Contents

Certificate ii

Declaration by the Candidate iii

Copyright Transfer Certificate iv

Acknowledgements v

Contents ix

List of Figures xiii

List of Tables xv

Abbreviations xvii

Symbols xxi

Preface xxii

1 Introduction 1
- 1.1 Major Issues . . . 2
- 1.2 Objectives . . . 4
- 1.3 Contributions . . . 7
  - 1.3.1 Ego Network Based Community Detection . . . 7
  - 1.3.2 Agglomerative Fuzzy Community Detection . . . 8
  - 1.3.3 Community Detection in Multiple Featured Networks . . . 9
  - 1.3.4 Applicability of Communities to Predict Missing Links . . . 10
  - 1.3.5 Designing Validation Metrics and Evaluation Methodologies . . . 11

1.4 Thesis Organization 13

2 Survey on Community Detection, Validation and Applications 15
2.1 Community Detection Techniques 15
2.1.1 Hierarchical Approaches 16
2.1.2 Optimization Techniques 19
2.1.3 Spectral Clustering 21
2.1.4 Statistical Inference 22
2.1.5 Other Approaches 24
2.2 Validation Metrics 25
2.2.1 Accuracy Metrics 25
2.2.2 Quality Metrics 30
2.3 Datasets 32
2.3.1 Real-world Networks 33
2.3.2 Synthetic Networks 36
2.4 Evaluation Methodologies 36
2.5 Applications of Community Structure 38
2.5.1 Community Structure in Link Prediction 38
2.5.2 Community Structure in Information Diffusion 39
2.5.3 Community Structure in Recommendation Systems 39
2.6 Summary 40

3 Ego Network Based Community Detection 41
3.1 Introduction 41
3.1.1 Ego Network 42
3.1.2 Mutual Interest and Relationship 44
3.2 Proposed Approach 46
3.2.1 Expansion Phase 48
3.2.2 Dissolution Phase 49
3.3 Empirical Analysis 50
3.3.1 Experimental Setup 50
3.3.2 Analyzing Accuracy 51
3.3.3 Analyzing Quality 53
3.3.4 MCDM Ranking 58
3.3.5 Parametric Analysis 61
3.4 Computational Complexity Comparison 66
3.5 Conclusion 68

4 Fuzzy Agglomerative Community Detection 71
4.1 Introduction 71
4.2 Proposed Approach 73
4.2.1 New Anchor Selection 74

4.2.2 Membership degree computation . . . 75
4.2.3 Redundant and False Anchor Removal . . . 77
4.2.4 Fuzzy Agglomerative Community Detection . . . 78
4.3 Time Complexity Comparison . . . 82
4.4 Empirical Analysis . . . 84
4.4.1 Experimental Setup . . . 84
4.4.2 Performance Analysis . . . 86
4.5 Conclusion . . . 101

5 Community Detection using Particle Swarm Optimization 103
5.1 Introduction . . . 103
5.2 PSO with Cognitive Avoidance Mechanism . . . 104
5.3 Community Detection using PSOCA . . . 108
5.3.1 Multiple Featured Network Representation . . . 109
5.3.2 Particle Representation and Objective Function . . . 110
5.3.3 Proposed Algorithm . . . 112
5.4 Empirical Analysis . . . 113
5.4.1 Experimental Setup for PSOCA and PSO-CATV . . . 113
5.4.2 Analysis of Solution Quality . . . 116
5.4.3 Analysis of Convergence . . . 120
5.4.4 Analysis of Community Detection with PSOCA . . . 122
5.5 Conclusion . . . 125

6 Community-based Link Prediction 127
6.1 Introduction . . . 127
6.2 Proposed Approach . . . 128
6.3 Evaluation Strategy . . . 133
6.3.1 Performance Metrics . . . 134
6.3.2 Experimental Setup . . . 136
6.4 Result Analysis . . . 137
6.5 Conclusion . . . 146

7 Community Validation Metrics and Evaluation Methodologies 149
7.1 Introduction . . . 149
7.2 Proposed Quality Metrics . . . 152
7.2.1 Social Community Formation . . . 152
7.2.2 Unifiability . . . 153
7.2.3 Isolability . . . 155
7.2.4 Balanced Isolability and Unifiability . . . 157
7.3 Relative Inclination Towards Accuracy . . . 159
7.3.1 MCDM Process . . . 159
7.3.2 RITA Framework . . . 160

7.4 Visual Analysis Methodology 162
7.4.1 Point Dominance 162
7.4.2 Regression Line Dominance 163
7.4.3 Regression Line Shifting 165
7.4.4 Comparative Analysis Methodology 168
7.5 Theoretical Analysis of AVI, AVU and ANUI Metrics 171
7.6 Empirical Analysis of Proposed Metrics 174
7.6.1 Experimental Setup 174
7.6.2 Metric Competitiveness Analysis 174
7.6.3 Metric Characteristics 180
7.7 Analysis using Proposed Evaluation Methodologies 183
7.7.1 RITA Analysis 183
7.7.2 Visual Analysis 187
7.8 Conclusion 193

8 Concluding Remarks and Future Directions **195**
8.1 Summary of Contributions 195
8.2 Scope for Further Work 197

Bibliography **199**

A Proofs of Theorems **221**
A.1 Proof of Theorems in section 7.2 221
A.2 Proof of Theorems in section 7.4 224
A.3 Proof of Theorems in section 7.5 226

B List of Publications **233**