

Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs

Feng Li^a, Timon C. Du^{b,*}

^a School of Business Administration, South China University of Technology, China

^b Department of Decision Sciences and Managerial Economics, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:

Received 17 September 2009

Received in revised form 22 November 2010

Accepted 19 December 2010

Available online 1 January 2011

Keywords:

Online social network

Blog

Word-of-mouth marketing

Ontology

ABSTRACT

Online social blogs have gained popularity recently. They provide an effective channel for word-of-mouth (WoM) marketing to promote products or service. In WoM marketing, an opinion leader, who is normally more interconnected and has a higher social standing, can deliver product information, provide recommendations, give personal comments, and supplement professional knowledge that help companies to promote their products. Many theories have been put forward about social networks, but few address the issue of opinion leader identification. This study proposes a framework to identify opinion leaders using the information retrieved from blog content, authors, readers, and their relationships, which we call BARR for short. We first build ontology for a marketing product and then collect parameters from BARR to identify “hot topics” related to the product. These hot topics are then associated with information disseminators, or opinion leaders. Marketers can use BARR to track blogs written by opinion leaders and identify their opinions to form effective marketing strategies.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Social network analysis involves the study of the relationships among interacting individuals. It usually focuses on the network itself, rather than on the attributes of the participating individuals. Various theoretical concepts have been developed to look into issues surrounding social networks, such as social groups, isolation, popularity, liaison, prestige, balance, transitivity, cliques, subgroups, social cohesion, social position, social role, reciprocity, mutuality, exchange, influence, dominance, and conformity [44]. An important issue is the identification of opinion leaders. This involves the examination of directed relationships and the phenomena of centrality and prestige, which describe how important an actor is in a social network [14]. This is the focus of this study. More specifically, we investigate how an opinion leader influences others in an online social blog that provides an arena for a group of people who share common interests or have social ties to interact online.

An opinion leader is normally more interconnected and has a higher status, education, and social standing and thus ability to influence followers. Opinion leaders are important individuals in social networks because of their ability to informally influence the attitudes or behavior of others in a desired way with relatively high frequency [43]. In the business world, this influence can be put to

commercial use. Specifically, in word-of-mouth (WoM) marketing, an opinion leader can deliver product information, provide recommendations, give personal comments, and supplement professional knowledge to help a company to promote its products.

WoM is a type of viral marketing [30] and an informal way of exchanging information among consumers about the characteristics, usage, and ownership of particular products or services. It shifts communication from a company-to-customer mode to a customer-to-customer mode [45]. WoM has stronger credibility than other forms of marketing, as there is no direct connection between the information sender and the merchant, and thus the information given is considered to be subjective and independent. It may also be more persuasive, as the information sender may have a better understanding of the receiver.

Because customers initiate WoM communication, companies may want to monitor, manage, motivate, and enforce positive WoM to enhance a product's value. However, negative WoM also provides important information for diagnosing the gap between a product's values and customer expectations [36]. A company can examine comments, either positive or negative, that appear on online social blogs to help them to build up or repair customer relationships.

With recent advances in information technology, social networks are no longer limited to physical face-to-face platforms: online social networks have become a new media for WoM marketing. They have many advantages. For example, the participants can communicate in a one-to-one or many-to-many format. The distribution of messages is swift. Messages take a written form, and can be reviewed by audiences concurrently and on demand. Furthermore, the anonymity

* Corresponding author. Department of Decision Sciences and Managerial Economics, The Chinese University of Hong Kong, Hong Kong, Shatin, NT, Hong Kong. Tel.: +852 26098569.

E-mail addresses: fenglee@scut.edu.cn (F. Li), timon@cuhk.edu.hk (T.C. Du).

of the messages fortifies the weak ties among communicators to make the network an important source of information [6]. Moreover, online WoM is dynamic, and information can be appended and revised at any point. Online social blogs thus provide a good channel through which opinion leaders can exert an influence.

In this study, we propose a framework for opinion leader identification in online social blogs. The study uses an ontology retrieved from written content to identify opinion leaders. We analyze online social networks with structured written content rather than multimedia (video or photos) or unstructured text (SMS). The framework identifies not only opinion leaders, but also the “hot” blogs that they published. Marketers can use the framework to analyze blog content and take necessary marketing action.

The remainder of this paper is organized as follows. Section 2 reviews existing opinion leader identification and recommendation systems. Section 3 presents the BARR framework and Section 4 shows its application. We conclude the study in Section 5.

2. Opinion leader and recommendation system

A recommendation system is a two-way communication platform that delivers opinions from an informer to a recipient and allows the recipient to comment on the opinion. The effectiveness of the communication relies not only on the sender's expertise, but also on the receiver's expertise and the perceived risk [4]. Feedback offered through this platform can also be very influential. For example, an online review written by a reputed or a higher exposure participant on a platform such as Amazon can have a greater impact on product sales [8]. Sometimes the feedback is polarized, as consumers may only submit feedback when they have strong opinions – either positive or negative – on a specific issue, product, or service [19]. Thus, both the original contributor and reviewers can be opinion leaders [20]. The motivations to contribute to an online social network are various, and include the enjoyment of helping others, self-enhancement, and economic reward [41]. Another important point is that the impact of the messages delivered by opinion leaders is closely correlated with the transfer distance, that is, with the number of users through which the information passes [23,40].

Modeling or simulations are the tools commonly used to study online recommendation systems. Dellarocas [11] developed a multi-agent platform to simulate complex social network relationships and used software agents to imitate individual participants to strategically manipulate Internet opinion forums. Similarly, Collings et al. [10] used an agent-based model to simulate the diffusion of an innovative product or service within a population of consumers. The results showed that key individuals within the populations had profound effects on the extent of diffusion. Other examples of similar research include the assessment of the impact of public conservation campaigns on water demand [2] and a characteristics study of a customer network [48].

Centrality and prestige in a social network indicate how opinion leaders become influential through the relationships that they build [44]. In an online social network, this influence can be exerted through a recommendation system that aggregates recommendations for participants. The recommendation system provides a platform for opinion leaders to influence followers. This reflects a natural social process. One of the most successful recommendation methods in online platforms is the completion of recommendations by the collaborative filtering of the data of a large number of participants [9,50]. There is no specific opinion leader during the filtering process. However, the collection of a profile of participant interests to make more effective recommendations may raise concerns about personal privacy intrusion [32]. This concern can be eased by developing personal recommenders that store personal information locally or only share it in encrypted form [27].

A recommendation system can be further developed into a reputation system that collects information on the past behavior of participants to indicate their skill or honesty in recommending [33]. This improves trust among the participants, as can be clearly seen in the case of eBay. WoM represents the successful application of recommendation systems [32]. It has been noted that the influence of WoM is stronger in the early stage of a product's life cycle [18], and that the influence of negative WoM on product judgment is greater than that of positive WoM [5]. This indicates that marketers should disseminate positive information to consumers early in the product life cycle.

Opinion leaders are usually innovative when a social system favors change, and may become adaptors when they are benefited and visible. When they are early adaptors, they also mediate the adoption of new products [22,37]. However, few studies have focused on opinion leader identification on an online platform. This may be because identification needs to consider the semantic level of the message, the relationships among and the profiles of the platform participants [34,42], and the reliability of the message [13]. In this study, we use the system presented in [12] to extract ontology and use it to identify “hot topics” and the associated information disseminator, or opinion leader. This serves as a tool for the acquisition of semantic information from Web pages. We then use the extracted information to analyze certain factors, such as “relationships.” In computer science, the identification of hot topics is called topic detection and tracking (TDT). This process aims to find and trace topics from chronologically ordered news or stories [1,47]. Topics are specific words that increase rapidly in frequency in a short time [16,39]. Here, we use TDT analysis to understand messages and retrieve information.

3. The opinion leader identification framework

We propose the BARR framework to identify an opinion leader. We refer to the suggestion in [46] that the interactions in an online social network are affected by shared information, involvement, and relationships. As shown in Fig. 1, our framework considers the multifaceted associations between bloggers, including the factors of blog content, author properties, reader properties, and the relationship between the author and readers, based on a pre-specified topic to identify opinion leaders. These factors are explained in detail in the following.

- (1) Blog content. This is content that determines the popularity of blogs. In an online social network, written blog content is a major source for understanding messages, writing style, and tone of communication [7]. Such content can provide information on the quality or price of a product [26], and also criticisms, recommendations, warnings, compliments, or complaints [38]. As the messages in blogs are in document format, they can be readily subjected to content analysis [24]. Additional information, such as the number of views and feedback (a vote or a comment), can then be used to measure the impact and popularity of a blog. Comments may also be made in response to other comments, which gives rise to more intensive interaction.
- (2) Author properties. The reputation of an author in a specific profession affects the effectiveness of WoM marketing and can make an author an opinion leader [4,17]. An author becomes a domain expert based on experience and knowledge [28]. It is thus possible to use the expertise of an author as a determinant of his or her influence. Opinion leaders are normally of central importance to the group of followers that they attain [46]. They are influential, knowledgeable, communicable, respective, and innovative. The popularity of an author can be measured by

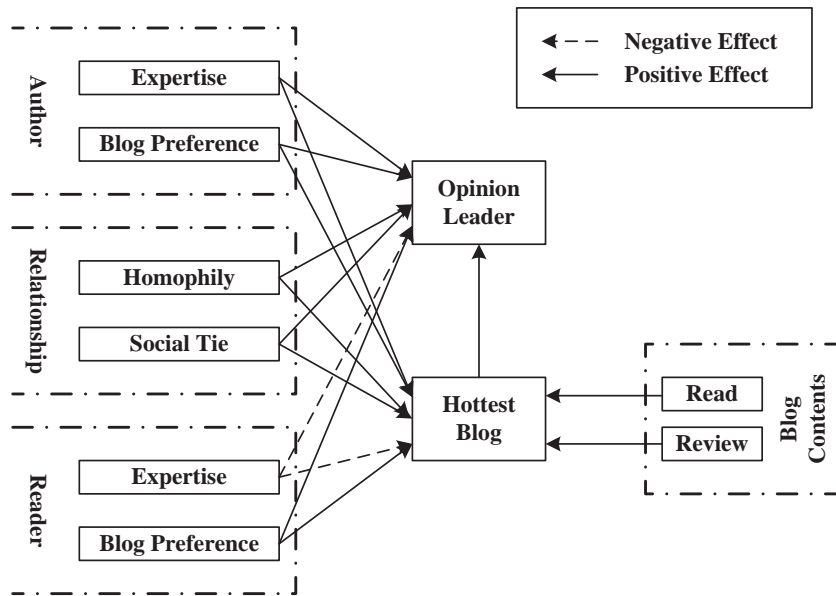


Fig. 1. BARR framework for the identification of blog opinion leaders.

blog preference. For example, the registration of a large number of readers as “friends” of a blog is likely to be a strong indication of preference. This preference will be stronger if an author has the ability to collect information, provide information to readers, and correspond with them frequently.

- (3) Reader properties. The properties of readers or recipients are also factors that affect the effectiveness of WoM marketing in various ways. For example, the knowledge of a reader may not be a good indicator of the popularity of a blog author, as a very knowledgeable but solitary reader will only look for highly relevant information and may not be willing to be involved in discussions with others. This may prevent a blog from being registered as a “hot” blog. To determine the expertise of a reader, we thus use the comments that the reader makes on his or her own blog. The blog preferences of a reader’s own blog B can be used to indicate whether or not the author of the original blog A is an opinion leader, as a reader with a high preference for a blog will be more involved in blog discussions. The presence of such a reader thus makes blog A more influential. We use the number of “friends” and the level of involvement in discussions to measure the blog preference of a reader.
- (4) Relationship. The relationship between the author of a blog and the blog’s readers is a perceptible indicator of the effectiveness of WoM marketing. However, the relationship is complex and cannot be measured simply by hyperlinks, as suggested in [35]. We use the intimacy (social ties) and similarity (homophily) between two parties to measure relationship. The strength of social ties indicates the perceived intimacy between reader and author [6], and can be determined by the affiliation and frequency of the interaction between them. When author and reader interact more frequently, they can be considered to have stronger ties. It is assumed that an author is more persuasive to a reader when the ties between them are stronger, as they will be more willing to share opinions openly.

The similarity between an information source and an information seeker, which is called homophily [25], is another measure of relationship. The theory of homophily has been extensively applied in the context of personal influence. It has been found that the greater the homophily between communicators, the more persuasive the message of the communicator [17]. In this study, it is assumed that the

greater the homophily, the greater the attention a reader will pay to a blog.

The flow of the opinion leader identification framework is shown in Fig. 2. It comprises five stages: (1) keyword blog search, (2) ontology extraction, (3) ontology-assisted extraction, (4) hot blog identification, and (5) opinion leader identification. The inputs are keywords, blog portals, and web sources. We illustrate each stage as follows.

- (1) Keyword blog search. This stage employs user-defined keywords to locate web pages in blog portals. This can be carried out using commercial search engines, such as Google.
- (2) Ontology extraction. This stage extracts ontology from the web pages and deposits it into the ontology base using the framework proposed in [12], in which an ontology is formed from websites based on an analysis of web page structure and hyperlinks. Note

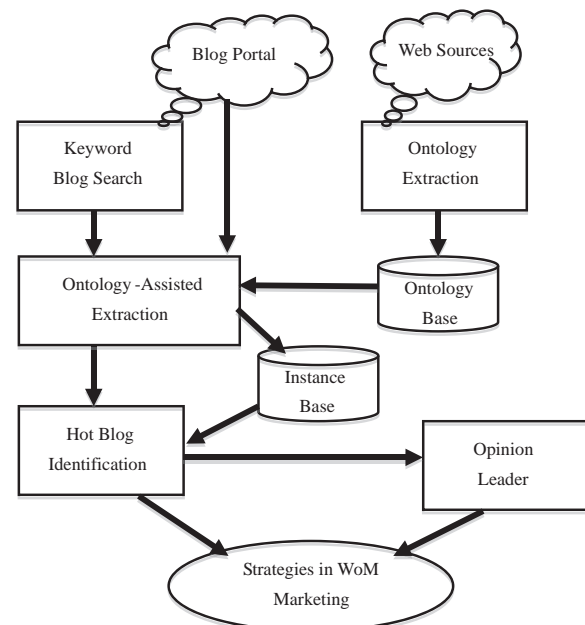


Fig. 2. System flow of the BARR framework.

that ontology has been successfully applied to many knowledge manage domains such multilingual knowledge management [3,29], knowledge representation [49], and multiple foci creation [31]. The framework determines the frequency of words in a document (first removing stop words or stemming words), and calculates their information entropy. Words with entropy values that are higher than a predefined threshold are saved to build the domain ontology. The entropy is defined as

$$Entropy(word) = - \frac{\frac{frequency(word)}{NumOfWords} \cdot \ln\left(\frac{frequency(word)}{NumOfWords}\right)}{\ln(NumOfPages)}, \quad (1)$$

where $frequency()$ means the frequency with which a word appears in the input source, $NumOfWords$ is the total number of words, and $NumOfPages$ is the number of documents in the input source. An ontology engineer then finalizes the ontology based on the system recommendations, and the ontology is deposited in the ontology base.

- (3) Ontology-assisted extraction. This stage builds ontology instances based on the ontology in the ontology base. It identifies (a) instances of a certain concept “blog” from the selected blogs, (b) instances of the concept “blogger” (either authors or readers) from the association of “blogger” and “blog,” and (c) the relationship instances between bloggers using the information from the first two parts.
- (4) Hot blog identification. As previously mentioned, centrality and prestige in a social network can be used to indicate how an opinion leader gains influence over others through the relationships that he or she builds. Centrality indicates how a point (blog) can constitute the center of other points (a hot blog), whereas prestige measures how an author becomes influential though publishing blogs (which is discussed in greater depth in stage 5). This stage uses four sources of information to identify hot blogs: blog content, author properties, reader properties, and the relationship between author and readers. The following 11 parameters are collected from the four sources.
 - (a) Blog content. The popularity of a blog is ascertained from the number of visits (the number of clicks on the web page) (F1), the number of reviews (the number of comments) (F2), and the rank (F3).
 - (b) Author. We determine an author's expertise (F4) and blog preference in terms of the number of blogs (F5) and number of comments (F6). To calculate the expertise of an author, we form a vector space based on the output of the ontology extraction module $V_0 = (tf_1^0, tf_2^0, \dots, tf_n^0)$. The element in each vector is the frequency of the words in the web pages. We then use the vector to analyze the blog content of each author for $V_i = (tf_1^i, tf_2^i, \dots, tf_n^i), i \neq 0$.

The cosine between the vectors obtained for an author and the extracted ontology then indicates the degree of expertise of the author. This approach is commonly used in artificial intelligence for case-based reasoning.

$$sim(V_0, V_i) = \frac{\sum_{k=1}^n (tf_k^0 \times tf_k^i)}{\sqrt{\sum_{k=1}^n (tf_k^0)^2} \times \sqrt{\sum_{k=1}^n (tf_k^i)^2}} \quad (2)$$

A larger value between 0 and 1 indicates greater expertise. For example, if the vector of A and the vector of the ontology are both (1,1), then the cosine of the two vectors is $= \frac{1 \times 1 + 1 \times 1}{\sqrt{1^2 + 1^2} \times \sqrt{1^2 + 1^2}} = \frac{2}{\sqrt{2} \times \sqrt{2}} = 1$. In contrast, if the vector of B is (1,0) but the ontology vector is (0,1), then

the cosine of the two vectors is 0, that is, $= \frac{1 \times 0 + 0 \times 1}{\sqrt{1^2 + 0^2} \times \sqrt{0^2 + 1^2}} = \frac{0}{\sqrt{1} \times \sqrt{1}} = 0$, which means that the expertise of the two entities is different. The blog preference of an author is calculated from the number of blogs and comments on other blogs published by the author. The number of comments n_i is normalized to n'_i , ranging between [0 and 1], using sigmoid curve transformation.

$$n'_i = 1 / \left(1 + e^{-\frac{n_i - \bar{n}}{\bar{n}}} \right), \quad (3)$$

where \bar{n} is the average of n_i , that is, $\bar{n} = \sum_{i=0}^k n_i / k$. The initial difference grows approximately exponentially and then slows as saturation begins.

- (a) Readers. The measurements adopted for readers are similar to those used for authors, that is, expertise (F7), number of blogs (F8), and number of comments (F9).
- (b) Relationship. We consider both homophily (F10) and tie strength (F11). As discussed, homophily indicates the degree of similarity of subjects [17], whereas tie strength shows the closeness of the parties involved [6]. As with the calculations of expertise, homophily is determined by forming a vector space with keywords, where only words with information entropy values higher than a predefined threshold are included in the vector. The cosine of the vectors between author and readers is then taken as the homophily between them. For example, the preference of a blogger can be (music, sport, book, travel), which is converted into the vector (1,0,1,1). If another blogger has a vector of (1,1,0,1), then the homophily between them can be calculated as

$$\frac{1 \times 1 + 0 \times 1 + 1 \times 0 + 1 \times 1}{\sqrt{1^2 + 0^2 + 1^2 + 1^2} \times \sqrt{1^2 + 1^2 + 0^2 + 1^2}} = \frac{1 + 0 + 0 + 1}{\sqrt{3} \times \sqrt{3}} = \frac{2}{3} = 0.6667.$$

There are four levels of tie strength: strangers (no relationship between bloggers, tie strength = 0), friends (set as friends on the blog website, tie strength = 0.25), good friends (share comments on the blog, tie strength = 0.5), and buddies (listed as friends and also share comments, tie strength = 1.0). Note that it is not possible to guarantee that a friend has read a blog from the information provided by websites. However, a comment posted to a blog by a “friend” can be used as an indication that the friend has read the blog. Different scores are assigned the four levels of tie strength based on the available information. Further discussion of the social ties between blog readers and authors can be found in [46]. As hot blog selection is a multi-attribute decision problem, we refer to the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), which summarizes the Euclidean distance between measurements and the ideal solution, to determine the popularity of a blog [21]. For all of the parameters except for F7 (which is assumed to be negatively associated), larger values indicate a more popular blog. For a positive ideal solution, the i vector [index]

$$Z_i^+ \text{ is defined as } Z_i^+ = \begin{cases} \max(F_j^i) & i \neq 7 \\ \min(F_j^i) & i = 7 \end{cases}, (i = 1, \dots, 11) \text{ where } F_j^i \text{ is}$$

the value of the i parameter. For a negative ideal solution, the i index

$$Z_i^- \text{ is defined as } Z_i^- = \begin{cases} \min(F_j^i) & i \neq 7 \\ \max(F_j^i) & i = 7 \end{cases}, (i = 1, \dots, 11). \text{ We then}$$

calculate the similarity between the ideal solutions of the blogs

$$c_j = s_j^- / (s_j^+ + s_j^-), \quad (4)$$

where $s_i^- = \sqrt{\sum_{i=1}^{11} (Z_i^- - F_j^i)^2}$ is the Euclidean distance between blog j and the negative ideal solution, and $s_i^+ = \sqrt{\sum_{i=1}^{11} (Z_i^+ - F_j^i)^2}$ is the

Euclidean distance between blog j and the positive ideal solution. The conventional TOPSIS measures the overall performance without considering the weights of the individual indices. Hence, it may identify a blog as a hot blog when the author has a high level of expertise but a low blog preference. To overcome this, a coefficient of dispersion is introduced to measure the deviation of the parameters

$$c_j' = (1 - SF_j^i / \bar{F}_j^i) \cdot [s_j^- / (s_j^+ + s_j^-)], \quad (5)$$

where \bar{F}_j^i is the average of the 11 parameters $\bar{F}_j^i = \left(\sum_{i=1}^{11} F_j^i \right) / 11$ and SF_j^i is the standard deviation of the parameters $SF_j^i = \left(\sqrt{\sum_{i=1}^{11} (F_j^i - \bar{F}_j^i)^2} \right) / 10$. As F_7 is a cost index (and is negatively associated), we use $1 - F_7^i$ as a substitute for F_7^i to give $\bar{F}_j^i = \left(\sum_{i=1}^6 F_j^i + (1 - F_7^i) + \sum_{i=8}^{11} F_j^i \right) / 11$. Finally, the blog with the greatest related similarity c' is identified as a hot blog. In this way, both the overall blog similarity and the individual parameter similarity are taken into consideration to improve the reliability of the results.

- (5) Opinion leader identification. A blogger cannot be considered to be influential if he or she has only published a few blogs. Thus, both quality and quantity are taken into consideration in the identification of opinion leaders. We formulate the influence V_i by

$$V_i = (\omega_1 \cdot v_{quantity} + \omega_2 \cdot v_{quality}) / (\omega_1 + \omega_2) \quad (6)$$

where $v_{quantity}$ is the quantity and $v_{quality}$ is the quality of blogs. Both are normalized by Eq. (3). The quality of blogs is measured by the largest c_j' of the blogs and is normalized by

$$x' = \frac{(x - \min(X))}{(\max(X) - \min(X))}. \quad (7)$$

The weights of ω_1 and ω_2 indicate the importance of the quantity and quality factors to an application, and are determined by the analyst to fit the circumstances.

4. Demonstration

To demonstrate the application of the BARR framework, we begin by showing the existence of hot topics in blogs. We then identify hot topics and use them to locate opinion leaders. We built and compiled a prototype using a Java platform (J2SE Development Kit 5.0) for demonstration purposes. The Xerces2 Java parser 2.5.0 plug-in was used to form and parse well-formed Web documents. We used the MySQL 4.1 database server and the graphical output JFreeChart.

We first entered the keyword “Apple iPhone” into a search engine and located 815 blogs from MySpace.com. Among these, 300 URLs were found, of which 294 were accessible. By analyzing the links, we identified 259 bloggers and 311 blogs. The general information that bloggers gave on MySpace was used to determine groups and personal networks. Using the same iPhone topic for illustration, Fig. 3 presents the relationship between bloggers in terms of groups, networks, and general interests using the Fruchterman–Reingold

algorithm [15] in Pajekman (a program for the analysis and visualization of large networks, available at <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>) to present the relationship among the vertices (bloggers). The dots in the figure represent a blogger and the lines linking the dots indicate that two bloggers belong to the same (a) group or (b) network, or have the same (c) interests. The plots show that the bloggers involved in discussions of the same topic do not belong to a specific group, network, or interest. Rather, they share their ideas with one another based on common hobbies. Thus, to

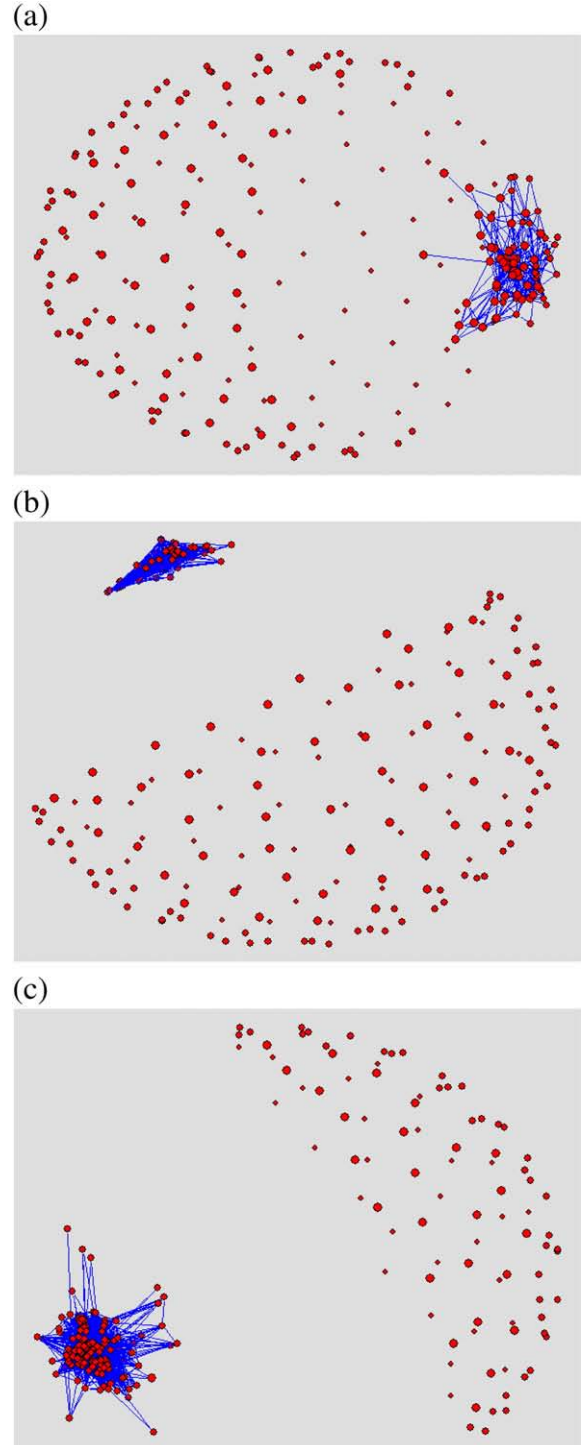


Fig. 3. Relationship between bloggers with similar (a) groups, (b) networks, and (c) interests.

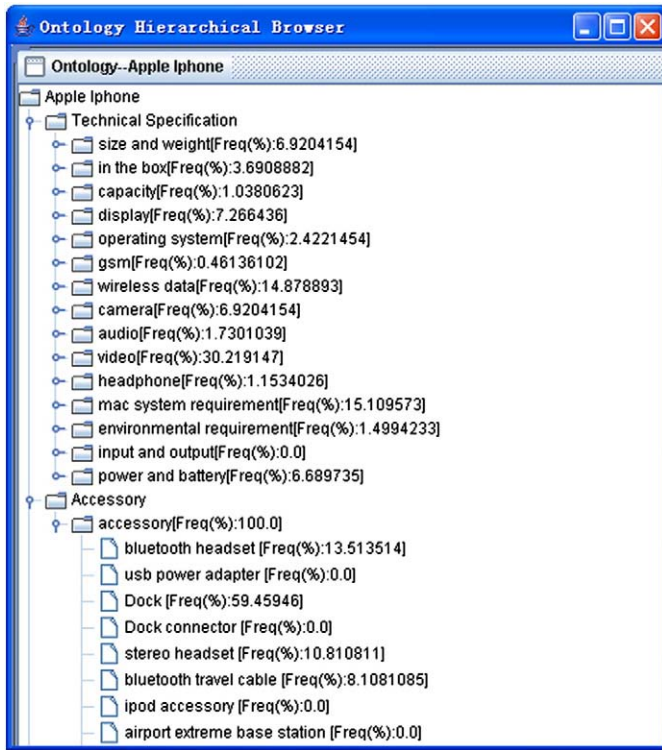


Fig. 4. Vector values of the official iPhone website.

identify an opinion leader, we need to delve further into the topics discussed.

We then built ontology of “Apple iPhone” retrieved from the web pages of the official Apple website (<http://www.apple.com/iphone>) using the approach proposed in [12]. This provided important

information for further analysis. We first calculated the frequency of words in a document and the information entropy of the words (calculated using Eq. (1)). Words with higher entropy values than the predefined threshold were then used to build the domain ontology. The ontology included information about components, technical specifications, and accessories, among others. We used the ontology to search for the frequency with which the keywords were used in

blogs $freq(term_j) = \frac{count(term_j)}{\sum_i count(term_i)} \times 100\%$. The keywords and their frequencies were then turned into vectors. The official iPhone website was converted into a vector for various items, such as camera, audio, and USB power adapter, as shown in Fig. 4. Each blog was then converted into a vector using a similar process.

In terms of technical specifications, the most commonly discussed topic was video function (28.82%). In the accessory category, the most popular topic was the iPhone Dock (59.46%), followed by headsets (13.51%).

The process of acquiring the 11 parameters can be summarized as follows.

- (1) Use a search engine to locate the URL of blogs from MySpace that discuss the iPhone.
- (2) Use the URLs to retrieve blogs that discuss the iPhone from the web pages (one page may have many blogs.)
- (3) Calculate the parameters for each blog. (Note that MySpace provides information on the number of comments (F2) and rank (F3), but not the number of clicks on a web page (F1)). Obtain the value of F4 (author) from the ratio of the vectors of a blog and the official iPhone website.
- (4) Obtain information about an author, such as general interests, groups, networks, and friends, from the website that corresponds to the URL in the blog. The information about the total number of blogs published by the same author (F5) and the number of comments made on other blogs (F6) comes from the same source.



Fig. 5. Rank of bloggers when $\omega_1 = 0.1$ and $\omega_2 = 0.9$, where the X coordinate represents blogger ID and the Y coordinate represents the influence score.

Table 1
Influence factors of opinion leaders (quality and quantity).

Blogger		Number of comments	Influence factor (TOPSIS value)
Blogger ID	Blog ID		
A	1	38	0.7874
	2	25	0.7689
B	3	24	0.7360
	1	100	0.7955
C	1	591	0.7889
	2	423	0.7885

- (5) Trace the URL of the author of a comment (defined as a reader of a blog) to obtain F7, F8, and F9 using a similar process to that in step 4.
- (6) Obtain F10 by comparing the value of the blog content of the author with that of the readers using Eq. (2).
- (7) Repeat steps 2 to 6 for all URLs to calculate the parameters.
- (8) The analysis in step 7 also acquires information on networking among the bloggers. Each pair of relationships is then used to measure the tie strength (F11).

These parameters are then employed to identify hot blogs using the TOPSIS to determine the popularity of a blog [21]. We first calculate Z_i^+ and Z_i^- for all of the parameters except for F1, and use the Z_i^+ of those 10 parameters to form a vector and Z^+ (the positive ideal solution), and Z_i^- to form a vector $-Z^-$ (the negative ideal solution). We then calculate the Euclidean distance between the vector of each blog to Z^+ and to Z^- , that is, s^+ and s^- . \bar{F}_j^i and SF_j^i are obtained from the 10 parameters for each blog. Finally, we use the coefficient of dispersion from Eq. (5) to measure the deviation of the parameters c_j^i for each blog. The hottest blog is the blog with the highest score.

To identify an opinion leader, we calculate the influence V_i by considering both the quality ω_1 and quantity ω_2 of blogs published by an author. Fig. 5 presents an example in which ω_1 and ω_2 are set at 0.1 and 0.9, which means that the quality is considered to be significantly more important than the quantity. The influence is normalized to be between [0 and 1]. For example, in Table 1, Blogger A published three popular blogs, Blogger B had the most popular blog, and Blogger C wrote two blogs that attracted the greatest amount of feedback. Blogger A is considered the opinion leader when the weights are set at $\omega_1 = 0.1$ and $\omega_2 = 0.9$.

5. Performance evaluation

The proposed BARR framework is compared with the conventional TOPSIS, using the Euclidean distance between measurements and the ideal solution to determine the popularity of a blog, to identify hot blogs. Fig. 6 shows the values for the 10 parameters F2 to F11 (the number of clicks on the web page, F1, was not provided by MySpace)

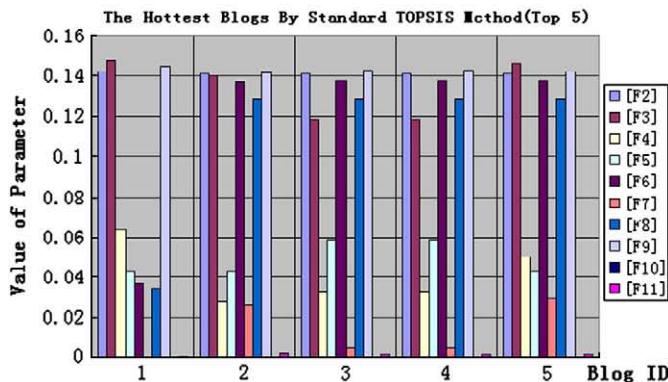


Table 2
Determination of opinion leaders using different weights.

	Blog ID	TOPSIS value	Number of comments
Google	B _G	0.4441	5
Accessible Blog	B _C	0.7889	591
BARR	B _T	0.7955	100

Table 3
Ranking of opinion leaders based on different weights.

(ω_1, ω_2)	Rank of blogger									
(0.1, 0.9)	1	2	3	4	5	6	7	8	9	10
(0.2, 0.8)	3	18	19	5	20	21	22	11	6	7
(0.3, 0.7)	10	18	19	4	20	21	22	11	5	6

of the five most popular blogs. It is clear that the values acquired with BARR are more evenly distributed than those obtained with the TOPSIS after transformation by sigmoid curve (Eq. (3)). This indicates that the introduction of the coefficient of dispersion (Eq. (5)) by considering the weights of the individual indices balances the contributions of the various parameters.

We then further compare the identification of popularity by the BARR framework, a search engine, and a blog web site. Table 2 shows three blogs identified as the most popular blogs. B_G is the blog listed top out of 815 searched results from Google, whereas B_C is the blog with the most comments among the 311 blogs downloaded (mentioned in the previous section). Finally, B_T is the blog identified with BARR. B_T has the highest TOPSIS score and B_G has the lowest. A closer look at the content shows that B_T specifically discusses the issue of iPhone pricing and quality, whereas both B_C and B_G are regular blogs that only briefly mention the iPhone. B_C has the greatest number of comments because the author claims to be a celebrity, which attracts more comments on the blog.

The weight ratio between the quality ω_1 and quantity ω_2 can be a decisive factor in determining an opinion leader. Table 3 shows how the determination of different parameters changes the influence ranking of a blogger. For example, when the quality draws more attention (with values of 0.1, 0.2, or 0.3), the rank of the original opinion leader drops from 1 to 3 and then to 10, respectively.

6. Conclusion

With the recent rapid development of online social networks, WoM has become an attractive marketing tool. However, to use this tool efficiently, it is important to identify the opinion leaders in such networks. This study proposes a framework for the identification of opinion leaders in online blogs. Using information on blog content, author properties, reader properties, and their relationship as anchors

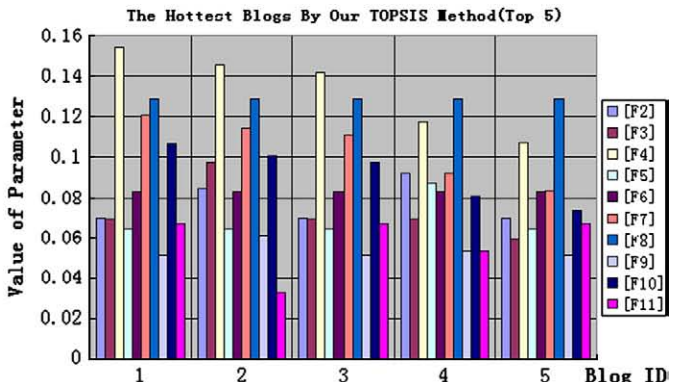


Fig. 6. Hot blog identification with the BARR framework and conventional TOPSIS.

for identification, we identify hot blogs in an online social blog platform and then use this data to identify opinion leaders, or the authors of hot blogs. We identify hot blogs by comparing the relevant identifying factors in an ontology built for selected topics concerning WoM. The identification of hot blogs and opinion leaders allows marketers to trace published blogs to ascertain whether the opinion therein supports their product or service. Action can then be taken to either promote products or services if the opinion is positive, or to repair customer relationships if the opinion is negative.

References

- [1] J. Allan, J. Carbonell, G. Doddington, J. Yamron, Y. Yang, Topic detection and tracking pilot study: final report, Proceedings of DARPA Broadcast News Transcription and Understanding Workshop, 1998, pp. 194–218.
- [2] I.N. Athanasiadis, P.A. Mitkas, Social influence and water conservation: an agent-based approach, *Computing in Science and Engineering* 7 (1) (2005) 65–70.
- [3] Segev Aviv, Avigdor Gal Enhancing portability with multilingual ontology-based knowledge management, *Decision Support Systems* 45 (3) (2008) 567–584, June.
- [4] H.S. Bansal, P.A. Voyer, Word-of-mouth processes within a services purchase decision context, *Journal of Service Research* 3 (2) (2000) 166–177.
- [5] P.F. Bone, Word-of-mouth effects on short-term and long-term product judgments, *Journal of Business Research* 23 (3) (1995) 213–223.
- [6] J.J. Brown, P.H. Reingen, Social ties and word-of-mouth referral behavior, *Journal of Consumer Research* 14 (3) (1987) 350–362.
- [7] H. Chan, Word-of-mouth communication: a conceptual framework and empirical tests, Proceedings of the Society for Consumer Psychology, Society for Consumer Psychology, 2001, pp. 142–143.
- [8] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: online book reviews, *Journal of Marketing Research* 43 (3) (2006) 345–354.
- [9] Y.H. Cho, J.K. Kim, S.H. Kim, A personalized recommender system based on web usage mining and decision tree induction, *Expert Systems with Applications* 23 (3) (2002) 329–342.
- [10] D. Collings, A.A. Reeder, I. Adjali, P. Crocker, M.H. Lyons, Agent based customer modelling: individuals who learn from their environment, Proceedings of the 2000 Congress on Evolutionary Computation, 2000, pp. 1492–1497.
- [11] C. Dellarocas, Strategic manipulation of Internet opinion forums: implications for consumers and firms, *Management Science* 52 (10) (2006) 1577–1593.
- [12] T.C. Du, F. Li, I. King, Managing knowledge on the web-extracting ontology from HTML web, *Decision Support Systems* 47 (2009) 319–331.
- [13] H. Endo, M. Noto, A word-of-mouth information recommender system considering information reliability and user preferences, Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 2003, pp. 2990–2995.
- [14] L.C. Freeman, Centrality in social networks: conceptual clarification, *Social Networks* 1 (3) (1979) 215–239.
- [15] T.M.J. Fruchterman, E.M. Reingold, Graph drawing by force-directed placement, *Software- Practice and Experience* 21 (11) (1991) 1129–1164.
- [16] T. Fujiki, T. Nanno, Y. Suzuki, M. Okumura, Identification of bursts in a document stream, Proceedings of the First International Workshop on Knowledge Discovery in Data Streams, 2004.
- [17] M.C. Gilly, J.L. Graham, M.F. Wolfinbarger, L.J. Yale, A dyadic study of interpersonal information search, *Journal of the Academy of Marketing Science* 26 (2) (1998) 83–100.
- [18] D. Godes, D. Mayzlin, Using online conversations to study word of mouth communication, *Marketing Science* 23 (4) (2004) 545–560.
- [19] N. Hu, P.A. Pavlou, J. Zhang, Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication, Proceedings of the 7th ACM Conference on Electronic Commerce, 2006, pp. 324–330.
- [20] N. Hu, L. Liu, J. Zhang, Analyst forecast revision and market sales discovery of online word of mouth, Proceedings of the 40th Hawaii International Conference on System Sciences, 2007.
- [21] C.L. Huang, K. Yoon, Multi-attribute decision making: methods and applications, a state of art survey, Springer-Verlag, New York, 1981.
- [22] Y. Jin, P. Bloch, G.T. Cameron, A comparative study: does the word-of-mouth communications and opinion leadership model fit epinions on the Internet? Proceedings of the Hawaii International Conference on Social Sciences, 2002.
- [23] D. Kempe, J. Kleinberg, E. Tardos, Influential nodes in a diffusion model for social networks, Proceedings of the 32nd International Colloquium on Automata, Languages and Programming, 2005, pp. 1127–1138.
- [24] K. Krippendorff, Content analysis: an introduction to its methodology, second edition Sage Publications, California, 2004.
- [25] P. Lazarsfeld, R.K. Merton, Friendship as a social process: a substantive and methodological analysis, in: M. Berger, T. Abel, C.H. Page (Eds.), *Freedom and control in modern society*, Van Nostrand, New York, 1954, pp. 18–66.
- [26] W.G. Mangold, F. Miller, G.R. Brockway, Word-of-mouth communication in the service marketplace, *Journal of Services Marketing* 13 (1) (1999) 73–89.
- [27] B.N. Miller, J.A. Konstan, J. Riedl, PocketLens: toward a personal recommender system, *ACM Transactions on Information Systems* 22 (3) (2004) 437–476.
- [28] A.A. Mitchell, P.A. Dacin, The assessment of alternative measures of consumer expertise, *Journal of Consumer Research* 23 (3) (1996) 219–239.
- [29] D.E. O'Leary, A multilingual knowledge management system: a case study of FAO and WAICENT, *Decision Support Systems* 45 (3) (June 2008) 641–661.
- [30] J.E. Phelps, R. Lewis, L. Mobilio, D. Perry, N. Raman, Viral marketing or electronic word-of-mouth advertising: examining consumer responses and motivations to pass along email, *Journal of Advertising Research* 44 (4) (2004) 333–348.
- [31] T.S. Raghu, A. Vinze, A business process context for Knowledge Management, *Decision Support Systems* 43 (3) (April 2007) 1062–1079.
- [32] P. Resnick, H.R. Varian, Recommender systems, *Communications of the ACM* 40 (3) (1997) 56–58.
- [33] P. Resnick, K. Kuwabara, R. Zeckhauser, E. Friedman, Reputation systems, *Communications of the ACM* 43 (12) (2000) 45–48.
- [34] Y. Sekiguchi, H. Kawashima, H. Okuda, M. Oku, Topic detection from blog documents using users' interests, Proceedings of the 7th International Conference on Mobile Data Management, 2006.
- [35] X. Song, Y. Chi, K. Hino, B.L. Tseng, Identifying opinion leaders in the blogosphere, Proceedings of the sixteenth ACM conference on information and knowledge management, 2007, pp. 971–974.
- [36] M.R. Subramani, B. Rajagopalan, Knowledge-sharing and influence in online social networks via viral marketing, *Communications of the ACM* 46 (12) (2003) 300–307.
- [37] T. Sun, S. Youn, G. Wu, M. Kuntaraporn, Online word-of-mouth (or mouse): an exploration of its antecedents and consequences, *Journal of Computer-Mediated Communication* 11 (4) (2006) 1104–1127.
- [38] J.E. Swan, R.L. Oliver, Postpurchase communications by consumers, *Journal of Retailing* 65 (4) (1989) 516–533.
- [39] Y. Takama, A. Matsumura, T. Kajinami, Visualization of news distribution in blog space, Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2006, pp. 413–416.
- [40] S. Takeuchi, J. Kamahara, S. Shimojo, H. Miyahara, Human-network-based filtering: the information propagation model based on word-of-mouth communication, Proceedings of 2003 Symposium on Applications and the Internet, 2003, pp. 40–47.
- [41] Y. Tong, X. Wang, H.H. Teo, Understanding the intention of information contribution to online feedback systems from social exchange and motivation crowding perspectives, Proceedings of the 40th Hawaii International Conference on System Sciences, 2007.
- [42] T.M. Tsai, C.C. Shih, S.T. Chou, Personalized blog recommendation using the value, semantic, and social model, Proceedings of the 3rd International Conference on Innovations in Information Technology, 2006, pp. 1–5.
- [43] M.P. Venkatraman, Opinion leaders, adopters, and communicative adopters: a role analysis, *Psychology and Marketing* 6 (1) (1989) 51–68.
- [44] S. Wasserman, K. Faust, *Social network analysis: methods and applications*, Cambridge University Press, New York, 1994.
- [45] R.A. Westbrook, Product/consumption-based affective responses and postpurchase processes, *Journal of Marketing Research* 24 (3) (1987) 258–270.
- [46] C.C. Yang, T.D. Ng, Terrorism and crime related weblog social network: link, content analysis and information visualization, Proceedings of 2007 IEEE International Conference on Intelligence and Security Informatics, 2007, pp. 55–58.
- [47] Y. Yang, J. Carbonell, R. Brown, T. Pierce, B.T. Archibald, X. Liu, Learning approaches for detecting and tracking news events, *IEEE Intelligent Systems* 14 (4) (1999) 32–43.
- [48] T. Yoshida, M. Hasegawa, T. Gotoh, H. Iguchi, K. Sugioka, K. Ikeda, Consumer behavior modeling based on social psychology and complex networks, Proceedings of the 9th IEEE International Conference on E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services, 2007, pp. 493–494.
- [49] Chen Yuh-Jen, Development of a method for ontology-based empirical knowledge representation and reasoning, *Decision Support Systems* 50 (1) (2010) 1–20, December.
- [50] G. Zacharia, A. Moukas, P. Maes, Collaborative reputation mechanisms for electronic marketplaces, *Decision Support Systems* 29 (4) (2000) 371–388, December.



Feng Li received his BS and MS degrees in control science and engineering in 1997 and 2000, respectively, and a PhD in systems engineering in 2004 from the Huazhong University of Science and Technology, China. He is currently a lecturer at the School of Business Administration at the South China University of Technology, China. His research interests include decision support systems, complex adaptive systems, and artificial intelligence.



Timon C. Du received his BS degree in Mechanical Engineering from the National Chung-Hsing University, Taiwan. He obtained his MS and PhD degrees in Industrial Engineering from the Arizona State University, USA. Dr. Du is a Professor at the Chinese University of Hong Kong. His research interests include e-business, data mining, collaborative commerce, and the semantic web. He has published papers in many leading international journals, including *Decision Support Systems*, *IEEE Transactions on Knowledge and Data Engineering*, *Communications of the ACM*, *IIE Transactions*, and *Information and Management*.