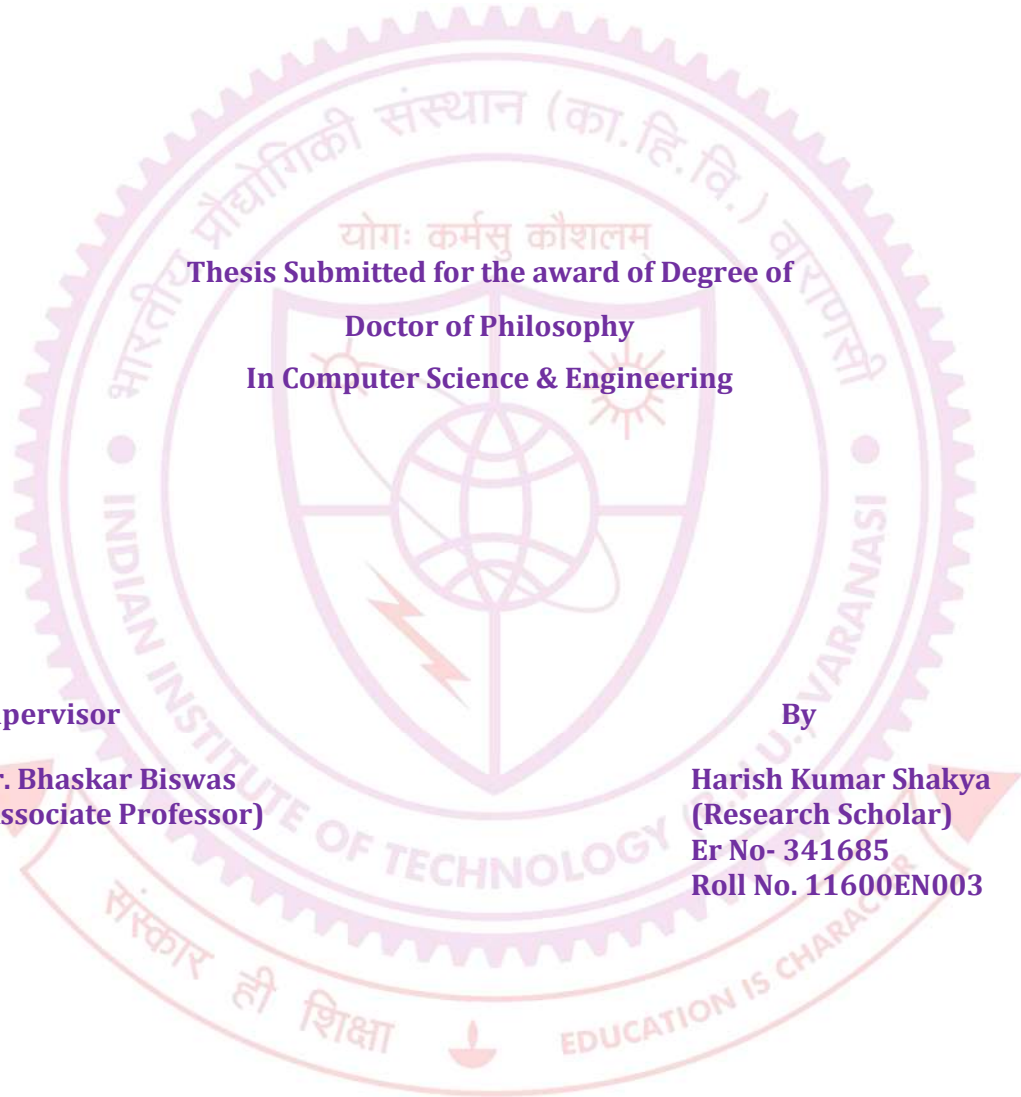


**Copyright © Indian Institute of Technology
(Banaras Hindu University), Varanasi, India, 2018
All rights reserved**

Community Detection in Social Networks Using Improvised Evolutionary Algorithms



योग: कर्मसु कौशलम्
Thesis Submitted for the award of Degree of
Doctor of Philosophy
In Computer Science & Engineering

Supervisor

Dr. Bhaskar Biswas
(Associate Professor)

By

Harish Kumar Shakya
(Research Scholar)
Er No- 341685
Roll No. 11600EN003

Department of Computer Science & Engineering
Indian Institute of Technology
(Banaras Hindu University)
Varanasi-221005
INDIA

2018



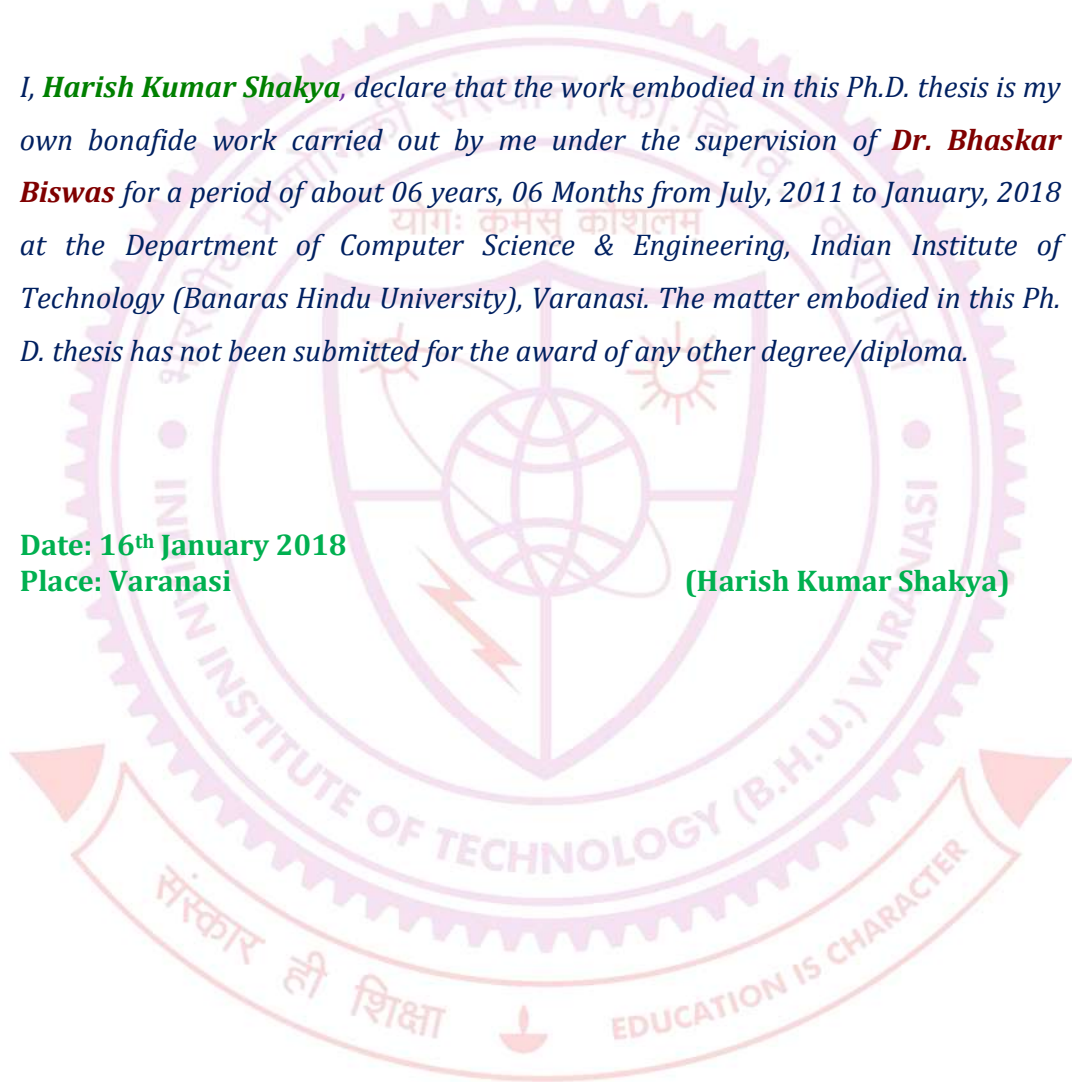
Undertaking from the Candidate

I, **Harish Kumar Shakya**, declare that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of **Dr. Bhaskar Biswas** for a period of about 06 years, 06 Months from July, 2011 to January, 2018 at the Department of Computer Science & Engineering, Indian Institute of Technology (Banaras Hindu University), Varanasi. The matter embodied in this Ph. D. thesis has not been submitted for the award of any other degree/diploma.

Date: 16th January 2018

Place: Varanasi

(Harish Kumar Shakya)





ANNEXURE- E

CANDIDATE'S DECLARATION

I, **Harish Kumar Shakya**, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of **Dr Bhaskar Biswas** for a period of about 06 years 06 Months from July, 2011 to January, 2018 at the Department of Computer Science & Engineering, Indian Institute of Technology (Banaras Hindu University), Varanasi. The matter embodied in this Ph.D. thesis has not been submitted for the award of any other degree/diploma.

I declare that I have faithfully acknowledged, given credit to and referred to the research workers wherever their works have been cited in the text and the body of the thesis. I further certify that I have not willfully lifted up some other's work, para, text, data, results, etc. reported in the journals, books, magazines, reports, dissertations, theses, etc., or available at web-sites and included them in this Ph.D. thesis and cited as my own work.

Date: 16th January 2018

Place: Varanasi

(Harish Kumar Shakya)

Certificate from the Supervisor/Co-supervisor/Head of the Department

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

(Dr. Bhaskar Biswas)
Supervisor

(Prof. Rajeev Srivastva)
Head of the Department



ANNEXURE- F

**COURSE/COMPREHENSIVE EXAMINATION/PRE-SUBMISSION SEMINAR
COMPLETION CERTIFICATE**

This is to certify that **Mr. Harish Kumar Shakya**, a bonafide research scholar of this Department, has satisfactorily completed the course work/comprehensive examination. The pre-submission seminar has successfully completed on 29.07.2017, entitled "Community Detection in Social Networks using Improvised Evolutionary Algorithms" which is a part of his Ph.D. programme.

Date: 16th January 2018

Place: IIT BHU, Varanasi

(Head of the Department)



Department of Computer Science & Engineering
Indian Institute of Technology
(Banaras Hindu University)
Varanasi - 221 005

ANNEXURE- G
COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: “Community Detection in Social Networks using Improved Evolutionary Algorithms”

Candidate’s Name: Mr. Harish Kumar Shakya

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University), Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the Ph.D. degree.

Date: 16th January 2018

(Harish Kumar Shakya)

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author’s personal use provided that the source and the Institute’s/University’s copyright notice are indicated.

Acknowledgements

I would like to take the opportunity to thank everybody who inspired me, helped me and contributed directly or indirectly in realizing this thesis. Predominantly, my sincere thanks and regards goes to my thesis supervisor **Dr. Bhaskar Biswas** for his support and guidance during the execution of research. It would not have been possible to successfully carry out this work without his directions, advices and suggestions.

I would like to express my sincere thanks for **Prof. Rajeev Srivastava**, Head, Computer Science and Engineering Department, for his kindness and valuable support in carrying out the research. I express my sincere thanks to the faculty members and staffs of the department, **Prof. A. K. Tripathi, Prof. K. K. Shukla, Prof. R. B. Mishra, Prof. S. K. Singh, Dr. Vinayak Srivastava, Dr. Bhaskar Biswas, Dr. Hari Prabhat Gupta** and **Dr Tanima Dutta** for their extensive and inspiring guidance throughout this tenure.

I specially thank the members of my Research Progress Evaluation Committee, **Prof. K. K. Shukla, Prof. Rajeev Srivastava, Prof. Neeraj Sharma**, and DPGC Convener **Dr. Bhaskar Biswas** for their invaluable suggestions regarding the thesis.

I wish to express my deep sense of thanks to **Mr. Kuldeep Singh, Mr. Dharmendra Prasad Mahato, Mr. Gyanendra Singh** and **Ms. Roli**, for being great friends and the best advisers I could ever have. Their advice, encouragement and critics were always the source of inspiration. This dissertation would not have been possible without their invaluable suggestions and persistent help. I wish to extend my gratitude towards all my lab mates **Mr. Ajay Singh, Mr. Shashank Singh, Mr. Sumit Jaiswal**, and **Mr. Anupam Biswas** whose company helped me a lot.

I would wish to thank all my family members who continuously supported me throughout my life. My parents did a lot of hard work to support the logistics for my studies and always encouraged me in every decision that I took. I would like to thank my Father, who was also my childhood teacher and is always there to stands by my side. I owe my thanks to my younger brother **Mr. Ravi** and sisters **Ms. Hema & Mrs. Nisha** who help me fulfill my obligation to take good care of our parents when I am far from home. I thank all of them for the sacrifice they made during this hard-working, yet fun-filled period in my life.

I would like to extend a warm thank to my wife *Deepshikha* and my loving son *Nishchhal* for all the sacrifice they have done for me in order to succeed throughout my life particularly during the completion of this.

I take this opportunity to sincerely acknowledge the University Grants Commission, Government of India for providing Ph.D. Fellowship for financial assistance. I sincerely thank all of them who contributed in for helping me to see the light at the end of every scary tunnel during my Ph.D.

Finally, I am grateful to God for his blessings and for giving me the strength to persevere throughout the long and arduous journey towards a Ph.D.


-Harish Kumar Shakya

LIST OF CONTENTS

CHAPTER I: INTRODUCTION	1
1.1. BACKGROUND	2
1.2. MOTIVATION FOR WORK	2
1.3. MAJOR ISSUES	3
1.4. OBJECTIVES OF THE THESIS	3
1.5. SCOPE OF THE THESIS	3
1.6. COMMUNITY DETECTION USING Fuzzy Based Community Detection COMMUNITY DETECTION USING ORGANIZATION	3
LITERATURE SURVEY	3
COMMUNITY DETECTION	3

TABLE OF CONTENTS

CERTIFICATE.....	ii
DECLARATION BY THE CANDIDATE.....	v
COPYRIGHT TRANSFER CERTIFICATE.....	vii
ACKNOWLEDGEMENT.....	viii
TABLE OF CONTENTS.....	x
LIST OF FIGURES.....	xv
LIST OF TABLES.....	xviii
LIST OF ABBREVIATIONS.....	xx
LIST OF SYMBOLS.....	xxii
PREFACE.....	xxiii
CHAPTER 1: INTRODUCTION.....	1
1.1. BACKGROUND.....	1
1.2. MOTIVATION FOR WORK.....	2
1.3. MAJOR ISSUES.....	3
1.4. OBJECTIVES OF THE THESIS.....	4
1.5. CONTRIBUTIONS.....	6
1.5.1. COMMUNITY DETECTION USING GENETIC ALGORITHM WITH OBL.....	7
1.5.2. FUZZY BASED COMMUNITY DETECTION.....	8
1.5.3. COMMUNITY DETECTION USING DIFFERENTIAL EVOLUTION ALGORITHM.....	9
1.6. THESIS ORGANIZATION.....	10
CHAPTER 2: LITERATURE SURVEY.....	12
2.1. ELEMENTS OF COMMUNITY DETECTION.....	12

2.1.1. COMPUTATIONAL COMPLEXITY	12
2.1.2. COMMUNITIES	13
2.1.3. PARTITIONS	14
2.2. COMMUNITY DETECTION TECHNIQUES.....	14
2.2.1. HIERARCHICAL APPROACHES	15
2.2.1.1. Agglomerative Techniques	16
2.2.1.2. Divisive Techniques.....	17
2.2.2. OPTIMIZATION TECHNIQUES	17
2.2.2.1. Modularity Maximization	18
2.2.2.2. Heuristic Approaches.....	18
2.2.3. SPECTRAL CLUSTERING.....	20
2.2.4. STATISTICAL INFERENCE.....	20
2.2.4.1. Generative models	20
2.2.4.2. Information theoretic approach.....	21
2.2.5. OTHER APPROACHES.....	21
2.3. VALIDATION METRICS	22
2.3.1. ACCURACY METRICS.....	23
2.3.1.1. ARI.....	23
2.3.1.2. NMI.....	24
2.3.1.3. Purity.....	25
2.3.1.4. F-Measure	26
2.3.1.5. Entropy.....	27
2.3.2. QUALITY METRICS	27
2.3.2.1. Modularity.....	27
2.3.2.2. Coverage	28
2.3.2.3. External Density.....	28
2.3.2.4. Fuzzy Modularity.....	28
2.4. DATASETS	29
2.4.1. REAL-WORLD NETWORKS	29
2.4.1. SYNTHETIC NETWORKS.....	32
2.5. EVALUATION METHODOLOGIES	32
2.6. APPLICATIONS OF COMMUNITY STRUCTURE	33
2.6.1. COMMUNITY STRUCTURE IN LINK PREDICTION.....	33
2.6.2. COMMUNITY STRUCTURE IN INFORMATION DIFFUSION	33

2.6.3. COMMUNITY STRUCTURE IN RECOMMENDATION SYSTEMS.....	34
2.7. SUMMARY	34
CHAPTER 3: GENTEIC ALGORITHM WITH OBL	35
3.1. INTRODUCTION.....	35
3.2. MODIFIED CROSSOVER OPPOSITION BASED GENETIC ALGORITHM (MCOBGA).....	36
3.2.1. NETWORK MODULARITY OBJECTIVE FUNCTION	36
3.2.2. INDIVIDUAL ENCODING	36
3.2.3. POPULATION INITIALIZATION	37
3.2.4. MODIFIED CROSSOVER OPERATION	38
3.2.5. MUTATION OPERATION.....	40
3.2.6. PROPOSED ALGORITHM DESCRIPTION	40
3.3. EXPERIMENTAL DESCRIPTION	41
3.3.1. EXPERIMENTAL ANALYSIS	42
3.3.2. ACCURACY AND QUALITY MEASURE	44
3.4. REGENERATIVE GENETIC ALGORITHM	46
3.5. RESULT AND DISCUSSION	47
3.5.1. EXPERIMENTAL DESCRIPTION	47
3.5.2. EXPERIMENTAL ANALYSIS	48
3.5.2.1. Zachary karate club.....	49
3.5.2.2. Dolphin Sociality	50
3.5.2.3. American College Football	52
3.5.2.4. Books on US Politics	54
3.6. CONCLUSION OF THE CHAPTER	55
CHAPTER 4: FUZZY BASED COMMUNITY DETECTION	55
4.1. INTRODUCTION.....	56
4.2. PROPOSED FGA(GENETIC ALGORITHM WITH FUZZY CONCEPT) APPROACH.....	56
4.2.1 VALUATION FUNCTIONS.....	56
4.2.1.1. Modularity.....	56
4.2.1.2. Normalized Mutual Information.....	57
4.2.1.3. Omega-Index.....	57
4.2.1.4. Simple Modularity And Fuzzy Modularity.....	57
4.2.1.5. Zhang Fuzzy Modularity.....	57
4.2.1.6. Liu Fuzzy Modularity	58
4.3. EXPERIMENTAL WORK	58

4.4. EXPERIMENTAL ANALYSIS	59
4.4.1. STRIKE DATASET	59
4.4.2. KARATE CLUB DATASET	60
4.4.3. DOLPHIN DATASET	60
4.5. MODIFIED GAFCD.....	64
4.5.1. EXPERIMENTAL RESULT & ANALYSIS	65
4.6. COMPARED PERFORMANCE OF COMMUNITY DETECTION ALGORITHMS.....	69
4.6.1. PERMANENCE BASED VERTEX REPLICATION	70
4.6.2. EXPERIMENTAL RESULT & ANALYSIS	71
4.7. CONCLUSION OF THE CHAPTER	74
CHAPTER 5: COMMUNITY DETECTION USING DIFFERENTIAL EVOLUTION ALGORITHM.....	76
5.1. INTRODUCTION.....	76
5.2. DE WITH MULTIPLE OBJECTIVE FUNCTIONS	77
5.2.1. EXPERIMENTAL DESCRIPTION	78
5.3. EXPERIMENTS PERFORM AND RESULTS DISCUSSION	79
5.3.1. EVALUATION METHOD.....	82
5.3.1.1. Accuracy Measure	83
5.3.1.2. Quality Measure.....	83
5.3.1.3. Qualitative Measure	84
5.3.2. MCDM RANKING RELATED SETTING.....	85
5.3.2.1. Measuring Accuracy	85
5.3.2.2. Measuring Quality	86
5.3.2.3. Value based Analysis.....	86
5.3.3. MCDM RANKING.....	89
5.4. VERTEX SIMILARITY BASED DIFFERENTIAL EVOLUTION	93
5.4.1. PROPOSED WORK DESCRIPTION.....	93
5.4.2. EXPERIMENTAL RESULT.....	95
5.5. TOURNAMENT & OPPOSITION LEARNING BASED DE.....	98
5.5.1. EXPERIMENTAL ANALYSIS & RESULT DISCUSSION.....	99
5.5.1.1. Modularity with Number of Iteration graph Analysis.....	100
5.5.1.2. Accuracy & Quality metric values Analysis.....	102
5.5.1.3. MCDM Rank Analysis	105
5.6. CONCLUSION OF THE CHAPTER	109

CHAPTER 6: CONCLUSION AND SCOPE FOR FUTURE WORK	111
6.1. CONCLUDING REMARKS	111
6.2. SCOPE FOR FUTURE WORKS	114
REFERENCES.....	115
LIST OF PAPERS PUBLISHED/PRESENTED/COMMUNICATED.....	131
COPIES OF MANUSCRIPTS/ REPRINTS OF THE PAPERS COMMUNICATED/ ACCEPTED/ PUBLISHED.....	133
PERSONAL PROFILE	

LIST OF FIGURES

Figure 1.1: A schematic representation of a network with community structure	2
Figure 2.1: Categories of communities.....	13
Figure 2.2: Type of community Detection Techniques	15
Figure 3.1: Example of matrix encoding	37
Figure 3.2: Calculating opposition of a point	38
Figure 3.3: Schematic of crossover operation.....	39
Figure 3.4: Q-fun value (on Y axis) vs no. of iteration (on X axis) for comparison between SGA and MCOBGA for strike dataset.....	43
Figure 3.5: Q-fun value (on Y axis) vs no. of iteration (on X axis) for comparison between SGA and MCOBGA for karate club dataset.	43
Figure 3.6: Q-fun value (on Y axis) vs no. of iteration (on X axis) for comparison between SGA and MCOBGA for Dolphin sociality dataset.....	43
Figure 3.7 Q-fun value (on Y axis) vs no. of iteration (on X axis) for comparison between SGA and MCOBGA for American college football dataset.	43
Figure 3.8: MCDM ranking acquired by each algorithm in Real world known network for Dolphin dataset with Variation of accuracy contribution.	46
Figure 3.9: MCDM ranking acquired by each algorithm in Real world known network for Football dataset with Variation of accuracy contribution.....	46
Figure 3.10 MCDM ranking acquired by each algorithm in Real world known network for Karate dataset with Variation of accuracy contribution.....	46
Figure 3.11: MCDM ranking acquired by each algorithm in Real world known network for Strike dataset with Variation of accuracy contribution.....	46
Figure 3.12 Accuracy of SGA and RGA running on artificial random networks.....	47
Figure 3.13 Q-value of SGA and RGA running on karate club dataset	49
Figure 3.14 Q-value of SGA and RGA running on Dolphin dataset.....	50
Figure 3.15 Q-value of SGA and RGA running on American College Football dataset	51
Figure 3.16 Q-value of SGA and RGA running on Books on US Politics dataset	53
Figure 4.1: Community-based graphical representations for Strike dataset	59
Figure 4.2: Community-based graphical representations for Karate club dataset	60
Figure 4.3: Community-based graphical representations for Dolphin dataset.....	60

Figure 4.4: MCDM ranking graphs for Strike datasets with variation of accuracy and quality	62
Figure 4.5: MCDM ranking graphs for Karate club datasets with variation of accuracy and quality.....	63
Figure 4.6: MCDM ranking graphs for Dolphin datasets with variation of accuracy and quality.....	63
Figure 4.7: Dolphin, $Q=0.5285$, $c=5$	66
Figure 4.8: Karate, $Q=0.4449$, $c=4$	66
Figure 4.9: Jazz, $Q=0.4452$, $c=4$	67
Figure 4.10: Metabolic, $Q=0.4447$, $c=9$	67
Figure 4.11: Block diagram of proposed Method	71
Figure 4.12: Community based graphical representation for datasets.....	72
Figure 4.13: Accuracy based comparison between overlapping and non-overlapping communities for datasets.....	73
Figure 5.1: NMI values represent with number of iterations using Strike Dataset.....	79
Figure 5.2: NMI values represent with number of iterations using Karate club Dataset	80
Figure 5.3: NMI values represent with number of iterations using Dolphin Dataset	81
Figure 5.4: NMI values represent with number of iterations using Football team dataset....	82
Figure 5.5: MCDM ranking graph for Strike dataset with Variation of accuracy involvement	91
Figure 5.6: MCDM ranking graph for Karate club Dataset with Variation of accuracy involvement.....	91
Figure 5.7: MCDM ranking graph for Dolphin Dataset with Variation of accuracy involvement.....	92
Figure 5.8: MCDM ranking graph for Football club Dataset with Variation of accuracy involvement.....	92
Figure 5.9: Fraction of vertices correctly classified by VSDE	96
Figure 5.10: Fraction of vertices correctly classified by DECD.....	96
Figure 5.11: Q-fun value (on Y axis) vs. no. of iteration (on X axis) for comparison between SDE, OBDE, TDE and TOBDE for Strike dataset.....	101
Figure 5.12: Q-fun value (on Y axis) vs. no. of iteration (on X axis) for comparison between SDE, OBDE, TDE and TOBDE for Karate club dataset.....	101
Figure 5.13: Q-fun value (on Y axis) vs. no. of iteration (on X axis) for comparison between SDE, OBDE, TDE and TOBDE for Dolphin dataset	102

LIST OF TABLES

<i>Table 2.1 : Contingency table Here, each n_{ij} denotes the number of nodes in common between communities C_i and R_j: $n_{ij} = C_i \cap R_j$.....</i>	<i>23</i>
<i>Table 2.2: Contingency table.....</i>	<i>24</i>
<i>Table 3.1: Parameters of MCOBGA.....</i>	<i>41</i>
<i>Table 3.2: Specifications of real world datasets used in experiments.....</i>	<i>42</i>
<i>Table 3.3: Average Q-function values of SGA, OBGA and MCOBGA running on 4 real social networks.....</i>	<i>43</i>
<i>Table 3.4: Accuracy and quality metric values in various datasets whose ground truth communities are known.....</i>	<i>44</i>
<i>Table 3.5: Accuracy and quality based performance for different algorithms.....</i>	<i>45</i>
<i>Table 3.6: MCDM ranking score obtained with 75% accuracy and 25% quality Higher score indicates more inclination of algorithm towards accuracy.....</i>	<i>45</i>
<i>Table 3.7: Average Q-values and NMI-values for karate club dataset.....</i>	<i>48</i>
<i>Table 3.8: Average Q-values and NMI-values for Dolphin dataset.....</i>	<i>49</i>
<i>Table 3.9: Average Q-values and NMI-values for American College Football dataset.....</i>	<i>51</i>
<i>Table 3.10: Average Q-values and NMI-values for Books on US Politics dataset.....</i>	<i>52</i>
<i>Table 3.11: Average Q-values of SGA, FN, RGA, GN and TGA running on and for four dataset.....</i>	<i>53</i>
<i>Table 4.1: Value Description of Parameters.....</i>	<i>59</i>
<i>Table 4.2: Accuracy and quality metric values for various datasets.....</i>	<i>61</i>
<i>Table 4.3: MCDM ranking score obtained for different datasets.....</i>	<i>61</i>
<i>Table 4.4: Values of different parameters.....</i>	<i>64</i>
<i>Table 4.5: Description of the used Datasets.....</i>	<i>65</i>
<i>Table 4.6: Compared Performance of Community Detection Algorithms.....</i>	<i>68</i>
<i>Table 4.7: Constraint of the Experiment.....</i>	<i>70</i>
<i>Table 4.8: Function values for disjoint community structures.....</i>	<i>71</i>
<i>Table 4.9: Function values for overlapping community structures.....</i>	<i>71</i>
<i>Table 5.1: Parameter of experiments.....</i>	<i>77</i>
<i>Table 5.2: Colour code of Algorithm for experiments.....</i>	<i>78</i>
<i>Table 5.3: Details of real world datasets.....</i>	<i>78</i>
<i>Table 5.4: DE with Multiple variations using Strike Dataset.....</i>	<i>87</i>

<i>Table 5.5: DE with Multiple variations using Karate Club Dataset</i>	87
<i>Table 5.6: DE with Multiple variations using Dolphin Dataset</i>	88
<i>Table 5.7: DE with Multiple variations using Football Dataset</i>	88
<i>Table 5.8: MCDM ranking score obtained with 75% accuracy and 25% quality, higher score indicates more inclination of algorithm towards accuracy</i>	89
<i>Table 5.9: Parameters of VSDE</i>	95
<i>Table 5.10: Experimental results of the Zachary's karate club network and the American college football network, N_{pr} is the average number of communities; Q_{avg} and Q_{bst} are the average and best values of modularity Q, respectively; and Acc_{avg} and Acc_{bst} are the average and best accuracy, respectively.</i>	98
<i>Table 5.11: Description of the Parameters</i>	99
<i>Table 5.12: Accuracy & Quality metric values for Strike data set whose ground truth communities are known</i>	103
<i>Table 5.13: Accuracy & Quality metric values for Dolphin Dataset whose ground truth communities are known</i>	103
<i>Table 5.14: Accuracy & Quality metric values for Dolphin Dataset whose ground truth communities are known</i>	104
<i>Table 5.15: Accuracy & Quality metric values for American Football club Dataset whose ground truth communities are known</i>	104
<i>Table 5.16: MCDM ranking score obtained with 75% Accuracy and 25% Quality. Higher score indicate more inclination of algorithm towards accuracy</i>	108

ix Algorithm
 Algorithm
 Algorithm based Fuzzy Community
 and Relativity and Quantum Chemistry

LIST OF ABBREVIATIONS

ACO	Ant Colony Optimization
ANS	Anticlockwise Negative Shift
ANUI	Average Normalized Unifiability and Isolability
APS	Anticlockwise Positive Shift
ARI	Adjusted Rand Index
AUC	Area Under the receiver operating characteristic Curve
AVI	Average Isolability
AVU	Average Unifiability
CDF	Cumulative Density Function
CEW	Community-based Edge Weight
CFGFC	Fuzzy Center-based Graph Clustering
CLP	Community-based Link Prediction
CNS	Clockwise Negative Shift
CPS	Clockwise Positive Shift
EA	Evolutionary Algorithm
EC	Edge Centrality
EOA	Evolutionary Optimization Algorithm
ERW-Kpath	Edge Random Walk k-Path
ExtD	External Density
FastU	Fast Unfolding
FMM/H2	Fuzzy Modularity Maximization/Heuristic 2
FN	False Negative
FP	False Positive
FGA	Fuzzy Genetic Algorithm
GA	Genetic Algorithm
GAFCD	Genetic Algorithm based Fuzzy Community Detection
GR-QC	General Relativity and Quantum Cosmology

GALS	Genetic Algorithm With Local Search
HC-PIN	Hierarchical Clustering-Protein Interaction Networks
HEP-TH	High Energy Physics - Theory
JDE	Janez Differential Evolution
LFR	Lancichinetti, Fortunato and Radicchi
MCOBGA	Modified Crossover and Opposition Based Genetic Algorithm
MGAFCD	Modified Genetic Algorithm for Fuzzy Community Detection
MSFCM	Multi-cut Spectral Fuzzy C-Means
MCDM	Multiple Criterion Decision Making
NL	Neutral Line
NMI	Normalized Mutual Information
NoC	Number of Communities
NSGAP	Node Similarity Based Genetic Algorithm with Permanence Concept
OBGA	Opposition Based Genetic Algorithm
OBDE	Opposition Based Differential Evolution
RandW	Random Walk
RL	Regression Line
RGA	Regenerative Genetic Algorithm
RS	Ranging Score
RW	Random Walk
SGA	Simple Genetic Algorithm
SCAN	Structural Clustering Algorithm for Networks
SIS	Susceptible-Infected-Susceptible
SLR	Simple Linear Regression
TN	True Negative
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
TP	True Positive
TDE	Tournament Based Differential Evolution
TOBDE	Tournament and Opposition Based Differential Evolution
VSDE	Vertex Similarity Based Differential Evolution
VGA	Vertex Similarity Based Genetic Algorithm

LIST OF SYMBOLS

C	detected community structure
R	real community structure
C_i	community i
α	threshold for Reachability
β	threshold for Isolability
η	ego network
μ	mixing parameter
ψ	membership threshold
U	partition matrix
U^+	extended partition matrix
t_{max}	maximum number of iterations
δ_i	size of community that include node i
$\sigma_{st}(ei j)$	number of shortest paths between node i and j
$\sigma_{st}^k(ei j)$	number of paths having at least k length
T_s	number of spanning trees
$\delta(u, v)$	strength of connection between any two nodes u and v

Preface

Complex networks such as social networks exhibit disproportionate connections among different nodes, resulting in densely interconnected groups of nodes. These highly connected groups of nodes within the network are referred as communities, which have significant role in understanding and uncovering various functional properties of the system. Today, introduction of social networking applications into every area of our lives makes social network analysis an important research area. An important property of networks/graphs modeling complex systems is the property of community structure, in which nodes are joined together in tightly knit groups (communities or clusters), between which there are only looser connections. The problem of detecting and extracting communities from such graphs has been the subject of intense investigations in recent years. This problem is very hard and not yet satisfactorily solved. Identification of communities has grown as one of the major research topics in social network analysis.

In this thesis, we explore and work on this community detection problem. We frame the problem as an optimization problem and hence explore the use of Evolutionary Algorithm i.e. Genetic Algorithms (GAs) & Differential Evolution (DE) in solving the same. We have studied, analyzed and implemented several existing algorithms including standard ones and GA-based ones. The standard algorithms include the Girvan-Newman (GN) Algorithm, FN, and the Label Propagation Algorithm by Raghavan et al. while the GA-based is Tasgin et al. algorithm, Vertex similarity based MENSGA, GAFCG, GALS, TGA, CCGA, LGA, . We have also designed a new GA-based algorithm for the problem. For this, a modified Genetic Algorithm of which chromosome structure and genetic operators are modified to find communities in social networks is used. This modified Genetic Algorithm can be used without giving proposed community number at the initialization with OBL (opposition based learning) and it runs faster compared to other Genetic Algorithm methods. Additionally, we did some other modification in Genetic algorithm just like name as MCOBGA, RGA, FGA, MGAFCG, and NSGAP.

We present a comparative performance (accuracy + quality) analysis of these algorithms (new + existing) to gain insights into the problem and reveal the advantages of our proposed algorithm over existing algorithms. We have also created some artificial datasets (based on

standard existing algorithms like the one for LFR graphs) for the purpose of the analysis and have acquired some real-world datasets (like Zachary's karate club network, American Football club, Strike dataset, Lusseau's network of bottlenose dolphins, etc.) too.

First two proposed chapter, we worked on the Genetic algorithm for the disjoint and overlapped community detection in social networks. We proposed a opposition based learning concept with genetic algorithm and called a MCOBGA, in this algorithm we find the disjoint communities for different datasets. Next one is the RGA; in this method we employed the regeneration of population behalf of mutation operation. In this algorithm, we update the convergence rate and efficiency and verified the number of artificial and real world datasets.

Another proposed chapter, we worked on the fuzzy community detection with the help of genetic algorithm. In FGA, we employed the fuzzy modularity concept (Liu & Zhang) and Omega for the overlapped community in social network. Similarly other one is modified the existing algorithm GAFCD; that is detect the crisp and fuzzy community both. So-that we modified this algorithm and create a new comparative method. In the last section of this proposed chapter, NSGAP means node similarity based genetic algorithm with permanence concept for the fuzzy and crisp community detection in social network.

In the last proposed chapter, we worked on the Differential evolution algorithm for the community detection in social network and find the best DE version of the different situation and the various types of datasets. We employed the tournament method; opposition based learning concept and vertex similarity concepts with DE algorithm. We created the TOBDE, OBDE, TDE, VSDE and the DE with Multiple objective functions.