

A diffusion mechanism for social advertising over microblogs

Yung-Ming Li^{*}, Ya-Lin Shiu

Institute of Information Management, National Chiao Tung University, Hsinchu, 300, Taiwan

ARTICLE INFO

Article history:

Received 7 October 2010
Received in revised form 13 February 2012
Accepted 28 February 2012
Available online 5 March 2012

Keywords:

Social media
Online advertising
Microblog
Diffusion mechanism
Influence model
Preference analysis

ABSTRACT

Social media have increasingly become popular platforms for information dissemination. Recently, companies have attempted to take advantage of social advertising to deliver their advertisements to appropriate customers. The success of message propagation in social media depends greatly on the content relevance and the closeness of social relationships. In this paper, considering the factors of user preference, network influence, and propagation capability, we propose a diffusion mechanism to deliver advertising information over microblogging media. Our experimental results show that the proposed model could provide advertisers with suitable targets for diffusing advertisements continuously and thus efficiently enhance advertising effectiveness.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, social media, such as Facebook, Twitter and Plurk, have flourished and raised much attention. Social media provide users with an excellent platform to share and receive information and give marketers a great opportunity to diffuse information through numerous populations. An overwhelming majority of marketers are using social media to market their businesses, and a significant 81% of these marketers indicate that their efforts in social media have generated effective exposure for their businesses [59]. With effective vehicles for understanding customer behavior and new hybrid elements of the promotion mix, social media allow enterprises to make timely contact with the end-consumer at relatively low cost and higher levels of efficiency [52]. Since the World Wide Web (Web) is now the primary message delivering medium between advertisers and consumers, it is a critical issue to find the best way to utilize on-line media for advertising purposes [18,29].

The effectiveness of advertisement distribution highly relies on well understanding the preference information of the targeted users. However, some implicit personal information of users, particularly the new users, may not be always obtainable to the marketers [23]. As users know more about their friends than marketers, the relations between the users become a natural medium and filter for message diffusion. Moreover, most people are willing to share their information with friends and are likely to be affected by the opinions

of their friends [35,45]. Social advertising is a kind of recommendation system, of sharing information between friends. It takes advantage of the relation of users to conduct an advertising campaign. In 2010, eMarketer reported that 90% of consumers rely on recommendations from people they trust. In the same time, IDG Amplify indicated that the efficiency of social advertising is greater than the traditional advertising. It seems that social advertising has become an important advertising model for marketers. In reality, one of the most popular microblog websites, Twitter, announced an innovative advertising model, called “Promoted Tweets”, in April 2010. It makes tweets as advertisements, which are distinctive from both traditional keyword based advertisements and recent social advertisements. They measure the advertising performance and payment of sponsored tweets by “resonance”—the interactions between users and sponsored tweets such as retweet, reply, or bookmarking [29].

In practice, the advertisers should disseminate advertising messages by information-sharing between people and increase the resonance so as to widen the coverage (spread of social advertisements) and keep the advertisement alive. However, currently, they still lack an appropriate advertising mechanism which helps marketers to diffuse their advertisements effectively and improve resonance among users. Besides, the existent sharing mechanisms have a problem of excess sharing between friends. For example, a broadcast-to-all approach forcing users to share information with all of their friends can be executed by the system with small cost and likely adopted by social media platforms. However, this approach will cause a negative impression if friends are not interested in the received advertisements. Although its diffusion coverage could be larger, the number of receivers unhappy with the spammed advertisements also significantly increases. As a result, social spam has become a severe problem

^{*} Corresponding author. Tel.: +886 3 5712121; fax: +886 3 5723792.

E-mail addresses: yml@mail.nctu.edu.tw (Y.-M. Li), ayaa.iim97g@mail.nctu.edu.tw (Y.-L. Shiu).

confronted by users of social media. Sharing information over the network can improve people's reputation and develop their social capital [65]. However, sending too many unsolicited and irrelevant messages to friends will make them feel uncomfortable and even harm the development of social capital.

To address these issues, in this research, we design a diffusion support mechanism, which considers the factors of user preference, network influence, and propagation strength, to increase effectiveness of advertisements and support users to share information with appropriate friends. The proposed mechanism allows us to identify social advertisement endorsers with strong propagation capability in advertisement delivery and provides a suitable repost list of friends for each endorser. We validate our mechanism by conducting experiments in Plurk, one of the most popular microblog services. Our experimental results show that the proposed model could enhance the efficiency and effectiveness of advertising campaigns in terms of advertisement relevance, resonance, and coverage.

The rest of the paper is organized as follows: Section 2 reviews related works; Section 3 presents the research methodology and the framework of our system; Section 4 describes the experiments with detailed data collection and data analysis; the experimental results and evaluations are discussed in Section 5; and, finally, Section 6 concludes this study and offers suggestions for further research.

2. Related literature

2.1. Social media

Social media are Internet platforms designed to disseminate information or messages through social interactions, using highly accessible and scalable publishing techniques. Social media is composed of content (information) and social interaction interface (intimate community engagement and social viral activity). With its emerging trend and promising popularity, researchers have put academic efforts into analyzing the characteristics and functionalities of social media. Social media provide an unprecedented study opportunity for researchers [43]. For example, Krishnamurthy et al. [42] identify distinct characteristics of social media by users' behavior and relationships between users to explore miscellaneous insights into social media. Kaplan and Haenlein [31] examine the challenges and opportunities of social media and recommend a set of ten rules that companies should follow when developing their own social media strategy. Mangold et al. [52] derive hybrid elements of the promotion mix for marketing managers with a better understanding of social media, and propose a framework for incorporating it into their strategies.

Furthermore, to communicate effectively with customers, researchers analyzed marketing trends and social relations. For example, Gilbert and Karahalios [22] develop a predictive model that maps the data of social activity to tie strength so as to improve design elements of social media. Kim et al. [37] analyze the factors influencing the adoption of social media from the perspective of information needs in order to understand each user's behavior regarding information adoption. To better assess users' behavior, many researchers examine social influence, social interactions, and information diffusion in social media [17]. For example, Sun et al. [61] propose a recommendation framework to extract relevant emergency news feeds for user-based information diffusion. Kwak et al. [43] study the topological characteristics and analyze the state of information dissemination. In this paper, we study the design of diffusion mechanism for realizing social advertising in microblogs, which is an emerging research avenue for exploiting social media in the context of online advertising.

2.2. Online advertising

The issue of online advertising has aroused much academic interest and has been spotlighted for decades. Online advertising can

usually be categorized into two types: 1) targeted advertising, which delivers advertisements to a recipient based on the user's preference profiles; and 2) social advertising, which delivers the advertisements to influential users determined by social relationship [47]. Targeted advertising also can be considered as a kind of applications of recommender systems, which utilizes two main techniques: the content-based approach, and the collaborative-based approach to discover users' personal preferences [53]. The content-based approach uses users' previous preference profiles [2,44], while the collaborative-based approach uses general tastes of similar users' profiles [44,71]. However, both filtering approaches rely heavily on subject user ratings, making it hard to recommend new items to users when there are no related comments or rating records [23].

Customers' purchasing decisions likely be affected by buying experiences shared by other users [35,45]. Compared with traditional online advertising, social advertising is a form of advertising that addresses people as part of a social network and uses social relations and social influences between people to sell products or services [64,66]. In other words, social advertising use an indirect method, such as the word-of-mouth approach or an endorsement process, to disseminate advertisements [46]. An endorser is any individual with established recognition to represent with and a product [6,58]. To realize social advertising, we need to identify influential endorsers by using social network data and distributes appropriate advertisements in a social way (e.g. sharing between endorsers and their friends). Specifically, the social "distance" between users is measured and appropriate advertisements are distributed through users according to their social distance [28].

As electronic networks became more complex and dense, social network analysis was introduced to analyze complex networks such as inter-organizational systems (IOS) [10]. Social relationships and social interaction are powerful because they can act as trusted referrals and reinforce the fact that people influence people and have become the crucial components in social advertising [3]. Some researchers measure the influential strength by analyzing the number of network links and users' relationships and interaction in the network to identify the influential nodes for social advertising [48,66]. Therefore, studying social influence can help us better understand why certain information is transmitted faster than others and how we could help advertisers and marketers design more effective campaigns [8]. In this paper, considering the factors of user preference, network influence, and propagation capability, we propose a social diffusion mechanism to identify the appropriate endorsers with high prior propagation from the social network to deliver relevant advertisements widely.

2.3. Information diffusion

Many of the earliest ideas and models about information diffusion come from epidemiology. Research into the spread of epidemics has been conducted extensively over past years. Most of the proposed models suppose that populations are fully mixed: everyone has the same probability to be infected with the disease by any infected person. Nevertheless, the epidemic spreading concept may not completely suit the information diffusion on the social network. Xia et al. [70] indicate that spreading behavior is a non-uniform transmission. Different users' characteristics, preference, relationships, and actions on the network lead to individuals' various infection probability. Researchers analyze information diffusion in the social network based on the individual's characteristics. Some of their works are based on bond percolation, graph theory or a probabilistic model to extract the influential nodes, considering the aspect of dynamic characteristics, such as distance, time, and interactions [14,34,38,39]. Others exploit social network analysis techniques, to evaluate the influential nodes from the aspect of the node's structural position or a temporal notion of "node's distance", such as degree centrality, and closeness

centrality [40,41]. For example, Kossinets et al. [41] propose a framework for analyzing communication in networks, based on inferring the potential for information to flow between nodes in a university email network. By revealing influential factors and realizing the processes of the information diffusion, marketers can predict when and how the information spreads over social networks to maximize the expected spreading performance [27,33]. Leskovec et al. [46] empirically analyze the topological patterns of cascades in the context of a large product recommendation network. This has recently led to the adoption of viral marketing, where a seller attempts to artificially create word-of-mouth advertising among potential customers. In this paper, we consider both network influence and dynamic propagation factors to evaluate the diffusion capabilities of nodes in social networks.

The design of diffusion mechanism is conceptually similar to that of computer network multicast process. Multicast is a network technology for the delivery of information to a specific group using the most efficient strategy to deliver the messages over each link of the network [67]. In the context of computer network multicast routing, researchers mainly improve and reinforce packet routing in communication networks [16]. In the context of social networking, the links in social networks are formed by social relationships and interaction and researchers focus on the study of the issue: delivering the right information to the right nodes and spreading it widely. The goal can be achieved by implementing feasible approaches to discover the initial nodes and leveraging the social relationships to further diffuse the messages between users. The relation and preference relevance of the nodes in the social network significantly influence the effectiveness of information diffusion in social networks. Therefore, we apply the multicast concept to generate specific groups which are determined by the characteristics of nodes. In the paper, we develop a social endorser engine, which can identify the influential endorsers (users with high propagation capability), to support the information diffusion in the social networks.

2.4. Recommendation mechanisms

In this research, we intend to design a new endorser based advertisement delivery framework, which utilizes the characteristics of social media. The core endorser discovery mechanism in the proposed framework can be regarded as a recommender system, in which the outcome is a list of influential persons. Conventional recommendation methodologies include the content-based and collaborative filtering approaches. These recommendation approaches have to collect rating information of recommended items (e.g. articles, products, or advertisements, etc.) [4,12]. However, there might be some rating mechanisms provided in social media, but the mechanism used by users in rating is generally missing. Content based systems recommend an item based on a description of the item and a profile of the user's interests [71]. Although the details of systems vary, content-based recommendation systems share in common the means for describing the items to be recommended: a means for creating a user's profile that describes the types of items the user might be interested and a means for comparing items to the user profile to determine what to recommend. However, in our research, as we aim to find people who can diffuse messages to a wider range, this objective cannot be achieved by merely digging into user profiles. As for collaborative filtering, there is one important assumption: users who have similar preferences in the past are likely to have similar preferences in the future [15]. For recommendation purpose, we need the information of similarity in their preferences on a particular item. Nevertheless, we still cannot obtain the attributes (such as relationship and influence) to measure the diffusion power of individual users.

To resolve these issues, we incorporate the social network analysis (SNA) in the proposed framework. SNA has been applied in social

science for decades. SNA views the actor or node as the source of the action, and the connection or link as the relationship developed among nodes [63]. Researchers apply SNA to analyze the social network architecture and the individual's social environment, such as characterized structures, positions, and dyadic properties [20,54]. Recently, the development of recommender systems based on social networks becomes increasingly promising. For example, Kautz et al. [32] use a collaborative filtering method and SNA techniques to discover people with some expertise. By capturing the activity traces users leave on the social network, Liu and Maes [50] develop a recommender system to infer their everyday interests. However, these works do not fully utilize the power and characteristics of social media.

Our work differentiates from existing works in several aspects. First, from the perspective of methodology, we design a grand new framework comprehensively considering the components of user preference, social influence, and propagation tendency to discover users with strong diffusion power. Second, from the empirical perspective, we use microblogs as the data source rather than traditional web sites. Third, from the perspective of application context, we apply this recommendation mechanism to online social advertising so as to deliver advertisements effectively.

3. The system architecture

3.1. Social diffusion mechanism

We design a social diffusion mechanism to disseminate advertising information via social endorsers. The mechanism will identify the influential endorsers for delivering advertisements and generate a list of the next-stage targeted users for each endorser at the current stage, together with suitable paths for information diffusion. The users included in the recommended lists are those who have strong propagation capabilities in social networks, and are potentially willing to share the relevant information with all their friends. The process of our diffusion mechanism (AdPlurker) is shown in Fig. 1 and detailed as follows.

1. The endorser discovery engine is triggered to identify the influential users (referred to as social endorsers) who have high preference in the advertisements of the sponsors (user A in Fig. 1). The components of the endorser discovery engine are described in Subsection 3.2.
2. The system delivers relevant advertisements of the sponsors to identified social endorsers together with a recommended list of their friends with strong propagation capability for forwarding further advertisements.
3. After the social endorsers receive the advertisements, they share the received advertisements with their friends spontaneously (users B, C, D in Fig. 1) with the support of the recommended list of friends further generated by the endorser discovery engine.
4. The endorser discovery engine sends a corresponding list of friends to all the users who receive the advertisement respectively. The users receiving the advertisement become new social endorsers.
5. The social advertisement diffusion proceeds continuously by repeating steps 3–4.

Note that the proposed social diffusion mechanism is different from spamming. The friends selected by the endorser discovery engine are based on quantitative measurement of the factors, such as user preference, network influence, and propagation strength. The advertising message will be guided to the right people by the user's judgments and with the support of the system recommendation. If users deliver advertisements to their friends, it means that users also think their friends will like the advertisements. The mechanism takes advantage of content relevance and social relationships to

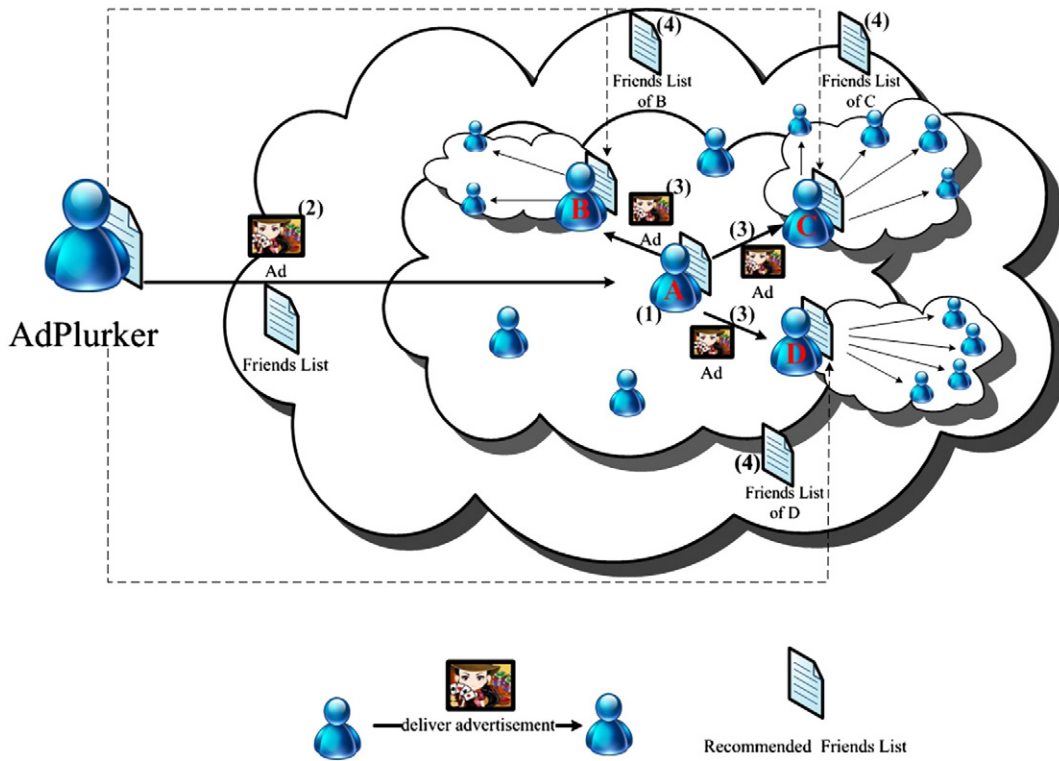


Fig. 1. Process of social diffusing mechanism.

reduce the negative impression of the advertisement and gain marketing effectiveness. It also suggests appropriate friends to users to share the information, which enhances the resonance and reduces the problem of social spam.

3.2. Endorser discovery engine

Social media provide us with the source data of individuals' preferences, social influence, and social interactions, which are openly

available. User preference is an important factor in targeted advertising. The social influence between users of social media happens when they are affected by others. It is likely that we are usually influenced by our friends or the people providing popular content. The social interaction is a crucial factor for empowering the social propagation. If a user interacts with someone frequently, to some extent there are more similarities and closeness between them. Therefore, we incorporate these components into our proposed social endorser discovery engine. Effective information diffusion on social networks is

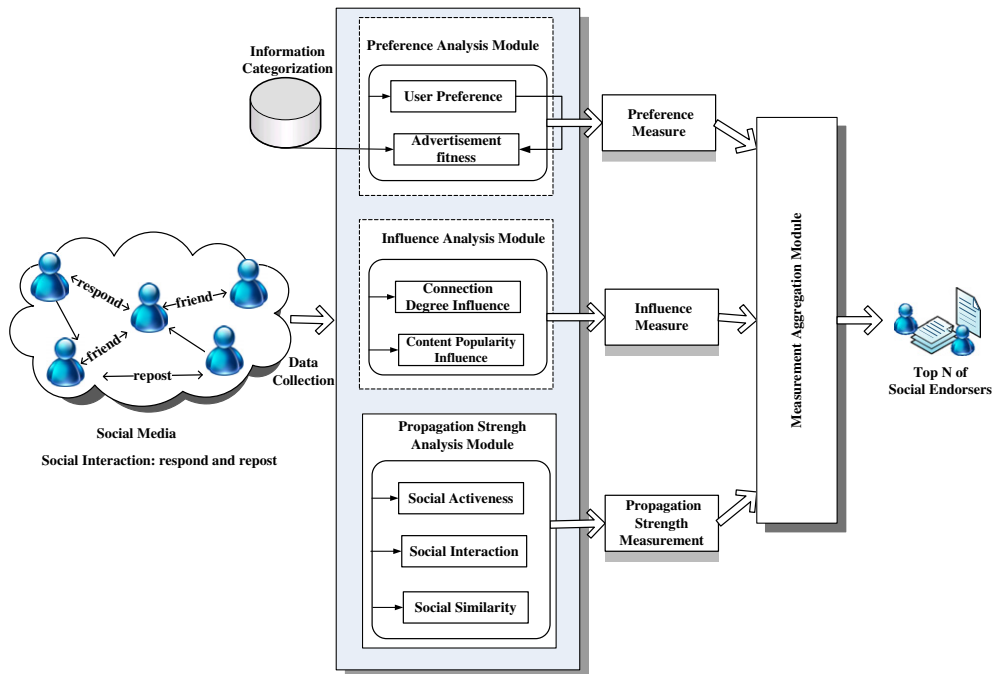


Fig. 2. Architecture of endorser discovery engine.

grounded in the relevance of individual preference and the closeness of social relations. Therefore, the main functionality of the proposed social endorser discovery engine is to identify the nodes with the strong propagation capabilities in disseminating relevant messages as widely as possible.

Fig. 2 shows the main components and procedures of the proposed endorser discovery engine. Our proposed system architecture includes four modules: “Preference Analysis” module; “Influence Analysis” module; “Propagation Strength Analysis” module; and “Measurement Aggregation” module. The details related to each module are described as follows.

3.2.1. Preference analysis module

3.2.1.1. User preference. User preference classification is an important issue in online marketing. By pre-classifying customers and offering personalized recommendation services or products it helps to improve customer satisfaction and target marketing. Personalized service requires understanding customers’ preferences to provide the right products to the right customers. Similarly, discovering a user’s preference is an important factor to be considered in identifying the influential social endorsers. By analyzing users’ preferences, we can better understand what kinds of information are suitable to deliver. Researchers focus on developing novel systems for product taxonomy [30] and analyzing customers’ histories, and present information on products to match customers’ preferences [9]. Kim et al. use a category tree to define users’ preferences and, based on users’ preference tree, to calculate their preference scores for personalizing advertisements [36]. A preference tree can reflect affinity levels among preference categories. In the preference analysis, we also adopt a tree structure to implement the classification of the preference domain. A tree-like structure is practically employed in much research, such as product taxonomy [2] and semantic similarity in taxonomy [55]. The preference tree of individual user is constructed based on a pre-defined category tree. A user’s preference can be collected by online questionnaires directly or collected indirectly from the user’s behavior, such as the fan page of particular brand he/she joins.

3.2.1.2. Advertisement fitness. For the purpose of evaluating advertisement matching, we establish the category tree of advertisement and use the same tree to define the user’s preference tree. Furthermore, we utilize a distance-based approach [69] to calculate the similarity between user preference tree and advertisement category tree. This approach proved to have better performance than other keyword similarity evaluation approaches [72]. We calculate the preference score to present the relevancy level (fitness degree) between the catalogs of an advertisement and a user’s preference. The preference score is measured based on the distance of the shortest path between the preference category of a user and the product category of an advertisement. If an advertisement belongs to two or more categories, the preference score will be the average value. Specifically, as shown in Fig. 3, assume C_1 and C_2 stand for the target category of the advertisement and the closest category of the user’s preference respectively and C_{fm} represents the first mutual parent node of C_1 and C_2 in a catalog tree. The fitness degree of the advertisements to a user can be calculated by the following formula:

$$Sim_p(C_1, C_2) = \frac{2D_{fm}}{D_1 + D_2 + 2D_{fm}}, \quad (1)$$

where D_1 (D_2) is the length of the path from C_1 (C_2) to C_{fm} and D_{fm} is the distance of the path from C_{fm} to the root node in the category tree.

3.2.2. Influence analysis module

Freeman et al. [21] indicate that degree centrality and betweenness centrality are important indicators for the influential leader

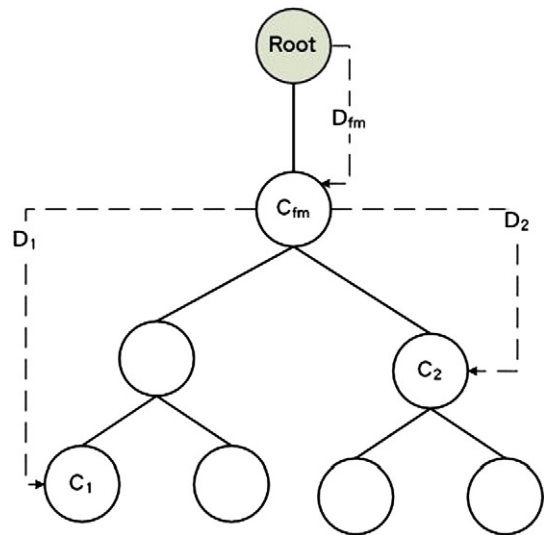


Fig. 3. Category tree.

discovery and group performance. Kiss and Bichler [40] adopt different centrality measures to select influencers in a customer network and compare the results of the diffusion of marketing messages. They found that the out-degree centrality is better than others on message diffusion. Also, Kwak et al. [43] make comparisons between in-degree approach and PageRank approach in the microblog environment. They found that the results of these two approaches are similar. Hence, we adopt degree centrality to evaluate users’ influence in the social network. Besides, we consider not only individual influence but also personal content influence in our system. Cha et al. [8] indicate that the content value of one’s posts is an important measure for evaluating influence. Therefore, we have connection degree influence and content popularity influence in the influence analysis module.

3.2.2.1. Connection degree influence. For the purpose of evaluating the relative importance of user position in the whole network, social network analysis is applied. Connection degree influence can be quantified by the degree centrality in a social network. Degree centrality is defined as the number of direct connections/links upon a node. Specifically, in-degree is a count of the number of connections directed to the node, and out-degree is the number of connections that the node directs to others. In this research, firstly, we consider the spammers or bots attempt to follow many people in order to gain attention. Secondly, some famous stars are followed by many fans, but they have little knowledge about the characteristics of their fans. That likely makes the distributed advertisement less relevant to the receivers. Neither the in-degree nor out-degree approach will appropriately catch the information of a user’s ability and incentive to effectively interact with other linked users in the social network. Therefore, we use mutual relationship (friendship) to measure the degree centrality as, in practice, the mutual degree represents the number of friends a user has. Mutual degree for user i is measured as:

$$Md(i) = \sum_{j=1}^n E_{ij}, \quad (2)$$

where E_{ij} is 1 if an edge from node i to j exists and an edge from node j to node i exists, too, otherwise it is 0.

3.2.2.2. Content popularity influence. Content popularity influence is used to evaluate the popularity of what a user posts. We measure the content degree of a user by the number of total responses and message reposts from people. We denoted $|\Phi|$ as the total number

of the elements in a set Φ . The formula for content popularity measure for a typical user i can be expressed as:

$$Cd(i) = \sum_{t=1}^T \frac{|\Phi_{response(i,t)}| + |\Phi_{repost(i,t)}|}{|\Phi_{post(i,t)}|}, \quad (3)$$

where $\Phi_{post(i,t)}$ stands for the set of the messages posted by user i at the time period t , $\Phi_{response(i,t)}$ represents the set of responses on user i 's posts, and $\Phi_{repost(i,t)}$ is the set of i 's posts which were reposted by others. T is the total number of time periods used to collect the messages. Denote $N_Md(i)$ as the normalized value of $Md(i)$ and $N_Cd(i)$ as the normalized value of $Cd(i)$. The network influence score of user i is the normalized value of summation $N_Md(i) + N_Cd(i)$.

3.2.3. Propagation strength analysis module

Besides the passive influence factors of connection degree and content popularity, the active propagation factors of user's capability and tendency in content diffusion should be considered.

3.2.3.1. Social activeness. Social activeness is used to calculate the activity intensity of a user. A user with higher activeness indicates that the user is engaged in greater information sharing and has a higher probability of transmitting the messages. We calculate the activeness of a user by the number of post records during a period of time in the social platform. The formula is defined as below:

$$Sa(i) = \frac{\sum_{t=1}^T |\Phi_{messages(i,t)}|}{T}, \quad (4)$$

where $\Phi_{messages(i,t)}$ is the total number of messages posted by user i at time period t .

3.2.3.2. Social similarity. Social similarity aims to measure the similarity degree between two people from implicit social structure and behavior, such as friend-in-common and content-in-common. The more friends-in-common of people generally reflects the higher connection level between them. Bagherjeiran and Parekh also discovered that friends tend to have similar interests; they are likely to be targeted with similar advertisements [3]. Therefore, if people have more common friends, their interests should be more similar. Under the conditions, when a person delivers a message to their friends, the possibility that friends are also interested in the message and repost it becomes higher. Denote $F(i)$ as a set of user i 's friends. The similarity of a friend-in-common between user i and friend j , is measured as:

$$Sim_{cf}(i,j) = \frac{|F(i) \cap F(j)|}{\max(|F(i)|, |F(j)|)}. \quad (5)$$

In addition, the more content-in-common posted by two people, the higher the similarity degree between them. Semantic analysis can be used to measure the social similarity in the aspect of content-in-comment and to discover the potential preference of users [5]. Specifically, traditional information retrieval (IR) technology can be exploited to analyze the semantics of content. To examine the semantic similarity among posts, we use the CKIP Chinese word segmentation system to parse and stem the crawled contents. All terms which were captured from the documents are used as the attributes of vector. We then apply the frequency-inverse document

frequency (TF-IDF) [57] weight analysis technique to measure how important a term is to a document in a collection or corpus. Fig. 4 shows the process of semantics similarity analysis.

The term frequency (TF) for term m in a post p is calculated as:

$$tf_{m,p} = \frac{freq_{m,p}}{\max_l(freq_{l,p})}, \quad (6)$$

where $freq_{m,p}$ is the raw frequency of term i appearing in post p and $\max_l(freq_{l,p})$ is the number of times the most frequent index term, l , appears in post j . The inverse document frequency (IDF) for term m is formulated as:

$$idf_m = \log \frac{N_p}{n_m}, \quad (7)$$

where N_p is the total number of posts and n_m is the number of posts in which term m appears. Then, the relative importance of term m to post p can be obtained by calculating:

$$W_{m,p} = tf_{m,p} * idf_m. \quad (8)$$

Finally, we measure the content similarity degree between people by a cosine similarity metric. The number of terms (keywords) selected for analyzing a person's document in a collection or corpus is denoted as M . In this research, M is the total number of unique terms captured in all documents. The similarity of corpus (content-in-comment degree) between user i and friend j is defined as:

$$Sim_{cc}(i,j) = \cos(\vec{i}_M, \vec{j}_M) = \frac{\vec{i}_M \cdot \vec{j}_M}{|\vec{i}_M| |\vec{j}_M|}, \quad (9)$$

where \vec{i}_M and \vec{j}_M are the document vectors in the M dimensional term space for user i and friend j .

Finally, the similarity score is measured by the sum of friend-in-common and content-in-comment scores. That is, social similarity between user i and friend j , is formulated as:

$$Ss(i,j) = Sim_{cf}(i,j) + Sim_{cc}(i,j). \quad (10)$$

3.2.3.3. Social interaction. Social interaction measure is different from social similarity measure since social interaction explicitly catches the factor of dynamic actions/interactions between people. It can be used to evaluate the intimacy between two users. For instance, it is common for a user to respond or repost someone's messages. It is reasonable to assume that the more interaction activities, the higher the probability of sharing information, as friends usually are interested in mutual messages. Given two users i and j , the social interaction score between them can be formulated as:

$$Si(i,j) = \sum_{t=1}^T \frac{|\Phi_{response(i,j,t)}|}{|\Phi_{response(i,t)}|} + \frac{|\Phi_{repost(i,j,t)}|}{|\Phi_{repost(i,t)}|}, \quad (11)$$

where $\Phi_{response(i,j,t)}$ is the set of responses generated by user j to user i 's posts and $\Phi_{repost(i,j,t)}$ is the set of reposts conducted by user j to user i 's posts.



Fig. 4. The process of semantics similarity analysis.

3.2.3.4. Social propagation. The social propagation score is used to evaluate the network diffusion strength of a user. The social propagation score of a user is calculated by aggregating the social propagation score and can be computed in a recursive way. To enhance the dissemination of advertising messages efficiently, it is important to pay attention to the next layer's social propagation. Though advertisers deliver advertisements to a social endorser with high social propagation capability, they can't ensure that the social endorser's friends will spread the advertisements with equal strength. Therefore, we suppose friends' social propagation capability would also affect a social endorser's propagation score, and include the friends' propagation score to evaluate the overall propagation score of a social endorser. In other words, individual's social propagation capability is aggregated recursively by their friends' social propagation capabilities. The social propagation score of user i is formulated as:

$$Sp(i) = \sum_{j \in F(i)} Sp(j) \cdot (Ss(i,j) + Si(i,j)), \quad (12)$$

where user j has not been visited and $Sp(j) = 1$ if $F(j) = \emptyset$. The termination condition holds when user j has no friends or all his/her friends have been visited. Denote $N_Sa(i)$ as the normalized value of $Sa(i)$ and $N_Sp(i)$ as the normalized value of $Sp(i)$. The propagation strength score of user i is the normalized value of the summation, $N_Sa(i) + N_Sp(i)$.

3.2.4. Measurement aggregation module

To evaluate the diffusion strength of a social endorser quantitatively, the preference measure, influence measure, and propagation strength measure, which are derived from the former analysis modules, need to be further aggregated with appropriate weighting distribution. Common approaches to deal with the uncertain weighting problem of parameter combination include a black-box method, such as artificial neural network (ANN), and white-box method, such as analytic hierarchy process (AHP) [19]. ANN can be trained to recognize and generalize the relationship between a set of inputs and outputs by nonlinear processes. However, it is difficult to acquire the proxy for an endorser's propagation capability, which is required in training data for learning ANN. Therefore, in this research, we choose AHP approach as our aggregation module. AHP is one well-known method to solve multi-criteria decision-making (MCDM) problems, which determinate the relative importance or weight of criteria by mathematical pair-wise comparison [7,25,56]. It has been applied extensively in many research fields, such as social network analysis [49], and recommendation systems [25,51].

Generally, there are three main stages involved in AHP. First of all, hierarchy construction building: users need to decompose their decision problem subproblems into multiple hierarchical levels. Secondly, pair-wise comparison: users have to compare each factor in the same level by the pair-wise comparison based on their own experience and knowledge. However, since the comparisons are carried out by personal or subjective judgments, some degree of inconsistency will occur. Therefore, users need to measure the consistency in order to ensure the judgments are consistent. The final stage, consistency verification, measures the degree of consistency among the pair-wise comparisons by computing the consistency ratio. This stage is also regarded as one of the greatest advantages of the AHP [24] and we leverage this benefit to adjust the weight to get better results.

4. Experiments

In this section, we apply the proposed mechanism in a microblog system to examine its effectiveness. Microblog service is one of the top platforms for social media marketing. Compared to traditional blogs, microblogs allow users to publish, and read and disseminate brief messages easily. Microblogs emphasize "what are you doing?"

that means that people's status has changed over time so they must keep friends updated frequently. These characteristics – brief messages, instant, easy to read, and easy to share – make microblogs a good platform to conduct social media marketing. Therefore, in this research, we apply and validate our proposed mechanism using microblog systems. For convenience and representativeness, we conduct our experiments in Plurk, a platform created in May 2008 and one of the most popular microblog services. Google Trend reported that the daily individual visitors of Plurk number more than the Twitter in Taiwan. Furthermore, according to Alexa's survey in 2010, 34.4% of Plurk's traffic comes from Taiwan. A user of Plurk (named as a Plurker) can connect with his/her friends via many functions such as updating instant messages, sharing images or video with friends and responding to friends' messages. A Plurker can also repost interesting messages (named as plurks) which was initially posted by his/her friends. In Plurk, the friendship relation needs mutual agreement of two users. Plurks can be broadcasted to all fans or sent privately to certain friends [68]. Notice that a fan in Plurks is similar to a follower in Twitter. While there does not exist "friend" relation in Twitter, it can be simulated by the relation of mutual signing up as a follower of each other. Therefore, the design concept of the presented framework can be applied to the Twitter.

Utilizing the available APIs, we develop a Plurk application named AdPlurker and invite users who are active and have used Plurk for a long time to join the experiments. AdPlurker is a software agent which will automatically reply a request according to the message it receives. To use AdPlurker service, users can simply add it as one of his/her friends. For an advertisement provided by a sponsor/merchant, AdPlurker will first identify the top- k most suitable users as the endorsers from the initial users, which are invited from the Plurk users. AdPlurker sends private messages with advertisements to the discovered social endorsers discovered. Social endorsers will first receive the messages, which include brief information about the advertisement, a relevant picture, and then receive a recommended list of users who are potentially interested in the advertisements and have prior propagation capabilities in their social network. When a social endorser delivers the advertisements to their friends, the system triggers the endorser discovery engine and sends a recommended list of friends to each endorser. The recommended list also displays friends with strong propagation capabilities in their network. Due to the 140-character limit, it is convenient for an endorser to distribute only one advertisement in a single message and the recommended friend list would be severed into multiple messages if the total length of the list exceeds the character limit.

Notice that Plurk does not provide an official replurk (repost) function to facilitate the sharing behavior. Therefore, we provide a replurk function for users. While a user posts our advertisements, AdPlurker will respond to the user's plurk with a replurk link immediately. The repost function is considered the feature that has made the microblog a new medium of information dissemination. The user can click the hyperlinks of brief information to get detailed information and the click-through record will be collected for further evaluation. Also, users can share the messages with their friends who are recommended by the system. The transmitted message records will be also collected. To prevent click fraud, we record only one click for the same advertisement for individual users.

4.1. Data description

4.1.1. Profile of participants

Until April 2011, there are 217 users (55% male, 45% female) aged between 20 and 50 participating in our experiment. To simulate a real network structure, our target users come from different career backgrounds, as shown in Fig. 5.

Spammers and bots are common phenomena in microblogs. To prevent spam accounts and bots from adding noise and bias in our

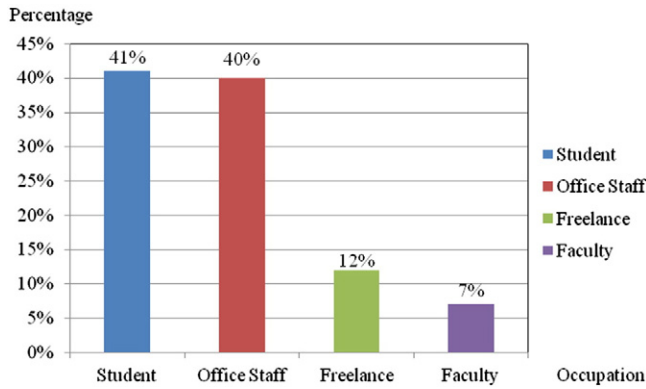


Fig. 5. Population of occupational distribution.

study, we removed spam accounts and bots reported by users in our experiment. There are 247,090 users and 1,969,253 plurks in our database. We collect data from the target users to the 3rd layers, as the degrees of separation are normally limited to the 3rd–4th layers [1,47], and crawl their plurks, responses, and interactions with friends that occurred within three months. In the Plurk, “become fans” is a function for users to follow others’ pages and also declare their preferences for information type. It is interesting to find that Plurk’s network has statistics (e.g. the distributions of the number of fan pages, friends, and reposts and responses of all plurks) which are nearly the power-law distributions [11,60]. 91.5% of the target users have less than 20 fan pages. 90.7% of the target users have less than 200 friends. 93.6% of plurks are shared among less than 15 users. The distributions of the number of friends, the number of reposts, and the number of responses in the Plurk’s network are following nearly the power-law distribution with properties explained in [11,60].

4.1.2. Profile of advertisements

In the experiment, a catalog tree with 18 leaf categories, belonging to six parent categories and three grandparent categories, is adopted. The category tree is constructed by combing the product classification of Yahoo and PChome, which are the two most famous online shopping sites in Taiwan. The affiliation of each category is illustrated in a tree structure as shown in Fig. 6. To simulate the diversity of advertisements used in the experiment, we took 216 advertisements sampled from multiple channels, which include Yahoo, PChome, KKBOX, and Engadget. We collected 18 various advertisements belonging to

the 18 leaf categories in a set of advertisements. In total, 12 sets of advertisements are used to compare the performance of 4 different advertising approaches (our social diffusion approach and 3 other different benchmark advertising approaches). There are 3 sets of advertisements use for the experiment for each advertising approach. Notice that in the experiment, a variety of advertisement is experimented in only one advertising approach to avoid the repetition effect that the selected users receive the same advertisement which was previously viewed.

4.2. Score calculation

After we constructed the social network structure of the participants and prepared the category tree for representing the user preference and advertisement type, we evaluated the scores of various analysis modules in the proposed diffusion mechanism and identified the influential users as the endorsers for social advertising. Notice that as the data and social networks in the Plurk evolve with time, the score data is updated on a regular time schedule.

4.2.1. Preference analysis calculation

In the preference module, to construct the personal preference category tree, we collect target users’ explicit preferences directly by online questionnaires. In addition, in order to better elicit users’ preferences, we also collect implicit preferences indirectly from their behaviors in Plurk – the well known accounts (fan pages) of particular brands (such as Dell, Taipei Walker and KKBOX,) a user joins. The categories of these well-known accounts (fan pages) are processed to match the classification of the category tree as shown in Fig. 6. Notice that although we need some human workloads in consolidating the product categories in two shopping sites and matching the category of fan page to the defined product category, the task is just a one-time pre-process work. By analyzing the preference category tree of each user, we can discover users’ profiles easily and precisely.

4.2.2. Influence analysis calculation

In the influence module, both the connection degree and content degree measures are taken into consideration. The friend links are usually the strongest links and imply the structural influence in the network. A user with high content popularity implies that they are influential in content, as a user becomes attention-grabbing when their plurks are often replurked and responded to by others. To avoid the different scale problem, we normalize both values to be in the range

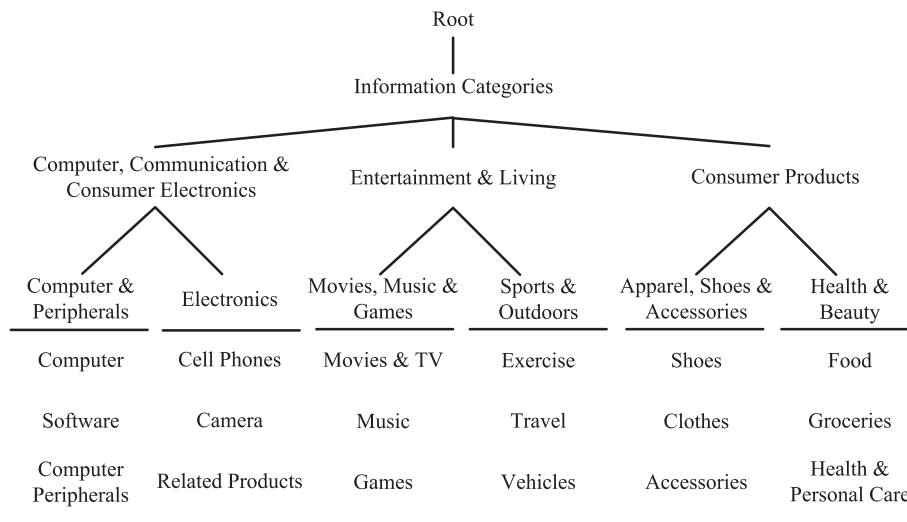


Fig. 6. The category tree of ADPlurker.

[0,1]. The normalized values of the connection degree and content degree are summed to obtain the influence score. Notice that the normalized value $N_V(i)$ of a measure value $V(i)$ is given by the following formula:

$$N_V(i) = \frac{V(i) - \min_j V(j)}{\max_j V(j) - \min_j V(j)}. \quad (13)$$

4.2.3. Propagation strength analysis calculation

The propagation strength analysis module is divided into three parts: 1) Social activeness score: We quantify users' occurring activities by the number of plurks during a month; 2) Social similarity score: We collect each user's friend list to compute the similarity degree of friend-in-common between users and crawled users' plurks during the past three months, to calculate the similarity degree of content-in-comment; 3) Social interaction score: We evaluate the intimacy between two users and collect the responded and reposted data between two users to evaluate the level of dynamic interactions between people. Notice that to get propagation strength score, all these values are also normalized to be in the range [0,1].

4.2.4. Measurement aggregation calculation

As mentioned in Section 3.2.4, in this research, we employed the AHP to calculate the weight combination of parameters.

4.2.4.1. Questionnaire survey. The questionnaire is constructed using the AHP concept. A nine-point scale is used for representing the relative importance (from 1, "equal importance", to 9 "absolute or extreme importance", in our questionnaire [51]. The expert questionnaire targets the following people: 1) professional lecturers; and 2) active users of social media. A total of 15 questionnaires were sent out, with 14 forms returned.

4.2.4.2. Weight determination. Four stages are involved in using AHP for attribute weighting calculation. Firstly, we establish a pair-wise comparison matrix formulated as:

$$R = \begin{bmatrix} a_{11} & a_{12} & a_{1N} \\ a_{21} & a_{22} & a_{2N} \\ a_{N1} & a_{N2} & a_{NN} \end{bmatrix}, \text{ where } a_{ij} = \begin{cases} a_{ij}^{-1} & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}. \quad (14)$$

Secondly, to derive the relative weights of these criteria from the comparison matrix, an Eigenvector method is used. We use Principal Eigen Value, which is obtained from the summation of products between each element of Eigenvector and the sum of columns of the matrix [62], to check the consistency of the experts' answer. Thirdly, consistency verification is conducted by measuring the degree of consistency among the pair-wise comparisons by the consistency ratio (CR) defined by $CR = CI/RI$, where CI is the consistency index used to measure deviation or the degree of consistency and RI is the random index used to measure the consistency verification. If the value of CR is smaller or equal to 10%, the inconsistency is acceptable. If the CR is greater than 10%, the subjective judgment needs further revision [56]. Finally, the weight determination of each evaluation factor is obtained by using 14 experts' opinions (Table 1).

From this AHP model, we find that each of the three factors have similar importance levels, while propagation plays the most important part in the advertising campaign, followed by the user's personal preference and then social influence.

Table 1
Measure and weight.

Measure	Preference	Influence	Propagation strength
Weight	0.340075768	0.325100803	0.33482343

4.3. Diffusion strategies

In this research, we compare our diffusion approach with the three other benchmark approaches, to evaluate the performance of the proposed system design. The four different approaches used in the experiments are described as follows.

1. In-degree: It is the most common approach used to evaluate the influence of a user in microblog systems (eg. the number of fans on Plurks or followers on Twitter). This approach is currently employed by many other third-part services, such as Twitterholic.com and Wefollow.com. In addition, Kwak et al. compared the in-degree approach with the PageRank approach in microblog and find two results to be similar [42]. Therefore, we chose the in-degree as our comparison.
2. Ratio-degree: The measure is the ratio between the number of a user's followers and the number of other people that the user follows. It was proposed by the Web Ecology Project applied in Twitter platform, which is an interdisciplinary research group based in Boston, Massachusetts. Owing to platform restriction, we could not obtain the data of the people that a user follows in Plurk. Therefore, we adopt friend relationships to replace the people that a user follows. We adjust it as the ratio between the number of a user's friends and the number of other people that a user follows.
3. Topic-influence: Discovering the topic-influential nodes for delivering advertising messages by taking advantage of the target advertising and social influence. In this approach, only preference and influence measures are used.
4. Our approach (Social diffusion): The approach we proposed in this study. We apply an analytic hierarchy process (AHP) to realize the final weight combinations of three components: preference measure, influence measure, and propagation strength measure.

According to the research, 75% reposts occurred in 24 h [42]. In our experiments, to reduce the errors due to time factor, we distributed the advertisements belonging to the same catalog to the users who were selected by the four approaches, and collected the feedback in one day. Then, we delivered the next set of advertisements of different catalogs to users. For each approach, we calculate the aggregated scores of the users and the top-k ranked users are selected in the list of recommended endorsers. In the experiment, considering the limitation of 140 characters in a microblog message, a list of top-5 influential users is recommended by these four diffusion strategies. Note that at each time period, four different advertisements belonging to the same product category were be used to evaluate performances of the four various advertising approaches. The information of the corresponding method used for a specific advertisement in the experiment is recorded in the database. However, when an endorser receives an advertisement, he/she does not know which advertising approach to be tested. All four advertising approaches are experimented concurrently. In total, we collected 54 sample performance records for each advertising approach.

5. Results and evaluations

In this section, we discuss the results of experimental study and some insights discovered from the observation and analysis. In order to evaluate the performance of our proposed mechanism, we used three performance indicators: click-through rate, repost rate, and exposure rate. Click-through rate encompasses practical statistics about advertising effectiveness [66]; Repost rate is an indicator about advertising resonance to evaluate the state of sharing. By analyzing repost rate, we could realize the resonance between users and advertisements [29]. Click-through rate and repost rate are the key measures in "promoted tweets", which is a new social advertisement platform proposed by Twitter. Furthermore, another type of performance that advertisers care about is the spreading coverage of

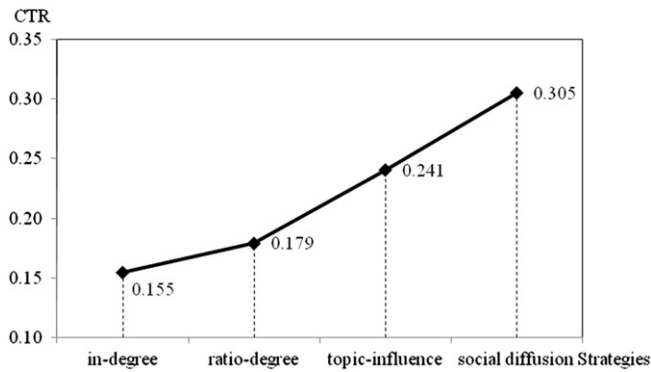


Fig. 7. Click-through rates (CTR) of four diffusion strategies.

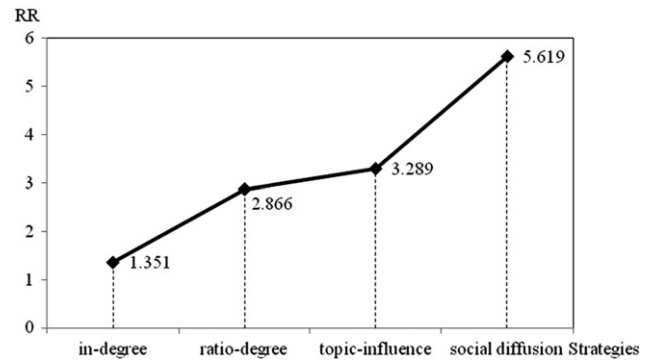


Fig. 8. Repost rates (RR) of four diffusion strategies.

advertisements. Therefore, we also compare the exposure rates (total unique receivers of the advertisements) of all four advertisement diffusion strategies.

5.1. Click-through rate of advertisements

N_c is the total number of clicks and N_d is total number of advertisements delivered. The click-through rate (CTR) is defined as:

$$CTR = N_c/N_d. \quad (15)$$

According to the collected dataset, the outcome is illustrated in Fig. 7. The advertisements of in-degree approach got a total of 0.155 CTR in 4141 delivery times. The ratio-degree approach got 0.179 CTR in 7954 deliveries. Topic-influence approach got 0.241 CTR in 9321 deliveries. Our social diffusion mechanism got 0.305 CTR in 14,918 deliveries. Our social diffusion approach has the highest result in CTR, followed by the topic-influence, ratio-degree, and in-degree approaches. Topic-influence is better than degree-based approaches, implying that the effectiveness of preference factor is significant. In addition, our social diffusion approach is greater than the topic-influence approach, which further indicates that the factor of propagation capability (social similarity and interactions) plays an important role in social advertising.

We further verify the statistical significance of our results. A paired sample t -test confirms the significant difference of the results (Table 2). At 95% significant level, all the test results show that our social diffusion approach is significantly different from other benchmark approaches. Therefore, it verifies that our proposed approach has the best performance, compared to other benchmark approaches.

5.2. Repost rate of advertisements

N_{repost} is the total number of reposts and N_d is total number of advertisements delivered. The repost rate (RR) formula is defined as:

$$RR = N_{repost}/N_d. \quad (16)$$

According to the collected dataset, the advertisements of in-degree approach got a total of 1.351 RR. Ratio-degree approach got 2.866 RR. The topic-influence approach got 3.289 RR. Our social

diffusion mechanism got 5.619 RR. The outcome is illustrated in Fig. 8. The RR results are consistent with CTR results.

A paired sample t -test (95% significant level) also confirms that our social diffusion approach is significantly different from other benchmark approaches (Table 3). The superiority of our approach is statistically verified.

As previously mentioned, “Resonance” is a crucial factor for social advertisement diffusion. Better click-through and repost rates give strong evidences that our proposed mechanism improves resonance among users and advertisements. Also, they show that the advertisements shared by endorsers are generally also interesting to their friends. Consequently, the risk of spamming friends is reduced.

5.3. Exposure rate of advertisements

Advertisers are concerned about the effective exposure for their advertisements. Therefore, the eventual spreading coverage of advertisements is an important indicator to evaluate the performance of an advertising campaign. $N_{receiver}$ is the total number of receivers and N_d is total number of advertisements delivered. The exposure rate (ER) is the average unique receiver per advertisement and formulated as:

$$ER = N_{receiver}/N_d. \quad (17)$$

Fig. 9 shows the exposure rate of the four diffusion strategies. The advertisements of in-degree approach got an average of 42.691 receivers per advertisement. The ratio-degree approach got an average of 82.000 receivers per advertisement. The topic-influence approach got on average 96.093 receivers per advertisement. Our proposed social diffusion mechanism got on average 153.794 receivers per advertisement. Table 4 shows the analysis of statistical significance of the results of our approach with other benchmark approaches. As we can observe, the proposed social diffusion approach is significantly different from topic-influence, ratio-degree and in-degree. Therefore, we could confidently infer that a social diffusion approach works better in propagating the advertisements.

Furthermore, we also compare the diffusion performance with respect to various types of advertisement. We compared the advertisements belonging to three grandparent categories in our category tree: “Computer, Communication and Consumer Electronics (3C)”, “Entertainment and Living” and “Consumer Products”. As shown in Figs. 10 and 11, the click-through rate (CTR) of each category is

Table 2
Statistical verification on CTR measurement.

Paired group		Mean	Std. deviation	Std. error mean	T	Sig. (2-tailed)
Social diffusion V.S.	Topic-influence	23.845	41.323	4.195	5.683	0.000
	Ratio-degree	32.298	42.116	4.276	7.553	0.000
	In-degree	40.371	42.483	4.313	9.359	0.000

Table 3
Statistical verification on RR measurement.

Paired group		Mean	Std. deviation	Std. error mean	T	Sig. (2-tailed)
Social diffusion V.S.	Topic-influence	2.329	4.317	0.438	5.315	0.000
	Ratio-degree	2.752	5.627	0.571	4.817	0.000
	In-degree	4.268	6.250	0.634	6.725	0.000

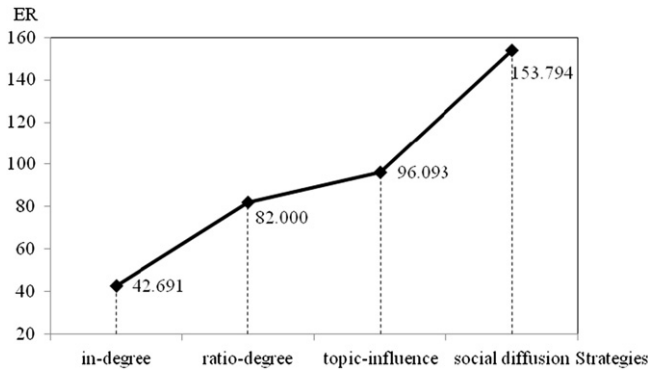


Fig. 9. Exposure rates (ER) of four diffusion strategies.

0.297, 0.373, and 0.268 respectively. The repost rate (RR) of each category is 4.310, 9.235, and 3.118 respectively. Entertainment and Living has better performance than the other two, in both CTR and RR. This phenomenon reveals that social advertising is more effective for the products of Entertainment and Living. That is, people tend to be affected by the opinions of their friends when selecting the products (e.g. movies/TV, music, and games) and services (e.g. sports, and outdoor pursuits) of this type.

5.4. Extended comparisons

In the subsection, we further verify the proposed framework by conducting more performance comparisons with respect to various aspects.

Firstly, we compare the performance of AHP factor weighing method with those of two other factor weighing approaches – equal weights (EW) and principal component analysis (PCA), which is a popular statistical method used to explain the covariance structure of data by means of a small number of components [13,26]. According to the result of PCA, the proposed three dimensions (preference, influence, and propagation) could explain the 92.417% covariance structure of data. The linearly combination coefficients of these three dimensions given by PCA are (0.641, 0.635, 0.502). The coefficients are used as weighting values in the evaluation process. The comparisons are depicted in Fig. 12. We can observe that the AHP method generates the best performance in advertisement CTR.

Secondly, we analyze the effects of various index combinations on the performance. Fig. 13 shows the performances of the methods using only two factors of the three indicators (preference analysis indicators (PA), influence analysis indicators (IA), and propagation strength analysis indicators (PSA)). This result reveals that our proposed mechanism, which includes three indexes, has the best performance in advertisement CTR.

Table 4
Statistical verification on ER measurement.

Paired group		Mean	Std. deviation	Std. error mean	T	Sig. (2-tailed)
Social diffusion V.S.	Topic-influence	57.701	71.083	7.217	7.995	0.000
	Ratio-degree	71.793	92.766	9.419	7.622	0.000
	In-degree	111.103	103.496	10.508	10.573	0.000

Thirdly, we examine the accuracy of a recommended friend list. As an endorser can self-select his/her friends who are suitable to receive the advertisement, the accuracy rate of the recommended list of friends which is defined the rate that users repost an advertising message to *all* the friends included in the list without any change. Fig. 14 compares the accuracy rate of the recommended list generated by different approaches. As we can observe, our proposed social diffusion approach has the highest performance. The results show that our proposed mechanism can effectively recommend a list of suitable friends to receive the advertisement and relieve a user’s burden to select the friends (84% accuracy rate). In the experiment, it took less than 10 s for the endorser discovery engine to generate a list of endorsers under the computing hardware – Win7 32 bits PC, 2.0 GB RAM, and 3.00 GHz CPU. Therefore, the computational time used in generating a list of recommended endorsers is practically feasible

Lastly, we compare the effectiveness of broadcast-to-all approach and other social advertising approaches. In the broadcast-to-all approach, an advertisement is sent to all the initial users, their friends, friends of their friends, and so on by the system, without any filtering process. As shown in Fig. 15. This score reveals that our social advertising has higher CTR (0.305). However, a broadcast-to-all approach has the lowest CTR (0.042).

Advertisers are also concerned out whether the users like the advertisements they deliver or not. Therefore, another popular measure for evaluating the effectiveness of each advertising method is to conduct a user evaluation about the relevance of the advertisements delivered. The underlying principle of relevance scale is that a higher score indicates a user has a higher level of interest to a received advertisement. The relevance measure helps us to evaluate the user satisfaction with the received advertisement. Therefore, we further evaluate the relevance levels of various advertising approaches. A Likert scale is used, with rating levels ranging from 1 to 5 (offensive: 1; irrelevant: 2; moderate: 3; interesting: 4; joyful: 5). Fig. 16 shows that although a broadcast-to-all approach may have higher coverage rate, the relevance of the advertisements received by the users is significantly low. Most of the receivers feel the advertisements delivered by the broadcast-to-all approach offensive or irrelevant.

6. Conclusion

With the prosperity of social media, marketers attempt to take advantage of social advertising to effectively disseminate advertisements. However, it lacks an effective online advertisement diffusion mechanism to help marketers increase the resonance and spreading coverage of advertising messages. Besides, users encounter the problem of social spam when sharing excessively with friends, which will cause a negative impression and even harm the development of social capital. Therefore, in this paper, utilizing social networks and content analysis techniques, we design an endorser-based diffusion mechanism, which considers extensively the factors of user preference,

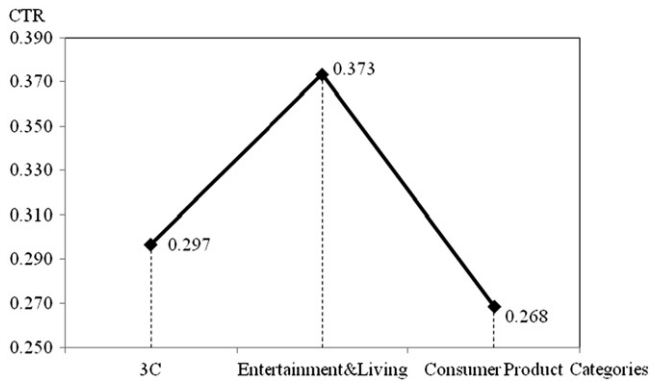


Fig. 10. CTRs of different categories of product.

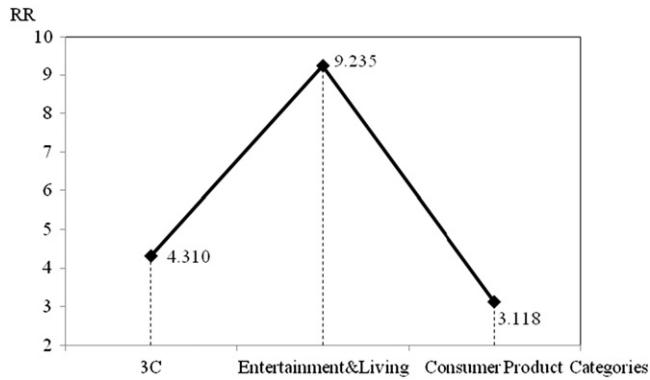


Fig. 11. RRs of different categories of product.

network influence, and propagation capability, for online advertisements over microblogs. The proposed mechanism is empirically verified on one microblog service website, Plurk, and experimental results show that our developed social diffusion mechanism outperforms other three benchmark approaches in the performance of advertisement relevance, resonance, and coverage.

6.1. Research contributions

The contributions and managerial implications of this paper are summarized as follows. Firstly, from the perspective of system innovation, while marketing on social media has become increasingly popular, little research has proposed a diffusion mechanism to study the online advertisements on social media. We are one of the pioneers to study a recommended list to indicate each intermediate node for information dissemination. Secondly, from the perspective of methodology, we not only consider the individual preference and

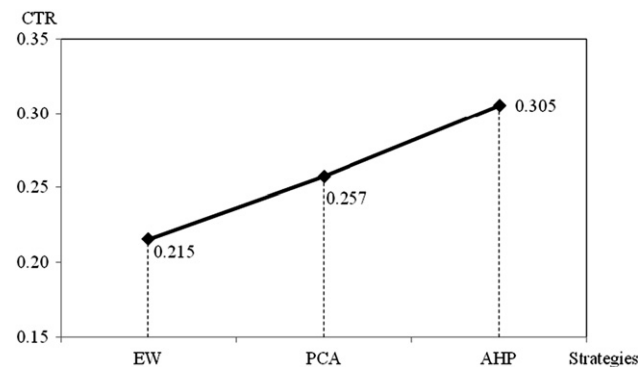


Fig. 12. CTRs of different weighting methods.

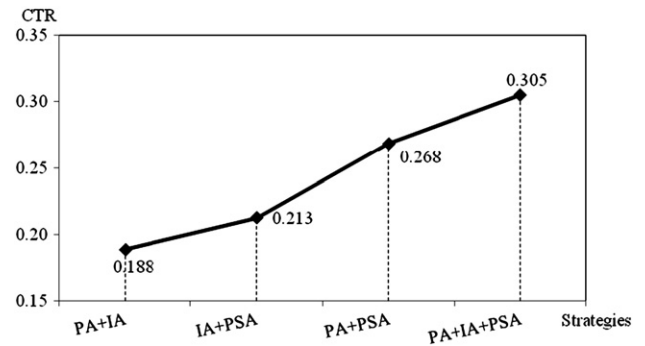


Fig. 13. CTRs of different methods containing different indicators.

network influence (link structure of relationship and content popularity), but also propagate (social activeness, social interactions, and social similarity) factors in the evaluation of nodes' diffusion capabilities to identify the people who can spread the advertising messages widely. Thirdly, from the perspective of performance, better click-through rate reflects that our mechanism can raise the visibility of advertising information. A higher repost rate indicates a higher exposure of the advertising and reveals that users are interested in the advertisement when shared by friends and are willing to share it with others. It also proves that our system can reduce the risk of spamming friends and improve resonance among users. Our proposed mechanism can widely extend the spreading coverage of advertisements and improve the resonance of advertisements. Lastly, from the perspective of practice, our empirical experiments show that social advertising is particularly effective in marketing goods and services such as movies/TV, music, games, sports, and outdoor pursuits. The proposed diffusion mechanism provides the advertisement sponsors with a powerful vehicle to conduct advertising diffusion campaigns successfully.

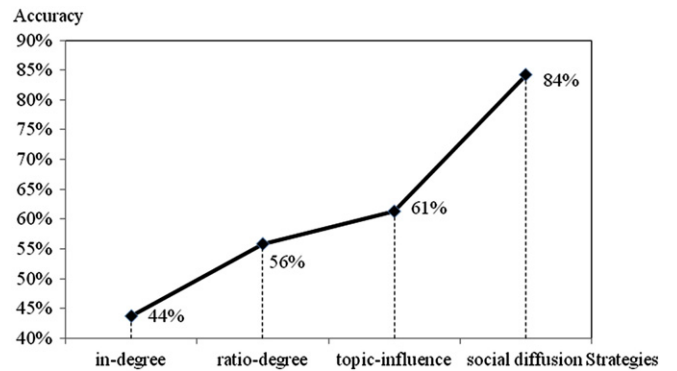


Fig. 14. Accuracy rate of recommended friend list.

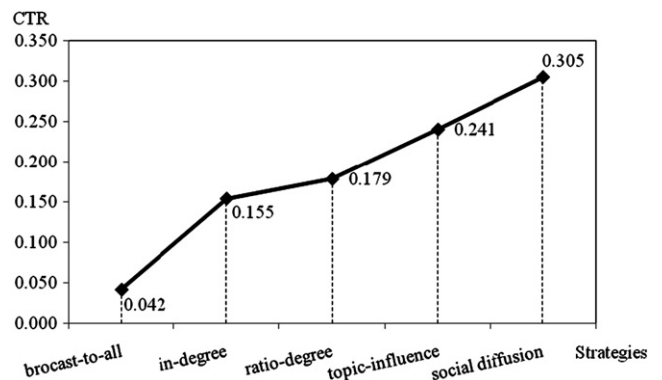


Fig. 15. CTRs of different advertising approaches.

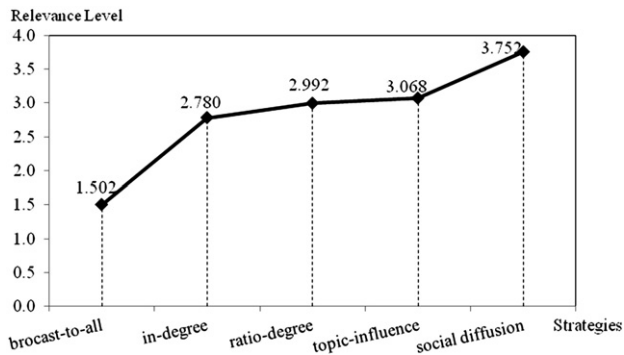


Fig. 16. Relevance levels of different advertising approaches.

6.2. Limitations and future studies

There are several limitations in this research. Firstly, the constructed social networks in the experiments only reflect a subset of an entire network as we couldn't look through all data of users due to the constraint of the experimental platform. For example, most social media provide privacy mechanisms for users; we can't collect data since we are not their friends or fans. Besides, only users themselves can obtain the list of people whom they follow; we can't retrieve the complete out-degree data to compare with the approach applied in the methodology. Secondly, the advertising efficiency might be affected by the format of the advertisements (e.g. size, position). However, the presentation of advertisements is restricted because of the limitation of 140 characters in our experimental platform. Thirdly, due to the scale of the experiment during limited time, we use only a category tree of 18 leaf categories and 12 sets of advertisements for the validation and comparison work. While the number of participants and varieties of advertisements are statistically explainable, an experiment on a larger scale would be helpful. Lastly, the purpose of the proposed social diffusion framework is to effectively distribute advertising information and avoid advertisement spamming. If the objective of advertisement diffusion is changed, the framework should be somewhat different. For example, a broadcasted-to-all approach may become a good alternative to reach more people.

There are some directions for future studies. First, the analysis modules included in our proposed mechanism can be further refined and improved. For example, user preference analysis may obtain a higher accuracy if a category ontology is constructed to analyze users' profile and discover users' preferences. We could improve the user preference analysis module by adopting advertising taxonomy and mapping with content analysis. Second, incorporating other tangible factors and dynamic modules into the mechanism may improve the quality of the system. For example, to enhance the advertisement diffusion efficiency, we could further design a monitor mechanism to adjust the recommended list or variation of users' preference after a time period. Third, we want to identify the shortest path of target node by recommending the intermediate nodes instead of recommending a specific group. Fourth, a successful social advertising system needs the cooperation and contribution of the participants. It is an interesting avenue to include an incentive mechanism in the system and analyze its impact. Finally, the impact of different types of social media can be further investigated. We can conduct the experiments on other social media to examine which type of platforms performs better.

Acknowledgment

This research was supported by the National Science Council of Taiwan (Republic of China) under the grant NSC 99-2410-H-009-035-MY2.

References

- [1] Y.Y. Ahn, S. Han, H. Kwak, S. Moon, H. Jeong, Analysis of topological characteristics of huge online social networking services, Proceedings of the 16th international conference on World Wide Web, New York, NY, USA, 2007, pp. 835–844.
- [2] A. Albadvi, M. Shahbazi, A hybrid recommendation technique based on product category attributes, Expert Systems with Applications 36 (9) (2009) 11480–11488.
- [3] A. Bagherjeiran, R. Parekh, Combining behavioral and social network data for on-line advertising, Proceedings of the 2008 IEEE International Conference on Data Mining Workshops, Pisa, Italy, 2008, pp. 837–846.
- [4] M. Balabanovic, Y. Shoham, Fab: content-based, collaborative recommendation, Communications of the ACM 40 (3) (1997) 66–72.
- [5] B. Berendt, R. Navigli, Finding your way through blogspace: using semantics for cross-domain blog analysis, Proceedings of the AAAI Spring Symposium 2006 (AAAI) on Computational Approaches to Analyzing Weblogs, Palo Alto, CA, 2006, pp. 1–8.
- [6] C.K. Hsu, D. McDonald, An examination on multiple celebrity endorsers in advertising, Journal of Product & Brand Management 11 (1) (2002) 19–29.
- [7] D. Cao, L.C. Leung, J.S. Law, Modifying inconsistent comparison matrix in analytic hierarchy process: a heuristic approach, Decision Support Systems 44 (4) (2008) 944–953.
- [8] M. Cha, H. Haddadi, F. Benevenuto, K.P. Gummadi, Measuring user influence in twitter: the million follower fallacy, Proceedings of the 4th International Conference on Weblogs and Social Media, Washington, 2010.
- [9] J.L. Chandon, M.S. Chtourou, D.R. Fortin, Effects of configuration and exposure levels on responses to web advertisements, Journal of Advertising Research 43 (2) (2003) 217–229.
- [10] L. Chi, C.W. Holsapple, C. Srinivasan, Competitive dynamics in electronic networks: a model and the case of interorganizational systems, International Journal of Electronic Commerce 11 (3) (2007) 7–49.
- [11] A. Clauset, C. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Review 51 (2007) 661–703.
- [12] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, M. Sartin, Combining content-based and collaborative filters in an online newspaper, ACM SIGIR '99 Workshop on Recommender Systems, 1999, pp. 60–64.
- [13] C. Croux, G. Haesbroeck, Principal components analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies, Biometrika 87 (2000) 603–618.
- [14] C. Danescu-Niculescu-Mizil, G. Kossinets, J. Kleinberg, L. Lee, How opinions are received by online communities: a case study on amazon.com helpfulness votes, Proc. 18th International World Wide Web Conference, 2009.
- [15] S. Debmath, N. Ganguly, P. Mitra, Feature weighting in content based recommendation system using social network analysis, Proceeding of the 17th International Conference on World Wide Web, 2008, pp. 1041–1042.
- [16] S.E. Deering, D.R. Cheriton, Multicast routing in datagram internetworks and extended LANs, ACM Transactions on Computer System 8 (2) (1990) 85–111.
- [17] S.A. Delre, W. Jager, T.H.A. Bijmolt, M.A. Janssen, Will it spread or not? The effects of social influences and network topology on innovation diffusion, Journal of Product Innovation Management 27 (2) (2010) 267–282.
- [18] T. Dinev, Q. Hu, A. Yayla, Is there an on-line advertisers' dilemma? A study of click fraud in the pay-per-click model, International Journal of Electronic Commerce 13 (2) (2008) 29–60.
- [19] R.F. Dyer, E.H. Forman, Group decision support with the analytic hierarchy process, Decision Support Systems 8 (2) (1992) 99–124.
- [20] L. Freeman, Centrality in social networks: conceptual clarification, Social Networks 1 (1979) 215–239.
- [21] L. Freeman, S. Borgatti, D. White, Centrality in valued graphs: a measure of betweenness based on network flow, Social Networks 13 (2) (1991) 141–154.
- [22] E. Gilbert, K. Karahalios, Predicting tie strength with social media, Proceedings of the 27th International Conference on Human Factors in Computing Systems, Boston, MA, USA, 2009, pp. 211–220.
- [23] S. Ha, Helping online customers decide through Web personalization, IEEE Intelligent Systems 17 (6) (2002) 34–43.
- [24] W. Ho, Integrated analytic hierarchy process and its applications: a literature review, European Journal of Operational Research 186 (1) (2008) 211–228.
- [25] Y. Huang, L. Bian, A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet, Expert Systems with Applications 36 (1) (2009) 933–943.
- [26] M. Hubert, P.J. Rousseeuw, K. Vanden Branden, ROBPCA: a new approach to robust principal components analysis, Technometrics 47 (2005) 64–79.
- [27] J.L. Iribarren, E. Moro, Information diffusion epidemics in social networks, URL: <http://arxiv.org/pdf/0706.0641>, Accessed: 04/29/2010.
- [28] J. Fiorillo, Twitter Advertising: Pay-Per-Tweet, URL: <http://www.wikinomics.com/blog/index.php/2009/04/22/twitter-advertising-pay-per-tweet/>, Access at: 04/12/2010.
- [29] K. Jason, Twitter Promoted Tweets, URL: <http://kottke.org/10/04/twitters-promoted-tweets>, Access at: 05/23/2010.
- [30] Y. Jiang, J. Shang, Y. Liu, Maximizing customer satisfaction through an online recommendation system: a novel associative classification model, Decision Support Systems 48 (3) (2010) 470–479.
- [31] A.M. Kaplan, M. Haenlein, Users of the world, unite! The challenges and opportunities of social media, Business Horizons 53 (1) (2010) 59–68.
- [32] H. Kautz, B. Selman, M. Shah, Combining social network and collaborative filtering, Communications of the ACM 40 (3) (1997) 63–65.

- [33] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network, Proc. 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2003.
- [34] D. Kempe, J. Kleinberg, E. Tardos, Influential nodes in a diffusion model for social networks, Proceedings of the International Colloquium on Automata, Languages and Programming (ICALP), ACM, New York, 2005, pp. 1127–1138.
- [35] Y. Kim, J. Srivastava, Impact of social influence in e-commerce decision making, Proceedings of the 9th International Conference on Electronic Commerce, ACM, New York, 2007, pp. 293–302.
- [36] J.W. Kim, K.M. Lee, M.J. Shaw, H.L. Chang, M. Nelson, R.F. Easley, A preference scoring technique for personalized advertisements on Internet storefronts, Mathematical and Computer Modelling 44 (1–2) (2006) 3–15.
- [37] Y. Kim, M. Kim, K. Kim, Factors Influencing the Adoption of Social Media in the Perspective of Information Needs, Proceedings of the iConference, 2010.
- [38] M. Kimura, K. Saito, R. Nakano, H. Motoda, Finding influential nodes in a social network from information diffusion, Social Computing and Behavioral Modeling (2009) 1–8.
- [39] M. Kimura, K. Saito, R. Nakano, Extracting influential nodes for information diffusion on a social network, In Proceedings of the 22nd national conference on artificial intelligence, Menlo Park, CA, 2007, pp. 1371–1376.
- [40] C. Kiss, M. Bichler, Identification of influencers – measuring influence in customer networks, Decision Support Systems 46 (1) (2008) 233–253.
- [41] G. Kossinets, J. Kleinberg, D. Watts, The structure of information pathways in a social communication network, Proc. 14th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2008.
- [42] B. Krishnamurthy, P. Gill, M. Arliitt, A few chirps about twitter, Proceedings of the First Workshop on Online Social Networks Seattle, WA, USA, 2008, pp. 19–24.
- [43] H. Kwak, C. Lee, H. Park, S. Moon, What is Twitter, a social network or a news media? Proceedings of the 19th International World Wide Web (WWW) Conference, Raleigh, North Carolina, USA, 2010, pp. 591–600.
- [44] C. Lam, SNACK: incorporating social network information in automated collaborative filtering, Proceedings of the 5th ACM conference on Electronic Commerce, 2004, pp. 254–255.
- [45] M. Lee, C. Cheung, C. Sia, K. Lim, How positive informational social influence affects consumers' decision of Internet shopping, Proceedings of the 39th Annual Hawaii International Conference on System Sciences, IEEE, Washington DC, 2006, pp. 115–124.
- [46] J. Leskovec, A. Singh, J. Kleinberg, Patterns of influence in a recommendation network, Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2006.
- [47] Y.M. Li, C.W. Chen, A synthetical approach for blog recommendation: combining trust, social relation, and semantic analysis, Expert Systems with Applications 36 (3) (2009) 6536–6547.
- [48] Y.M. Li, N.J. Lien, An endorser discovering mechanism for social advertising, Proceedings of the 11th International Conference on Electronic Commerce, Taipei, Taiwan, 2009, pp. 125–132.
- [49] J. Liebowitz, Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations, Journal of Knowledge Management 9 (1) (2005) 76–86.
- [50] H. Liu, P. Maes, InterestMap: harvesting social network profiles for recommendations, Proceedings of IUI Beyond Personalization, 2005.
- [51] D.R. Liu, Y.Y. Shih, Integrating AHP and data mining for product recommendation based on customer lifetime value, Information Management 42 (3) (2005) 387–400.
- [52] G. Mangold, D. Faulds, Social media: the new hybrid element of the promotion mix, Business Horizons 52 (4) (2009) 357–365.
- [53] P. Massa, B. Bhattacharjee, Using trust in recommender systems: an experimental analysis, iTrust 2004, 2004, pp. 221–235.
- [54] M.E.J. Newman, A measure of between-ness centrality based on random walks, Social Networks 27 (1) (2005) 39–54.
- [55] P. Resnik, Semantic similarity in a taxonomy: an information-based measure and its application to problems of ambiguity in natural language, Journal of Artificial Intelligence Research 11 (1999) 95–130.
- [56] T.L. Saaty, The Analytic Network Process: Decision Making With Dependence and Feedback, RWS Publication, Pittsburgh, 2001.
- [57] G. Salton, Developments in automatic text retrieval, Science 253 (1991) 974–980.
- [58] D. Seno, B.A. Lukas, The equity effect of product endorsement by celebrities: a conceptual framework from a co-branding perspective, European Journal of Marketing 41 (1/2) (2007) 121–134.
- [59] M. Stelzner, Social Media Marketing Industry Report 2009, URL: <http://marketingwhitepapers.s3.amazonaws.com/sms09/SocialMediaMarketingIndustryReport.pdf>, Accessed: 06/17/2010.
- [60] A.T. Stephen, O. Toubia, Explaining the power-law degree distribution in a social commerce network, Social Networks 31 (4) (2009) 262–270.
- [61] A.R. Sun, J. Cheng, D.D. Zeng, A novel recommendation framework for a micro-blogging based on information diffusion, Proceeding of the 19th Workshop on Information Technologies and Systems, Phoenix, Arizona, US, 2009, pp. 199–216.
- [62] K. Teknomo, Analytic Hierarchy Process (AHP) Tutorial, URL: <http://people.revoledu.com/kardi/tutorial/AHP/>, Accessed: 02/19/2010.
- [63] J.C. Wang, C.C. Chiu, Recommending trusted online auction sellers using social network analysis, Expert Systems with Applications 34 (3) (2008) 1666–1679.
- [64] K. Wang, E.T.G. Wang, C.K. Farn, Influence of Web advertising strategies, consumer goal-directedness, and consumer involvement on Web advertising effectiveness, International Journal of Electronic Commerce 13 (4) (2009) 67–96.
- [65] M. Wasko, S. Faraj, Why should I share? Examining social capital and knowledge contribution in electronic networks of practice, MIS Quarterly 29 (2005) 35–57.
- [66] C. Wen, B.C.Y. Tan, K.T.T. Chang, Advertising effectiveness on social network sites: an investigation of tie strength, endorser expertise and product type on consumer purchase intention, Proceedings of the International Conference on Information Systems (ICIS), Phoenix, Arizona, United States, 2009, pp. 15–18.
- [67] Wikipedia, Multicast, URL: <http://en.wikipedia.org/wiki/Multicast>, Accessed: 09/28/2010.
- [68] Wikipedia, Plurk, URL: <http://en.wikipedia.org/wiki/Plurk>, Accessed: 05/04/2011.
- [69] Z. Wu, M. Palmer, Verbs semantics and lexical selection, Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics, Las Cruces, New Mexico, 1994, pp. 133–138.
- [70] C.Y. Xia, Z.X. Liu, Z.Q. Chen, Z.Z. Yuan, Spreading behavior of SIS model with non-uniform transmission on scale-free networks, The Journal of China University of Posts and Telecommunication 16 (1) (2009) 27–31.
- [71] W.S. Yang, J.B. Dia, Discovering cohesive subgroups from social networks for targeted advertising, Expert Systems with Applications 34 (3) (2008) 2029–2038.
- [72] S.T. Yuan, C. Cheng, Ontology-based personalized couple clustering for heterogeneous product recommendation in mobile marketing, Expert Systems with Applications 26 (4) (2004) 461–476.



Yung-Ming Li is an Associate Professor at the Institute of Information Management, National Chiao Tung University in Taiwan. He received his Ph.D. in Information Systems from the University of Washington. His research interests include network science, Internet economics, and business intelligence. His research has appeared in IEEE/ACM Transactions on Networking, European Journal of Operational Research, Decision Support Systems, International Journal of Electronic Commerce, Information Technology and Management, Electronic Commerce Research and Applications, ICIS, WTIS, among others.



Ya-Lin Shiu is a software engineer at Chunghwa Telecom Laboratories in Taiwan. She received her M.S. degree from the Institute of Information Management, National Chiao Tung University in Taiwan and B.S. degree in Information Management from the National University of Kaohsiung, Taiwan. Her research interests focus on electronic commerce and social media.