

MULTIOBJECTIVE EVOLUTIONARY ALGORITHM WITH GRAPHLET
MEASURE FOR DETECTION OF COMPLEX TOPOLOGICAL AND
BIOLOGICAL COMMUNITY STRUCTURE

DHUHA ABDULHADI ABDULJABBAR

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

MARCH 2021

DEDICATION

To my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake.

To my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

To my husband, who has been a constant source of support and encouragement during the challenges of doctoral study and life.

To my beloved sons, for their constant love and assistance.

I dedicate this thesis ...

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Associate Professor Dr. Roselina Sallehuddin, for encouragement, guidance, critics and friendship. I am also very thankful to my co-supervisor Professor Dr Siti Zaiton Mohd Hashim for her guidance, advices and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

I would also like to express my sincere gratitude and thanks to my beloved parents for their support and encouragement. Special and deepest thanks express to all my family member for their love, patience and continued support.

My thanks also extends to all my colleagues and others who have provided assistance at various occasions. Department of Digital Services (UTMDigital) in Universiti Teknologi Malaysia also deserve special thanks for their assistance in supporting and providing facilities and services of high performance computing.

ABSTRACT

In the past few years, community structure detection has garnered much attention due to its significant role in analyzing complex network structures and functions. Detecting natural divisions in complex networks is proved to be an extremely Non-deterministic Polynomial-time hard (NP-hard) problem, which has been solved using evolutionary computation methods. Despite many efforts to design an effective community structure formula, the definition is still general, depending solely on the nodes' intra- and inter-connections. It lacks complete reflection of inherent topological properties, such as graphlet measure in terms of graphlet degree signatures and signature similarities, that can accurately detect complex communities' structure such as topological and biological community. The research proposes a new method termed MOEA_CGN (MultiObjective Evolutionary Algorithm based on Cooperation between Graphlet-based measure and Neighborhood relations) to improve the detection quality of complex topological and biological community structure in terms of accuracy and velocity. Thus, the contribution of this study is summarized in threefold. First, a new multiobjective optimization function is proposed to tackle the issue of a community structure definition. Second, a heuristic mutation operator is designed to enhance MOEA_CGN performance to accurately detect complex topological community structure by tackling the resolution limit problem. Third, the heuristic mutation operator is improved to make the MOEA_CGN method identify and detect complex biological community structure accurately by tackling the heterogeneity issue of real-world networks. Systematic experiments on different real-world networks from various domains and synthetic networks of different complexities have demonstrated the proposed method's effectiveness and robustness to define the community detection problem. Specifically, the proposed method has achieved detection reliability with an average improvement of 6.7% in detecting complex topological communities and 9.17% in detecting complex biological communities compared with the state-of-the-art benchmark studies. Moreover, the proposed MOEA_CGN method has demonstrated its ability to detect the optimal community structures in most complex networks faster than the competent multiobjective models based on community detection.

ABSTRAK

Dalam beberapa tahun kebelakangan ini, pengesanan struktur komuniti telah mendapat banyak perhatian kerana peranannya yang signifikan dalam menganalisis struktur dan fungsi rangkaian yang kompleks seperti komuniti topologi dan biologi. Mengesan pembahagian semula jadi dalam rangkaian kompleks terbukti menjadi masalah tegar polinomial-masa tidak deterministik (NP-tegar), yang telah diselesaikan dengan menggunakan kaedah pengkomputeran evolusi. Walaupun terdapat banyak usaha untuk merancang formula struktur komuniti yang berkesan, definisi ini masih umum, bergantung sepenuhnya pada hubungan intra dan inter nod. Ia kekurangan gambaran lengkap mengenai sifat topologi yang wujud, seperti ukuran grafik, dalam erti kata tandatangan darjah graflet dan persamaan tandatangan, yang dapat mengesan struktur komuniti yang kompleks dengan tepat. Penyelidikan ini mencadangkan kaedah baru yang dikenali sebagai MOEA_CGN (Algoritma Evolusi Multi Objektif berdasarkan Kerjasama antara Graphlet dan hubungan Kejiranan) untuk meningkatkan kualiti pengesanan struktur komuniti topologi dan biologi yang kompleks dari segi ketepatan dan halaju. Oleh itu, sumbangan kajian ini diringkaskan dalam tiga elemen berikut. Pertama, fungsi pengoptimuman multiobjektif baru dicadangkan untuk mengatasi masalah definisi struktur komuniti. Kedua, pengendalian mutasi heuristik dirancang untuk meningkatkan prestasi MOEA_CGN dalam mengesan struktur komuniti topologi yang kompleks dengan menangani masalah had penyelesaian. Ketiga, pengendali mutasi heuristik diperbaiki untuk menjadikan kaedah MOEA_CGN dapat mengenalpasti dan mengesan struktur komuniti biologi yang kompleks dengan tepat dalam menangani masalah heterogenitas rangkaian dunia nyata. Eksperimen sistematik pada rangkaian dunia nyata yang berbeza dari pelbagai domain dan rangkaian sintetik dengan kerumitan yang berbeza telah menunjukkan keberkesanan dan keteguhan kaedah yang dicadangkan untuk menentukan masalah pengesanan komuniti. Secara khusus, kaedah yang dicadangkan telah mencapai kebolehpercayaan pengesanan dengan peningkatan purata 6.7% dalam mengesan komuniti topologi kompleks dan 9.17% dalam mengesan komuniti biologi yang kompleks berbanding dengan kajian penanda aras terkini. Tambahan pula, kaedah MOEA_CGN yang dicadangkan telah menunjukkan kemampuannya untuk mengesan struktur komuniti yang optimum di rangkaian yang paling kompleks dengan lebih cepat daripada model multiobjektif yang kompeten berdasarkan pengesanan komuniti.

TABLE OF CONTENTS

	TITLE	PAGE
	DECLARATION	iii
	DEDICATION	iv
	ACKNOWLEDGEMENT	v
	ABSTRACT	vi
	ABSTRAK	vii
	TABLE OF CONTENTS	viii
	LIST OF TABLES	xiii
	LIST OF FIGURES	xvi
	LIST OF ABBREVIATIONS	xxii
	LIST OF SYMBOLS	xxv
CHAPTER 1	INTRODUCTION	1
	1.1 Introduction	1
	1.2 Problem Background	3
	1.3 Problem Statement	7
	1.4 Research Goal	9
	1.5 Research Objectives	9
	1.6 Research Scope	10
	1.7 Research Significance	11
	1.8 Thesis Organization	13
CHAPTER 2	LITERATURE REVIEW	15
	2.1 Introduction	15
	2.2 Community Detection in Complex Networks	16
	2.2.1 Concepts of Graph Theory	18
	2.2.2 Graphlet Degree Signatures	20

2.2.3	Preliminary Definitions of Community Detection in Complex Networks	22
2.3	Community Detection Methods	23
2.4	Community Detection based on Evolutionary Single (and Multi) Objective Algorithms	24
2.5	Single Objective versus Multiobjective	27
2.6	Systematic Literature Review	29
2.6.1	MOEA for Detection Complex Topological Community Structure	44
2.6.2	MOEA for Detection Complex Biological Community Structure	52
2.7	Community Detection Gaps and Challenges	61
2.7.1	Community Structure Definition	62
2.7.2	Resolution Limit Problem	63
2.7.3	Heterogeneity of real-world networks	64
2.8	Discussion on Related Work	64
2.9	Multiobjective Evolutionary Algorithm Based on Decomposition (<i>MOEA/D</i>)	70
2.10	Summary	73
CHAPTER 3	RESEARCH METHODOLOGY	75
3.1	Introduction	75
3.2	Research Framework	76
3.3	Dataset Planning and Compiling	78
3.3.1	Complex Networks with Known Ground-Truth	79
3.3.1.1	Synthetic-benchmark networks	79
3.3.1.2	Real-complex networks	80
3.3.1.3	Biological networks	82
3.3.2	Real-Complex Networks with Unknown Ground-Truth	83
3.4	Pre-Processing Modules Phase	85
3.5	The Proposed Method <i>MOEA_CGN</i> for Community Detection	86
3.5.1	Solution Representation and Initialization	90
3.5.2	Fitness Computation	91

3.5.2.1	Extraction of graphlet degree signatures and signature similarities	92
3.5.2.2	New objective functions	94
3.5.3	Applying Evolutionary Operators	96
3.5.3.1	Crossover	96
3.5.3.2	Mutation	97
3.5.4	Enhancement of the <i>MOEA_CGN</i> Method	98
3.5.4.1	New heuristic mutation operator	98
3.5.4.2	Improvement of the heuristic mutation operator	102
3.5.5	Update Processes	103
3.6	Evaluation Measures	104
3.6.1	Evaluation Measures for Complex Networks with Known Ground-Truth	105
3.6.1.1	Synthetic-benchmark networks and real-complex networks	105
3.6.1.2	Biological networks	106
3.6.2	Evaluation Measures for Real-Complex Networks with Unknown Ground-Truth	108
3.7	State-of-the-art Benchmark Studies	108
3.8	Summary	109
CHAPTER 4	MOEA_CGN TO DETECT COMPLEX TOPOLOGICAL COMMUNITIES	111
4.1	Introduction	111
4.2	MOEA_CGN Method	112
4.3	Results for the Synthetic Benchmark Networks	114
4.3.1	GN Networks	114
4.3.2	LFR Networks	116
4.4	Results for the Real-Complex Networks with Known Ground-Truth	118
4.5	Results for the Real-Complex Networks with Unknown Ground-Truth	127
4.6	Summary	129

CHAPTER 5	ENHANCED MOEA_CGN WITH NEW HEURISTIC MUTATION OPERATOR	131
5.1	Introduction	131
5.2	Enhanced MOEA_CGN Method with New Heuristic Mutation Operator	132
5.3	Results for the Synthetic Benchmark Networks	134
5.3.1	GN Networks	134
5.3.2	LFR Networks	140
5.4	Results for the Real-Complex Networks with Known Ground-Truth	147
5.5	Results for the Real-Complex Networks with Unknown Ground-Truth	153
5.6	Comparing Results with a State-of-the-art Benchmark Study	156
5.7	Partitions at Different Hierarchical Levels	160
5.8	Summary	167
CHAPTER 6	MOEA_CGN WITH IMPROVED HEURISTIC MUTATION OPERATOR BASED ON NODES' SIGNATURE DISTANCE TO ACCURATELY DETECT BIOLOGICAL COMMUNITIES	169
6.1	Introduction	169
6.2	MOEA_CGN with Improved Heuristic Mutation Operator	170
6.3	Results of <i>MOEA_CGN</i> on PPI Networks with Different Mutation Versions	172
6.4	Comparing Results with Four MOO Models	176
6.5	Comparing Results with State-of-the-art Benchmark Studies	182
6.6	Negative Control Analysis of MOEA_CGN Method	183
6.7	An Illustrative Example	186
6.8	Summary	188
CHAPTER 7	CONCLUSION AND FUTURE WORK	191
7.1	Introduction	191
7.2	Conclusion	192

7.2.1	Summary on Formulation of the MOEA_CGN method with Two New Objective Functions	193
7.2.2	Summary on Formulation of the New Heuristic Mutation Operator to Enhance MOEA_CGN Performance	194
7.2.3	Summary on Improvement of the Heuristic Mutation Operator	194
7.3	Future Work and Recommendations	195
REFERENCES		197
LIST OF PUBLICATIONS		215

LIST OF TABLES

TABLE NO.	TITLE	PAGE
Table 2.1	Summarization of multiobjective community detection studies using EA & GA	31
Table 2.2	Summarization of single objective community detection studies using EA & GA	33
Table 2.3	Summarization of single objective and multiobjective community detection studies using MA	35
Table 2.4	Summarization of single objective and multiobjective community detection studies using PSO algorithm	37
Table 2.5	Summarization of single objective community detection studies using DE algorithm	38
Table 2.6	Summarization of single objective community detection studies using CA	38
Table 2.7	Summarization of single objective and multiobjective community detection studies using FA	39
Table 2.8	Summarization of single objective community detection studies using BA	39
Table 2.9	Summarization of single objective and multiobjective community detection studies using ACO/BCO algorithms	39
Table 2.10	Summarization of single objective and multiobjective community detection studies using CS algorithm	40
Table 2.11	Summary of some MOO related works for CD	66
Table 2.12	Summary of the proposed study structure	69
Table 3.1	Summary of parameters used to generate synthetic-benchmark networks (GN_{128_γ} and LFR_{5000_γ})	80
Table 3.2	Topological characteristics of the four real-complex networks with known ground-truth	82

Table 3.3	Topological characteristics of the PPI networks with known generic complexes	83
Table 3.4	Topological characteristics of real- complex networks with unknown ground-truth	85
Table 4.1	Detection reliability in terms of NMI_{avg} and $WNMI_{avg}$ for the five competent MOO models when using the traditional neighbor mutation operator with $p_m = 0.2$ over the $GN_{128}_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ synthetic benchmark networks	115
Table 4.2	Detection reliability in terms of NMI_{avg} and $WNMI_{avg}$ for the five competent MOO models when using the traditional neighbor mutation operator with $p_m = 0.2$ over the $LFR_{5000}_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ synthetic benchmark networks	117
Table 4.3	Detection reliability (NMI_{avg} , $WNMI_{avg}$ values) for the four complex-real networks with known true-divisions	119
Table 4.4	Detection reliability results in terms of Q_{max} and Q_{avg} over the complex-real networks with unknown real-partitions. Among the different models, the best Q_{max} and Q_{avg} values are presented in bold font	128
Table 5.1	Detection reliability in terms of NMI_{avg} , $WNMI_{avg}$, and detection velocity (expressed within the parentheses) for the five competent MOO models when applying the suggested heuristic mutation operator ($p_{hm} = \{0.5, 0.8\}$) over the $GN_{128}_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ synthetic benchmark networks	136
Table 5.2	Detection reliability in terms of NMI_{avg} , $WNMI_{avg}$ for the five competent MOO models when applying the suggested heuristic mutation (with $p_{hm} = \{0.5, 0.8\}$) over the $LFR_{5000}_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ synthetic benchmark networks	143
Table 5.3	Detection reliability (NMI_{avg} , $WNMI_{avg}$ values) and detection velocity (expressed within the parentheses) for the four complex-real networks with known true-divisions	148

Table 5.4	Detection reliability (in terms of Q_{avg} values) for complex-real networks with unknown real-partitions. Among the different models and for all heuristic mutation settings $p_{hm} = \{0.5, 0.8\}$, the best result is shown in bold font	155
Table 5.5	Detection reliability (in terms of NMI_{avg} and Q_{avg} values) over the LFR_1000 $_{\gamma \in [0.0, 0.5]}$ synthetic benchmark networks. The best NMI_{avg} and best Q_{avg} are given in bold font	157
Table 5.6	Detection reliability (in terms of NMI_{avg} and Q_{avg} values) over the complex-real networks. The best NMI_{avg} and best Q_{avg} are given in bold font	159
Table 6.1	Detection reliability (in terms of recall, precision and F-measure values) for PPI_D1 and PPI_D2 networks (when $\sigma_{OS} = 0.2$)	183
Table 6.2	Detection reliability in terms of recall, precision and F-measure for MOEA_CGN method on PPI_D1 network with $p_{ihm} = 0.5$ versus $\sigma_{OS} = 0.4$, when increasing the percentage of randomly added / removed edges	184

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
Figure 2.1	An undirected graph of 4 vertices, 5 edges and its associated adjacency matrix	19
Figure 2.2	An undirected graph of 4 vertices, 5 edges and its associated incidence matrix	20
Figure 2.3	(a) Graphlets of size up to five nodes (Pržulj, 2007). (b) A clarification of how the degree of an individual node v (top leftmost) generalized into its respective signature or GDV, like orbit 3 of G_2 represents the triangle graphlet (2 nd top); orbit 6 of G_4 represents a graphlet of 3-crossing lines (3 rd top), and orbit 8 of G_5 represents the square graphlet (top rightmost). The values of $GDV(v)$ of 73 coordinates are given in the table below the panels	21
Figure 2.4	Graphical examples of community structure in social complex networks (a: Slashdot news review site, b: Epinions consumer review site) that appearing on average somewhat homogeneous. Adapted from: Lancichinetti <i>et al.</i> , (2010)	45
Figure 2.5	Graphical examples of complexes in biological networks, the clusters of PPIs of two organisms: a: <i>D. melanogaster</i> (fruit fly), and b: <i>H. sapiens</i> (man), the larger the community, the less tree-like it is. Adapted from: Lancichinetti <i>et al.</i> , (2010)	53
Figure 2.6	The illustration of MOEA/D general framework. The middle column represents the basic processes of MOEA/D. Both sides depict the extensions of the processes including initialization, objective functions, updating rules, and solution selection	73
Figure 3.1	The general framework of the research	77
Figure 3.2	The flowchart of the MOEA_CGN method for community structure detection in complex networks. The method is based on MOEA/D algorithm	89
Figure 3.3	An illustrative example of locus-based representation scheme	90

Figure 3.4	Graphlets from two-to-five nodes with their respective automorphism orbits (Pržulj, 2007)	93
Figure 3.5	An illustration of how to generalize the degree of node v (top leftmost) into its respective signature or GDV, such that orbit 3 of G_2 represents the triangle graphlet (2 nd top), and orbit 8 of G_5 represents the square graphlet (top rightmost). The values of 73 coordinates of GDV (v) are shown in the table	93
Figure 3.6	An illustrative example of the uniform crossover operator	97
Figure 3.7	An illustrative example of the random neighbor mutation operator	97
Figure 4.1	Detection reliability (<i>left: NMI_{avg}, right: $WNMI_{avg}$</i>) for the five MOO models when using the traditional neighbor mutation operator with $p_m = 0.2$ over the $GN_{128_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}}$ synthetic benchmark networks	115
Figure 4.2	Detection reliability (<i>left: NMI_{avg}, right: $WNMI_{avg}$</i>) for the five competent MOO models when using the traditional neighbor mutation operator with $p_m = 0.2$ over the $LFR_{5000_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}}$ synthetic benchmark networks	118
Figure 4.3	Experiment results for Zachary's network. (a) true community structure; (b) communities detected by $\Phi_{MOEA-CGN}$, where $NMI, WNMI = 0.8372$; (c) communities detected by $\Phi_{MOEA-CGN}$, where $NMI, WNMI = 0.8365$	121
Figure 4.4	Experiment results for Dolphins network. (a) true communities detected by $\Phi_{MOEA-CGN}$ in 19 runs, where $NMI, WNMI = 1$, (b) communities detected by $\Phi_{MOEA-CGN}$ in one run, where $NMI, WNMI = 0.887$	122
Figure 4.5	Experiment results for Krebs' books network. (a) true communities, (b) communities detected by Φ_{MOCD} model with best NMI value, (c) communities detected by $\Phi_{MOEA-CGN}$ model	124
Figure 4.6	Experiment results for American College Football network. (a) true communities, (b) communities detected by the proposed $\Phi_{MOEA-CGN}$ model with best NMI value, (c) communities detected by the	126

	proposed $\Phi_{MOEA-CGN}$ model with best WNMI value	
Figure 5.1	Detection reliability (<i>left: NMI_{avg}, right: $WNMI_{avg}$</i>) for the five competent MOO models when applying the suggested heuristic mutation operator with $p_{hm} = \{0.5\}$ over the $GN_{128_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}}$ synthetic benchmark networks	135
Figure 5.2	Detection reliability (<i>left: NMI_{avg}, right: $WNMI_{avg}$</i>) for the five competent MOO models when applying the suggested heuristic mutation operator with $p_{hm} = \{0.8\}$ over the $GN_{128_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}}$ synthetic benchmark networks	135
Figure 5.3	Detection reliability for the <i>MOGA_Net</i> model over the $GN_{128_{\gamma \in [0.1, 0.2, 0.3, 0.4, 0.5]}}$ networks, when using the traditional neighbor mutation operator (<i>MOGA_Net</i>), and when applying the suggested heuristic mutation (at $p_{hm} \in \{0.5, 0.8\}$) with the model	137
Figure 5.4	Detection reliability for the <i>MOCD</i> model over the $GN_{128_{\gamma \in [0.1, 0.2, 0.3, 0.4, 0.5]}}$ networks, when using the traditional neighbor mutation operator (<i>MOCD</i>), and when applying the suggested heuristic mutation (at $p_{hm} \in \{0.5, 0.8\}$) with the model	137
Figure 5.5	Detection reliability for the <i>MODPSO</i> model over the $GN_{128_{\gamma \in [0.1, 0.2, 0.3, 0.4, 0.5]}}$ networks, when using the traditional neighbor mutation operator (<i>MODPSO</i>), and when applying the suggested heuristic mutation (at $p_{hm} \in \{0.5, 0.8\}$) with the model	138
Figure 5.6	Detection reliability for the <i>Net_Degree (Fk)</i> model over the $GN_{128_{\gamma \in [0.1, 0.2, 0.3, 0.4, 0.5]}}$ networks, when using the traditional neighbor mutation operator (<i>Net_Degree (Fk)</i>), and when applying the suggested heuristic mutation (at $p_{hm} \in \{0.5, 0.8\}$) with the model	138
Figure 5.7	Detection reliability for the suggested <i>MOEA_CGN</i> model over the $GN_{128_{\gamma \in [0.1, 0.2, 0.3, 0.4, 0.5]}}$ networks, when using the traditional neighbor mutation operator (<i>MOEA-CGN</i>), and when	139

	applying the suggested heuristic mutation (at $p_{hm} \in \{0.5, 0.8\}$) with the model	
Figure 5.8	Detection reliability (left: NMI_{avg} , right: $WNMI_{avg}$) for the five competent <i>MOO</i> models when applying the heuristic mutation operator (with $p_{hm} = 0.5$) over the $LFR_5000_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ synthetic benchmark networks	141
Figure 5.9	Detection reliability (left: NMI_{avg} , right: $WNMI_{avg}$) for the five competent <i>MOO</i> models when applying the heuristic mutation operator (with $p_{hm} = 0.8$) over the $LFR_5000_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ synthetic benchmark networks	141
Figure 5.10	Detection reliability for the <i>MOGA_Net</i> model in the $LFR_5000_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ networks, when using the traditional neighbor mutation operator (<i>MOGA_Net</i>), and when applying the suggested heuristic mutation ($p_{hm} \in \{0.5, 0.8\}$) with the model	144
Figure 5.11	Detection reliability for the <i>MOCD</i> model in the $LFR_5000_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ networks, when using the traditional neighbor mutation operator (<i>MOCD</i>), and when applying the suggested heuristic mutation ($p_{hm} \in \{0.5, 0.8\}$) with the model	145
Figure 5.12	Detection reliability for the <i>MODPSO</i> model in the $LFR_5000_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ networks, when using the traditional neighbor mutation operator (<i>MODPSO</i>), and when applying the suggested heuristic mutation ($p_{hm} \in \{0.5, 0.8\}$) with the model	145
Figure 5.13	Detection reliability for the <i>Net_Degree(Fk)</i> model in the $LFR_5000_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ networks, when using the traditional neighbor mutation operator (<i>Net_Degree(Fk)</i>), and when applying the suggested heuristic mutation ($p_{hm} \in \{0.5, 0.8\}$) with the model	146
Figure 5.14	Detection reliability for the <i>MOEA_CGN</i> model in the $LFR_5000_{\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}}$ networks, when using the traditional neighbor mutation operator (<i>MOEA_CGN</i>), and when applying the suggested heuristic mutation ($p_{hm} \in \{0.5, 0.8\}$) with the model	146

Figure 5.15	Detection reliability (left: NMI_{avg} , right: $WNMI_{avg}$) for the five competent MOO models over the complex-real networks (top: Zachary's network, bottom: Dolphins network) with different mutation probabilities, $p_m = 0.2$ for traditional mutation and $p_{hm} = \{0.5, 0.8\}$ for heuristic mutation	149
Figure 5.16	Detection reliability (left: NMI_{avg} , right: $WNMI_{avg}$) for the five competent MOO models over the complex-real networks (top: Books network, bottom: Football network) with different mutation probabilities, $p_m = 0.2$ for traditional mutation and $p_{hm} = \{0.5, 0.8\}$ for heuristic mutation	150
Figure 5.17	Detection reliability (Q_{avg}) for the 5 MOO models on the complex-real networks with unknown real-partitions at different mutation probabilities, $p_m = 0.2$ for traditional mutation and $p_{hm} = \{0.5, 0.8\}$ for heuristic mutation	154
Figure 5.18	Pareto Front of the Zachary network with 15 nondominated solutions obtained by the MOEA_CGN model. NMI values were also presented for each solution	161
Figure 5.19	Four different topological partitions of the Zachary network, where (a) two clusters; (b) three clusters; (c) four clusters and (d) five clusters. The vertices were drawn with different colors in the various clusters	161
Figure 5.20	Pareto Fronts for three real networks obtained by MOEA_CGN model, where (a) the Bottlenose Dolphins network, (b) the Krebs' Books network, (c) the American College Football network	163
Figure 5.21	Four different topological partitions of the Dolphins network, where (a) two clusters; (b) three clusters; (c) four clusters and (d) five clusters. The vertices were drawn with different colors in the various clusters	163
Figure 5.22	Four different topological partitions of the SFI network, where (a) 13 clusters; (b) 14 clusters; (c) 15 clusters and (d) 16 clusters. The vertices were drawn with different colors in the various clusters	164
Figure 5.23	Four different topological partitions of the Flight MH370 network, where (a) 11 clusters; (b) 12 clusters; (c) 13 clusters and (d) 14 clusters. The	166

vertices were drawn with different colors in the various clusters

Figure 6.1	Detection reliability of protein complexes in <i>PPI_D1</i> network in terms of recall, precision and F-measure versus different σ_{OS} values for the proposed method at different mutation versions	174
Figure 6.2	Detection reliability of protein complexes in <i>PPI_D2</i> network in terms of recall, precision and F-measure versus different σ_{OS} values for the proposed method at different mutation versions	175
Figure 6.3	Detection reliability of protein complexes in <i>PPI_D1</i> network in terms of recall, precision and F-measure for the five <i>MOO</i> models at different mutation versions when $\sigma_{OS} = \{0.2, 0.4\}$	177
Figure 6.4	Detection reliability of protein complexes in <i>PPI_D1</i> network in terms of recall, precision and F-measure for the five <i>MOO</i> models at different mutation versions when $\sigma_{OS} = \{0.6, 0.8\}$	178
Figure 6.5	Detection reliability of protein complexes in <i>PPI_D2</i> network in terms of recall, precision and F-measure for the five <i>MOO</i> models at different mutation versions when $\sigma_{OS} = \{0.2, 0.4\}$	180
Figure 6.6	Detection reliability of protein complexes in <i>PPI_D2</i> network in terms of recall, precision and F-measure for the five <i>MOO</i> models at different mutation versions when $\sigma_{OS} = \{0.6, 0.8\}$	181
Figure 6.7	Detection reliability in terms of recall, precision and F-measure for MOEA_CGN method on <i>PPI_D1</i> network versus different σ_{OS} values, when increasing the percentage of randomly added edges	185
Figure 6.8	Detection reliability in terms of recall, precision and F-measure for MOEA_CGN method on <i>PPI_D1</i> network versus different σ_{OS} values, when increasing the percentage of randomly removed edges	185
Figure 6.9	An illustrative example of complex detection in <i>PPI_D1</i>	188

LIST OF ABBREVIATIONS

ACO	-	Ant Colony Optimization
BA	-	Bat Algorithm
BCO	-	Bee Colony Optimization
BIAs	-	Bio-Inspired Approaches
CA	-	Cultural Algorithm
CAEA	-	Conical Area Evolutionary Algorithm
CC-GA	-	Clustering Coefficient-based Genetic Algorithm
CD	-	Community Detection
CO	-	Conductance
CPL	-	Characteristic Path Length
CR	-	Cut Ratio
CS	-	Cuckoo Search Algorithm
CS	-	Community Score
DE	-	Differential Evolution
DIMMOEA/D	-	Discrete Inverse Modelling-based Multi-Objective Evolutionary Algorithm with Decomposition
DPSO	-	Discrete Particle Swarm Optimization
EAs	-	Evolutionary Algorithms
EP	-	Exterior Population
EX	-	Expansion
FA	-	Firefly Algorithm
GA	-	Genetic Algorithm
GACD	-	Genetic Algorithm for Community Detection
GDV	-	Graphlet Degree Vector
GN	-	Girvan and Newman
GO	-	Gene Ontology
HTO	-	Heuristic Turbulence Operator
ID	-	Internal Density
KKM	-	Kernel K-Means
LFR	-	Lancichinetti-Fortunato-Radicchi

LMOEA	-	Local information based Multi-Objective Evolutionary Algorithm
MA	-	Memetic Algorithm
MCL	-	Markov Cluster algorithm
MCODE	-	Molecular Complex Detection
MDCL	-	Multi-objective Discrete Cuckoo search algorithm with Local search
MDPSO	-	Modified Discrete Particle Swarm Optimization
MGOC	-	Multiobjective Gene Ontology-based Clustering
MIPS	-	Munich Information Center for Protein Sequences
MOACO	-	Multi-Objective Ant Colony Optimization
MOACO/D-Net	-	Multi-Objective Ant Colony Optimization algorithm based on Decomposition
MOCD	-	Multi-Objective Community Detection Algorithm
MODAPROC	-	Multi-Objective Disease-Associated Protein Complex
MODPSO	-	MultiObjective Discrete Particle Swarm Optimization
MOEA/D	-	MultiObjective Evolutionary Algorithm Based on Decomposition
MOEA_CGN	-	MultiObjective Evolutionary Algorithm based on Cooperation between Graphlet-based measure and Neighborhood relations
MOEAs	-	MultiObjective Evolutionary Algorithms
MOEPGA	-	MultiObjective Evolutionary Programming Genetic Algorithm
MOGA_Net	-	MultiObjective Genetic Algorithms for networks
MOO	-	MultiObjective Optimization
MOP	-	Multiobjective Optimization Problem
MOPSO-Net	-	Multi-Objective Particle Swarm Optimization
NC	-	Normalized Cut
NMI	-	Normalized Mutual Information
NP-hard	-	Non-deterministic Polynomial-time hard
NNIA	-	Nondominated Neighbor Immune Algorithm
NRA	-	Negative Ratio Association

NS	-	Neighborhood Similarity
NSGA-II	-	Nondominated Sorting Genetic Algorithm II
S	-	Signature Similarity
OS	-	Overlapping Score
PCACD	-	Parallel Conical Area Community Detection
PF	-	Pareto Optimal Front
PGP	-	Pretty-Good Privacy
PPI	-	Protein-Protein Interaction network
PROCOMOSS	-	Protein Complex Detection using Multi-objective Evolutionary Approach based on Semantic Similarity
PSO	-	Particle Swarm Optimization
QDM-PSO	-	Quantum-based Discrete Multi-Objective Particle Swarm Optimization
RC	-	Ratio Cut
RMOEA	-	Reduction based- MultiObjective Evolutionary Algorithm
RNSC	-	Restricted Neighborhood Search Clustering
SCF	-	<i>Scaled cost function</i>
SOO	-	Single Objective Optimization
TS	-	Topological Similarity
US	-	United States
WNMI	-	Weighted Normalized Mutual Information
WOCDA	-	Whale Optimization-based Community Detection Algorithm
WOS	-	Web of Science
X	-	Crossover

LIST OF SYMBOLS

A	-	Adjacency or connection matrix
C	-	Cluster (group of vertices)
\mathcal{C}	-	Detected partition
\mathcal{C}^*	-	True partition
E	-	Set of edges
I	-	Individual
G	-	Undirected graph
K	-	Number of clusters
m	-	Number of features
$m(G)$	-	Volume of graph
$m(v)$	-	Degree of any vertex
$m(C)$	-	Intra-cluster connections (Volume of cluster C)
$\bar{m}(C)$	-	Inter-cluster connections
N	-	Population size
n	-	Number of nodes
$n(C)$	-	Cardinality of cluster
\mathbb{P}	-	Population (set of individual)
P	-	Current solution
p_c	-	Probability of crossover
p_{hm}	-	Probability of the proposed heuristic mutation
p_{ihm}	-	Probability of the proposed improved heuristic mutation
p_m	-	Probability of traditional mutation
Q	-	Modularity
r	-	Random number
s	-	Size of neighbors
gen	-	Generation
V	-	Set of nodes
\mathcal{N}	-	Complex network
z^*	-	Best reference point

\mathcal{F}	-	Fitness function
Ω	-	Solutions space
γ	-	Mixing parameter
λ	-	Weight vector
Γ	-	Decoding function
α	-	Random number
c	-	Confusion matrix
\mathfrak{g}	-	Set of neighbors
Λ	-	Set of weight vectors
X	-	Decision variables vector
X^*	-	Best solution
k	-	No. of objectives
Q	-	Modularity
Con	-	Contribution of a node to cluster
CC	-	Closeness centrality of a node to cluster
$maxgen$	-	Maximum number of generations
I'	-	Individual after applying heuristic mutation operator
$S(u, v)$	-	Signature similarity between two nodes' signature
$D(u, v)$	-	Distance between two nodes' signature

CHAPTER 1

INTRODUCTION

1.1 Introduction

Nowadays, many real complex systems from various fields of sociology, biology, engineering, and communication can be modelled and studied as complex networks of connected communities. In fact, complex networks are an effective formalism in representing the relationships among objects composing many real world systems. Protein interaction networks, neural networks, collaboration networks, world-wide-web networks, Internet network, communication networks, paper citation networks, transport networks, biological and metabolic networks are some examples of complex real-life networks (Pizzuti, 2018; Cheng *et al.*, 2018; Pourkazemi and Keyvanpour, 2017). In general, complex networks can be represented as an undirected graph included all the given system's entities, wherein the vertices of the network correspond to the objects of the system, and the edge (or link) between a pair of objects corresponds to the activities between them. For instance, in biological networks, Protein-Protein Interaction (PPI) network is shaped by considering the collaboration between proteins, wherein the proteins represented the network cardinality (nodes), and the functional or physical interactions between proteins represented the network volume (edges) (Elmsallati *et al.*, 2016; Guzzi and Milenković, 2017). In social complex networks, on the other hand, the objects or entities of the network can be people, and the edges of the network match the various social relations between people like friendship, familiarities, and everyday interests, etc.

Due to increasing the appearance of such complex systems and many other real networked systems, this has fostered the desire to solve and analyze them into organized-topological structures; under the name of modules, clusters, or communities using different mechanisms, in order to provide some organizational principles and fundamentals for real-world networks (Žalik and Žalik, 2018a; Pizzuti, 2018).

Detecting hidden communities in such complex networks has considered a crucial and proceeding challenge since lots of networked problems such as clustering, aligning and searching for relationships are corresponding to the subgraph isomorphism problem known in the literature as NP-hard (Non-deterministic Polynomial-time hard) that lately witnessed a considerable interest (Said *et al.*, 2018; Rao *et al.*, 2018; Elmsallati *et al.*, 2016; Cook, 1971). Research on complex networks analysis and community detection is of a prime interest and has a multidisciplinary nature coming from the graph theory, physics, statistics, and data mining. The problem has been widely addressed as a data analysis problem where Girvan and Newman formally stated it as a community structure detection problem or shortly, Community Detection (CD) problem (Girvan and Newman 2002; Newman and Girvan, 2004). Community detection, or community structure mining in complex networks, means partitioning a given complex network represented as an undirected sparse graph, into communities or groups of vertices (or nodes) having dense intra connections among them and sparse interconnections with the nodes of other clusters.

Thus, a community detection algorithm should be able to divide a complex network into communities with relatively sparse inter-connections and relatively dense intra-connections. This problem may not be accurately solved by using traditional methods due to its complexity and intractable computations where it has been classified as a highly combinatorial optimization NP-hard problem (Gach and Hao, 2012), so it solved using evolutionary computation methods, and thereby it is modelled in the literature as an optimization problem. In the last decade, multiobjective evolutionary algorithms have given a significant contribution in detecting complex community structure and outperforming the single objective evolutionary algorithms. A main characteristic of the multiobjective approach is that the set of Pareto optimal solutions reveals the hierarchical organization of the network. This peculiarity gives a great chance to analyze the network at various hierarchical levels and study communities with different modular levels. Hence, the decision-makers can make their decisions easily (Pizzuti, 2018; Pourkazemi and Keyvanpour, 2017).

All earlier community detection- based evolutionary algorithms focused on solving complex community detection by optimizing neighborhood relations in terms

of intra- and inter- community connections. But most of them, if not all, ignored employing graphlet measure in terms of nodes signature and nodes signature similarity along with the neighborhood relations in defining community structure in complex networks. Solving complex community detection problem by considering cooperation between graphlet-based measures and neighborhood relations has a significant impact in identifying ambiguous nodes to more suitable communities. Since it greatly helps to investigate the entire area of the node's vicinity by looking at all types of graphlet that a node can touch, and hence, the detection accuracy will be improved notably. In this study, a new multi objective method is proposed termed *MOEA_CGN* (MultiObjective Evolutionary Algorithm based on Cooperation between Graphlet-based measure and Neighborhood relations) to detect the structure of topological and biological communities more accurately than the earlier methods. The purpose of this chapter is to provide an overview of the research, including problem statement, the research goal, the objectives, the scope, and the significance of the research.

1.2 Problem Background

The last decade has witnessed the emergence of several novel evolutionary computation-based community detection approaches. In which several heuristics have been customized and developed using various types of optimizing objective functions, which either be single objective optimization or multiobjectives optimization. Based on the term of the optimization functions, only one objective function is optimized in the single objective optimization techniques. In contrast, many objective functions are simultaneously optimized in MultiObjective Optimization (MOO) techniques. Here, a few recent studies of MultiObjective Evolutionary Algorithm (MOEA) have been reviewed that achieved prominent performance in addressing the community detection problem and proved their ability to overcome the drawbacks of the single-objective evolutionary algorithm and the other traditional techniques.

The use of multiobjective frameworks to define the community structure in complex networks was first proposed by Pizzuti in 2009, 2012. The researcher used an NSGA-II framework for formulating *MOGA – Net*, a multiobjective model to

maximize two objective functions, namely, *community score* and *community fitness*. At the same year, Shi *et al.* (2012) defined the community structure as a multiobjective minimization model, called MOCD, wherein the modularity function (Q) was split into two terms. Based on the formulations given by Shi *et al.* (2012) and Pizzuti (2012), the emphasis was on neighborhood relations in terms of intra- community connections while the influence of the inter-community connections is either implicitly or indirectly optimized (Attea *et al.*, 2016). The connections between the network nodes is one of the key concepts to define community detection problem, so it will be right to separate those connections that link set of nodes within a single cluster or community explicitly from those that link nodes with different clusters or communities by considering both inter- and intra- cluster connections and nodes' degree generalization (i.e., graphlet measures) to define community structure effectively.

Gong *et al.* (2012) followed a similar method to that of Shi *et al.* (2012) by splitting the *modularity density* formula into two terms, namely, Negative Ratio Association (*NRA*) and Ratio Cut (*RC*), to detect community structure in complex networks. Accordingly, they formulated the problem as a multiobjective minimization model, called *MOEA/D_Net* (multiobjective evolutionary algorithm with decomposition). After that, Gong *et al.* (2014) reformulated *MOEA/D_Net* model into an effective model termed *MODPSO* (multiobjective discrete particle swarm optimization based on decomposition), by changing the first objective from *NRA* into the Kernel K-Means (*KKM*), while *RC* remained as a second objective. Although the proposed models by Gong *et al.* (2012) and (2014) explicitly emphasized the impact of neighborhood relations in terms of intra- and inter- community connections, but their models did not able to detect the ideal community structure that describing main characteristics of all networks and their relationships in the best way. To effectively address the issue of community structure definition, this study has proposed a new multiobjective model based on cooperative between graphlet-based measures (in terms of nodes signature and nodes signature similarity) and neighborhood relations (in terms of intra- and inter- community connections).

On the other hand, in biological network, community or cluster structure definition has also been formulated by considering the topological objectives based on neighborhood relations only, which impacted on the developed methods' performance and restricted them from detecting a more accurate structure of complex biological community. Cao *et al.* (2015) proposed a new clustering algorithm termed *MOEPGA* (MultiObjective Evolutionary Programming Genetic Algorithm). In *MOEPGA*, several topological properties of the network, i.e. size, *CPL* (characteristic path length) and density have been integrated to detect meaningful protein complexes and conducting a systematic analysis to the problem with regards identifying complexes of protein in the benchmark PPI networks. Bandyopadhyay *et al.* (2015) introduced a multiobjective formulation to address the problem by optimizing three distinct criteria in a maximization language. The framework included two topological features of PPI networks that were formulated into two different quality functions which are the contribution of a node to cluster and closeness centrality of a node to cluster. Moreover, the semantic similarity of GO (Gene Ontology) was used as a third objective function. Ray *et al.* (2016) developed a method called *MODAPROC* (**M**ulti-**O**bjective **D**isease-**A**ssociated **P**rotein **C**omplex) based on a modified framework of *NSGA – II*. Three objective functions, namely, disease association, complex density, and outward interaction, were optimized in *MODAPROC* to discover and find disease-specific gene complexes. A parallel optimization-based multiobjective method called *DCRS* was proposed by Sharma and Bhattacharyya (2018). The aim of developing *DCRS* was to detect protein complexes and identify their quality using the *NSGA – II* framework. The method's name, *DCRS*, was derived from the four objective functions used in the system, which are density, contribution, reachability contribution, and semantic similarity.

Recently, several methods have been proposed to address community detection problem in complex networks effectively by employing the objective functions of either Pizzuti (2012), Shi *et al.* (2012), Gong *et al.* (2012), or Gong *et al.* (2014); and developing a local search operator to enhance the detection accuracy. Sani *et al.* (2018) introduced a new method named *MOACO* (Multi-Objective Ant Colony Optimization) to address the problem by applying Pareto theory (i.e., Pareto dominance and Pareto Archive) and Pearson Correlation measure as heuristic information along with the two

objectives of Pizzuti (2012) (community score and community fitness). Rahimi *et al.* (2018) developed the *MOPSO – Net* method (Multi-Objective Particle Swarm Optimization) based upon a modified version of PSO that revolved around changing the particles' movement in the swarm. The technique used two competing objectives of Gong *et al.* (2014) (*KKM and RC*). Ji *et al.* (2019) proposed a multiobjective community detection method termed *MOCD – ACO* that combined the MultiObjective Evolutionary Algorithm based-Decomposition (*MOEA/D*) along with the local heuristic operator of *ACO* (Ant Colony Optimization). In the *MOCD – ACO*, two fitness functions of Gong *et al.* (2012) (*NRA and RC*) were simultaneously optimized. Besides, the search range was expanded by integrating a local search operator named weighted simulated annealing into the framework. As a result, the combined technique improved the search process by avoiding premature convergence during the optimization process and achieved better performance than other competent algorithms. Zou *et al.* (2019) proposed *DIMMOEA/D* (Discrete Inverse Modelling-based Multi-Objective Evolutionary Algorithm with Decomposition), a new algorithm to detect the embedded community structure in real and benchmark complex networks. By using the network topology properties, an inverse modelling strategy was applied to sample the area of objective space and produce a new offspring; and to measure the fitness quality of each generated solution, the authors used two conflicting criteria of Gong *et al.* (2012) (*NRA and RC*). Mu *et al.* (2019) developed *MOACO/D – Net* (Multi Objective Ant Colony Optimization algorithm based on Decomposition) to address main difficulties in the *ACO* algorithm based community detection problem. *MOACO/D – Net* solved the problem by simultaneously minimizing two objectives of Gong *et al.* (2012) (*RC and NRA*) in the optimization process. New solutions were constructing by the ants using the new proposed transition probability model. Then, an improvement operation was performed to enhance each solution based on the strong communities' concept.

Despite the increasing interest in developing effective local search operators based on general neighborhood relations, most of the developed methods are still suffering from the resolution limit problem and restricting the solution from going deep into the network structure to find accurate clusters. To address resolution limit

problem effectively, this study proposed a new heuristic mutation operator based on cooperation between nodes signature similarity and neighborhood relations.

It is worth to note that all the above explored studies have mainly interested on addressing community detection problem in either: social complex networks with known ground-truth and real-complex networks with unknown ground-truth, or in PPI networks with known generic complexes. This means that there is a need to develop a new method addresses the heterogeneity issue of real-world networks to detect more accurate topological and biological community structure in complex networks derived from various domains.

To this end, the research aims to propose *MOEA_CGN* method to improve the predictive power of the evolutionary-based community detection algorithm. By addressing community detection challenges (which is community structure definition issue, resolution limit problem, and heterogeneity issue of real-world networks) and transcend the limitations in the earlier studies through exploiting neighborhood relations along with strong-theoretically grounded topological properties derived from a network (i.e., graphlet-based measures) in two main components of the adopted optimization method which are the objective functions and mutation operator.

1.3 Problem Statement

Community detection in complex networks is an ill-defined problem as there are no typical accepted protocols on the fundamental ingredients such as defining community itself (Fortunato and Hric, 2016; Javed *et al.*, 2018). The broad issue for the community detection problem in complex networks is the community structure definition. What this means in the context of an evolutionary algorithm is the development of a meaningful mathematical formula (objective functions) that effectively defines the topological structure of the hidden community in the complex networks. Many ideas have emerged to define the community structure, which in turn led to different definitions of single objective or multiobjective optimization functions. Most of the studies, if not all, defined the community structure in a complex network

by solely considering the general neighborhood relations of the complex network and lacked complete reflection of inherent topological properties (i.e., graphlet-based measures in terms of nodes' signature and nodes' signature similarity) that can address the issue of a community structure definition effectively. Another limitation is that most of the studies aspired to the traditional or randomize explorative property only. They lacked to design any other heuristic evolutionary operator based on strong, systematic topological properties (i.e., graphlet-based measures in terms of nodes' signature and nodes' signature similarity) to avoid resolution limit problem that restrict solutions from going deep into the network structure to find the precise complexes and accurately representing the domain knowledge. The growth of real-world networks from different domains (i.e., social and biological) is inevitable. But these real networks are characterized by topological heterogeneity and differed in some networked characteristics (volume and cardinality) depending on their domains, and this considered another discrete issue for the CD algorithm.

In summary, in this study, three main issues and challenges facing the community detection problem in complex networks have been considered, namely community structure definition issue, resolution limit problem, and heterogeneity issue of real-world networks. In view of that, this research has been guided by the following main research question:

How to design a multiobjective evolutionary algorithm based on cooperation between graphlet-based measure and neighborhood relations (*MOEA_CGN*) for detection of complex topological and biological community structure accurately?

This question will be fulfilled by answering detailed questions:

1. How a *MOEA_CGN* method with two new objective functions can be designed to define complex topological community structure?
2. How to improve *MOEA_CGN* method by making it more likely to converge to more accurate solutions? In other words, how to enhance *MOEA_CGN* method with a new heuristic mutation operator to avoid resolution limit problem?

3. How to make the *MOEA_CGN* method regarded a reliable and dependable method in detecting complex biological communities accurately?

These questions can be answered by designing a CD method that invests the robust topological properties of networks and enhances the search properties in terms of intensification and diversification, as these properties considered essential to improve the ability of the adopted searching algorithm and detecting a more accurate community structure.

1.4 Research Goal

The goal of the study is to propose a new method termed *MOEA_CGN* (**M**ulti**O**bjective **E**volutionary Algorithm based on **C**ooperation between **G**raphlet-based measure and **N**eighborhood relations) to detect topological and biological community structure in complex networks. *MOEA_CGN* method takes into account addressing three main issues, namely, community structure definition issue, resolution limit problem, and heterogeneity issue of real-world networks, to improve the quality of community detection in terms of accuracy and velocity.

1.5 Research Objectives

The following are the objectives of the proposed study to detect topological and biological community structure in complex networks derived from different domains:-

1. To propose *MOEA_CGN* method with two new objective functions based on cooperation between neighborhood relations and strong, systematic topological measures (i.e., graphlet-based measures) to address the issue of community structure definition.

2. To enhance *MOEA_CGN* method with a new mutation operator based on a new qualitative definition of the topological similarity measure to address resolution limit issue.
3. To improve the proposed mutation operator based on the distance between the nodes' signature to make *MOEA_CGN* dependable method in detecting complex biological communities accurately.

1.6 Research Scope

The scope of the research is listed as below:

- i. This study has focused on detection topological and biological communities in complex networks derived from different domains.
- ii. The datasets used in this study are:
 - Synthetic-benchmark networks: five networks of GN: Girvan Newman, (2002) and sixteen networks of LFR: Lancichinetti-Fortunato-Radicchi, (2008) have been used.
 - Real-complex networks with known ground-truth: four standard real-complex networks of various sizes have been used.
 - Real-complex networks with unknown ground-truth: eight standard real-complex networks of various sizes have been used
 - Biological networks with known ground-truth: two Protein-Protein Interaction networks (PPI) have been used.
- iii. This study has proposed *MOEA_CGN* method based on *MOEA/D* (multiobjective evolutionary algorithm based on decomposition) framework of Zhang and Li (2007). *MOEA/D* is considered a powerful search and optimization mechanism as it providing an effective method to solve several NP-hard optimization problems. At the meanwhile, it never needs a high level of mathematical knowledge, and its computational complexity is low,

compared to other methods (Gong *et al.*, 2012; Cheng *et al.*, 2018; Zhang and Li, 2007; Zou *et al.*, 2019, Ying *et al.*, 2019). These features have made MOEA/D framework appropriate for solving the CD problem when a multiobjective evolutionary algorithm is needed.

- iv. The performance is measured according to detection velocity (with regards to the maximum number of generations required to get the optimal solution), and detection reliability (with regards to the average of the best value of the following measures: Modularity (Q), Normalized Mutual Information (NMI), Weighted Normalized Mutual Information ($WNMI$), Recall, Precision, and F-measure).
- v. The proposed method is simulated using Matlab R2016b. The experiments over small networks have been performed on a PC with an IntelR CoreTM i7-7700HQ CPU, 16 GB RAM under Windows 10 x64. While the experiments over large networks have been performed on a High Performance Computers (HPC) with 32-core 2.6GHz Intel Xeon E7-8895 server.

1.7 Research Significance

Community detection mainly aims at discovering the inherent structure of community within network or providing useful information on the organization of a network. Community detection allows us to focus on the areas having some degree of independence within a graph. It helps in the classification of the vertices according to their role in the communities to which they belong. For instance, it may be possible to distinguish entities fully embedded within groups from those on the boundary of the group that may act as an intermediary between units. Such boundary nodes can play an essential role in installing the units together and in the procedures of deploying the processes over the network (Chakraborty *et al.*, 2017; Fortunato and Hric, 2016). The growing demand for developing effective frameworks to discover such community structure in complex networks comes from its contribution to many significant vital applications. In social networks, for instance, organizations or individuals are bonded through several social contacts, profiles, or familial. Thus, it can be said that social

modularity represented detection a group of individuals linked with dense social ties. Detecting social modularity is useful in social network analysis, network summarization and privacy, i.e. member grouping and group segmentation (Karataş and Şahin, 2018, Javed *et al.*, 2018). In general, the presence of community structure in complex networks made available an assortment of applications for example targeted marketing, grouping social media subscribers for good advertising in addition to facilitating recommendations for readers, discovering fraudulent sites, discovering fraudulent communication networks activities, designing network protocols in delay tolerant networks, link prediction, refactoring software packages, dimensional reduction in pattern recognition, recommendation systems, skills acquisition in robotics (Karataş and Şahin, 2018; Rahimi *et al.*, 2018) and tracking bitcoin users (Remy *et al.*, 2017). With the increasing demand for these and many other real-world applications, the community structure seeks to model the basic features of these complex networked-systems by determining sub-structures and their topological features.

In biological networks, cell activity in PPI networks can be comprehended by analyzing those proteins that are organized as separate and interactable units. Hence, the PPI modularity refers to detection a set of proteins that physically or functionally interact to achieve a specific function. Additionally, detection of protein complexes in PPI networks can help in many significant therapeutic purposes, describing the evolutionary orthology signal (Jancura *et al.*, 2012), detecting cancer, and predicting the biological functions of yet unrecognized proteins (Pizzuti, 2014). On the other hand, a change in the pattern of proteins interaction is associated with many diseases and identification of such interactions through detecting protein complexes can aid in disease diagnosis (Lei *et al.*, 2018; Bandyopadhyay *et al.*, 2015). In fact, the network can be better understood by knowing the organization of the objects; due to it will provide better information which may be missed by considering the network as a whole.

1.8 Thesis Organization

This thesis is organized into 7 chapters. Chapter 2 reviews some basic concepts, the related works of this study and discusses the gaps that led to formulating the research problem. Chapter 3 presents the methodology that this research followed as well as exploring and discussing the details of the framework components and the datasets being used. The experimental results of the proposed MOEA_CGN method based on general MOEA/D framework are presented in Chapter 4. Chapter 5 presents the experimental results of the proposed *MOEA_CGN* method with a new heuristic mutation operator. In Chapter 6, the experimental results of the proposed *MOEA_CGN* method with an improved heuristic mutation operator are presented in order to detect complex biological communities accurately. Finally, Chapter 7 concludes the research, future works and some suggestions are provided in this chapter.

REFERENCES

- Adamcsek, B., Palla, G., Farkas, I. J., Derényi, I., and Vicsek, T. (2006). CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics*, 22(8), 1021-1023.
- Agrawal, R., 2011. Bi-objective community detection (BOCD) in networks using genetic algorithm. *Proceedings of 4th International Conference IC3 2011 India Contemporary Computing Springer*, 2011. 5-15.
- Al-Taie, M. Z., Shamsuddin, S. M., & Ahmad, N. B. (2014). Flight MH370 community structure. *Int. J. Advance. Soft Comput. Appl*, 6(2), 33487-2742.
- Amiri, B., Hossain, L., & Crawford, J. (2012, July). A multiobjective hybrid evolutionary algorithm for clustering in social networks. *In Proceedings of the 14th annual conference companion on Genetic and evolutionary computation* (1445-1446). ACM.
- Amiri, B., Hossain, L., Crawford, J.W. and Wigand, R.T., 2013. Community detection in complex networks: Multi-objective enhanced firefly algorithm. *Knowledge-Based Systems*, 46, 1-11.
- Arnau, V., Mars, S., and Marín, I. (2005). Iterative cluster analysis of protein interaction data. *Bioinformatics*, 21(3), 364-378.
- Attea, B. A., & Abdullah, Q. Z. (2018). Improving the performance of evolutionary-based complex detection models in protein-protein interaction networks. *Soft Computing*, 22(11), 3721-3744.
- Attea, B. A., Hariz, W. A., & Abdulhalim, M. F. (2016). Improving the performance of evolutionary multi-objective co-clustering models for community detection in complex social networks. *Swarm and Evolutionary Computation*, 26, 137-156.
- Babers, R., & Hassanien, A. E. (2017). A nature-inspired metaheuristic cuckoo search algorithm for community detection in social networks. *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, 8(1), 50-62.
- Bandyopadhyay, S., Ray, S., Mukhopadhyay, A., and Maulik, U. (2015). A multiobjective approach for identifying protein complexes and studying their

- association in multiple disorders. *Algorithms for Molecular Biology*, 10(1), 1-15.
- Barabasi, A. L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2), 101-113.
- Bello-Orgaz, G., & Camacho, D. (2014, July). Evolutionary clustering algorithm for community detection using graph-based information. In *Evolutionary Computation (CEC), 2014 IEEE congress on* (930-937). IEEE.
- Boguná, M., Pastor-Satorras, R., Díaz-Guilera, A., & Arenas, A. (2004). Models of social networks based on social distance attachment. *Physical review E*, 70(5), 056122.
- Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., and Wagner, D. (2008). On modularity clustering. *IEEE transactions on knowledge and data engineering*, 20(2), 172-188.
- Brohée, S. and van Helden, J. (2006). Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics*, 7(1), 1-19.
- Cai, Q., Gong, M., Ma, L., Ruan, S., Yuan, F., & Jiao, L. (2015). Greedy discrete particle swarm optimization for large-scale social network clustering. *Information Sciences*, 316, 503-516.
- Cai, Q., Ma, L., Gong, M., and Tian, D. (2016). A survey on network community detection based on evolutionary computation. *International Journal of Bio-Inspired Computation*, 8(2), 84-98.
- Cai, X., Shi, Y., Zhu, Y., Qiao, Y., & Hu, F. (2017, October). An Algorithm Q-PSO for Community Detection in Complex Networks. In *Distributed Computing and Applications to Business, Engineering and Science (DCABES), 2017 16th International Symposium on* (76-79). IEEE.
- Cao, B., Luo, J., Liang, C., Wang, S., & Song, D. (2015). Moepga: A novel method to detect protein complexes in yeast protein-protein interaction networks based on multiobjective evolutionary programming genetic algorithm. *Computational biology and chemistry*, 58, 173-181.
- Chakraborty, T., Dalmia, A., Mukherjee, A., & Ganguly, N. (2017). Metrics for community analysis: A survey. *ACM Computing Surveys (CSUR)*, 50(4), 54.
- Chen, B., & Wu, F. X. (2013). Identifying protein complexes based on multiple topological structures in PPI networks. *IEEE Transactions on Nanobioscience*, 12(3), 165-172.

- Chen, B., Chen, L., & Chen, Y. (2012, July). Detecting community structure in networks based on ant colony optimization. In *International Conference on Information & Knowledge Engineering* (247-253).
- Chen, Z., Liu, F., Gao, C., Li, X., & Zhang, Z. (2017, July). An Enhanced Particle Swarm Optimization Based on Physarum Model for Community Detection. In *International Conference in Swarm Intelligence* (99-108). Springer, Cham.
- Cheng, F., Cui, T., Su, Y., Niu, Y., and Zhang, X. (2018). A local information based multi-objective evolutionary algorithm for community detection in complex networks. *Applied Soft Computing*, 69, 357-367.
- Chira, C., & Gog, A. (2011, May). Collaborative community detection in complex networks. In *International Conference on Hybrid Artificial Intelligence Systems* (380-387). Springer, Berlin, Heidelberg.
- Chira, C., Gog, A., & Iclănzan, D. (2012, June). Evolutionary detection of community structures in complex networks: A new fitness function. In *Evolutionary Computation (CEC), 2012 IEEE Congress on* (1-8). IEEE.
- Cho, Y. R., Hwang, W., Ramanathan, M., and Zhang, A. (2007). Semantic integration to identify overlapping functional modules in protein interaction networks. *BMC bioinformatics*, 8(1), 1-13.
- Chunyu, W., & Yun, P. (2015). Discrete bat algorithm and application in community detection. *The Open Cybernetics & Systemics Journal*, 9, 967-972.
- Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6), 066111.
- Coello, C.A.C., Van Veldhuizen, D.A. and Lamont, G.B., (2002). *Evolutionary algorithms for solving multi-objective problems* (Vol. 242). New York: Kluwer Academic.
- Cook, S. A. (1971, May). The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing* (pp. 151-158). ACM.
- Coscia, M., Giannotti, F., and Pedreschi, D. (2011). A classification for community discovery methods in complex networks. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 4(5), 512-546.
- de Andrade, L. P., Espíndola, R. P., & Ebecken, N. F. F. (2014, June). Community detection by an efficient ant colony approach. In *International Symposium on Methodologies for Intelligent Systems* (1-9). Springer, Cham.

- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. A. M. T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation*, 6(2), 182-197.
- Del Ser, J., Lobo, J. L., Villar-Rodriguez, E., Bilbao, M. N., & Perfecto, C. (2016, July). Community detection in graphs based on surprise maximization using firefly heuristics. In *Evolutionary Computation (CEC), 2016 IEEE Congress on* (2233-2239). IEEE.
- Deng, K., Zhang, J. P., & Yang, J. (2015). An Efficient Multi-Objective Community Detection Algorithm In Complex Networks. *Tehnicki vjesnik/Technical Gazette*, 22(2):319-328.
- Ding, C., *et al.* (2001). A MinMaxCut spectral method for data clustering and graph partitioning. *Proc. IEEE Int'l Conf. Data Mining. 2001*;107–14.
- Elmsallati, A., Clark, C., and Kalita, J. (2016). Global alignment of protein-protein interaction networks: A survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(4), 689-705.
- Firat, A., Chatterjee, S., & Yilmaz, M. (2007). Genetic clustering of social networks using random walks. *Computational Statistics & Data Analysis*, 51(12), 6285-6294.
- Folino, F., and Pizzuti, C. (2010, August). A multiobjective and evolutionary clustering method for dynamic networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on* (256-263). IEEE.
- Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3), 75-174.
- Fortunato, S., and Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, 659, 1-44.
- Gach, O. and Hao, J.K. (2012, September). A memetic algorithm for community detection in complex networks. In *International Conference on Parallel Problem Solving from Nature* (327-336). Springer, Berlin, Heidelberg.
- Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., ... and Edlmann, A. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440(7084), 631.
- Georgii, E., Dietmann, S., Uno, T., Pagel, P., and Tsuda, K. (2009). Enumeration of condition-dependent dense modules in protein interaction networks. *Bioinformatics*, 25(7), 933-940.

- Girvan, M. and Newman, M.E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12), 7821-7826.
- Gleiser, P. M., & Danon, L. (2003). Community structure in jazz. *Advances in complex systems*, 6(04), 565-573.
- Gog, A., Dumitrescu, D. and Hirsbrunner, B. (2007, September). Community detection in complex networks using collaborative evolutionary algorithms. *In European Conference on Artificial Life* (886-894). Springer, Berlin, Heidelberg.
- Gong, M., Cai, Q., Chen, X., and Ma, L. (2014). Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition. *IEEE Transactions on Evolutionary Computation*, 18(1), 82-97.
- Gong, M., Cai, Q., Li, Y., & Ma, J. (2012a, June). An improved memetic algorithm for community detection in complex networks. In *Evolutionary Computation (CEC), 2012 IEEE Congress on* (1-8). IEEE.
- Gong, M., Fu, B., Jiao, L., & Du, H. (2011). Memetic algorithm for community detection in networks. *Physical Review E*, 84(5), 056101.
- Gong, M., Jiao, L., Du, H., and Bo, L. (2008). Multiobjective immune algorithm with nondominated neighbor-based selection. *Evolutionary Computation*, 16(2), 225-255.
- Gong, M., Ma, L., Zhang, Q., & Jiao, L. (2012c). Community detection in networks by using multiobjective evolutionary algorithm with decomposition. *Physica A: Statistical Mechanics and its Applications*, 391(15), 4050-4060.
- Guerrero, M., Montoya, F. G., Baños, R., Alcayde, A., & Gil, C. (2017). Adaptive community detection in complex networks using genetic algorithms. *Neurocomputing*, 266, 101-113.
- Gulbahce, N., & Lehmann, S. (2008). The art of community detection. *BioEssays*, 30(10), 934-938.
- Guo, Y., Li, X., Tang, Y., & Li, J. (2017). Heuristic Artificial Bee Colony Algorithm for Uncovering Community in Complex Networks. *Mathematical Problems in Engineering*, 2017, 4143638, 1-12.
- Guoqiang, C. and Xiaofang, G. (2010, December). A genetic algorithm based on modularity density for detecting community structure in complex networks.

- In *Computational Intelligence and Security (CIS), 2010 International Conference on* (151-154). IEEE.
- Guzzi, P. H., & Milenković, T. (2017). Survey of local and global biological network alignment: the need to reconcile the two sides of the same coin. *Briefings in bioinformatics*, 19(3), 472-481.
- Hafez, A. I., Al-Shammari, E. T., ella Hassanien, A., and Fahmy, A. A. (2014a). Genetic algorithms for multi-objective community detection in complex networks. In *Social Networks: A Framework of Computational Intelligence* (145-171). Springer, Cham.
- Hafez, A. I., Ghali, N. I., Hassanien, A. E., & Fahmy, A. A. (2012, November). Genetic algorithms for community detection in social networks. In *Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on* (460-465). IEEE.
- Hafez, A. I., Zawbaa, H. M., Hassanien, A. E., & Fahmy, A. A. (2014b). Networks community detection using artificial bee colony swarm optimization. In *Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014* (229-239). Springer, Cham.
- Halalai, R., Lemnaru, C. and Potolea, R. (2010, August). Distributed community detection in social networks with genetic algorithms. In *Intelligent Computer Communication and Processing (ICCP), 2010 IEEE International Conference on* (35-41). IEEE.
- Handl, J., & Knowles, J. (2007). An evolutionary approach to multiobjective clustering. *IEEE transactions on Evolutionary Computation*, 11(1), 56-76.
- Haque, M. N., Mathieson, L., & Moscato, P. (2017, November). A memetic algorithm for community detection by maximising the connected cohesion. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (1-8). IEEE.
- Hassan, E. A., Hafez, A. I., Hassanien, A. E., & Fahmy, A. A. (2015, June). A discrete bat algorithm for the community detection problem. In *International Conference on Hybrid Artificial Intelligence Systems* (188-199). Springer, Cham.
- He, D., Liu, J., Liu, D., Jin, D., & Jia, Z. (2011, July). Ant colony optimization for community detection in large-scale complex networks. In *Natural Computation (ICNC), 2011 Seventh International Conference on* (Vol. 2, 1151-1155). IEEE.

- He, D., Liu, J., Yang, B., Huang, Y., Liu, D., & Jin, D. (2012). An ant-based algorithm with local optimization for community detection in large-scale networks. *Advances in Complex Systems*, 15(08), 1250036.
- He, D., Wang, Z., Yang, B., & Zhou, C. (2009, November). Genetic algorithm with ensemble learning for detecting community structure in complex networks. In *Computer Sciences and Convergence Information Technology, 2009. ICCIT'09. Fourth International Conference on* (702-707). IEEE
- Honghao, C., Zuren, F., & Zhigang, R. (2013, June). Community detection using ant colony optimization. In *Evolutionary Computation (CEC), 2013 IEEE Congress on*(3072-3078). IEEE.
- Jancura, P., Mavridou, E., Carrillo-de Santa Pau, E., and Marchiori, E. (2012, June). A methodology for detecting the orthology signal in a PPI network at a functional complex level. In *BMC bioinformatics*, vol. 13, no. 10, 1-13. BioMed Central, 2012.
- Javadi, S. H., Khadivi, S., Shiri, M. E., & Xu, J. (2014, August). An ant colony optimization method to detect communities in social networks. In *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (200-203). IEEE Press.
- Javed, M. A., Younis, M. S., Latif, S., Qadir, J., and Baig, A. (2018). Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications*, 108, 87-111.
- Ji, P., Zhang, S., & Zhou, Z. (2019). A decomposition-based ant colony optimization algorithm for the multi-objective community detection. *Journal of Ambient Intelligence and Humanized Computing*, 11(1), 173-188..
- Jia, G., Cai, Z., Musolesi, M., Wang, Y., Tennant, D.A., Weber, R.J., Heath, J.K. and He, S. (2012). Community detection in social and biological networks using differential evolution. In *International conference on learning and intelligent optimization* (71-85). Springer, Berlin, Heidelberg..
- Jin, D., He, D., Liu, D., & Baquero, C. (2010, October). Genetic algorithm with local search for community mining in complex networks. In *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on* (Vol. 1, 105-112). IEEE.

- Ju, Y., Zhang, S., Ding, N., Zeng, X. and Zhang, X. (2016). Complex network clustering by a multi-objective evolutionary algorithm based on decomposition and membrane structure. *Scientific reports*,6(1), 1-13.
- Kadry, S., & Al-Taie, M. Z. (2014). Social network analysis: An introduction with an extensive implementation to a large-scale online network using Pajek. Bentham Science Publishers.
- Karataş, A., & Şahin, S. (2018, December). Application areas of community detection: A review. In *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)* (65-70). IEE.
- Kaur, S., Singh, S., Kaushal, S., & Sangaiah, A. K. (2016). Comparative analysis of quality metrics for community detection in social networks using genetic algorithm. *Neural Network World*, 26(6), 625.
- Khan, B. S., and Niazi, M. A. (2017). Network community detection: a review and visual survey. *arXiv preprint arXiv:1708.00977*.
- King, A. D., Pržulj, N., and Jurisica, I. (2004). Protein complex prediction via cost-based clustering. *Bioinformatics*, 20(17), 3013-3020.
- Kuchaiev, O., Milenković, T., Memišević, V., Hayes, W., & Pržulj, N. (2010). Topological network alignment uncovers biological function and phylogeny. *Journal of the Royal Society Interface*, 7(50), 1341-1354.
- Lancichinetti, A., Fortunato, S., and Kertész, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3), 033015.
- Lancichinetti, A., Fortunato, S., and Radicchi, F. (2008). Benchmark graphs for testing community detection algorithms. *Physical review E*, 78(4), 046110.
- Lancichinetti, A., Kivela, M., Saramaki, J., & Fortunato, S. (2010). Characterizing the community structure of complex networks. *PloS one*, 5(8), e11976.
- Le, B. D. (2018). *Community detection in complex networks*. PhD Dissertation, University of Adelaide, School of Computer Science, 2018.
- Lei, X., Wang, S., and Pan, L. (2018). Identifying Essential Proteins in Dynamic PPI Network with Improved FOA. *International Journal of Computers Communications and Control*, 13(3), 365-382.
- Li, J. and Song, Y. (2013). Community detection in complex networks using extended compact genetic algorithm. *Soft computing*, 17(6), 925-937.

- Li, L., Jiao, L., Zhao, J., Shang, R. and Gong, M. (2017a). Quantum-behaved discrete multi-objective particle swarm optimization for complex network clustering. *Pattern Recognition*, 63, 1-14.
- Li, S., Chen, Y., Du, H., & Feldman, M. W. (2010). A genetic algorithm with local search strategy for improved detection of community structure. *Complexity*, 15(4), 53-60.
- Li, X. L., Foo, C. S., Tan, S. H., & Ng, S. K. (2005). Interaction graph mining for protein complexes using local clique merging. *Genome Informatics*, 16(2), 260-269.
- Li, Y., Liu, G., & Lao, S. Y. (2013). A genetic algorithm for community detection in complex networks. *Journal of Central South University*, 20(5), 1269-1276.
- Li, Y.H., Wang, J.Q., Wang, X.J., Zhao, Y.L., Lu, X.H. and Liu, D.L. (2017b). Community detection based on differential evolution using social spider optimization. *Symmetry*, 9(9), 183.
- Li, Z., Zhang, S., Wang, R. S., Zhang, X. S., and Chen, L. (2008). Quantitative function for community detection. *Physical review E*, 77(3), 036109.
- Lipczak, M. and Milios, E. (2009, July). Agglomerative genetic algorithm for clustering in social networks. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation* (1243-1250). ACM.
- Liu, C., & Liu, Q. (2018). Community Detection Based on Differential Evolution Using Modularity Density. *Information*, 9(9), 218.
- Liu, D., Jin, D., Baquero, C., He, D., Yang, B., & Yu, Q. (2013). Genetic algorithm with a local search strategy for discovering communities in complex networks. *International Journal of Computational Intelligence Systems*, 6(2), 354-369.
- Liu, H., Yang, F., & Liu, D. (2016, July). Genetic algorithm optimizing modularity for community detection in complex networks. In *2016 35th Chinese Control Conference (CCC)*(1252-1256). IEEE.
- Liu, J., & Zeng, J. (2010, September). Community detection based on modularity density and genetic algorithm. In *Computational Aspects of Social Networks (CASoN), 2010 International Conference on* (29-32). IEEE.
- Liu, J., Zhong, W., Abbass, H.A. and Green, D.G. (2010, July). Separated and overlapping community detection in complex networks using multiobjective

- evolutionary algorithms. In *Evolutionary Computation (CEC), 2010 IEEE Congress on* (1-7). IEEE.
- Liu, R., Liu, J., & He, M. (2019). A multi-objective ant colony optimization with decomposition for community detection in complex networks. *Transactions of the Institute of Measurement and Control*, 41(9), 2521-2534.
- Liu, S., & Li, Z. (2017, February). A modified genetic algorithm for community detection in complex networks. In *Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), 2017 International Conference on* (1-3). IEEE.
- Lusseau, D. (2003). The emergent properties of a dolphin social network. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(suppl_2), S186-S188
- Ma, L., Gong, M., Liu, J., Cai, Q., & Jiao, L. (2014b). Multi-level learning based memetic algorithm for community detection. *Applied Soft Computing*, 19, 121-133.
- MacKay, D. J. (2003). *Information theory, inference, and learning algorithms* (Vol. 7). Cambridge: Cambridge university press.
- Mazur, P., Zmarzłowski, K. and Orłowski, A.J. (2010). Genetic algorithms approach to community detection. *Acta Physica Polonica-Series A General Physics*, 117(4), 703.
- Mewes, H.W., Frishman, D., Güldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Weil, B. (2002). MIPS: a database for genomes and protein sequences. *Nucleic Acids Res* 30(1):31–34.
- Milenković, T., & Pržulj, N. (2008). Uncovering biological network function via graphlet degree signatures. *Cancer informatics*, 6, CIN-S680.
- Milenković, T., Ng, W. L., Hayes, W., & Pržulj, N. (2010). Optimal network alignment with graphlet degree vectors. *Cancer informatics*, 9, CIN-S4744.
- Moradi, M., & Parsa, S. (2019). An evolutionary method for community detection using a novel local search strategy. *Physica A: Statistical Mechanics and its Applications*, 523, 457-475.
- Mu, C., Zhang, J., & Jiao, L. (2014, July). An intelligent ant colony optimization for community detection in complex networks. In *Evolutionary Computation (CEC), 2014 IEEE Congress on* (700-706). IEEE.

- Mu, C., Zhang, J., Liu, Y., Qu, R., & Huang, T. (2019). Multi-objective ant colony optimization algorithm based on decomposition for community detection in complex networks. *Soft Computing*, 23(23), 12683-12709.
- Mu, C.H., Xie, J., Liu, Y., Chen, F., Liu, Y. and Jiao, L.C., 2015. Memetic algorithm with simulated annealing strategy and tightness greedy optimization for community detection in networks. *Applied Soft Computing*, 34, 485-501.
- Mukhopadhyay, A., Ray, S., and De, M. (2012). Detecting protein complexes in a PPI network: a gene ontology based multi-objective evolutionary approach. *Molecular BioSystems*, 8(11), 3036-3048
- Naeni, L. M., Berretta, R., & Moscato, P. (2015). MA-Net: A reliable memetic algorithm for community detection by modularity optimization. In *Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems, Volume 1* (311-323). Springer, Cham.
- Naruchitparames, J., Gunes, M. H., and Louis, S. J. (2011, June). Friend recommendations in social networks using genetic algorithms and network topology. In *Evolutionary Computation (CEC), 2011 IEEE Congress on* (2207-2214). IEEE.
- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), 8577-8582.
- Newman, M. E. (2006b). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3), 036104.
- Newman, M. E., and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2), 026113.
- Oti, M., Snel, B., Huynen, M. A., & Brunner, H. G. (2006). Predicting disease genes using protein–protein interactions. *Journal of medical genetics*, 43(8), 691-698.
- Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043), 814.
- Pei, P., and Zhang, A. (2005). A two-step approach for clustering proteins based on protein interaction profile. In *Proceedings/IEEE Computational Systems Bioinformatics Conference, CSB. IEEE Computational Systems Bioinformatics Conference* (Vol. 2005, No. 1544467, 201). NIH Public Access.

- Pizzuti, C. (2008a, September). GA-NET: A genetic algorithm for community detection in social networks. In *International Conference on Parallel Problem Solving from Nature* (1081-1090). Springer, Berlin, Heidelberg.
- Pizzuti, C. (2008b, July). Community detection in social networks with genetic algorithms. In *Proceedings of the 10th annual conference on Genetic and evolutionary computation*(1137-1138). ACM.
- Pizzuti, C. (2009b, November). A multi-objective genetic algorithm for community detection in networks. In *Tools with Artificial Intelligence, 2009. ICTAI'09. 21st International Conference on* (379-386). IEEE.
- Pizzuti, C. (2012). A multiobjective genetic algorithm to find communities in complex networks. *IEEE Transactions on Evolutionary Computation*, 16(3), 418-430.
- Pizzuti, C. (2014). Computational intelligence for community detection in complex networks and bio-medical applications. PhD Dissertation, Radboud Universiteit Nijmegen [Sl: sn].
- Pizzuti, C. (2018). Evolutionary computation for community detection in networks: a review. *IEEE Transactions on Evolutionary Computation*, 22(3), 464-483.
- Pizzuti, C., and Rombo, S. (2012). Experimental evaluation of topological-based fitness functions to detect complexes in PPI networks. In *Proceedings of the 14th annual conference on Genetic and evolutionary computation* (193-200). ACM.
- Pizzuti, C., and Rombo, S. E. (2014a). Algorithms and tools for protein-protein interaction networks clustering, with a special focus on population-based stochastic methods. *Bioinformatics*, 30(10), 1343-1352.
- Pizzuti, C., and Rombo, S. E. (2014b). An evolutionary restricted neighborhood search clustering approach for PPI networks. *Neurocomputing*, 145, 53-61.
- Pons, P., and Latapy, M. (2006). Computing communities in large networks using random walks. *J. Graph Algorithms Appl.*, 10(2), 191-218.
- Pourkazemi, M. and Keyvanpour, M.R. (2017). Community detection in social network by using a multi-objective evolutionary algorithm. *Intelligent Data Analysis*, 21(2), 385-409.
- Pržulj, N. (2007). Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2), e177-e183.
- Pujol, J. M., Béjar, J., and Delgado, J. (2006). Clustering algorithm for determining community structure in large networks. *Physical Review E*, 74(1), 016107.

- Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., and Parisi, D. (2004). Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9), 2658-2663.
- Rahimi, S., Abdollahpouri, A., & Moradi, P. (2018). A multi-objective particle swarm optimization algorithm for community detection in complex networks. *Swarm and Evolutionary Computation*, 39, 297-309.
- Ramadan, E., Naef, A., and Ahmed, M. (2016). Protein complexes predictions within protein interaction networks using genetic algorithms. *BMC bioinformatics*, 17(7), 269.
- Rao, A. K. G., Janikow, C., Bhatia, S., and Climer, S. (2018). Efficient Reduced-Bias Genetic Algorithm (ERBGA) for Generic Community Detection Objectives. *MWAIS 2018 Proceedings*, 32.
- Ray, S., De, M., & Mukhopadhyay, A. (2012). A multiobjective go based approach to protein complex detection. *Procedia Technology*, 4, 555-560.
- Ray, S., Hossain, A., and Maulik, U. (2016, January). Disease associated protein complex detection: a multi-objective evolutionary approach. In *Microelectronics, Computing and Communications (MicroCom), 2016 International Conference on* (1-6). IEEE.
- Remy, C., Rym, B., & Matthieu, L. (2017, November). Tracking bitcoin users activity using community detection on a network of weak signals. In *International conference on complex networks and their applications* (166-177). Springer, Cham.
- Romano, S., Bailey, J., Nguyen, V., and Verspoor, K. (2014, January). Standardized mutual information for clustering comparisons: One step further in adjustment for chance. In *International Conference on Machine Learning* (143-1151).
- Rossetti, G., and Cazabet, R. (2018). Community discovery in dynamic networks: a survey. *ACM Computing Surveys (CSUR)*, 51(2), 35.
- Sadi, S., Etaner-Uyar, S., & Gündüz-Öğüdücü, Ş. (2009). Community detection using ant colony optimization techniques. In *15th International Conference on Soft Computing* (206-213).
- Said, A., Abbasi, R. A., Maqbool, O., Daud, A., and Aljohani, N. R. (2018). CC-GA: A clustering coefficient based genetic algorithm for detecting communities in social networks. *Applied Soft Computing*, 63, 59-70.

- Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., & Eisenberg, D. (2004). The database of interacting proteins: 2004 update. *Nucleic acids research*, 32(suppl_1), D449-D451.
- Sani, N. S., Manthouri, M., & Farivar, F. (2018). A multi-objective ant colony optimization algorithm for community detection in complex networks. *Journal of Ambient Intelligence and Humanized Computing*, 1-17.
- Schaeffer, S. E. (2007). Graph clustering. *Computer science review*, 1(1), 27-64.
- Schwefel, H. P. P. (1993). *Evolution and optimum seeking: the sixth generation*. John Wiley & Sons, Inc., New York, NY, USA, 1993.
- Shang, R., Bai, J., Jiao, L., & Jin, C. (2013). Community detection based on modularity and an improved genetic algorithm. *Physica A: Statistical Mechanics and its Applications*, 392(5), 1215-1231.
- Sharma, P., and Bhattacharyya, D. (2018). DCRS: A Multi-objective Protein Complex Finding Method. In *Proceedings of the International Conference on Computing and Communication Systems* (801-809). Springer, Singapore.
- Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature genetics*, 31(1), 64.
- Shi, C., Wang, Y., Wu, B. and Zhong, C. (2009a, February). A new genetic algorithm for community detection. In *International Conference on Complex Sciences* (1298-1309). Springer, Berlin, Heidelberg.
- Shi, C., Yan, Z., Cai, Y., and Wu, B. (2012). Multi-objective community detection in complex networks. *Applied Soft Computing*, 12(2), 850-859.
- Shi, C., Yan, Z., Pan, X., Cai, Y., & Wu, B. (2011b, June). Multi-objective decisionmaking in the detection of comprehensive community structures. In *Evolutionary Computation (CEC), 2011 IEEE Congress on* (1489-1495). IEEE.
- Shi, C., Yan, Z., Wang, Y., Cai, Y. and Wu, B. (2010a). A genetic algorithm for detecting communities in large-scale complex networks. *Advances in Complex Systems*, 13(01), 3-17.
- Shi, C., Yu, P.S., Cai, Y., Yan, Z. and Wu, B. (2011, October). On selection of objective functions in multi-objective community detection. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (2301-2304). ACM

- Shi, C., Yu, P.S., Yan, Z., Huang, Y. and Wang, B. (2014). Comparison and selection of objective functions in multiobjective community detection. *Computational Intelligence*, 30(3), 562-582.
- Shi, C., Zhong, C., Yan, Z., Cai, Y. and Wu, B. (2010b, July). A multi-objective approach for community detection in complex network. In *Evolutionary Computation (CEC), 2010 IEEE Congress on* (1-8). IEEE.
- Shi, J., and Malik, J. (2000). Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8), 888-905.
- Shi, Z., Liu, Y. and Liang, J. (2009b, November). PSO-based community detection in complex networks. In *2009 Second International Symposium on Knowledge Acquisition and Modeling* (114-119). IEEE.
- Shirjini, M. F., Farzi, S., & Nikanjam, A. (2020). MDPCluster: a swarm-based community detection algorithm in large-scale graphs. *Computing*, 102, 893–922 (2020).
- Song, A., Li, M., Ding, X., Cao, W., & Pu, K. (2016). Community detection using discrete bat algorithm. *IAENG International Journal of Computer Science*, 43(1), 37-43.
- Song, X., Ji, J., Yang, C., & Zhang, X. (2014, July). Ant colony clustering based on sampling for community detection. In *Evolutionary Computation (CEC), 2014 IEEE Congress on*(687-692). IEEE.
- Spirin, V., and Mirny, L.A. (2003). Protein complexes and functional modules in molecular networks. *PNAS*, 100:12123–12128.
- Srinivas, N., and Deb, K. (1994). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, 2(3), 221-248.
- Sun, H., Ma, S., & Wang, Z. (2017, December). A community detection algorithm using differential evolution. In *Computer and Communications (ICCC), 2017 3rd IEEE International Conference on* (1515-1519). IEEE.
- Tasgin, M., Herdagdelen, A., & Bingol, H. (2007). Community detection in complex networks using genetic algorithms. *arXiv preprint arXiv:0711.0491*.
- Van Steen, M. (2010). *Graph theory and complex networks. An introduction*. Maareen van Steen, January 2010, vol. 144.
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., and Bork, P. (2002). Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, 417(6887), 399-403.

- Wang, H., Azuaje, F., Bodenreider, O., & Dopazo, J. (2004, October). Gene expression correlation and gene ontology-based similarity: an assessment of quantitative relationships. In *2004 Symposium on Computational Intelligence in Bioinformatics and Computational Biology* (25-31). IEEE.
- Wang, L., Liang, Y., Tian, Q., Song, C., & Wu, Z. (2014a). A Community Detection Method Based On Multi-Objective Optimization Method. *Applied Mechanics & Materials* (Vol. 571, 177-182). Trans Tech Publications Ltd.
- Wang, L., Liang, Y., Tian, Q., Song, C., & Wu, Z. (2014b). An Effective Community Detection Method Based on Improved Genetic Algorithm. *Applied Mechanics & Materials* (Vol. 568, 852-857). Trans Tech Publications Ltd
- Wang, S., Gong, M., Shen, B., Wang, Z., Cai, Q., & Jiao, L. (2015c, May). Deep community detection based on memetic algorithm. In *Evolutionary Computation (CEC), 2015 IEEE Congress on* (648-655). IEEE.
- Wang, Z., Zhao, X., Wen, P., Xue, J., & Hu, C. (2016). Community Detection in Complex Networks Using Improved Artificial Bee Colony Algorithm. In *2016 International Conference on Communications, Information Management and Network Security* (283-288). Atlantis Press.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684), 440.
- Wu, H., Sorathia, V., and Prasanna, V. K. (2013). When diversity meets speciality: Friend recommendation in online social networks. *Human*, 1(1), pp-52.
- Wu, P., and Pan, L. (2015). Multi-objective community detection based on memetic algorithm. *PloS one*, 10(5), e0126845.
- Xiaodong, D., Cunrui, W., Xiangdong, L. and Yanping, L. (2008, June). Web community detection model using particle swarm optimization. In *Evolutionary Computation, 2008. CEC 2008.(IEEE World Congress on Computational Intelligence). IEEE Congress on* (1074-1079). IEEE.
- Xu, B., Qi, J., Zhou, C., Hu, X., Xu, B., & Sun, Y. (2015). Hybrid self-adaptive algorithm for community detection in complex networks. *Mathematical Problems in Engineering*, 2015, 273054, 1-12.
- Yang, X. S. (2014). *Nature-inspired optimization algorithms*. Elsevier, London.
- Yuanyuan, M., and Xiyu, L. (2018). Quantum inspired evolutionary algorithm for community detection in complex networks. *Physics Letters A*, 382(34), 2305-2312.

- Ying, W., Jalil, H., Wu, B., Wu, Y., Ying, Z., Luo, Y., & Wang, Z. (2019). Parallel conical area community detection using evolutionary multi-objective optimization. *Processes*, 7(2), 111.
- Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, , 33(4), 452-473.
- Zadeh, P. M., & Kobti, Z. (2015b, June). Community detection in social networks by cultural algorithm. In *Collaboration Technologies and Systems (CTS), 2015 International Conference on* (319-325). IEEE.
- Zadeh, P.M. and Kobti, Z. (2015a). A multi-population cultural algorithm for community detection in social networks. *Procedia Computer Science*, 52, 342-349.
- Zaki, N., Berengueres, J., and Efimov, D. (2012). Detection of protein complexes using a protein ranking algorithm. *Proteins: Structure, Function, and Bioinformatics*, 80(10), 2459-2468.
- Žalik, K. R. (2019). Evolution Algorithm for Community Detection in Social Networks Using Node Centrality. In *Intelligent Methods and Big Data in Industrial Applications* (73-87). Springer, Cham.
- Žalik, K. R., & Žalik, B. (2018b). Node attraction-facilitated evolution algorithm for community detection in networks. *Soft Computing*, 23(15), 6135-6143.
- Žalik, K. R., and Žalik, B. (2017). Multi-objective evolutionary algorithm using problem-specific genetic operators for community detection in networks. *Neural Computing and Applications*, , 30(9), 2907-2920.
- Žalik, K. R., and Žalik, B. (2018a). Memetic algorithm using node entropy and partition entropy for community detection in networks. *Information Sciences*, 445, 38-49.
- Zhang, N., and Bilsland, E. (2011). Contributions of *Saccharomyces cerevisiae* to understanding mammalian gene function and therapy. In *Yeast Systems Biology*, vol 759, (501-523). Humana Press.
- Zhang, Q., and Li, H. (2007). MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on evolutionary computation*, 11(6), 712-731.
- Zhang, X., Zhou, K., Pan, H., Zhang, L., Zeng, X., & Jin, Y. (2018). A Network Reduction-Based Multiobjective Evolutionary Algorithm for Community

- Detection in Large-Scale Complex Networks. *IEEE transactions on cybernetics*, 50(2), 703-716.
- Zhang, Y., Lin, H., Yang, Z., Wang, J., Li, Y., & Xu, B. (2013). Protein complex prediction in large ontology attributed protein-protein interaction networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 10(3), 729-741.
- Zhang, Y., Liu, Y., Li, J., Zhu, J., Yang, C., Yang, W., & Wen, C. (2020). WOCDA: A whale optimization based community detection algorithm. *Physica A: Statistical Mechanics and its Applications*, 539, 122937.
- Zhou, X., Liu, Y. and Li, B. (2016). A multi-objective discrete cuckoo search algorithm with local search for community detection in complex networks. *Modern Physics Letters B*, 30(07), 1650080.
- Zhu, Y. (2013). *Community detection in complex networks*. PhD Dissertation, University of New Mexico, Albuquerque, New Mexico.
- Zhu, X., & Wang, B. (2010, December). Community mining in complex network based on parallel genetic algorithm. *In Genetic and Evolutionary Computing (ICGEC), 2010 Fourth International Conference on (325-328)*. IEEE.
- Zou, F., Chen, D., Huang, D. S., Lu, R., & Wang, X. (2019). Inverse modelling-based multi-objective evolutionary algorithm with decomposition for community detection in complex networks. *Physica A: Statistical Mechanics and its Applications*, 513, 662-674.
- Zou, F., Chen, D., Li, S., Lu, R., and Lin, M. (2017). Community detection in complex networks: Multi-objective discrete backtracking search optimization algorithm with decomposition. *Applied Soft Computing*, 53, 285-295.

LIST OF PUBLICATIONS

Journal with Impact Factor

1. **Abduljabbar, D. A.**, Hashim, S. Z. M., & Sallehuddin, R. (2020). An improved multi-objective evolutionary algorithm for detecting communities in complex networks with graphlet measure. *Computer Networks*, 169, 107070. <https://doi.org/10.1016/j.comnet.2019.107070>. **(Q1, IF: 3.03)**
2. **Abduljabbar, D. A.**, Hashim, S. Z. M., & Sallehuddin, R. (2020). Nature-inspired optimization algorithms for community detection in complex networks: a review and future trends. *Telecommunication Systems*, 1-28. <https://doi.org/10.1007/s11235-019-00636-x>. **(Q3, IF: 1.707)**

Indexed Journal

1. **Abduljabbar, D. A.**, Hashim, S. Z. M., & Sallehuddin, R. (2019). An Enhanced Evolutionary Algorithm with Local Heuristic Approach for Detecting Community in Complex Networks. *Journal of Theoretical and Applied Information Technology*, 97(20). <http://www.jatit.org/volumes/Vol97No20/18Vol97No20.pdf>. **(Indexed by SCOPUS)**

Indexed Conference Proceedings

1. **Abduljabbar, D. A.**, Hashim, S. Z. M., & Sallehuddin, R. (2020, January). An Enhanced Evolutionary Algorithm for Detecting Complexes in Protein Interaction Networks with Heuristic Biological Operator. In *International Conference on Soft Computing and Data Mining* (334-345). Springer, Cham. https://doi.org/10.1007/978-3-030-36056-6_32. **(Indexed by SCOPUS)**
2. **Abduljabbar, D. A.**, Hashim, S. Z. M., & Sallehuddin, R. (2019). An Evolutionary Algorithm for Community Detection Using an Improved Mutation Operator. In *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)* (406-410). IEEE. 10.1109/ICECOS47637.2019.8984551. **(Indexed by WOS)**