, *Artificial Intelligence in Medicine* **54** (1), 63-71, 2012.
177. Weston, J. et al. Semi-supervised protein classification using cluster kernels, *Bioinformatics* **21**, 3241-3247, 2005.
178. Kall, L. et al. Semi-supervised learning for peptide identification from shotgun proteomics datasets, *Nature Methods* **4**, 923-925, 2007.
179. Ernst, J. et al. A semi-supervised method for predicting transcription factor-gene interactions in Escherichia coli, *Plos Computing Biology* **4** (3), e1000044, 2008.
180. Bair, E., & Tibshirani, R. Semi-supervised methods to predict patient survival from gene expression data, *Plos Biology* **2**, 511-522, 2004.
181. Koestler, D. C. et al. Semi-supervised recursively partitioned mixture models for identifying cancer subtypes, *Bioinformatics* **26**, 2578-2585, 2010.
182. Steinfeld, I. et al. Clinically driven semi-supervised class discovery in gene expression data, *Bioinformatics* **24**, 190-197, 2008.
183. Kuncheva, L., & Whitaker, C. Measures of diversity in classifier ensembles and their relationship with ensemble accuracy, *Machine Learning* **2** (51), 181-207, 2003.
184. McLachlan, G. J., Do, K.-A., & Ambrose, C. *Analyzing microarray gene expression data*, Wiley, 2004.

185. Yeung, K. Y., Haynor, D. R., & Ruzzo, W. L. Validating clustering for gene expression data, *Bioinformatics* **17** (4), 309-318, 2001.
186. Halkidi, M. On clustering validation techniques, *Journal of Intelligent Information Systems* **17**, 107-145, 2001.
187. Santos, J. M., & Embrechts, M. On the Use of the Adjusted Rand Index as a Metric for Evaluating Supervised Classification, in Proceedings of the 19th International Conference on Artificial Neural Networks: Part II (ICANN '09), Springer-Verlag Berlin, Heidelberg, 175-184.
188. Hubert, L. J., & Arabie, P. Comparing partitions, *Journal of Classification* **2** (1), 193-218, 1985.
189. Batistakis, Y., Halkidi, M., & Vazirgiannis, M. Cluster validity methods: Part I, *Sigmod Record* **31** (12), 2002.
190. Fowlkes, E., & Mallows, C. A method for comparing two hierarchical clusterings, *Journal of the American Statistical Association* **78** (383), 553-569, 1983.
191. van Rijsbergen, C. J. *Information Retrieval*, Butterworth-Heinemann, Newton, MA, USA, 1979.
192. Gibbons, F., & Roth, F. Judging the Quality of Gene Expression Based Clustering Methods using Gene Annotation, *Genome Research* **12**, 1574–1581, 2002.
193. Handl et al. Computational Cluster Validation in Post-Genomic Data Analysis, *Bioinformatics* **21** (15), 3201-3212, 2005.
194. Sharan, R., Maron-Katz, A., & Shamir, R. CLICK and EXPANDER: A System for Clustering and Visualizing Gene Expression Data, *Bioinformatics* **19** (14), 1787–1799, 2003.
195. Rousseeuw, P. J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis, *Journal of Computational and Applied Mathematics* **20**, 153-165, 1987.
196. Davies, D., & Bouldin, D. A cluster separation measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1** (2), 1979.
197. Dunn, J. C. Well Separated Clusters and Fuzzy Partitions, *Journal on*

- Cybernetics* **4**, 95-104, 1974.
198. Tavazoie, S. et al. Systematic Determination of Genetic Network Architecture, *Nature Genetics* **22** (3), 281-285, 1999.
199. Wagner, A. How the global structure of protein interaction networks evolves, *The Royal Society of Biological Sciences* **270**, 457–466, 2003.
200. Nepusz, T., Yu, H., & Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks, *Nature Methods* **9** (5), 471-472, 2012.
201. Lin, C. et al. Clustering Methods in a Protein–Protein Interaction Network, in *Knowledge Discovery in Bioinformatics: Techniques, Methods and Application*, edited by Hu, X., & Pan, Y., John Wiley & Sons, Inc, 2006, 319-355.
202. Palzkill, T. *Proteomics*, Springer 2002 edition, 2002.
203. Waksman, G. *Protein Reviews*, Springer, 2005.
204. Peink, J., & Alber, M. Improved multifractal box-counting algorithm, virtual phase transitions, and negative dimensions, *Physical Review E*. **57** (5), 5489-5493, 1998.
205. Pellegrini, M. et al. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles, *Proceedings of National Academy of Sciences, USA* **96** (8), 4285-4288, 1999.
206. Qi, Y., Klein-Seetharaman, J., & Bar-Joseph, Z. A mixture of feature experts approach for protein-protein interaction prediction, *BMC Bioinformatics* **8** (Suppl 10: S6), 2007.
207. Rao, M. A., & Srinivas, J. *Neural Networks: Algorithms and Applications*, Alpha Science International Ltd, 2003.
208. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. Learning representations by back-propagating errors, *Nature* **323**, 533-536, 1986.
209. Hu, X. H., & Pan, Y. *Knowledge Discovery in Bioinformatics: Techniques, Methods, and Applications*, Wiley, 2007.
210. King, A., Przulj, N., & Jurisica, I. Protein complex prediction via cost-based clustering, *Bioinformatics* **20** (17), 3013–3020, 2004.
211. Blatt, M., Wiseman, S., & Domany, E. Superparamagnetic clustering of data,

- Physics Review Letters* **76** (18), 3251-3254, 1996.
212. Girvan, M., & Newman, M. Community Structure in social and biological networks, *Proceedings of the National Academy of Sciences, USA (PNAS)* **99** (12), 7821-7826, 2002.
213. Newman, M. E. J. Fast algorithm for detecting community structure in networks, *Physical Review E* **69** (6), 66-133, 2004.
214. Cho, Y. R. et al. Semantic integration to identify overlapping functional modules in protein interaction networks, *BMC Bioinformatics* **8** (265), 1-13, 2007.
215. Bader, G., & Hogue, C. An automated method for finding molecular complexes in large protein interaction networks, *BMC Bioinformatics* **4** (2), 2003.
216. Altaf-Ul-Amin, M. et al. Development and implementation of an algorithm for detection of protein complexes in large interaction networks, *BMC Bioinformatics* **7**, 207, 2006.
217. Macropol, K., Can, T., & Singh, A. K. RRW: repeated random walks on genome-scale protein networks for local cluster discovery, *BMC Bioinformatics* **10**, 283, 2009.
218. van Dongen, S. *Graph Clustering by Flow Simulation*, University of Utrecht, 2000.
219. Palla, G. et al. Uncovering the overlapping community structure of complex networks in nature and society, *Nature* **435** (7043), 814-818, 2005.
220. Adamcsek, B. et al. CFinder: locating cliques and overlapping modules in biological networks, *Bioinformatics* **22** (8), 1021–1023, 2006.
221. Li, X. L., Foo, C. S., & Ng, S. K. Discovering protein complexes in dense reliable neighborhoods of protein interaction networks, in Proceedings of Computational Systems Bioinformatics Conference (CSB'07), IEEE, 157–168.
222. Wu, M. et al. A Core-Attachment based Method to Detect Protein Complexes in PPI Networks, *BMC Bioinformatics* **10**, 169, 2009.
223. Jung, S. H. et al. Protein complex prediction based on simultaneous protein interaction network, *Bioinformatics* **26** (3), 385-391, 2010.

224. Ozawa, Y. et al. Protein complex prediction via verifying and reconstructing the topology of domain-domain interactions, *BMC Bioinformatics* **11** (350), 2010.
225. Cannataro, M., Guzzi, P. H., & Veltri, P. IMPRECO: Distributed prediction of protein complexes, Future Generation Computer Systems, *The International Journal of Grid Computing-Theory Methods and Applications* **26** (3), 434-440, 2010.
226. Brohee, S., & van Helden, J. Evaluation of clustering algorithms for protein-protein interaction networks, *BMC Bioinformatics* **7** (488), 2008.
227. Vlasblom, J., & Wodak, S. Markov Clustering versus Affinity Propagation for the partitioning of protein interaction graphs, *BMC Bioinformatics* **10** (99), 2009.
228. Li, X. L. et al. Computational approaches for detecting protein complexes from protein interaction networks: A Survey, *BMC Genomics* **11**(Suppl 1) (S3), 2010.
229. Srihari, S., & Leong, H. W. A survey of computational methods for protein complex prediction from protein interaction networks, *Journal of Bioinformatics Computational Biology* **11** (1230002), 2013.
230. Liu, G., Wong, L., & Chua, H. N. Complex discovery from weighted PPI networks, *Bioinformatics* **25** (15), 1891-1897, 2009.
231. Nagi, S., Bhattacharyya, D. K., & Kalita, J. K. Subspace Clustering in Gene Expression Data Analysis: A Survey, in in *Machine Intelligence: Recent Advances*, Narosa Publishing, Delhi, Tezpur University, 211-219.
232. Ahmed, H. et al. Module Extraction from Subspace Co-expression Networks, *Network Modeling Analysis in Health Informatics and Bioinformatics* **1** (4), 183-195, 2012.
233. Kiliç, C., & Tan, M. Positive unlabeled learning for deriving protein interaction networks, *Network Modeling Analysis in Health Informatics and Bioinformatics* **1** (3), 87-102, 2012.
234. Boström, H., Johansson, R., & Karlsson, A. On evidential combination rules for ensemble classifiers, in 11th International Conference on Information Fusion (ICIF '08), IEEE, Cologne, 1-8.
235. Polikar, R. Ensemble learning, *Scholarpedia* **1** (4), 2776, 2008.

236. Ho, T. K. The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20** (8), 832-844, 1998.
237. Skurichina, M., & Duin, R. P. W. Bagging, Boosting and the Random Subspace Method for Linear Classifiers, *Pattern Analysis and Applications* **5**, 121–135, 2002.
238. Tax, D. M. J., Duin, R. P. W., & van Breukelen, M. Comparison Between Product and Mean Classifier Combination Rules, in Proceedings 1st International Workshop on Statistical Pattern Recognition (IAPR '97), 165-170.
239. Duin, R. P. W., & Tax, D. M. J. Experiments with Classifier Combining Rules, in *Multiple Classifier Systems (Proc. First International Workshop, MCS 2000, Cagliari, Italy, June 2000)*, edited by Kittler, J., & Roli, F., Lecture Notes in Computer Science, Berlin, 2000, 16-29.
240. Kittler, J. et al. On Combining Classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20** (3), 226-239, 1998.
241. Tax, D. M. J. et al. Combining multiple classifiers by averaging or by multiplying?, *Pattern Recognition* **33** (9), 1475–1485, 2000.
242. Tax, D. M. J., & Duin, R. P. W. Using two-class classifiers for multiclass classification, in Proceedings of International Conference on Pattern Recognition (ICPR2002), IEEE, Quebec City, Canada.
243. Jain, A. K., Duin, R. P. W., & Mao, J. Statistical Pattern Recognition: A Review, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22** (1), 4-37, 2000.
244. Tsymbal, A., Pechenizkiy, M., & Cunningham, P. Diversity in search strategies for ensemble feature selection, *Information Fusion* **6**, 83-98, 2005.
245. Statnikov, A., Wang, L., & Aliferis, C. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification, *BMC Bioinformatics* **9**, 319, 2008.
246. Curtis, R. K., Oresic, M., & Vidal-Puig, A. Pathways to the analysis of microarray data, *TRENDS in Biotechnology* **23** (8), 429-435, 2005.
247. Bertoni, A., Folgieri, R., & Valentini, G. Bio-molecular cancer prediction with

- random subspace ensembles of Support Vector Machines, *Neurocomputing* **63** (C), 535-539, 2005.
248. Re, M., & Valentini, G. Simple ensemble methods are competitive with state-of-the-art data integration methods for gene function prediction, *Machine Learning in Systems Biology* **8**, 98-111, 2010.
249. Guan, Y. et al. Predicting gene function in a hierarchical context with an ensemble of classifiers, *Genome Biology* **S2** (9), 2008.
250. Cesa-Bianchi, N., Re, M., & Valentini, G. Synergy of multi-label hierarchical ensembles, data fusion, and cost-sensitive methods for gene functional inference, *Machine Learning* **1** (88), 209-241, 2012.
251. Brown, M. et al. Knowledge-based analysis of microarray gene expression data by using support vector machines, *Proceedings of National Academy of Science, USA* **97** (1), 262–267, 2000.
252. Guyon, I. et al. Gene selection for cancer classification using support vector machines, *Machine Learning* **46** (1-3), 389–422, 2002.
253. Yeang, C. et al. Molecular classification of multiple tumor types, *Bioinformatics* **17** (Suppl-1), 316–322, 2001.
254. Li, J., & Liu, H. Ensembles of cascading trees, in 3rd IEEE International Conference on Data Mining (ICDM'03), IEEE, Florida, 585–588.
255. Li, J. et al. Discovery of significant rules for classifying cancer diagnosis data, *Bioinformatics* **19** (Suppl 2ii), 93–102, 2003.
256. Abbott, D. W. Comparison of data analysis and classification algorithms for automatic target recognition, in IEEE International Conference on Systems, Man and Cybernetics (ICSMC 1994), IEEE, San Antonio, TX, 902 - 907.
257. Dettling, M. BagBoosting for tumor classification with gene expression data, *Bioinformatics* **20** (18), 3583–3593, 2004.
258. Li, J., & Liu, H. Kent ridge bio-medical dataset repository, Available at <http://levis.tongji.edu.cn/gzli/data/mirror-kentridge.html>, 2002.
259. The Gene Ontology Consortium Creating the Gene Ontology resource: design and implementation, *Genome Res* **11** (8), 1425-33, 2001.

260. DeRisi, J. L., Iyer, V. R., & Brown, P. O. Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale, *Science* **278** (5338), 680–686, 1997.
261. Chu, S. et al. The Transcriptional Program of Sporulation in Budding Yeast, *Science* **282** (5389), 699–705, 1998.
262. Eisen, M. B. et al. Cluster analysis and display of genome-wide expression patterns, *Proceedings of the National Academy of Sciences* **95**, 14863-14868, 1998.
263. Bhattacharjee, V. et al. Neural Crest and Mesoderm Lineage-Dependent Gene Expression in Orofacial Development, *Differentiation* **75** (5), 463–477, 2007.
264. Acharya, S., Thadisina, Y., & Saha, S. Multi-objective clustering of tissue samples for cancer diagnosis, in International Conference on Advances in Computing, Communications and Informatics (ICACCI, 2014), IEEE, New Delhi, 1059-1064.
265. Lord, P. W. et al. Investigating Semantic Similarity Measures across the Gene Ontology: The Relationship between Sequence and Annotation, *Bioinformatics* **19**, 1275-1283, 2003.
266. Couto, F. M., Silva, M. J., & Coutinho, P. *Implementation of a Functional Semantic Similarity Measure between Gene-Products*, Department of Informatics, University of Lisbon, Lisbon, 2003.
267. Azuaje, F., & Bodenreider, O. Incorporating Ontology-Driven Similarity Knowledge into Functional Genomics: An Exploratory Study, in Proceedings of IEEE Fourth Symposium on Bioinformatics and Bioengineering (BIBE 2004), IEEE, Taichung, Taiwan, 317 - 324.
268. Sevilla, J. L. et al. Correlation between Gene Expression and GO Semantic Similarity, *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* **2** (4), 330-337, 2005.
269. Hartigan, J. A., & Wong, M. A. A K-means Clustering Algorithm, *Applied Statistics* **28**, 100–108, 1979.
270. Dopazo, J., & Carazo, J. M. Phylogenetic Reconstruction using a Growing Neural Network that Adopts the Topology of a Phylogenetic Tree, *Journal of*

- Molecular Evolution*, 226–233, 1997.
271. Herrero, J., Valencia, A., & Dopazo, J. A Hierarchical Unsupervised Growing Neural Network for Clustering Gene Expression Patterns, *Bioinformatics* **17** (2), 126–36, 2001.
272. Fraley, C., & Raftery, A. E. Model-based Clustering, Discriminant Analysis, and Density Estimation, *Journal of the American Statistical Association* **17**, 126–136, 2001.
273. Wang, H., Azuaje, F., & Bodenreider, O. An ontology-driven clustering method for supporting gene expression analysis, Computer-based Medical Systems, in Proceedings of the 18th Symposium on Computer-based Medical Systems, IEEE, 389 - 394.
274. Lord, P. W. et al. Semantic similarity measures as tools for exploring the gene ontology, *Pacific Symposium on Biocomputing, PubMed* **8**, 601-612, 2003.
275. Schlicker, A. et al. A new measure for functional similarity of gene products based on gene ontology, *BMC Bioinformatics* **7**, 302, 2006.
276. Rada, R. et al. Development and application of a metric on semantic nets, *IEEE Transaction on Systems, Man and Cybernetics* **19** (1), 17-30, 1989.
277. Wu, Z., & Palmer, M. S. Verb semantics and lexical selection, in Proceedings of the 32nd. Annual Meeting of the Association for Computational Linguistics (ACL 1994), Association for Computational Linguistics Stroudsburg, PA, USA, 133-138.
278. Pekar, V., & Staab, S. Taxonomy learning: factoring the structure of a taxonomy into a semantic classification decision, in Proceedings of the 19th international conference on Computational linguistics (COLING '02), Morristown, NJ, USA, Taipei, Taiwan, 786-792.
279. Yu, H. et al. Broadly predicting specific gene functions with expression similarity and taxonomy similarity, *Gene* **352**, 75-81, 2005.
280. Cheng, J. et al. A knowledge-based clustering algorithm driven by gene ontology, *Journal of Biopharmaceutical Statistics* **14**, 687-700, 2004.
281. Wu, H. et al. Prediction of functional modules based on comparative genome analysis and gene ontology application, *Nucleic Acids Research* **33**, 2822-2837,

- 2005.
282. Wu, X. et al. Prediction of yeast proteinprotein interaction network: insights from the gene ontology and annotations, *Nucleic Acids Research* **34**, 2137-2150, 2006.
283. Couto, F. M., Silva, M. J., & Coutinho, P. Measuring semantic similarity between gene ontology terms, *Data Knowl Eng* **61**, 137-152, 2007.
284. Resnick, P. Using information content to evaluate semantic similarity in a taxonomy, in Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI '95), Morgan Kaufmann Publishers Inc. San Francisco, CA, Montreal, 448-453.
285. Couto, F. M., Silva, M. J., & Coutinho, P. Semantic similarity over the gene ontology: Family correlation and selecting disjunctive ancestors, in Proc. of the ACM Conference in Information and Knowledge Management (CIKM '05), ACM New York, NY, 343-344.
286. Lin, D. An information-theoretic definition of similarity, in Proc. of the 15th International Conference on Machine Learning, Morgan Kaufmann, San Francisco, CA, 296-304.
287. Jiang, J., & Conrath, D. Semantic similarity based on corpus statistics and lexical taxonomy, in Proceedings of the 10th International Conference on Research on Computational Linguistics, Taiwan, 19-33.
288. Wang, J. Z. et al. A new method to measure the semantic similarity of GO terms, *Bioinformatics* **23** (10), 2007.
289. Othman, R., Deris, S., & Ilias, R. A genetic similarity algorithm for searching the gene ontology terms and annotating anonymous protein sequences, *Journal of Biomedical Information* **23**, 529-538, 2007.
290. Yu, G. et al. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products, *Bioinformatics* **26** (7), 976-978, 2010.
291. Zheng, H., Azuaje, F., & Wang, H. seGOsa: software environment for Gene Ontology-driven similarity assessment, in Proceedings of the International Conference on Bioinformatics and Biomedicine (BIBM '10), IEEE , 539-542.
292. Li, J. et al. DOSim: an R package for similarity between diseases based on

- disease ontology, *BMC Bioinformatics* **12** (266), 2011.
293. Resnick, P. Semantic Similarity in a Taxonomy: An Information Based Measure and Its Application to Problems of Ambiguity in Natural Language, *Journal of Artificial Intelligence Research* **11**, 95-130, 1999.
294. Needleman, S. B., & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins, *Journal of Molecular Biology* **48**, 443-453, 1970.
295. Smith, T. F., & Waterman, M. S. Identification of common molecular subsequences, *Journal of Molecular Biology* **147**, 195-197, 1981.
296. Altschul, S. F. et al. Basic Local Alignment Search Tool, *Journal of Molecular Biology* **215**, 403-410, 1990.
297. Naldi, M. C., Carvalho, A. C. P. L. F., & Campello, R. J. G. B. Cluster ensemble selection based on relative validity indexes, *Journal of Data Mining and Knowledge Discovery - Springer* (2), 259-289, 2013.
298. Datta, S., & Datta, S. Comparisons and validation of statistical clustering techniques for microarray gene expression data, *Bioinformatics* **19** (4), 459-466, 2003.
299. Datta, S., & Datta, S. Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes, *BMC Bioinformatics* **7** (397), 2006.
300. Nagi, S., & Bhattacharyya, D. K. Classification of microarray cancer data using ensemble approach, *Network Modeling Analysis in Health* **2** (3), 159-173, 2013.
301. Gentleman, R. C., Carey, V. J., & Bates, D. M. Bioconductor: open software development for computational biology and bioinformatics, *Genome Biology* **5** (10), R80, 2004.
302. Larkin, M. et al. Clustal W and Clustal X Version 2.0, *Bioinformatics* **23** (21), 2947-2948, 2007.
303. Wang, H. et al. Gene Expression Correlation and Gene Ontology-Based Similarity: An Assessment of Quantitative Relationships, in Proceedings of the Computational Intelligence in Bioinformatics and Computational Biology (CIBCB '04), IEEE, 25-31.

304. van 't Veer, L. J. Gene expression profiling predicts clinical outcome of breast cancer, *Nature* **415**, 530-536, 2002.
305. Jaskowiak, P. A., Campello, R. J. G. B., & Costa, I. G. Proximity Measures for Clustering Gene Expression Microarray Data: A Validation Methodology and a Comparative Analysis, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **10** (4), 845-857, 2013.
306. Su, A. I. et al. Large-Scale Analysis of the Human and Mouse Transcriptomes, *Proceedings of the National Academy of Sciences, USA* **99** (7), 4465-4470, 2002.
307. Lee, H. K. et al. Coexpression Analysis of Human Genes across Many Microarray Data Sets, *Genome Research* **14**, 1085-1094, 2004.
308. Stuart, J. M. et al. A Gene- Coexpression Network for Global Discovery of Conserved Genetic Modules, *Science* **302** (5643), 249-255, 2003.
309. Chenna, R. et al. Multiple Sequence Alignment with Clustal Series of Programs, *Nucleic Acids Research* **31** (13), 3497-3500, 2003.
310. Hartwell, L. H. et al. Clustering algorithm based graph connectivity, *Nature* **402**, C47-C52, 1999.
311. Pereira, J. B., Enright, A. J., & Ouzounis, C. A. Detection of functional modules from protein interaction networks, *Proteins: Structure, Function, and Bioinformatics* **20**, 49-57, 2004.
312. Tornow, S., & Mewes, H. W. Functional modules by relating protein interaction networks and gene expression, *Nucleic Acids Research* **31** (21), 6283-6289, 2003.
313. Hung, I. H. et al. Biochemical characterization of the Wilson disease protein and functional expression in the yeast *Saccharomyces cerevisiae*, *Journal of Biological Chemistry* **272**, 21461-21466, 1997.
314. Hung, M. C., & Link, W. Protein localization in disease and therapy, *Journal of Cell Science* **124**, 3381-3392, 2011.
315. Payne, A. S., Kelly, E. J., & Gitlin, J. D. Functional expression of the Wilson disease protein reveals mislocalization and impaired copper-dependent trafficking of the common H1069Q mutation, *Proceedings in National Academy of Sciences, USA* **95**, 10854-10859, 1998.

316. Tanaka, A. R. et al. Effects of mutations of ABCA1 in the first extracellular domain on subcellular trafficking and ATP binding/hydrolysis, *Journal of Biological Chemistry* **278**, 8815-8819, 2003.
317. Rives, A. W., & Galitski, T. Modular organization of cellular networks, *Proceedings of the National Academy of Sciences, USA* **100** (3), 1128-1133, 2003.
318. Arnau, V., Mars, S., & Marin, I. Iterative cluster analysis of protein interaction data, *Bioinformatics* **21** (3), 364-378, 2005.
319. Pei, P., & Zhang, A. A two-step approach for clustering proteins based on protein interaction profiles, in In IEEE International Symposium on Bioinformatics and Bioengineering (BIBE'2005), IEEE, 201-209.
320. Lubovac, Z., Gamalielsson, J., & Olsson, B. Combining functional and topological properties to identify core , *Proteins: Structure, Function, and Bioinformatics* **64**, 948-959, 2006.
321. Frey, B. J., & Dueck, D. Clustering by passing messages between data points, *Science* **315**, 972-976, 2007.
322. Chua, H. N. et al. Using indirect protein-protein interactions for protein, in Proceedings of Computational Systems Bioinformatics Conference (CSB'07), IEEE, 97-109.
323. Georgii, E. et al. Enumeration of condition-dependent dense modules in protein interaction networks, *Bioinformatics* **25** (7), 933-940, 2009.
324. Ruan, J., & Zhang, W. Identifying network communities with a high resolution, *Physical Review E* **77** (1), 016104, 2008.
325. Kovacs, I. A. et al. Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics, *PLoS ONE* **5** (9), 2010.
326. Cho, Y. -R., Hwang, & Zhang, A. Identification of overlapping functional modules in protein interaction networks: Information flow-based approach, in Proceedings of the Sixth International Conference on Data Mining-Workshops (ICDMW'06), IEEE, 147-152.
327. Hwang, W. et al. A novel functional module detection algorithm for protein-

- protein interaction networks, *Algorithms for Molecular Biology* **1** (24), 2006.
328. Mewes, H. W. et al. MIPS: Analysis and Annotation of Proteins from whole Genomes, *Nucleic Acids Research* **32**, 41-44, 2004.
329. Leung, H. C. et al. Predicting Protein Complexes from PPI Data: A Core-Attachment Approach, *Journal of Computational Biology* **16** (2), 133-144, 2009.
330. Xenarios, I. et al. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions, *Nucleic Acids Research* **30**, 303-305, 2002.
331. Krogan, N. et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*, *Nature* **440** (7084), 637-643, 2006.
332. Pereira-Leal, J. B., Enright, A. J., & Ouzounis, C. A. Detection of functional modules from protein interaction networks, *PROTEINS Structure Function and Bioinformatics* **54**, 49-57, 2004.
333. Huh, W. K. et al. Global analysis of protein localization in budding yeast, *Nature* **425**, 686-691, 2003.
334. Coleman, T. F., & More, J. J. Estimation of sparse Jacobian matrices and graph coloring problems, *SIAM Journal on Numerical Analysis* **20**, 187-209, 1983.
335. Hong, E. L. et al. Gene Ontology annotations at SGD: new data sources and annotation methods, *Nucleic Acids Research* **36** (Database issue), 577-581, 2008.
336. Boulesteix, A.-L. Over-optimism in Bioinformatics Research, *Bioinformatics* **26** (3), 437-439, 2010.
337. Gavin, A. C. et al. Proteome survey reveals modularity of the yeast cell machinery, *Nature* **440** (7084), 631-636, 2006.
338. Jansen, R., & Gerstein, M. Analyzing protein function on a genomic scale: the importance of gold-standard positives and negatives for network prediction, *Current Opinion in Microbiology* **7**, 535-545, 2004.
339. Friedel, C. C., Krumsiek, J., & Zimmer, R. Bootstrapping the interactome: unsupervised identification of protein complexes in yeast, *Journal of Computational Biology* **16** (8), 971-987, 2009.

340. Krumsiek, J., Friedel, C. C., & Zimmer, R. ProCope - protein complex prediction and evaluation, *Bioinformatics* **24** (18), 2115-6, 2008.
341. Ashburner, M. et al. Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium, *Nature Genetics* **25**, 25–29, 2000.
342. Bihai, Z. et al. Construction of Uncertain Protein-Protein Interaction Networks and Its Applications, in *Bioinformatics Research and Applications*, edited by Eulenstein, O., Janies, D., & Schwartz, D., LNCS, Springer, Berlin Heidelberg, 2013, 286-297.
343. Boyle, E. I. et al. GO: TermFinder-open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes, *Bioinformatics* **20** (18), 3710-3715, 2004.
344. Yang, P. et al. A Review of Ensemble Methods in Bioinformatics, *Current Bioinformatics* **4** (5), 296-308, 2010.
345. Greene, D. et al. Ensemble Nonnegative Matrix Factorization, *Bioinformatics* **24**, 1722–1728, 2008.
346. Ou-Yang, L. et al. Protein Complex Detection via Weighted Ensemble, *PLoS ONE* **8** (5), e62158, 2013.
347. Zhang, Y. et al. A graph-based cluster ensemble method to detect protein functional modules from multiple information sources, in In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (BCB '12), ACM, New York, NY, USA, 567-569.
348. Srihari, S., & Leong, H. W. Employing functional interactions for the characterization and detection of sparse complexes from yeast PPI networks, *International Journal of Bioinformatics Research and Application* **8** (3/4), 286-304, 2012.
349. Von Heijne, G. The Membrane Protein Universe: What's out there and why bother?, *Journal of Internal Medicine* **261** (6), 543-557, 2007.
350. Daley, D. O. The Assembly of Membrane Proteins into Complexes, *Current Opinions in Structured Biology* **8** (4), 420-424, 2008.