

Contents

List of Tables	xi
List of Figures	xiii
1 Introduction	1
1.1 Background	2
1.1.1 The online marketing industry	3
1.1.2 The information diffusion through social networks problem	4
1.1.3 Related work	9
1.2 Research hypothesis	15
1.3 Thesis objectives	15
1.3.1 General objective	15
1.3.2 Specific objectives	15
1.4 Expected results	15
1.5 Structure of this report	16
2 Diffusion: General Background	17
2.1 The Diffusion of Innovations	18
2.1.1 The importance of imitation	18
2.1.2 A unified framework	22
2.1.3 The Bass Model	28
2.2 Epidemiology: the Diffusion of Pathogens	31
2.2.1 The SIR model	31
2.3 The voter model	35
2.4 The importance of the network structure	39
3 Decision Making Models	43
3.1 Perceptual choice: The leaky, competing accumulator model	43
3.1.1 The equations governing the LCA model	44
3.1.2 An analysis of the LCA dynamics	46
3.2 The logit model	50
4 Information retrieval for Preference Extraction	52
4.1 Some background on IR	53
4.1.1 Boolean models	55
4.1.2 Vector models	56
4.1.3 Probabilistic models and Bayesian networks	57

4.2	The latent Dirichlet allocation	58
4.2.1	Generative process	59
4.2.2	Joint and marginal distributions	60
4.2.3	Inference and parameters estimation	61
4.2.4	Use of LDA in this work	64
5	Description of the models	65
5.1	Forum-Agent system framework	65
5.1.1	Forum structure	66
5.1.2	Users behavior	67
5.2	Decision-making models of OSN users	69
5.2.1	Main model: LCA-based and random	69
5.2.2	Benchmark: a logit-based random model	71
5.2.3	Benchmark: a purely random model	71
5.2.4	Benchmark: a deterministic model	72
5.3	Discussion	72
6	Methodological framework	74
6.1	Experimental data set	75
6.1.1	The Plexilandia web forum	75
6.1.2	Available data	76
6.2	Data processing	77
6.2.1	Posts processing	78
6.2.2	Users processing	78
6.3	OSN simulation	81
6.4	Results analysis	84
6.4.1	Contents generation analysis	85
6.4.2	Graph generation analysis	86
6.4.3	Contents variance temporal analysis	90
6.5	Discussion	91
7	Results	92
7.1	Contents generation	92
7.1.1	Average weekly MAPE	93
7.1.2	Average monthly MAPE	95
7.2	Graph generation	97
7.2.1	Average weekly F-measure	97
7.2.2	Average monthly F-measure	97
7.2.3	Average global F-measure	97
7.3	Contents variance	99
7.4	Discussion	100
8	Conclusions	102
A	Notes on Bass' paper	105
A.1	Calculation of the expected time to purchase	105
A.2	Errors in page 223	106

B	Deduction of logit probabilities	107
C	Contents generation results	109
D	Graph generation results	118
E	Contents variance	155
F	Bibliography	160