# Contents


**1 Introduction** **1**

1.1 Introduction . . . 1

1.1.1 Protein Protein Interaction Data Analysis . . . 2

1.1.2 Gene Expression Data Analysis . . . 2

1.1.3 Data mining . . . 3

1.2 Issues and Challenges . . . 4

1.3 Research Objectives . . . 5

1.4 Research Contributions . . . 6

1.4.1 Protein complex extraction techniques that show consistent performance across a range of yeast datasets as well as human dataset. . . . 6

1.4.2 Quality protein complex detection methods and disease associated complex inferencing . . . 6

1.4.3 A gene module extraction technique to find interesting behavior of genes involved during disease progression . . . 7

1.4.4 A subspace based gene module extraction technique to study the progression of disease from normal to diseased state . . . 7

1.5 Organization of the Thesis . . . 7

**2 Background** **9**

2.1 Introduction . . . 9

2.1.1 Central Dogma . . . 9

2.1.2 Protein Protein Interaction Data . . . 11

2.1.2.1 PPI Detection Methods . . . 12

2.1.2.2 Types of PPI data . . . 14

2.1.2.3 Protein protein interaction data analysis . . . 14

2.1.3 Gene Expression Data . . . 15

2.1.3.1 Gene Expression Data Measurement Techniques . . . 16

2.1.3.2 Types of gene expression data . . . 17

2.1.3.3 Gene Expression Data Analysis . . . 18

2.1.4 Data Mining in PPI and GE Analysis . . . 18

2.1.5 PPI and Gene Expression Data Analysis Tools . . . 20

2.1.6 Datasets Used . . . 22
2.1.6.1 PPI dataset . . . 22
2.1.6.2 Gene Expression Disease datasets . . . 24
2.1.7 Biological Knowledge bases . . . 24
2.1.7.1 PPI Gold Standards . . . 25
2.1.7.2 Co-localization Datasets . . . 25
2.1.7.3 Gene Ontology . . . 26
2.1.7.4 GeneCard . . . 26
2.1.8 Data cleaning . . . 26
2.1.9 Performance Indices . . . 26
2.1.9.1 Precision, Recall and F-measure . . . 27
2.1.9.2 Positive Predictive Value, Sensitivity and Accuracy 28
2.1.9.3 Co-localization score . . . 29
2.1.9.4 p-value . . . 30
2.1.10 Discussion . . . 30

**3 Protein Complex Finding Methods 31**
3.1 Introduction . . . 31
3.2 Related Work . . . 32
3.3 Motivation . . . 34
3.4 Contributions . . . 36
3.5 Protein complex finding based on topological information: CNCM . 37
3.5.1 Proposed Method : CNCM . . . 38
3.5.2 Experimental Results . . . 41
3.5.3 Discussion . . . 50
3.6 Finding Protein Complexes using Multi objective Approach : DCRS 50
3.6.1 Proposed Method:DCRS . . . 50
3.6.2 Experimental Results . . . 58
3.6.3 Discussion . . . 59
3.7 Conclusion . . . 60

**4 Protein Complex Finding Methods: An Application to Alzheimer's Disease 62**
4.1 Introduction . . . 62
4.2 Related Work . . . 62
4.3 Protein complex finding based on topological and functional information: CSC . . . 65
4.3.1 Proposed algorithm: CSC . . . 66
4.3.2 Experimental Results . . . 70
4.3.3 CSC: An Application to Alzheimer's Disease . . . 73

4.3.4 Discussion . . . . . 80
4.4 Protein complex finding using semi-supervised technique: ComFiR . 81
4.4.1 Proposed Method : ComFiR . . . . . 83
4.4.2 Experimental Results . . . . . 87
4.4.3 Discussion . . . . . 97
4.5 Conclusion . . . . . 97

**5 Gene Module Extraction and Analysis-An Application to Breast Cancer 99**
5.1 Introduction . . . . . 99
5.2 Related Works . . . . . 102
5.3 Proposed Method . . . . . 103
5.3.1 Preprocessing . . . . . 103
5.3.2 Gene Co-expression network construction and module extraction . . . . . 104
5.4 Experimental Results . . . . . 107
5.4.1 Parameter tuning . . . . . 108
5.4.2 Comparison with existing work . . . . . 109
5.4.3 Pathway identification from module members . . . . . 110
5.4.4 Role of common genes during disease progression . . . . . 110
5.4.5 Adaption of gene expression and role of common pathways during disease progression . . . . . 115
5.4.6 Neutrality of few causal genes during module formation in metastasis stage . . . . . 116
5.4.7 Expression pattern of module members . . . . . 117
5.5 Conclusion . . . . . 119

**6 Subspace Module Extraction and Analysis-An application to Parkinson's Disease 120**
6.1 Introduction . . . . . 120
6.2 Motivation . . . . . 122
6.3 Related Works . . . . . 122
6.4 Proposed Method . . . . . 124
6.4.1 Preprocessing . . . . . 124
6.4.2 Multi-edge network construction . . . . . 125
6.4.3 Calculating the topological subspace overlap matrix . . . . . 126
6.4.4 Network module extraction . . . . . 127
6.5 Experimental Results . . . . . 131
6.5.1 Comparison with existing work based on p-value . . . . . 132
6.5.2 Comparison with biclustering techniques . . . . . 134
6.5.3 Module correspondence between the two stages . . . . . 135

6.5.3.1 Topological aspect of module correspondence . . . 135
6.5.3.2 Module correspondence from pathway point of view 136
6.6 Identification of interesting biomarkers . . . 137
6.6.1 Expression pattern of new suspected genes . . . 147
6.7 Hub gene analysis in modules based on centrality . . . 147
6.8 Discussion . . . 150

**7 Conclusion** **152**
7.1 Concluding Remarks . . . 152
7.2 Future Works . . . 154

**A Appendix** **174**