

Contents

Certificate	iii
Declaration by the Candidate	v
Copyright Transfer Certificate	vii
Acknowledgements	ix
Preface	xiii
Contents	xvii
List of Figures	xix
List of Tables	xxi
Abbreviations	xxii
Symbols	xxv
1 Introduction	1
1.1 Challenges and Issues	3
1.1.1 Major Challenges	4
1.1.2 Major Issues	5

1.2	Objectives	7
1.3	Contributions	10
1.3.1	Bio-inspired Optimization based Influence Maximization	11
1.3.2	Context-aware Influence Maximization	12
1.3.3	Multiple Influence Maximization in Multiple Featured Networks	13
1.3.4	Applicability of Information Diffusion to Predict Missing Links	14
1.4	Thesis Organization	14
2	Literature Review	17
2.1	Information Diffusion Models (IDMs/DMs)	19
2.2	Problem Hardness	24
2.3	Overview of IM Approaches	25
2.3.1	The Greedy Framework	26
2.3.2	Taxonomy of Existing IM Approaches	28
2.3.3	Simulation based IM Approaches	28
2.3.4	Approximate Scoring based IM Approaches	31
2.3.5	Sampling based IM Approaches	36
2.3.6	Summary and Discussion.	40
2.4	Context-aware Influence Maximization	44
2.4.1	Location-aware Influence Maximization	45
2.4.2	Topic-aware Influence Maximization	47
2.4.3	Time-aware Influence Maximization	51
2.4.4	Competitive Influence Maximization	53
2.4.5	Dynamic Influence Maximization	56
2.5	Performance Metrics	60
2.6	Datasets	65
2.7	Application of Influence Maximization	69
2.7.1	Influence Maximization in Events	70
2.7.2	Influence Maximization in Recruitment	72
2.7.3	Influence Maximization in Social Media Population Screening	73

2.7.4	Others	74
2.8	Business Implications of Influence Maximization	75
2.9	Summary	77
3	Influence Maximization using Learning-automata based Particle Swarm Optimization	79
3.1	Introduction	79
3.1.1	Discrete Particle Swarm Optimization	81
3.1.2	Learning Automata	82
3.2	Proposed Approach	83
3.3	Algorithm	89
3.3.1	Applying the Algorithm	93
3.3.2	Complexity Analysis	97
3.4	Empirical Analysis	99
3.4.1	Experimental Setup	99
3.4.2	Parametric Analysis	100
3.4.2.1	The size of the population and the number of iterations	100
3.4.2.2	The inertia weight and learn factors	102
3.4.2.3	The reward and penalty parameter	102
3.4.3	Analyzing Quality	103
3.4.4	Analyzing Efficiency	105
3.4.5	Statistical Test	110
3.5	Conclusion	115
4	Community-based Context-aware Influence Maximization	119
4.1	Introduction	119
4.1.1	Data Model	121
4.1.2	Context-aware Information Diffusion Models	122
4.2	Proposed Approach	123
4.2.1	Algorithm Overview	124
4.2.2	Algorithm Description	126
4.2.2.1	Preprocessing	126
4.2.2.2	Finding non-desirable nodes	129

4.2.2.3	Selecting seed nodes	130
4.2.2.4	Computing spread of nodes	130
4.3	Algorithm	132
4.3.1	Applying the Algorithm	136
4.3.2	Complexity Analysis	138
4.4	Empirical Analysis	139
4.4.1	Experimental Setup	139
4.4.2	Parameter Setting	140
4.4.3	Varying Δ	141
4.4.4	Analyzing Quality	144
4.4.5	Analyzing Efficiency	147
4.5	Conclusion	150
5	Multiple Influence Maximization across Multiple Social Networks	153
5.1	Introduction	153
5.1.1	Graph Notations	156
5.1.2	Influence Propagation Model	156
5.2	Proposed Approach	158
5.2.1	Identification of Overlapping Users	159
5.2.2	Node Alignment Process	160
5.2.3	Graph Coupling Scheme: Direct Linkage	161
5.2.4	Multiple Influence Maximization	163
5.2.4.1	Generation of product diffusion graph	164
5.2.4.2	Finding non-candidate nodes	164
5.2.4.3	Identifying most influential user for product P_i	165
5.2.4.4	Identifying most influential user among all product diffusion graph	165
5.2.4.5	Influence estimation	166
5.3	Algorithm	168
5.3.1	Applying the Algorithm	172
5.3.2	Complexity Analysis	175
5.4	Empirical Analysis	176

5.4.1	Experimental Setup	176
5.4.2	Analyzing Coupling Schemes	177
5.4.3	Advantages of using IM2	178
5.4.4	Advantages of using MIM	180
5.4.5	Advantages of using MIM2	183
5.4.5.1	Influence spread	184
5.4.5.2	Running time	184
5.5	Conclusion	184
6	Link Prediction using Information Diffusion Perspective	187
6.1	Introduction	187
6.2	Proposed Approach	192
6.2.1	Identification of Community Structure	192
6.2.1.1	Partition phase	193
6.2.1.2	Combination phase	194
6.2.2	Incorporation of Community Importance	195
6.2.3	Computation of Likelihood Score	195
6.3	Algorithm	198
6.3.1	Applying the Algorithm	199
6.3.2	Complexity Analysis	201
6.4	Empirical Analysis	203
6.4.1	Performance Metrics	204
6.4.2	Methods to Compare	206
6.4.3	Parameter Analysis with Community Detection Performance Metrics	215
6.4.3.1	Average isolability	215
6.4.3.2	External density	217
6.4.3.3	Coverage	218
6.4.3.4	Cluster count	219
6.4.3.5	Modularity	220
6.4.4	Parameter Analysis with Link Prediction Performance Metrics	221
6.4.4.1	AUPR	226
6.4.4.2	Recall	226

6.4.4.3	AUC	227
6.4.4.4	Precision	228
6.4.5	Performance Analysis	229
6.4.5.1	AUPR	229
6.4.5.2	Recall	235
6.4.5.3	AUC	235
6.4.5.4	Precision	236
6.4.6	Statistical Test	236
6.4.6.1	Friedman test: non-parametric analysis	238
6.4.6.2	Friedman Conover test: Post hoc analysis	239
6.5	Conclusion	240
7	Concluding Remarks and Future Directions	243
7.1	Summary of Contributions	243
7.2	Scope for Future Work	246
	Bibliography	253
A	Proofs of Theorems	289
A.1	Proof of Theorems in Section 4.1	289
A.2	Proof of Theorems in Section 5.1	290
B	List of Publications	293