Targeting Key Influentials for Direct Marketing Activities in Social Networks: Methodical Progress and an Application

Martin Klaus, Jörg Schwerdtfeger and Ralf Wagner
SVI Endowed Chair for International Direct Marketing
DMCC - Dialog Marketing Competence Center
University of Kassel, Germany
{mklaus,rwagner}@wirtschaft.uni-kassel.de
jschwerdtfeger@uni-kassel.de

Abstract
Social networks and virtual communities provide credible information about customers’ opinions and facilitate innovative customer dialogues. We build upon Kosinets’ netnography to extract consumer insights from such services. To cope with the overwhelming amount of data recorded by these services, we combine netnography with innovative social network analysis (SNA). We propose a new criterion to assess the number of influentials which should be involved in communication measures or monitored to assess consumers’ perceptions. Various previous attempts to utilize particular facets of networks or communities for direct marketing have lacked a suitable management framework to guide practitioners. Addressing this gap, we introduce a management cycle of gathering insights and implementing activities.

Keywords: Online community, social network analysis, word-of-mouth communication
Targeting Key Influentials for Direct Marketing Activities in Social Networks: Methodical Progress and an Application

Introduction

Online communities like the blogosphere comprise both user generated content (UGC) and an interaction network structure. In the blogosphere, weblogs are the knots, and their links to other weblogs, comments, posts, blogrolls and trackbacks are the edges (Leskovec, 2007). For this paper, we focus on the blogosphere because it is open and freely accessible to everyone without registration or login (Efimova, 2003). Direct marketers aim to utilize virtual networks and communities for communication with their prospective customers or even for selling their products and services directly rather than limiting their activities to monitoring customer insights. Net citizens frequently insist on individual interactions and ignore digital mass communication measures. As a consequence, Web 2.0 communication is expensive, although it may well become highly effective through viral effects. An impressive number of 126 million weblog users (BlogPulse, 2010) poses a major challenge for direct marketers as no structure method exists to cope with analysis and communication tasks related to such numbers. Although progress has been made to categorize the blogs with respect to their influence in the sub-communities (Klaus and Wagner, 2009, Klaus et al., 2009), practitioners are still unsure how many influentials should be monitored. This paper proposes an approach of reducing the mass of data and gaining the important consumer insights from only a few influential actors of online social networks. By seizing the challenge of identifying the relevant knots in a communication network, the methodology covers two distinct direct marketing tasks: first, the identification of information and opinion aggregating individuals, and second the identification of the opinning that might trigger viral marketing effects. This paper aims to:

- Propose a criterion to assess the number of the most relevant knots that should be considered.
- Demonstrate the extraction of specific “consumer insights” concerning a chosen topic from selected central weblogs.
- Propose a management cycle connecting the stages of insight gathering and communicating to a recursive work flow.

This will guide practitioners in selecting and analyzing social networks as well as in monitoring the impact of their direct marketing measures. We demonstrate the practicability of this approach using the example of a weblog community discussing mobile communication and cellular phones.

In the following section, we give a brief overview of related work. The methodology is outlined in Section 3. Section 4 provides a brief data description, and Section 5 covers both the results and the discussion. Finally, conclusions are drawn in Section 6.

Related Work

The structure of the Internet and social networks have been studied for many years (e.g., Albert et al., 1999; Barabási, 2007, Milgram, 1967), but the upsurge of communities in the WWW within the past ten years has increasingly attracted the attention of marketing scholars and practitioners (Zhang et al., 2007). The usability of already established social networks for marketing and market research, or as a supplemental instrument for sales forces, has already been investigated. Some of the most recent publications meet the claim of Subramani and
Rajagopalan (2003) to overcome the limitations of descriptive accounts of particular initiatives and advice based on anecdotal evidence when considering digital word-of-mouth phenomena. Recent research focuses on quantitative measures related to the results of the activities. These include the adoption of new ideas, products or opinions (Cheung et al., 2008; Ma et al., 2008; De Bruyn and Lilien, 2008), the impact on purchase probabilities (East et al., 2008) and reputation-related issues (Helm, 2000; Reichheld, 2003).

Weblogs are of particular interest to marketers because of the high variety of topics discussed within the blogosphere (Glance et al., 2004). The thematic context reconnaissance and the analysis of weblog entries are researched mainly by Nanno et al. (2004), Qamra et al. (2006), Anjewierden et al. (2005), Avesani et al. (2005), Berendt and Navigli (2006), Gill (2004) and Adamic and Glance (2005). Structural characteristics of the blogosphere are researched, among others, by Chin and Chignell (2005), Tseng et al. (2005), Merelo-Guervos et al. (2004) and Leskovec et al. (2007). These studies, however, are restricted either to:

- considering topic detection and their relation digital communication behavior or
- position in the communication network.

To have a comprehensive understanding, both aspects need to be integrated in the analysis framework (Klaus and Wagner, 2009, Klaus et al., 2009).

### Methodology

The approach to quantitatively reduce the data mass of social network information and afterwards qualitatively gain the aggregated consumer insights of the community uses the characteristics of the Internet and the blogosphere.

First, the structure is scale free, which allows us even to analyze smaller networks like thematically differentiated communities with only 100-500 users (Barabási, 2007; Ravid and Rafeli, 2004). Second, Broder et al., 2000 have shown in their research that the linking structure of web pages draws topic clusters which they call thematically unified clusters (TUCs). These clusters have a highly connected core in which the information of the community is accumulated (Dill et al., 2001). Thus a number of linked weblogs related to a topic draw clusters in which those with very high centrality, if there are any, accumulate the information to the topic. Thus, it is not necessary to read all the weblogs of the network, but only the central ones to access nearly all the information. Third, the distribution of links in the blogosphere follows the long tail phenomenon (Barabási, 2007). The long tail distribution emerges if a majority of individuals have a relatively small number of links to other users. A few users (called hubs) have a high number of link relations to other individuals (Pennock et al., 2002). This means that there are weblogs which have higher centrality and aggregate the information of the TUCs. This approach hints at the need for identifying the relevant individuals.

We identify the hubs, weblogs with a promising position, by one of the following three centrality measures: degree, closeness and betweenness, which have been proposed for the identification of “important” individuals (Everett et al., 1999; Klaus and Wagner, 2009).

The degree centrality is deemed to be the dimension of possible communication activity within the network. Thus, we assess how applicative profiles are in order to start canvassing on these profiles with a high degree of centrality.

The closeness centrality is deemed to be the dimension of independence from other profiles because the closer the centrality of a profile is, the more direct connections are linked to it. Considering the distance from one profile to all the other profiles in the graph, the closeness centrality indicates how fast a marketing communication measure could spread through the network, starting from one profile.
The betweenness centrality assesses the opportunities for controlling the communication process. In this way, communication from these profiles can be monitored and assessed by marketers with a view to influencing them as they wish.

Once the most central hubs for the three centrality measures are identified, we extract subnetworks from the original community network, which include the hubs, all weblogs they are directly connected with, called alteri, and the connections between the alteri. These subnetworks lead to a relation between the number of chosen hubs and the number of alteri from the whole community network the hubs reach directly.

In addition, we use netnography (Kozinets, 2002) to analyze and interpret the content data from the identified hubs to gain key insights following the four steps of the methodology:

1. Making a cultural entrée
2. Gathering and analyzing data
3. Conducting ethical research
4. Providing opportunities for culture member feedback

Finally, this paper’s approach is summarized in a management cycle to give an overview of how to use the methodology in research and practice.

**Application Domain and Data Description**

We chose the often discussed topic of “mobile communication and phones” in the English-speaking blogosphere to demonstrate exemplarily this paper’s approach. Blog URLs related to this topic were collected with weblog search engines like blogsearch.google.com and technorati.com, using key terms like “cell phone”, “mobile phone” and “mobile communication”. Not only general topic-related blogs were searched, but also blogs related to mobile phone brands and products. In total, 499 blog URLs were collected. The data were crawled with the SocSciBot crawler by Thelwall (2004, p. 245ff.). After small subnetworks, which could be identified as blog spam farms, and isolates were excluded from the network data, the final blog community network concerning the topic of mobile communication and phones included 415 English-speaking weblogs with 2,838 edges between them. This weblog network has a density \( d(\text{whole net}) = 0.016\% \) and it is the database for all the following analyses.

**Results and Discussion**

The structural analysis of the community with 499 weblogs, dealing with the topic of mobile communication and phones showed first that the community exhibited the characteristics of the long tail phenomenon and was scale free. Thus, it was reasonable to detect hubs in the network with the assumption that these hubs accumulated the information of the network.

The following table gives an overview of the results of the analysis using the three centrality measures, their subnetworks and the relation between the hubs and the alteri which can be reached by the hubs.

<table>
<thead>
<tr>
<th></th>
<th>Whole community Network</th>
<th>Degree-Subnetwork</th>
<th>Betweenness-Subnetwork</th>
<th>Closeness-Subnetwork</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knots</td>
<td>415</td>
<td>266</td>
<td>245</td>
<td>253</td>
</tr>
<tr>
<td>Number of chosen central knots</td>
<td>–</td>
<td>12</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Reached percentage of the main network</td>
<td>–</td>
<td>57.63%</td>
<td>55.55%</td>
<td>55.55%</td>
</tr>
<tr>
<td>Edges</td>
<td>2,838</td>
<td>2,421</td>
<td>2,351</td>
<td>2,365</td>
</tr>
<tr>
<td>Density</td>
<td>0.016</td>
<td>0.034</td>
<td>0.04</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 1: Overview of the SNA results for the weblog network and the built hub subnetworks
The number of hubs from which to extract a subnetwork was determined by the gap criteria (Tibshirani et al., 2001) method which chooses the point where the relation between the number of hubs and the number of new alteri becomes the first time one. This means we chose the number of hubs at the point where one more hub to build the subnetwork would not add even one more alteri only the this hub is connected to in the new subnetwork.

For example concerning the degree centrality for active marketing in the network with the gap criteria, the 12 weblogs with the highest degree centrality need to be selected to extract the subnetwork from them, their alteri and the relations among them. Theses 12 hubs reach 266 weblogs by one direct link and their subnetwork encompasses 57.63% of the whole community network. The density of 0.034, which is twice as high as the density of the whole community network, shows that a more strongly connected core is identified, which implies that the information is accumulated on the hubs. Similar results were produced in the analysis for the betweenness and closeness centrality, as shown in Table 1.

Next, we used netnography to analyze exemplarily the content of the 12 degree central hubs for the past three months from the day of the structural analysis backwards to gain consumer insights. In total, 71 consumer insights about products and services from 25 different companies were found for five categories: hardware, software, hard- & software, accessories, studies and discussion topics. All insights contained at least one positive or negative valued proposition or gave new ideas, approaches or critiques.

It should be pointed out that 61.97% of the consumer insights (CI) deal with one of the four leading brands: Nokia (15 CIs), Apple (14 CIs), HTC (8 CIs) and Samsung (7 CIs).

For each of the three centrality measures, there are only a certain number of possibilities for marketing actions in the social network of the blogosphere:

*active interaction with customers: degree centrality*
  - react on posts and comments
  - integrate a corporate weblog in the community

*monitoring: betweenness centrality*
  - subscribe for RSS Feeds
  - read weblogs

*classic online advertisement: closeness centrality*
  - banners and widgets on weblogs
  - ads in comments
  - become listed in blogrolls of hubs
  - direct contact with hub weblogs
  - offer samples and incentives to hub bloggers

Finally, the approach of this paper can be summarized in a management cycle, shown in Figure 1. The cycle is composed of seven stages which must be processed consecutively. First, a type of social network like the blogosphere should be chosen as a database. Also, at the same stage, a topic should be defined. At first, all the data, and afterwards only the new, added data need to be crawled. Continuing into the third stage, one or more social network measures will be considered where each measure is related with a different type of marketing activity, as explained above. Using each measure, the most influencing knots of the network can be determined. This helps to reduce the data considerably for the next step, where netnography is used. This methodology requires reading all the content from the influential actors to gain consumer insights. Depending on the measures, which were chosen beforehand, the next stage challenges the execution of a related marketing activity.
All the activities have to be controlled. Once they run well, the cycle continues by just updating the new, added content periodically and then checking if this leads to any changes in the measures or brings out new insights. If the data do not bring out enough or only bad results, the social network and/or the topic can be expanded or changed. Following this management cycle might help businesses to generate structured procedures to use consumer-generated content combined with the structure of online networks to start very promising direct marketing activities.

**Conclusion**

In this paper, we aim to extend the methodology for revealing the systematic patterns of influencers and recipients in online social networks. We analyzed the characteristics of the Internet and the blogosphere to show that they are well suited for digital word-of-mouth communication and viral marketing activities. Moreover, we outlined three measures for the degree of centrality of an individual and the related interpretation. In the case of our weblog network, we found that involving about 8-12 hubs (1.9-2.8%) would be sufficient to bring up to 57.63% of the whole weblog community into contact with the marketing communication. The hubs were examined to gain consumer insights about brands, products and topics from which marketing action alternatives were finally derived. The selection of important weblogs out of the whole network meant an expedient reduction of burdens in the subsequent netnographical analysis, whereby marketers can reach more customers with less effort in marketing activities. The stepwise procedure is assembled in the management cycle. This emphasizes the benefit of complementing the qualitative analysis with a quantitative social network analysis. Continuing working in this research area, we need to evaluate two mode networks with additional information describing the actors in order to provide researchers with even more informative clustering results.
References


