

Contents

<i>Preface</i>	<i>page xiii</i>
1 Introduction	1
1.1 Rapid Growth of Protein–Protein Interaction Data	1
1.2 Computational Analysis of PPI Networks	3
1.2.1 Topological Features of PPI Networks	4
1.2.2 Modularity Analysis	5
1.2.3 Prediction of Protein Functions in PPI Networks	6
1.2.4 Integration of Domain Knowledge	7
1.3 Significant Applications	7
1.4 Organization of this Book	9
1.5 Summary	10
2 Experimental Approaches to Generation of PPI Data	11
2.1 Introduction	11
2.2 The Y2H System	11
2.3 Mass Spectrometry (MS) Approaches	13
2.4 Protein Microarrays	15
2.5 Public PPI Data and Their Reliability	15
2.5.1 Experimental PPI Data Sets	15
2.5.2 Public PPI Databases	16
2.5.3 Functional Analysis of PPI Data	17
2.6 Summary	20
3 Computational Methods for the Prediction of PPIs	21
3.1 Introduction	21
3.2 Genome-Scale Approaches	21
3.3 Sequence-Based Approaches	25
3.4 Structure-Based Approaches	26
3.5 Learning-Based Approaches	27
3.6 Network Topology-Based Approaches	29
3.7 Summary	32

4	Basic Properties and Measurements of Protein Interaction Networks	33
4.1	Introduction	33
4.2	Representation of PPI Networks	33
4.3	Basic Concepts	34
4.4	Basic Centralities	35
4.4.1	Degree Centrality	35
4.4.2	Distance-Based Centralities	35
4.4.3	Current-Flow-Based Centrality	37
4.4.4	Random-Walk-Based Centrality	40
4.4.5	Feedback-Based Centrality	41
4.5	Characteristics of PPI Networks	44
4.6	Summary	49
5	Modularity Analysis of Protein Interaction Networks	50
5.1	Introduction	50
5.2	Useful Metrics for Modular Networks	51
5.2.1	Cliques	51
5.2.2	Cores	51
5.2.3	Degree-Based Index	52
5.2.4	Distance (Shortest Paths)-Based Index	53
5.3	Methods for Clustering Analysis of Protein Interaction Networks	53
5.3.1	Traditional Clustering Methods	54
5.3.2	Nontraditional Clustering Methods	55
5.4	Validation of Modularity	56
5.4.1	Clustering Coefficient	56
5.4.2	Validation Based on Agreement with Annotated Protein Function Databases	57
5.4.3	Validation Based on the Definition of Clustering	59
5.4.4	Topological Validation	60
5.4.5	Supervised Validation	61
5.4.6	Statistical Validation	61
5.4.7	Validation of Protein Function Prediction	62
5.5	Summary	62
6	Topological Analysis of Protein Interaction Networks	63
	<i>With Woo-chang Hwang</i>	
6.1	Introduction	63
6.2	Overview and Analysis of Essential Network Components	64
6.2.1	Error and Attack Tolerance of Complex Networks	64
6.2.2	Role of High-Degree Nodes in Biological Networks	67
6.2.3	Betweenness, Connectivity, and Centrality	69
6.3	Bridging Centrality Measurements	73
6.3.1	Performance of Bridging Centrality with Synthetic and Real-World Networks	75
6.3.2	Assessing Network Disruption, Structural Integrity, and Modularity	77

6.4	Network Modularization Using the Bridge Cut Algorithm	84
6.5	Use of Bridging Nodes in Drug Discovery	87
6.5.1	Biological Correlates of Bridging Centrality	88
6.5.2	Results from Drug Discovery-Relevant Human Networks	92
6.5.3	Comparison to Alternative Approaches: Yeast Cell Cycle State Space Network	94
6.5.4	Potential of Bridging Centrality as a Drug Discovery Tool	95
6.6	PathRatio: A Novel Topological Method for Predicting Protein Functions	97
6.6.1	Weighted PPI Network	97
6.6.2	Protein Connectivity and Interaction Reliability	98
6.6.3	PathStrength and PathRatio Measurements	99
6.6.4	Analysis of the PathRatio Topological Measurement	100
6.6.5	Experimental Results	101
6.7	Summary	108
7	Distance-Based Modularity Analysis	109
7.1	Introduction	109
7.2	Topological Distance Measurement Based on Coefficients	109
7.3	Distance Measurement by Network Distance	112
7.3.1	PathRatio Method	112
7.3.2	Averaging the Distances	113
7.4	Ensemble Method	114
7.4.1	Similarity Metrics	115
7.4.2	Base Algorithms	116
7.4.3	Consensus Methods	116
7.4.4	Results of the Ensemble Methods	118
7.5	UVCLUSTER	118
7.6	Similarity Learning Method	120
7.7	Measurement of Biological Distance	124
7.7.1	Sequence Similarity-Based Measurements	124
7.7.2	Structural Similarity-Based Measurements	125
7.7.3	Gene Expression Similarity-Based Measurements	127
7.8	Summary	128
8	Graph-Theoretic Approaches to Modularity Analysis	130
8.1	Introduction	130
8.2	Finding Dense Subgraphs	130
8.2.1	Enumeration of Complete Subgraphs	130
8.2.2	Monte Carlo Optimization	131
8.2.3	Molecular Complex Detection	132
8.2.4	Clique Percolation	133
8.2.5	Merging by Statistical Significance	134
8.2.6	Super-Paramagnetic Clustering	136
8.3	Finding the Best Partition	137
8.3.1	Recursive Minimum Cut	137
8.3.2	Restricted Neighborhood Search Clustering (RNSC)	138

8.3.3	Betweenness Cut	140
8.3.4	Markov Clustering	140
8.3.5	Line Graph Generation	143
8.4	Graph Reduction-Based Approach	144
8.4.1	Graph Reduction	144
8.4.2	Hierarchical Modularization	146
8.4.3	Time Complexity	147
8.4.4	k Effects on Graph Reduction	147
8.4.5	Hierarchical Structure of Modules	149
8.5	Summary	150
9	Flow-Based Analysis of Protein Interaction Networks	152
9.1	Introduction	152
9.2	Protein Function Prediction Using the FunctionalFlow Algorithm	153
9.3	CASCADE: A Dynamic Flow Simulation for Modularity Analysis	155
9.3.1	Occurrence Probability and Related Models	156
9.3.2	The CASCADE Algorithm	158
9.3.3	Analysis of Prototypical Data	160
9.3.4	Significance of Individual Clusters	162
9.3.5	Analysis of Functional Annotation	164
9.3.6	Comparative Assessment of CASCADE with Other Approaches	169
9.3.7	Analysis of Robustness	175
9.3.8	Analysis of Computational Complexity	175
9.3.9	Advantages of the CASCADE Method	176
9.4	Functional Flow Analysis in Weighted PPI Networks	177
9.4.1	Functional Influence Model	178
9.4.2	Functional Flow Simulation Algorithm	179
9.4.3	Time Complexity of Flow Simulation	180
9.4.4	Detection of Overlapping Modules	181
9.4.5	Detection of Disjoint Modules	189
9.4.6	Functional Flow Pattern Mining	191
9.5	Summary	198
10	Statistics and Machine Learning Based Analysis of Protein Interaction Networks	199
	<i>With Pritam Chanda and Lei Shi</i>	
10.1	Introduction	199
10.2	Applications of Markov Random Field and Belief Propagation for Protein Function Prediction	200
10.3	Protein Function Prediction Using Kernel-Based Statistical Learning Methods	207
10.4	Protein Function Prediction Using Bayesian Networks	211

10.5	Improving Protein Function Prediction Using Bayesian Integrative Methods	213
10.6	Summary	214
11	Integration of GO into the Analysis of Protein Interaction Networks	216
	<i>With Young-rae Cho</i>	
11.1	Introduction	216
11.2	GO structure	217
11.2.1	GO Annotations	217
11.3	Semantic Similarity-Based Integration	218
11.3.1	Structure-Based Methods	219
11.3.2	Information Content-Based Methods	220
11.3.3	Combination of Structure and Information Content	221
11.4	Semantic Interactivity-Based Integration	223
11.5	Estimate of Interaction Reliability	223
11.5.1	Functional Co-Occurrence	224
11.5.2	Topological Significance	225
11.5.3	Protein Lethality	226
11.6	Functional Module Detection	227
11.6.1	Statistical Assessment	227
11.6.2	Supervised Validation	229
11.7	Probabilistic Approaches for Function Prediction	231
11.7.1	GO Index-Based Probabilistic Method	231
11.7.2	Semantic Similarity-Based Probabilistic Method	235
11.8	Summary	241
12	Data Fusion in the Analysis of Protein Interaction Networks	243
12.1	Introduction	243
12.2	Integration of Gene Expression with PPI Networks	243
12.3	Integration of Protein Domain Information with PPI Networks	244
12.4	Integration of Protein Localization Information with PPI Networks	245
12.5	Integration of Several Data Sources with PPI Networks	247
12.5.1	Kernel-Based Methods	247
12.5.2	Bayesian Model-Based Method	249
12.6	Summary	249
13	Conclusion	251
	<i>Bibliography.</i>	255
	<i>Index</i>	273

Color plates follow page 82