Abstract

With the popularity of social networking sites (SNS) in this era of Web 2.0, increasingly more users are contributing their opinions about products and organizations. These online comments often have direct influence on consumers' buying decisions and the public's impressions of enterprises. As a result, enterprises have begun to use SNS to conduct targeted marketing and reputation management. As indicated from recent marketing research, the joint influence power of a small group of active users could have considerable impact on consumers' buying decisions and the public's perception of the enterprises. To help enterprises conduct cost-effective targeted marketing and reputation management, this paper illustrates a novel methodology that can effectively discover the most influential users from SNS. In particular, the general methodology of mining the influence network from SNS and the computational models of mathematical programming for discovering the user groups with the maximal joint influence power are proposed. The empirical evaluation with real data extracted from SNS shows that the proposed method can effectively identify the most influential groups when compared to the benchmark methods. This study opens the door to effectively conducting targeted marketing and enterprise reputation management on SNS.

Keywords: Social Network Mining, Web Mining, Online Marketing, e-Commerce, e-Business
Introduction

In Web 2.0 era, users are permitted to express their opinions on products through many channels such as online forums, shopping websites, blogs, and wikis. These opinions can influence other users’ buying decisions and their views on companies (Cheung et al. 2009; Chevalier and Mayzlin 2006; Dellarocas 2003; Hennig-Thurau and Walsh 2003; Koh et al. 2010; Mayzlin 2006; Park et al. 2007). Recently, an emerging channel of Social Networking Sites (SNS) such as Facebook, Twitter, and Epinions has attracted the attention from marketing practitioners and researchers. These sites not only permit users to express comments and opinions on products, people, organizations, and many other entities, but users are also able to build various social relationships. For example, on the Epinions site (Epinions.com 2010), one user can build a trust relationship with another by adding him or her to a trust list, or the user can block him or her with a block list. The site then shows the trusted users’ opinions at the top of the list. With these social relationships, opinions will have greater impact on users than those expressed on the other channels (such as shopping websites), because people always believe or accept more easily the opinions of those with whom they have social relationships (Golbeck 2005; Lu et al. 2010; Massa and Avesani 2007). In addition, the influence of opinions in SNS can be disseminated more widely than other channels. Thus, some users’ opinions captured at SNS can greatly influence other users’ buying decisions or their views on certain companies.

Many business entities have recently come to recognize this phenomenon, and some companies have begun to identify certain users to conduct online marketing and reputation management (Conlin and MacMillan 2009; Marks 2009; Miller and Dickson 2001). SNS are becoming an important platform for e-commerce and e-business. For companies to better utilize SNS for cost-effective, targeted marketing and reputation management, they must face an important issue: given a large number of SNS users and limited budget, how can we identify the users whose opinions are the most influential on the others’ actions? If the most influential group of users could be identified, companies could consume minimal resources to improve product sales and enhance their reputations.

Although there are many existing studies on measuring node importance in social network analysis (Wasserman and Faust 1994), as well as studies that explore the spread of influence in social networks (Kempe et al. 2003; Kempe et al. 2005), these works emphasize the importance of each node, without considering the joint influence power of a group of nodes. According to the latest findings from marketing research (Katona et al. 2007; Katona et al. 2010), if the customers are provided with positive information on products or enterprises by all related users in online communities, they may have a higher probability of purchasing such products or having positive perceptions of these enterprises. This is called as the joint influence power of a group of users. Previous research also indicates that the joint influence power of a small group of users could have considerable impact on a large number of consumers. Therefore, marketing personnel should identify the users who have great joint influence power in SNS, and find ways to encourage these users to express positive opinions about companies and the companies’ products through the strategy of targeted marketing. As a result, companies could maximally promote product sales and improve enterprises’ reputations through the joint influence power of the specific group of users.

Effectively discover the group of users with maximal joint influence power from the huge number of users in SNS has become the key issue for companies to conduct targeted marketing and reputation management. Although previous research has examined the problem of discovering a group of influential users, the heuristic method does not address the issue of identifying a group of with maximal joint influence power (Zhang et al. 2008). One of the weaknesses of the previous research is that the users are added into the target group one by one according to their attributes and the partial trust relationships of them, without considering the whole influence power of the target group on a whole, so this method usually discovers some target groups consisting of mutual influence users, which always have not most joint influence powers on other users (Kempe, et al. 2003). In contrast, our proposed method represents global influence relationships among all users as a directed graph, and uses Mathematical Programming as the computational apparatus to discover the group of users with maximal joint influence power. By considering the cost of marketing, the proposed models can discover the target group with arbitrary sizes. The empirical evaluation with real data from Epinions and Twitter websites shows that the proposed method can improve the performances greatly, compared to the two benchmark methods of discovering influential target groups.
In summary, the main contribution of our research is the development of a novel methodology for online targeted marketing and enterprise reputation management based on widely available data captured at SNS. In particular, the proposed computational method can discover the user group with maximally joint influence power in a cost-effective way; it also overcomes the disadvantages of the existing methods (Zhang et al. 2008), which only emphasize the importance of single user, or only consider the partial influence relationship of users. Our research opens the door to apply widely available data captured at SNS to conduct targeted marketing and enterprise reputation management in e-commerce and e-business.

This paper is organized as follows. Section 2 discusses related research work. Section 3 provides an overview of the proposed methodology for cost-effective targeted marketing and enterprise reputation management. Section 4 presents the computational models for discovering the most influential groups of users. Section 5 reports the experimental evaluation of the proposed method, and Section 6 summarizes our research work and discussed the directions of future work.

Related Work

Community Discovery

Much work has been done on community discovery in social networks (Leskovec et al. 2010). The main purpose has been to partition the nodes of a network into groups, such that the nodes in one group have better internal connectivity than external connectivity. Many principles regarding community discovery have been proposed, and they can be classified into two main categories: (1) multi-criterion scores, which combine the criteria of the number of edges inside groups and the number of edges crossing groups into a single objective function; and (2) single-criterion scores, which employs only one of the above two criteria. A number of algorithms for community discovery have been proposed. Some are based on graph partitions, and others are heuristic approaches. The two well-known graph-partitioning algorithms are the spectral-partitioning algorithm and the flow-based algorithm. Some studies infer web communities from link topology (Girvan and Newman 2009); other work analyzes the community structure and semantic community (Gibson et al. 1998; Xu and Chen 2005). In addition, the characters of the social network community are explored, such as small-world property, power-law degree distribution, and network transitivity (Chen et al. 2008; Flake et al. 2002; Zhou et al. 2006). Different with these works, the main purpose of our study is to discover a group of users with maximally joint influence power, not a group of users connecting closely with each other.

Node Importance

In social network analysis (Wasserman and Faust 1994), some metrics are proposed to measure the importance of one node. The typical measure is that of degree centrality, defined as the number of links incident upon a node. Persons who have more ties to others within a network may have more chances to influence others. In an undirected social network, centrality value is the number of connections one person has. With a directed social network, there are two ways to measure centrality: centrality based on out-degree and centrality based on in-degree. Other evolved centrality measures include closeness centrality, considering the distance factor, and betweenness centrality, considering the position factor. In addition, the importance of each reviewer in SNS is also identified by mining the linguistic features (Li et al. 2010). But these metrics only consider the importance of one node, and the joint influence power of a group of nodes is not accurately reflected by their degree (Kempe et al. 2003). In contrast, our study discovers the group of users with maximally joint influence power, to help companies in conducting online marketing and reputation management.

Spread of Influence

This problem was proposed initially by Pedro Domingos and Matt Richardson in (Domingos and Richardson 2001; Richardson and Domingos 2002), and various studies are followed to improve upon this work. The purpose of these studies is to explore the spread of influence in social networks, and to
identify influential users activating other users to buy products. In (Domingos and Richardson 2001; Richardson and Domingos 2002), a model based on the Markov random field is proposed for this problem. In (Even-Dar and Shapira 2007; Kempe et al. 2003; Kempe et al. 2005), the diffusion models, the linear threshold model, independent cascade model, and voter model, are adopted for describing the process. In (Ben-Zwi et al. 2009; Chen et al. 2009; Kimura et al. 2007), some efficient approximate algorithms are proposed for this problem. The influence of network structure is explored in (Galstyan et al. 2009); product buying under a competitive social network is analyzed in (Carnes et al. 2007) Also there are some work on diffusion and social influence in marketing research, for example, an approach is proposed to identify the specific users who most influence others’ activities for advertising targeting and retention efforts (Trusov et al. 2010); Nair et al. (2010) explore the impact of social interactions and measure the value of influential users in directing targeted sales; the market strategy for maximizing revenues is studied in (Hartline et al. 2008).

In contrast to these works, our study focuses on discovering a target group of users with the maximal joint influence, instead of analyzing the spread of influence. In addition, these existing studies mainly involve in the “binary-value” applications, such as buying or not, or adopting or not, while our work is more generic in that it can apply to ‘continues-value’ scenarios, such as reputation management, brand identification, and public opinion guidance, which need to change users’ impressions on enterprises gradually. For these applications, discovering a group of users with the maximal joint influence is more important, because their opinions can mostly improve enterprises’ images, strengthen brand recognitions, and change public opinions.

**Targeting Group**

A few studies exist about discovering a target group, which are similar to this study. In (Zhang et al. 2008), an algorithm based on the trust relationships between users is proposed to detect the influential target groups: The users are ranked according to the numbers of written reviews, and then the users and the trust relationships are sequentially added into the target group until the clustering coefficient of the target group is less than a threshold. But the clustering coefficient can only reflects the partial influence relationships that are added into the group, without considering the influence relationships outside the group. So this method fails to consider the influence power of the target group on the whole, and trends to discover some target groups consisting of mutual influence users, which always have not most joint influence powers on other users (Kempe et al. 2003). Different with their method, our proposed method mines the influence relationships from SNS, represents them in a directed graph, and discovers the target group of maximally joint influence power by exploring the all influence relationships with the models of Mathematical Programming. In (Chen et al. 2008), a method is proposed to find the core bloggers in a network community by considering certain factors such as the internal link, external link, and co-links. This work is mainly used to detect active bloggers. Different with it, our study focuses on discovering a target group with the maximal joint influence power.

In summary, our study is different from the existing ones in the following aspects:

In contrast to the existing studies that evaluate the importance of each individual node in a social network separately, our study is aimed at discovering a group of users with the maximal joint influence power. This is critically important to influencing users’ adoption of products and the reputation of companies in SNS, as shown in recent findings of marketing research (Katona et al. 2007; Katona et al. 2010). In addition, most existing studies often assume the relationships between users are symmetric and model the social network as an undirected graph. This is not true for some social relationships such as the trust relationship. In this study, we use a directed graph to model the asymmetric influence relationship.

**General Process for Discovering Target Groups**

**Influence Network**

In SNS, the influence relationships exist only among some users, and the influence strengths vary between different pairs of users. In addition, unlike friend relationships, influence relationships are
asymmetric. Here, we construct a influence network to describe the influence relationships between users in SNS. A influence network is represented as a directed graph. A simple influence network is shown in Fig. 1.

The influence network can be represented as \( G = (V, E, W) \), and includes the following components:

**Node** \( i \): represents one user, e.g., Nodes 1 to 20 in Fig. 1. The set of all nodes is marked as \( V = \{1, \ldots, n\} \).

**Link** \((i, j)\): represents the influence relationship from user \( i \) to \( j \). For example, the arc from Node 1 to Node 18 in Fig. 1 shows that the user “1” has influence on “18”. The set of all arcs is marked as \( E = \{(i, j) \in V \times V | i \neq j\} \).

**Weight** \( \omega_{ij} \): Each arc is set a weight \( \omega_{ij} \) representing the strength of influence, for example, the influence strength from Node 1 to 18 is 9. The set of all weights is marked as \( W = \{\omega_{ij} | (i, j) \in E\} \).

**General Process**

As we know, there are a lot of SNS, such as Facebook, MySpace, Twitter etc. The proposed method is generic in that it can be used for discovering the influential groups based on data captured at any SNS. The general process of this method is as follows:
Data Collection. Some information about user profiles and the influence relationships among users are collected from a SNS, including:

1) User profile information, such as user id, user role, etc.

2) Social relationships among users, such as friend relationship in Facebook and Myspace, following relationship in Twitter, and trust relationship in Epinions. Usually these social relationships can reflect the influence relationships among users very well, because users always easily believe or accept the opinions of those with whom they have social relationships (Golbeck 2005; Lu et al. 2010; Massa and Avesani 2007).

3) Ratings on user’s reviews. In some SNS, users can rate the opinions of other users. For example, in the Epinions website, users can rate opinions as five levels: “Not Helpful, Somewhat Helpful, Helpful, Very Helpful, Off Topic”, to express their favors on other users’ opinions. These ratings also indicate the influence relationships among users in some degree: if user A always gives high rating on user B’s reviews, user A tends to trust user B, and user B has great influence on user A very likely.

4) Interaction information. SNS, users can interact with each other in different ways such as comment, reply, like, retweet, etc. These interactions often contain information about users’ opinions or sentiments towards others. Such information can be mined using sentiment analysis techniques (Pang and Lee 2008) and reflect the influence relationships among users.

Influence Network Construction. The influence network can be constructed by analyzing the collected data. The users compose the node set. The influence relationships are built by analyzing social relationships, ratings, and interaction information. The influence strength can be calculated by integrating information including the social relationship, the number of positive and negative ratings and the number of supporting and opposing interactions among users. For different SNSs, site-specific information can be explored to construct the influence network. Taking a customer review site like Epinions as an example, the influence strength between two users can be calculated by considering the following four situations:

1) User \( u \) is in user \( v \)’s trust/friend list: \( v \) trusts \( u \), and \( w_{uv} = 1 \).

2) User \( u \) is in user \( v \)’s block list: \( v \) does not trust \( u \). \( w_{uv} = -1 \).

3) User \( v \) rates or comments on user \( u \)’s review: \( w_{uv} = \frac{1-e^{-x}}{1+e^{-x}} \),

where \( x = \) (sum of the scales of positive ratings or comments from \( v \) to \( u \)) - (sum of the scales of negative ratings or comments from \( v \) to \( u \))

4) Otherwise, \( w_{uv} = 0 \).
Thus, weight $w_{uv} \in [-1, +1]$ measures the influence of $u$ on $v$. The larger the weight, the more likely $u$ can influence $v$.

**Target Group Discovery.** Based on the constructed influence network, the target groups are discovered using the proposed models of mathematics programming. The detail method is presented in the following section.

**Computational Models for Discovering the Most Influential Groups**

Here, the joint influence power of one group is defined as the sum of the influence strengths of all users of this group on other users, and some models are proposed for discovering the target group with the maximal joint influence power using mathematical programming technology.

**Max-Influence Group Problems**

1) Basic Max-Influence Group (Basic-MIG)

In the influence network, one target group corresponds to a subset of nodes, and the joint influence power of this subset corresponds to the sum of the weights on the arcs from the nodes of this subset to other nodes. Thus, discovering the target group with the maximal joint influence power is to find a subset of nodes $S \subseteq V$, where the sum of the weights from the nodes in $S$ to other nodes is maximal. That is:

$$\max \sum_{i \in S, j \in \bar{S}} \omega_{ij}$$

subject to: $S \subseteq V$

here, $\bar{S}$ is the complementary set. In this formula, $\sum_{i \in S, j \in \bar{S}} \omega_{ij}$ represents the sum of the weights from the nodes in one subset $S$ to the nodes outside $S$; and the constraint $S \subseteq V$ makes sure the discovered nodes are a subset of the node set $V$.

2) Certain Size Max-Influence Group (CS-MIG)

As we know, the number of users in SNS is often quite large, and enterprises only have limited budgets for targeting customers. Therefore, many enterprises may desire to find a target group within a certain size, which can have the maximal joint influence power. Thus, we can modify the basic model by adding and other constraint that the group size is less than or equal to $t$. The new model is as follows:

$$\max \sum_{i \in S, j \in \bar{S}} \omega_{ij}$$

subject to: $S \subseteq V$

$$|S| \leq t$$

In this formula, the constraint $|S| \leq t$ makes sure the size of the subset $S$ is not greater than $t$.

**Models**

The above Basic-MIG and CS-MIG problems involve discovering a subset of nodes from a graph. Similar to the max-cut problem (Goemans and Williamson 1995), they are NP-hard problems. Recently, SemiDefinite Programming (SDP) (Todd 2003) is proven to be an efficient way of approximately solving this kind of problem (Goemans and Rendl 1997). Here, following the method of the max-cut problem, the
Basic-MIG and CS-MIG problems are formulated as integer programming models, which can be easily transformed into SDP and solved efficiently.

Here, \( n + 1 \) decision variables \( x_0, x_1, \ldots, x_i, \ldots, x_n \in \{+1, -1\} \) are defined; \( x_0 \) is a reference variable, which denotes being in the target group \( S \). If \( x_i = x_0 \), the node \( i \) is in \( S \); otherwise, the node \( i \) is not in \( S \). Then the Basic-MIG problem can be formulated as follows:

1) Basic-MIG Model:

\[
\begin{align*}
\max & \quad \frac{1}{4} \sum_{(i,j) \in E} \omega_{ij} (1 + x_0x_i - x_0x_j - x_ix_j) \\
\text{subject to:} & \quad x_i \in \{-1, 1\}, \ i = 0, 1, \ldots, n
\end{align*}
\]

Note that if node \( i \) is in \( S \) and \( j \) is not, then \( x_i = x_0 \neq x_j \), and thus the corresponding part in the objective function is \( \frac{1}{4} \omega_{ij} (1 + 1 - (-1) - (-1)) = \omega_{ij} \). Otherwise, if nodes \( i \) and \( j \) are both in the target group \( S \), then \( x_i = x_j = x_0 \), and the corresponding part in the objective function in (1) is \( \frac{1}{4} \omega_{ij} (1 + 1 - 1 - 1) = 0 \). Thus, \( \frac{1}{4} \sum_{(i,j) \in E} \omega_{ij} (1 + x_0x_i - x_0x_j - x_ix_j) \) is the sum of the weights from the nodes in \( S \) to other nodes, and this model captures the joint influence power of nodes in \( S \).

If the node \( i \) is in \( S \), then \( x_0x_i = +1 \); otherwise, \( x_0x_i = -1 \). So \( \sum_{i=1}^{n} x_0x_i = |S| - |\overline{S}| = |S| - (n - |S|) = 2|S| - n \); here \( n \) is the number of all nodes, thus \( |S| = \frac{1}{2} (\sum_{i=1}^{n} x_0x_i + n) \). For the CS-MIG problem, the size of the target group is less than or equal to \( t \), that is \( \frac{1}{2} (\sum_{i=1}^{n} x_0x_i + n) \leq t \). So the CS-MIG problem is modeled as follows:

2) CS-MIG Model:

\[
\begin{align*}
\max & \quad \frac{1}{4} \sum_{(i,j) \in E} \omega_{ij} (1 + x_0x_i - x_0x_j - x_ix_j) \\
\text{subject to:} & \quad x_i \in \{-1, 1\}, \ i = 0, 1, \ldots, n \\
& \quad \sum_{i} x_0x_i \leq 2t - n
\end{align*}
\]

In these three models, both the objective function and the constraints are linear in \( x_i, x_j \), so they can be easily relaxed as SDP problems. Let \( Y = xx^T \), the three models can be relaxed as the following SDP for efficiently solving (Goemans and Rendl 1997; Goemans and Williamson 1995):
1) Basic-MIG Relaxation:  
\[
\begin{align*}
\max & \frac{1}{4} \sum_{(i,j) \in E} \omega_{ij} (1 + Y_{ij} - Y_{ij} - Y_{ij}) \\
\text{subject to:} & \quad Y_{ii} = 1, \ i = 0, \ldots, n \\
& \quad Y \geq 0 \\
& \quad \sum_{i=1}^{n} Y_{0i} \leq 2t - n \\
& \quad \sum_{i=1}^{n} \sum_{j=1}^{n} Y_{ij} \leq (2t - n)^2 \\
& \quad Y \geq 0 
\end{align*}
\]

2) CS-MIG Relaxation:  
\[
\begin{align*}
\max & \frac{1}{4} \sum_{(i,j) \in E} \omega_{ij} (1 + Y_{ij} - Y_{ij} - Y_{ij}) \\
\text{subject to:} & \quad Y_{ii} = 1, \ i = 0, \ldots, n \\
& \quad Y \geq 0 \\
& \quad \sum_{i=1}^{n} Y_{0i} \leq 2t - n \\
& \quad \sum_{i=1}^{n} \sum_{j=1}^{n} Y_{ij} \leq (2t - n)^2 \\
& \quad Y \geq 0 
\end{align*}
\]

After finding the solutions of these SDP problems, the approximate solutions of the Basic-MIG and CS-MIG problems can be acquired by the randomized rounding procedure (Goemans and Williamson 1995): compute the Cholesky factorization of the SDP solution, and take the columns of Cholesky factorization as the vectors \( v_i \) (\( i = 0, \ldots, n \)), corresponding the nodes \( i \) (\( i = 0, \ldots, n \)); choose a random hyperplane through the origin, and partition the vectors \( v_i \) (\( i = 0, \ldots, n \)); if \( v_i \) (\( i = 1, \ldots, n \)) falls the same side with the vector \( v_0 \), the decision variable \( x_i \) takes the value of 1; otherwise -1. For the CS-MIG problem, additional processes are needed to assure the target group satisfying the size constraint. This randomized rounding algorithm can lead to a 0.79607 times the optimum value of the original problem (Goemans and Williamson 1995).

Indirect Influence:

In addition, the indirect social influence also plays very important role in people’s decision making (Denrell 2008; Joseph 2001). In SNS, a user can also influence or be influence by others through indirect relationships (e.g., a friend’s friends). In our proposed models, indirect influence can be taken into account via manipulating the weight matrix \( W \). In the two models, we can replace the weight matrix \( W \) by \( \overline{W} = W + \lambda \cdot W^2 \), where \( W^2 \) represents the indirect influence through one intermediate node, and \( \lambda \) is the ratio of indirect influence in the final weight set \( \overline{W} \). Similarly, the higher order indirect influence can be considered using this way, although it rarely occurs in reality.

Experimental Evaluation

Experimental Settings

Since discovering the target group with joint influence power is a NP-hard problem and there are no available benchmark datasets, it is infeasible to evaluate the proposed method by comparing the discovered target groups with the “Ground Truth” groups. Here, we adopt the similar approach as other studies (Kempe et al. 2003; Richardson and Domingos 2002) on the influence of social network. Our proposed method is compared with other benchmark methods. In this evaluation, the joint influence power of the discovered target group (the sum of the influence strengths of all users of this group on other users), is as the evaluation metric.

Benchmark Methods

In order to show the effectiveness of our proposed methods, we compare our methods with two heuristic methods for influential group discovery, the weighted degree centrality method and the revised clustering coefficient method (Zhang et al. 2008), are used to compare with the proposed methods. These two methods are commonly used ones and show good performances.
**Revised clustering coefficient method:** The method in (Zhang et al. 2008) is originally proposed for discovering target group with large influence power. It uses the number of written reviews and the trust relationships (similar to the influence relationships here) to discover the target group. Here we adapt this method to the influence network: First, users are ranked according to the number of their written reviews; then the users and their relationships are orderly added into the target group, and the clustering coefficient of the target group is calculated; The adding process continues until the clustering coefficient is less than a threshold. For the basic MIG problem, the clustering coefficient threshold is gradually reduced until the target group achieves the maximally influence power. For the CS-MIG problem, the clustering coefficient is set to make the target group reaches a certain size.

**Weighted degree centrality method:** This is the most commonly used method for measuring the importance of nodes in network (Barrat et al., 2004; Newman 2004; Opsahl et al., 2008). Some studies (Ten Kate et al. 2010; Wan and Tian 2010; Zhu et al. 2008) adopt this method for measuring the influence of nodes in social network analysis. This method has also been used for discovering the most influential group of users in a network (Domingos and Richardson 2001; Richardson and Domingos 2002). Here, the weighted degree centrality of user $i$ is calculated as the sum of the influence strengths of this user on others, $\sum_{j=V, j \neq i} \omega_{ij}$, and the users are ranked according to their weighted degree centralities, and the target group is chosen based on this ranking list. For the basic MIG problem, the top users are sequentially put into the group until the group reaches a maximally influence power; for the CS-MIG problem, the top $t$ users compose the target group.

For the basic MIG and CS-MIG problems, the joint influence powers of the target groups are calculated and compared with the two baseline methods. In this experiment, the DSDP MATLAB toolbox (DSDP 2006) is used to solve the relaxed SDP problems.

In order to fully evaluate the effectiveness of the proposed methods, we conducted experiments on two real-world datasets collected from two famous SNS’s: Epinions and Twitter. In this experimental evaluation, we only consider the effect of direct influence among users but no indirect influence, which will be investigated in our future work.

**Experiment Evaluation I: Epinions Dataset**

**Epinions Dataset**

The data set used in this evaluation is from the well-known Epinions website. Epinions is a user opinions website supporting social network functions. On this website, users can rate products and express their opinions, also users can build trust relationships with each others by adding them to their trust lists, or block some users. The system shows the trusted users’ opinions at the top of the list, so users are influenced mostly by their trusted users. For each product category, the website provides a ranking list of the top 1,000 users according to the total hits to member reviews. In this evaluation, we extracted information about these top-ranked users from 10 different product categories (listed in Table 1), including their rankings, their trust relationships, etc.

**Constructing the influence network:** For the extracted data of each product category, we built one influence network of 1,000 nodes(users). If user $i$ is trusted by user $j$, there exists an influence relationship from $i$ to $j$; here the influence strength is set as 1. A descriptive summary of the 10 influence networks from different categories is shown in Table 1. For example, the influence network of the Books category has the densest influence relationships: on average, each node has 32.5 influence relationships. In contrast, the influence network of the Computer Software category has relatively sparse influence relationships: on average, each node has only 4.7 influence relationships.
Table 1. Summarizing Descriptions of the Influence Networks of Epinions

<table>
<thead>
<tr>
<th>Node No.</th>
<th>Influence Relationship No.</th>
<th>Average Influence Relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books</td>
<td>1000</td>
<td>32502</td>
</tr>
<tr>
<td>Movies</td>
<td>1000</td>
<td>25274</td>
</tr>
<tr>
<td>Music</td>
<td>21842</td>
<td>21.8</td>
</tr>
<tr>
<td>Kids&amp;Family</td>
<td>19212</td>
<td>19.2</td>
</tr>
<tr>
<td>Home&amp;Garden</td>
<td>17904</td>
<td>17.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Node No.</th>
<th>Influence Relationship No.</th>
<th>Average Influence Relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronics</td>
<td>1000</td>
<td>10131</td>
</tr>
<tr>
<td>ComputerHardware</td>
<td>1000</td>
<td>8636</td>
</tr>
<tr>
<td>Sports&amp;Outdoors</td>
<td>6560</td>
<td>6.6</td>
</tr>
<tr>
<td>Autos&amp;Motorsports</td>
<td>5545</td>
<td>5.5</td>
</tr>
<tr>
<td>ComputerSoftware</td>
<td>4741</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Results & Discussions

Performance of Basic-MIG model

Figure 3 shows the joint influence powers and sizes of the discovered target groups with the Basic-MIG model and the benchmark methods in the 10 product categories.

It shows that the target groups discovered by the Basic-MIG model have larger joint influence powers than those discovered by the benchmark methods. We conducted paired t-tests between our method and benchmark methods, which shows the difference is statistically significant at p-value <.01. The Basic-MIG model achieves 30-50% more joint influence power than the weighted degree centrality method, and 60-70% more than the revised clustering coefficient method. The denser the influence relationship is, the larger the difference.

This result shows that the target group consisting of top weighted degree users is not the group with maximal joint influence power. The reason is that these top weighted degree users may be clustered, and only strongly influence each other, but they do not influence other users the most (Kempe et al. 2003). If choosing these users as the targets in online marketing, their opinions cannot generate the most influence other users’ decisions. The revised clustering coefficient method can discover the target group with a certain joint influence power, but its performance is less than that of the Basic-MIG model. Although the Basic-MIG model can discover the target groups with larger joint influence powers, the disadvantage of the Basic-MIG model is: the sizes of the target groups are larger than those of other benchmark methods, and targeting the large size of users may require a high cost for marketing actions.
Performance of CS-MIG model

Figure 4 shows the joint influence powers of the target groups with different sizes using the CS-MIG model and the benchmark methods in the Electric category. Due to the page limit, the results of other categories are omitted.

It shows that, under different group size limits (10 ~ 400), the CS-MIG model achieves greater joint influence powers than the two benchmark methods. Our paired t-tests show that the difference are statistically significant at \( p \)-value < 0.01. The average influence power of each member of CS-MIG model is also much greater than that of two benchmark methods. And the curve shows when the target group size is larger, the difference is more obvious. For the CS-MIG model, the joint influence power gradually grows with the size of the target group, and becomes stable when the size is greater than 250. The curve of the revised clustering coefficient also increases gradually, but the increase rate is much less than that of the CS-MIG model. For the curve of the weighted degree centrality, it increases at the beginning, and decreases quickly when the size is larger than 100. The reason is that the nodes with high weighted degree always are clustered and mutual influence each other, so that targeting all of them is unnecessary. If the group size is small, the weighted degree centrality method can discover a similar joint influence power target group to the CS-MIG model, since when the group size is small, the mutual influence of nodes in the group can be ignored. The CS-MIG model requires much more computation than the weighted degree centrality method because it needs to solve a SDP problem.

Experiment Evaluation II: Twitter Dataset

The purpose of this experiment is to evaluate the performance of the proposed models on the large-scale social network with the sparse influence relationships.

Twitter Dataset

The dataset in this evaluation is from the most popular microblog system, Twitter. In the Twitter website, users can follow other users they are interested in and they can also be followed by other users. The web interfaces of users will automatically show the “tweets” of those they follow, so they are most likely to be influenced by these users’ opinions. The following and followed relationships compose the social relationships on Twitter. When extracting the dataset from the Twitter website, two users are randomly chosen as the seeds, and the following users and the following relationships information are extracted in
the Breadth-First-Search way. In total, the dataset contains 5008 users and 14840 following relationships.

**Constructing the influence network:** In this experiment, we built an influence network of 5008 nodes (users). If user $i$ is followed by user $j$, there exists an influence relationship from $i$ to $j$; the influence strength is set as 1. The summary information on the constructed influence networks is shown in Table 2.

<table>
<thead>
<tr>
<th>Node No.</th>
<th>Influence Relationship No.</th>
<th>Average Influence Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>5008</td>
<td>14840</td>
<td>2.96</td>
</tr>
</tbody>
</table>

**Results on Twitter**

Figure 5 shows the joint influence power and average influence power of the target groups with different sizes discovered by the CS-MIG model and benchmark methods. It shows that the CS-MIG model can discover the target groups with larger joint influence power than the benchmark methods. Especially when the sizes of target groups are large, the improvement is much obvious (the difference is about 5%). But when the size is small, the CS-MIG model method achieves the similar performance as the benchmark methods. This experimental results also show the improvement achieved by the CS-MIG model in the sparse influence network is not as large as in the dense ones of the previous experiment. This is reasonable because in a sparse influence network, the joint influence powers of each groups differentiate less obviously compared to those in the dense influence network.

**Conclusions and Future Work**

The latest marketing research shows that the joint influence power of a group of users is a very important factor influencing other users’ buying decisions and their views on companies. A novel method is proposed in this study to help companies discover the target group with the maximal joint influence power. First, the influence relationships are mined from SNS, and represented as a directed graph. Then the proposed mathematics programming models, which can be transformed to SDP for solving efficiently, are used to discover the target group. The empirical evaluation with the data from real SNS indicates the proposed method achieves much better performances than the benchmark methods. The current work has several limitations and can be improved in the future. In our study, the joint influence power is
defined as the sum of influence strengths. For our future work, alternative model which defines the joint influence power in terms of the number of users will be considered. Moreover, in an SNS, the influence relationship may be negative. For example, if user A distrusts user B, B's opinions will have the opposite influence on user A. In the future, such negative influences will also be integrated in the models. In addition, the indirect influence will be considered in the optimization model and evaluated on the real data set.

References


