Social Network Discovery and Customer Network Value Assessment from Social Networking Websites[1]

Hong Guo, Praveen Pathak, Hsing Kenneth Cheng
Department of Information Systems and Operations Management
Warrington College of Business Administration, University of Florida
{guohong, praveen, hkcheng}@ufl.edu

Paper Summary

Submitted to The 2008 Winter Conference on Business Intelligence, University of Utah

1. Introduction

Recently, social networking websites have gained a lot of popularity. In the US market 6.5% of all Internet traffic is due to the top twenty social networking websites (Hitwise, February 2007). Aware of its growth potential, companies are beginning to utilize this new marketing channel – online social networks. In 2006, the online social network advertising spending was 350 million dollars and is predicted to grow to more than 2 billion dollars by the year 2010 (eMarketer).

This paper utilizes social networking websites to examine possible marketing strategies through online social networks. Towards the end of social network marketing design it is crucial to learn the underlying social networks. Social networking websites provide an integrated virtual venue for social activities and represent a rich data source for extracting social networks. This paper proposes a novel IR-based social network discovery method to identify the implicit social network. The obtained implicit social network is then incorporated with the explicit social network to form the composite social network as a representative of the underlying social network.

From the perspective of companies who are interested in marketing through online social networks, the discovered social network represents a potential customer network. Based on the information of social networks, companies can further develop the optimal marketing strategies. This paper proposes a simple formulation of social network marketing campaign based on the concept of customer's network value.

2. Research Questions

This paper addresses the following three research questions.

Research Question 1: How to discover the underlying social networks from social networking websites? As a data-intensive platform, social networking websites raise several challenges for social network discovery. First, data from social networking websites is on a very large scale. Each user has numerous friends. Some users even have millions of friends. Users update their profile regularly and communicate with each other frequently. Second, social networking websites involve heterogeneous data with different formats. For example, users' friend lists are structured data and contain information of static social links while blogs and comments are unstructured data written in natural language and contain information of

[1] The first part of this paper was presented at WITS 2007, Montreal, Canada.

dynamic communications among users. These facts make the discovery of the underlying social networks a very tough task.

Research Question 2: How to assess customer's network value based on social network information?
In the literature of customer's life time value to firms, most existing research treats customer's value independently. However, to some extent customers depend on each other to make purchase decisions. Customer's network value captures one form of such interdependence which intuitively is a quantitative measure of a customer's influence on other customers. This paper investigates how to assess customer's network value based on social network information.

Research Question 3: How to carry out social network marketing given the underlying social networks and each member's network value?
Based on the results of the previous two research questions, the ultimate goal of this study is to examine optimal social network marketing strategies.

3. Approach

To address the stated research questions different approaches are taken for different tasks. We apply information retrieval techniques to construct implicit social network while we use pattern matching to extract explicit social network from social networking websites. By combining explicit social network with implicit social network, we obtain the composite social network and use it to represent the underlying social network.

We then adopt concepts and methods from social network analysis to analyze the resulting social networks and compare their structural features. For customers' network value, we define *influence territory* for each customer as the network with all the nodes who are reachable to the focal node and all the edges on the paths to the focal node. Customers have and only have influence within their own influence territories. Then customers' network values can be calculated based on their influence territories. In the context of new product adoption, the network value of a customer is the number of purchases through the influence of the purchase of the focal customer.

Finally, we formulate a simple form of social network marketing campaign as a nonlinear optimization problem given the underlying social networks and each member's network value. The solution to the optimization problem characterizes the structure of optimal social network marketing strategies.

4. Main Findings and Contributions

Current Findings
We apply the proposed method to a large-scale social network obtained from the largest social networking website (MySpace) and find that the composite social network discovered is significantly different from the explicit social network. The explicit social network overestimates the cohesion of the network and underestimates the distances between users.

Expected Findings
Currently we are working on assessing customers' network values. We hypothesize that customers' network value in a social network follows a power-law distribution with only a few customers having high network value and most customers having low network value. The result of customer's network value distribution has important implications for social network marketing. Specifically, by identifying

the top ranked customers in terms of network value and initiating marketing campaign through them, firms are able to reach the desired potential customers at an affordable cost.

Contributions
This paper is among the first to apply information retrieval techniques to extract implicit social networks from unstructured data sources and evaluate the influences on the relationships in the networks. This paper also represents the first attempt to formulate social network marketing campaign based on customer's network value.

5. Current Status of Manuscript

We have completed data collection and social network construction. The resulting social network contains more than 10,000 individual nodes and more than 100,000 edges with estimated influence weight on each edge. Currently we are working on customer network value assessment and expect to finish it by the end of January 2008. We expect to write out a working paper by the end of February 2008.