



Developing a Mining Methodology for Customer Relationship Analytics

A thesis submitted in partial fulfillment of the requirements for the degree of
Ph. D. in Computer and Information Sciences
To
The Department of Information Systems Faculty of Computer and Information
Sciences, Ain Shams University

By

Nivin Atef Helal Amin

M.Sc. in Computer and Information Sciences (2008)
Ain Shams University

Under the supervision of

Prof. Dr. Mostafa Gadai-Haqq M. Mostafa

Computer Science Department, Faculty of Computer and Information Sciences,
Ain Shams University

Prof. Dr. Nagwa Lotfy Badr

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

Assoc. Prof. Rasha Mohammed Ismail

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

Acknowledgment

First of all, thanks to God for giving me the will and strength to finish this work.

To the soul of my unforgettable father who spent his entire life sacrificing for raising me. My sincere appreciation goes to my mum, husband, brothers and girls. Your perseverance and words of encouragement gave me the extra energy to see this work through.

I sincerely appreciate the valuable guidance and supervision I received from Prof. Dr. Mostafa Gadal-Haqq. His constructive criticism and advice gave me the needed drive to successfully complete this work

I wish to express a deep sense of gratitude and appreciation to Prof. Dr. Nagwa Badr for believing in me even more than I did and for her continuous positive encouragement.

I will be an ingrate without recognizing the valuable supervision from Dr. Rasha Ismail. Thanks a lot for being my mentor and my support all through the work.

Finally, I would like to thank my friends who trusted in my abilities.

Abstract

Online social networks emergence enables individuals to communicate and share their opinions and feedback. Social networks mining aims at applying mining techniques to online social networks to reveal interesting hidden patterns on human behavior and interaction. Organizations utilize social network mining to expand their market and to improve customer social relations.

In this work, we propose a novel social network mining approach for social customer relationship analysis. In our approach, we propose a community detection technique which benefits from the most influential users on the network. As communities tend to be formed around users of great influence on their peers, the proposed approach utilizes such influential users to build communities around them.

Moreover, we propose a new community detection algorithm that incorporates behavioral information attached to users in the social network. Using such behavioral data of nodes is for the aim of detecting communities that are closely mapped to the underlying behavioral communities in real social networks.

We use the behavioral data, namely, the actions done by users on their social network to propose a new similarity measure to measure the degree of similarity between users. Furthermore, the proposed algorithm uses the demographic data of users to enhance the quality of communities detected.

Experimental evaluation on two real social network datasets has been carried out and the results show that the proposed social network mining approach surpasses others in respect of all the evaluation measures used which indicates the ability of the proposed approach in identifying communities with high quality.

Publications



1. Nivin A. Helal, Rasha M. Ismail, Nagwa L. Badr, Mostafa G. M. Mostafa, *A Novel Social Network Mining Approach for Customer Segmentation and Viral Marketing*, WIREs Data Mining Knowl Discov 2016. doi: 10.1002/widm.1183 (ISI IF=1.759)
2. Nivin A. Helal, Rasha M. Ismail, Nagwa L. Badr, Mostafa G. M. Mostafa, *An Efficient Algorithm for Community Detection in Attributed Social Networks*, INFOS '16, May 09-11, 2016, 198-202.
3. Nivin A. Helal, Rasha M. Ismail, Nagwa L. Badr, Mostafa G. M. Mostafa, *LBCD: A Leader Based Community Detection Algorithm for Social Network Clustering*, Submitted to WIREs Data Mining Knowl Discov, May, 2016 (ISI IF=1.759).
4. Nivin A. Helal, Rasha M. Ismail, Nagwa L. Badr, Mostafa G. M. Mostafa, *Towards an Enhanced Social Network Mining Approach for Social Customer Relationship Analytics*, Submitted to Intelligent Data Analysis journal, June, 2016 (ISI IF= 0.631).
5. Nivin A. Helal, Rasha M. Ismail, Nagwa L. Badr, Mostafa G. M. Mostafa, *Social Network Mining and Analysis: Community Detection and Influential Analysis*, Submitted to the Asian Journal of Information Technology, July 2016. (SJR: 0.18)

Table of Contents

Abstract.....	I
Publications.....	II
Table of Contents	III
List of Figures.....	VI
List of Tables	VIII
List of Abbreviations	IX
Chapter 1 Introduction.....	1
1.1 Motivation.....	1
1.2 Objective	2
1.3 Thesis Organization	4
Chapter 2 Background	6
2.1 Social Networks	6
2.1.1 Social Network Representations.....	7
2.1.3 Characteristics of Social Networks	9
2.1.4 Static and Dynamic Social Networks	11
2.1.5 Social Network Mining and Analysis.....	12
2.2 Customer Relationship Management.....	16
2.3 CRM and Data Mining	18
2.4 From CRM to Social CRM.....	19
2.5 SCRM and Social Network mining	22
2.6 Summary	24
Chapter 3 Related Works.....	25
3.1 Community Detection.....	25
3.1.1 Global Community Detection.....	27

3.1.2 Local Based Community Detection.....	32
3.1.3 Attributed Community Detection	35
3.1.4 Overlapping Community Detection	37
3.1.5 Community Evaluation Measures	38
3.2 Influential Analysis.....	44
3.2.1 Influence Maximization.....	44
3.2.2 Direct Data Mining Approaches.....	46
3.3 Summary	46
Chapter 4 The Proposed Social Network Mining Approach for Customer Relationship Analytics	47
4.1 The Proposed Approach.....	48
4.2 The Proposed Architecture	49
4.2.1 Social Network Pre-processing	49
4.2.2 Community Detection based on Influence Propagation.....	50
4.2.3 Detected Communities with their Influential Users	57
4.3 Experimental Results and Discussion.....	58
4.3.1 Flixster Dataset	58
4.3.2 Experimental Setup.....	59
4.3.3 Experimental Results	59
4.4 The Proposed Similarity Measure	62
4.4.1 Experimental Results	62
4.5 Summary	69
Chapter 5 The Proposed Leader Based Community Detection Algorithms....	71
5.1 The Proposed Leader Based Community Detection Algorithm (LBCD)	72
5.1.1 Real Datasets	76
5.1.2 Experimental Setup and Results	77
5.2 The Proposed Leader Based Community Detection Algorithm for Attributed Network (aLBCD)	87

5.2.1 Problem Definition	87
5.2.2 Experimental Results	89
5.3 Summary	92
Chapter 6 Conclusion and Future Work.....	94
6.1 The Proposed Social Network Mining Approach for Customer Segmentation and Viral Marketing.....	94
6.2 The Proposed Leader Based Community Detection Algorithm.....	96
6.3 Leader Based Community Detection Algorithm for Attributed Social Networks	97
6.4 Future Work.....	98
References	100
Arabic Summary	111

List of Figures

Fig 1.1: Graphical Illustration for the Proposed Social Network Mining Approach	3
Fig 2.1: Undirected and Directed Graphs	8
Fig 2.2: Adjacency Matrix for the undirected social network	8
Fig 2.3: Communities in Zachary's Karate Club social network.	13
Fig 2.4: Classification framework for data mining techniques in CRM	19
Fig 2.5: Applying Social Network Mining for Social CRM	23
Fig 3.1: A simple graph with three communities	26
Fig 3.2: Graph partitioning	28
Fig 3.3: Hierarchical Clustering Dendogram	29
Fig 3.4: The Thresholding Step in the Hierarchical Diffusion algorithm	34
Fig 3.5: The Hierarchical Diffusion Algorithm	35
Fig 4.1: Proposed Architecture for Social Network Mining Approach	49
Fig 4.2: Community detection based on Influence Propagation	50
Fig 4.3: a) An Example of Social Graph and b) Log of Actions	51
Fig 4.4: (a) Propagation of action a (b) Propagation of action b.....	52
Fig 4.5: Compute Influence Matrix	55
Fig 4.6: Pseudocode for the Proposed Community Detection Algorithm	57
Fig 4.7: Social Network and Action Log Description	59
Fig 4.8: Average Conductance for the Proposed Algorithm vs. the Hierarchical Diffusion algorithm (Weight=Common Neighbors)	61
Fig 4.9: Average Separability for the Proposed Algorithm vs. the Hierarchical Diffusion algorithm	61
Fig 4.10: Average conductance for the proposed algorithm vs. the Hierarchical Diffusion algorithm	63
Fig 4.11: Average Separability for the Proposed Algorithm vs. the Hierarchical Diffusion algorithm	64
Fig 4.12: Average Conductance for the Proposed Algorithm over Both Similarity Measures (Weight=common neighbors and common actions).....	65
Fig 4.13: Average Separability for the Proposed Algorithm over Both Similarity Measures (Weight=Common neighbors and common actions).....	66
Fig 5.1: Illustrative example on Social Network Graph	73
Fig 5.2: The Proposed Leader Based Community Detection Algorithm	74

Fig 5.3a: No. of leaders at $\pi = 1, 2, 3, 4$ Weeks and $\psi = 2$	78
Fig 5.3b: No. of leaders at $\pi = 1, 2, 3, 4$ Weeks and $\psi = 3$	78
Fig 5.4a: No. of leaders at $\pi = 1, 2, 3, 4$ Weeks and $\sigma = 2$	78
Fig 5.4b: No. of leaders at $\pi = 1, 2, 3, 4$ Weeks and $\sigma = 3$	78
Fig 5.5: Comparing Accuracy for the proposed LBCD and the HDA	79
Fig 5.6: Comparing ARI for the proposed LBCD and the HDA	81
Fig 5.7(a): No. of leaders at $\pi = 10, 20, 30, 60$ days and $\sigma = 3$ Flixster Dataset	82
Fig 5.7(b): No. of leaders at $\pi = 10, 20, 30, 60$ days and $\psi = 3$ Flixster Dataset	82
Fig 5.8: Conductance for proposed LBCD algorithm and the HDA	82
Fig 5.9: Separability for proposed LBCD algorithm and the HDA	83
Fig 5.10: Proposed Attributed Leader Based Community Detection Algorithm	89
Fig 5.11: Comparing Accuracy for the proposed aLBCD, LBCD and the HDA	90
Fig 5.12: Comparing ARI for the proposed aLBCD, LBCD and the HDA	91

List of Tables

Table 3.1: Contingency Table	42
Table 4.1: Comparing the Three Quality Measures for the Proposed Algorithm and the Hierarchical Diffusion Algorithm (Weight=common neighbors)	62
Table 4.2: Comparing the Three Quality Measures for the Proposed Algorithm and the Hierarchical Diffusion Algorithm (Weight=common actions)	64
Table 4.3: Comparing the Three Quality Measures for the Proposed Algorithm over Both Similarity Measures	67
Table 4.4: Relative Community Sizes of the Identified Communities for the Proposed Algorithm.....	68
Table 5.1: Social Network Dataset	76
Table 5.2: Accuracy for the proposed LBCD and HDA	80
Table 5.3: ARI for the proposed LBCD and HDA	80
Table 5.4: Comparing the two quality measure for both the proposed LBCD algorithm and the hierarchical Diffusion Algorithm	83
Table 5.5: Accuracy for the proposed aLBCD, LBCD and the HDA	90
Table 5.6: ARI for the proposed aLBCD, LBCD and the HDA	92

List of Abbreviations

CRM	Customer Relationship Management
SCRM	Social Customer Relationship Management
ID	Internal Density
ARI	Adjusted Rand Index
NMI	Normalized Mutual Information
HAD	Hierarchical Diffusion Algorithm
LBCD	Leader Based Community Detection
aLBCD	Attributed Leader Based Community Detection

Chapter 1

Introduction

With the rapid development of information technology; organizations have a high growth rate in customer information. In addition to that they have recognized that customer information is one of their assets. Thus, the issues of Customer Relationship Management (CRM) have attracted many concerns nowadays and business operations have gradually turned from product-focused to customer-centric.

Customer Relationship Management (CRM) is defined in (Swift, 2001) as an “enterprise approach to understanding and influencing customer behavior through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability”.

By the emergence of social networks where customers can easily communicate, share information, and form new communities online; organizations could utilize social networks in establishing a fruitful relationship with their customers.

1.1 Motivation

Social networks can be used as a platform for large companies to build a fan base and as a channel to communicate with customers to expand their market and to improve customer social relations.

As customers have become the main actors on behalf of the companies, it is necessary for companies to monitor their social behavior pattern and understand their influence. It is crucial for companies to incorporate more analytical skills to analyze social networks content to effectively leverage their social marketing, customer segmentation and retaining tasks.

The rapid growth and widespread use of social networks provides a wealth of social data that necessitates the application of variety of data analytics techniques through social network mining. Applying mining techniques to online social networks can yield interesting hidden perspectives and patterns on human behavior and interaction. Mining techniques can be used on social networks to give more insights into understanding hidden behavior and patterns in social network data.

1.2 Objective

The main objective of this research is to develop an advanced mining approach to excavate hidden information from social networks. The approach combines several aspect of social network mining that can be applied to better leverage the processes of Social Customer Relationship Management (SCRM). Social network mining tasks include community detection, influential analysis, topic detection and monitoring individual behavior analysis, group behavior analysis, and recommending products via marketing research for businesses.

The proposed approach benefits from the task of discovering the influential nodes in the social network in the task of community detection. Communities are built around those influential nodes. The idea is based on the social observation that communities tend to be formed around users with great influence. Fig 1.1 shows a graphical illustration for the main idea of the proposed approach.

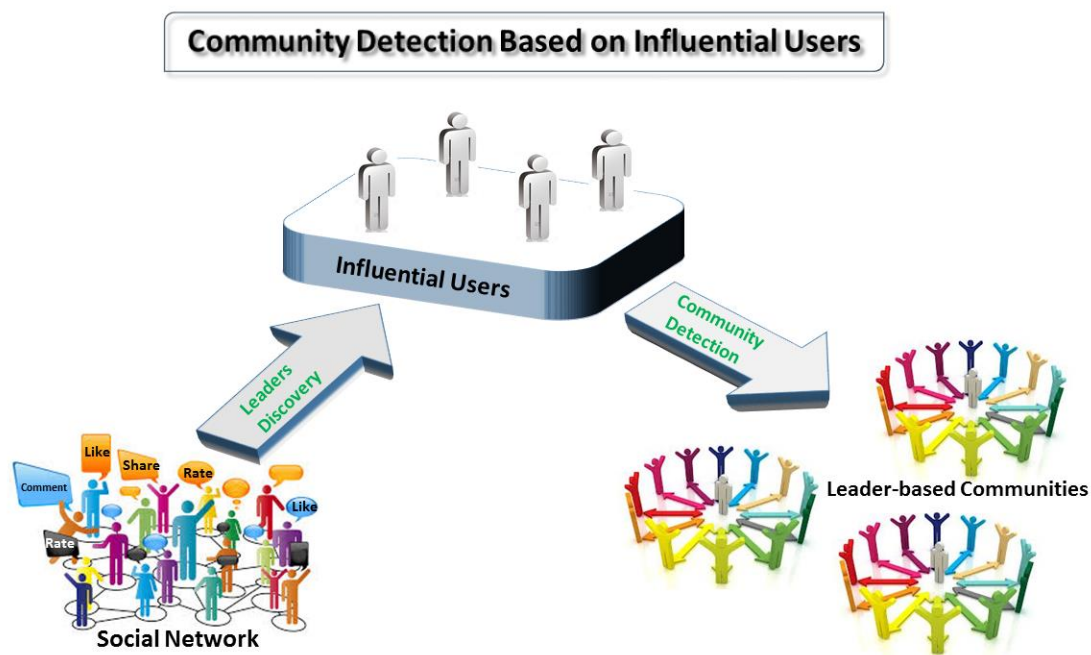


Fig 1.1: Graphical Illustration for the Proposed Social Network Mining Approach

We propose a promising community detection technique that incorporates available information related to nodes. The approach uses the behavior data of nodes for the aim of detecting communities that reveal common behavior among their members. As a result, the detected communities reflect real customers' communities for customer segmentation. In addition to the identified influential users that could be used for viral marketing.

Moreover, the proposed community detection technique incorporates the demographic data attached to nodes to enhance the quality of the communities detected from the perspective of exhibiting common properties among members of the same community.

1.3 Thesis Organization

This thesis is organized as follows:

Chapter 2: Social Networks Background

This chapter introduces an overview of social networks and main concepts regarding social network representations and characteristics. The chapter explores the main tasks of social network mining, namely, community detection, influential node analysis, link prediction, node classification and text mining. On the other hand, it introduces the concept of customer relationship management (CRM) and CRM dimensions. The concept of social CRM (SCRM) is also discussed in this chapter and how to apply the social network mining techniques to leverage the processes of SCRM.

Chapter 3: Related Works

This chapter explores the research done in the fields of community detection and influential node analysis. For community detection, the chapter presents the different approaches used to detect social network communities. In addition to the popular evaluation measures used to evaluate the quality of the communities. Regarding the influential node analysis, the discrete optimization and direct mining technique used to discover leaders are also introduced.