



Influential Users Detection in Online Social Networks

Nouran Ayman Roushdy Abd Al-Azim

A thesis submitted to the department of Information Systems,
Faculty of Computer and Information Sciences,
Ain Shams University.

In partial fulfilment of the requirements for the degree of Master of Science
in Computer and Information Sciences

Supervised by:

Prof. Tarek F. Gharib

Head of Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University

Assistant Prof. Mohamed Hamdy
Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University

Dr. Yasmine Afify
Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University

2020

Acknowledgements

I would like to thank Prof. Tarek F. Ghraib for his continuous guidance, motivation, support and for enlightening me the first glance of research.

I express my sincere gratitude to Dr. Mohamed Hamdy for his immense knowledge and fruitful supervision.

I would like to show my appreciation to my life-long role model Dr. Yasmine Afify for her continuous advices and for the sleepless nights.

I must express my very profound gratitude to papy and mummy for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis.

This accomplishment would not have been possible without them.

Mahy and Dido, you are the ones who pull me to earth when I feel so down, you are the ones who defend me, thank you, and I love you both.

I would like to show my full appreciation to my grandparents, I hope all of you are here today with me. Nanna and Daddi, you are my source of joy and delectation. Gedo and Toufa, you are my source of endurance. I am looking forward to making you proud. I love you unconditionally.

Mariam and Lobna, you are always here for my ups and downs. Our delightful moments are priceless. I love you.

FCIS-IT team, you are the most dependable, trustful and friendly team I have ever met. I want to express my sincere thanks to all of you.

Finally, I owe thanks to a very special person, my fiancé, Faysal for his continued and unfailing love, support and understanding during my pursuit of master degree that made the completion of thesis possible. You are always around at times I thought that it is impossible to continue, you helped me to keep things in perspective. I greatly value your contribution and deeply appreciate your belief in me.

Abstract

Social networks are considered one of the main merits of this era. People worldwide use this online platform to build their own social ties. The key feature behind the success of social networks is *microblogging*. This feature facilitates the interactions between people around the globe. People use social networking platforms to share their ideas, populate their beliefs and find other people with the same preferences.

Social network users tend to interact with each other by sharing, commenting and reacting to disseminated content. These interactions help in the content spread across the network. The dynamics of user interest in the disseminated content leads to the clustering of social network users to varying groups (communities) called “*interest groups*”. The analysis of users behaviour raises some crucial questions about who is responsible for content spread, the roles played by users in an interest group, the user rank based on his/her role and the rank of the interest group as a whole.

Our research objective is to propose ranking models that take into consideration the dynamic nature of social networks topology and the users interest to tackle the previously mentioned limitations. In order to achieve this objective, four models are proposed. First, Influence Ranking Model (IRM) which aims to rank all the social network users based on their interactivities. It introduces the usage of weighted and directed graph with the classic kshell

decomposition methodology. The uniqueness of the obtained ranking list of IRM is on average equals 1 and network coverage is improved by 0.3%.

Second, Interest Group Identification (IGI) model which aims to cluster users based on their interest in the disseminated content. The quality of separation of IGI reaches 0.923.

Third, Influence Propagation (IP) model which aims to identify the role played by each member in the interest group to spread content to other members. Moreover, a new role called “*ultimate observer*” is introduced. Then these roles are used to rank interest group members based on their contribution in content dissemination. The distinction of ranking of IP is on average equals 1 and its network coverage is competing with the benchmark approaches.

Finally, UltRank model which aims to rank the interest group as a whole to reach the goal of ultimate rank using a new reachability metric. It takes in consideration: 1.Distance from interest group to the other groups. 2.Size of the interest group compared to the size of all the reachable interest groups. 3.Number of reachable interest groups with respect to the number of all interest groups in social network. Meanwhile, new role called “*bridge nodes*” is presented. The ranking capability and network coverage results of UltRank is improved by 1.6% and 4% respectively compared to the benchmark approach. These promising results encourage the employment of the proposed models in different applications in social networks such as viral marketing, monitoring public opinion, event prediction and recommendation systems.

Contents

1	Introduction	17
1.1	Research Problem	19
1.2	Research Motivation	19
1.3	Research Objective	19
1.4	Research Contribution	20
1.5	Thesis Structure	22
2	Related Work	23
2.1	Community Detection in Social Networks	24
2.2	Ranking in Social Networks	27
2.2.1	User Ranking in Social Networks	27
2.2.2	Community Ranking in Social Networks	32
2.3	Summary	36
3	Influence Ranking Model for Social Network Users	39
3.1	Kshell Decomposition Methodology	40
3.2	Social Network Representation	42
3.3	Proposed Influence Ranking Model	43
3.3.1	Preliminaries	43
3.3.2	Proposed Influence Ranking Model Phases	45

3.3.3	Case Study	46
3.3.4	Pseudo Code and Complexity Analysis	47
3.4	Proposed Influence Ranking Model Evaluation and Discussion	50
3.4.1	Datasets	50
3.4.2	Benchmark Approach	50
3.4.3	Evaluation Metrics	51
3.4.4	Experimental Results	51
3.5	Insights on Results	55

4 Influence Propagation: Interest Groups and Node Ranking Models 57

4.1	Proposed Interest Group Identification and Influence Propagation Models	58
4.1.1	Dynamic Nature of Social Networks	59
4.1.2	Preliminaries	60
4.1.3	Modeling Interest Groups and Node Ranking	61
4.1.4	Proposed Interest Group Identification Model	64
4.1.5	Proposed Influence Propagation Model	64
4.1.6	Case Study	66
4.1.7	Pseudo Code and Complexity Analysis	68
4.2	Proposed Interest Group Identification and Influence Propagation Models Evaluation and Discussion	71
4.2.1	Datasets	71
4.2.2	Benchmark Approaches	73
4.2.3	Evaluation Metrics	73
4.2.4	Experimental Results	75
4.3	Insights on Results	85

5	Interest Group Ultimate Ranking Model	89
5.1	Proposed Interest Group Ultimate Ranking	90
5.2	Interest Groups Ultimate Ranking	91
5.2.1	Social Network Dynamics	91
5.2.2	Interest Groups Identification and Ranking	92
5.2.3	Preliminaries	93
5.2.4	Proposed Ultimate Ranking Model Phases	96
5.2.5	Case Study	98
5.2.6	Pseudo Code and Complexity Analysis	104
5.3	Proposed Ultimate Ranking Model Evaluation and Discussion	109
5.3.1	Datasets	109
5.3.2	Benchmark Approaches	111
5.3.3	Evaluation Metrics	111
5.3.4	Experimental Results	113
5.3.5	Experimental Results Discussion	121
5.3.6	Insights on Results	122
6	Conclusion	125
7	Future Work	129
	References	131

List of Figures

1.1	Social Network Users Ranking System Architecture	20
3.1	Kshell Decomposition Methodology Flowchart	41
3.2	Graph Representation of Rigid Relationships between SN Users	42
3.3	Graph Representation of Content Dissemination Direction between SN Users	43
3.4	Proposed IRM Pipeline	45
3.5	Graph Representation of SN, Nodes Represent Eight SN Users and Edges Represent Flow of Content in SN	46
3.6	Number of Nodes in Each of the Top 5 Shells (Bitcoin Alpha Trust Network)	52
3.7	Number of Nodes in Each of the Top 5 Shells (Bitcoin Alpha Trust Network)	52
3.8	IC Model for Advogato Trust Network	54
3.9	IC Model for Bitcoin Alpha Trust Network	54
4.1	Graph Representation of SN	59
4.2	Graphical Representation of Follow Relationship Between Nodes in SN	60
4.3	Graphical Representation of Content Propagation Between Nodes in SN	60

4.4	SN Nodes are Clustered as Three Interest Groups. This Kind of Node Clustering Guarantees The Content Dissemination in Each Group	62
4.5	The Separated Interest Groups out of $G(N, E)$	66
4.6	Network Sizes vs. Number of Interest Groups Under Pruning Factor k	78
4.7	The Effect of Using Multiple Number of Ultimate Observers on Ranking Success Factor	79
4.8	Comparison Between IP Model and Different Benchmark Approaches in Terms of Ranking List Distinction Using Monotonicity Relation $M(R)$	82
4.9	Comparison Between IP Model and Different Benchmark Approaches in Terms of Network Coverage Using Independent Cascade (IC) Model	84
5.1	Social Network Graphical Representation	91
5.2	Different Content Propagation Between SNs' Nodes	92
5.3	Interest Groups Identification in SN	94
5.4	Graph Summarization for the Extracted Interest Groups	95
5.5	SN Graphical Model Where Nodes and Edges Represent Users and Their Interactivities Respectively	99
5.6	SN Users Decomposition Based on Common Interest	100
5.7	Bridge Nodes Identification in SN	101
5.8	SN Graphical Model Where Nodes Represent Either Interest Groups as a Whole or Bridge Nodes While Edges Represent Their Interactivities	102
5.9	Silhouette Coefficient Ranges for Network Nodes	116

LIST OF FIGURES

5.10 Jaccard Coefficient vs. Jaccard Coefficient Distribution . . .	117
5.11 IC Model for UltRank Model vs. Benchmark Approaches . .	120

List of Tables

3.1	Application of IRM on SN Representation	47
3.2	Average Shell Load (ASL)	53
4.1	IP for Interest Group A	67
4.2	IP for Interest Group B	68
4.3	IP for Interest Group C	68
4.4	Topological Features of the Datasets	72
4.5	MAD Results for Two Datasets with Number of Interest Groups = 1	77
4.6	Silhouette Coefficient Results for Three Datasets with Num- ber of Interest Groups > 1	77
4.7	RSF for IP Model vs Different State of Art Approaches . . .	81
5.1	UltRank Weight for Each Interest Group in Summarized Graph	103
5.2	Exp(I) Datasets Topological Features	110
5.3	Exp(II) Datasets Topological Features	111
5.4	Kendall Tau for Proposed UltRank Model vs. Benchmark Approaches	113
5.5	Exp(II) Number of Interest Groups vs Pruning Factor K . . .	114
5.6	Silhouette Coefficient Results for Network Nodes	115
5.7	Average Silhouette Coefficient Results for Interest Groups .	117