Ads by Whom? Ads about What? Exploring User Influence and Contents in Social Advertising

Jaimie Y. Park
Division of Web Science and Technology, KAIST, Daejeon, Korea
jaimie@islab.kaist.ac.kr

Kyoung-Won Lee
Division of Web Science and Technology, KAIST, Daejeon, Korea
kyoungwon.lee@islab.kaist.ac.kr

Sang Yeon Kim
Department of Computer Science, KAIST, Daejeon, Korea
sangyeon@islab.kaist.ac.kr

Chin-Wan Chung
Division of Web Science and Technology & Department of Computer Science, KAIST, Daejeon, Korea
chungcw@kaist.edu

ABSTRACT

Despite the growing interest in using online social networking services (OSNS) for advertising, little is understood about what contributes to the social advertising performance. In this research, we pose following questions: How many clicks do social advertisements actually receive? What are the characteristics of the advertisements that receive many clicks? What factors contribute to the clicks on advertisements? In order to answer these questions, we collect data from AdbyMe, a social media advertisement platform that connects businesses, or advertisers, with users of online social network services. Businesses can reach a large target audience through AdbyMe users who publish the advertisements on their social networks. We analyze the factors that may affect the clicks on advertisements being published on OSNS. In particular, we look into the advertised contents as well as the characteristics of users who publish the advertisements. We find that the traditional advertisement content analysis alone cannot fully explain the effectiveness of social advertisements. More importantly, we discover that in a social advertising paradigm, social influence of a publisher has a strong impact on the number of clicks on the advertisements. Our findings suggest that considering both the advertised contents and the influence of advertising publishers allows better understanding of the social advertisement phenomenon.

COSN'13, October 7–8, 2013, Boston, Massachusetts, USA.

http://dx.doi.org/10.1145/2512938.2512950.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences

General Terms

Human Factors, Measurement

Keywords

Social Media Marketing; Social Network Advertising; Online Advertising; Consumer Behavior; Online Social Networks; Twitter; Content Analysis; User Influence

1. INTRODUCTION

Online social networking services (OSNS) such as Twitter and Facebook have faced a phenomenal growth in popularity. It has been reported that social networking sites have reached 82% of the world's online population, and nearly 19% of the time spent online is now spent on OSNS [9]. With the growing popularity, OSNS are used for various purposes across diverse application domains, one of which is marketing/advertising. OSNS are believed to be powerful marketing environment mainly because they consist of a tremendously large group of users worldwide who actively use the services on a daily basis, not to mention that the cost involved in connecting with these users and maintaining accounts on OSNS is fairly low. In addition, advertisements shared among social connections are expected to be persuasive because people are known to be susceptible to peer influence.

In order to leverage the marketing opportunities offered by OSNS, a new type of business platform named "social media advertising platform" has recently emerged. AdbyMe[1] is one of the many social media advertising platforms currently available on the Web. The goal of social media advertising platforms is to connect businesses, or advertisers, with

[1] AdbyMe. https://adby.me/

(a) Requested Advertisement

(b) Publishing the Advertisement

(c) Published Advertisement on Twitter

Figure 1: Social Advertising Process through AdbyMe

OSNS users who can act as social advertisement publishers. These advertisement publishers, also known as "social publishers", can be any individual who uses popular OSNS including Twitter and Facebook. In particular, OSNS users who wish to participate as publishers can register themselves on AdbyMe and specify which OSNS they will be using to publish advertisements.

Figure 1 describes the social advertising process through AdbyMe. Advertisers who wish to have their advertisements placed on OSNS make a request to AdbyMe, with a webpage specifying their products or services. Let us say COSN organizers want to advertise the upcoming conference and make a request to AdbyMe with a webpage as shown in Figure 1(a). After the request has been made, AdbyMe shows the list of advertisement requests on their system. Any user registered on AdbyMe can browse through this list and find out about the advertisements they can publish on their OSNS. Let us say a user named Twitter User decides to publish an advertisement requested by COSN and spread the information to his/her social network through Twitter. Twitter User is given an option to recreate his/her own advertisement content by typing in the slogan of his/her choice, as shown in Figure 1(b). While AdbyMe provides a default slogan, in most cases publishers choose to create a different slogan that can appeal to other Twitter users. As Figure 1(c) illustrates, a slogan is published on Twitter User's Twitter timeline, with a temporarily generated unique URL link to the advertised webpage shown in Figure 1(a). All the advertisements published through AdbyMe are associated with this clickable URL link. Whenever a friend on Twitter User's social network clicks on the link, Twitter User gets paid a fixed amount of money as a compensation for a successful delivery of an advertisement.

By providing a systematic way for general individuals to become engaged in publishing advertisements on OSNS, social media advertising platforms are attracting a growing number of new advertisers and social publishers. In this paper, we collect and analyze social advertisement data generated by users of AdbyMe and explore factors that affect the advertisement performance measured by clicks on the advertisements. We gathered the data consisting of information on the advertised contents, advertising publishers, and resulting performance of the advertisements in terms of clicks. Click is a direct measure of "attention" of the audience, and being able to capture the attention of an audience is an important preliminary step in advertising, as described in the AIDA - Attention, Interest, Desire, and Action - marketing model [24]. Our study aims to understand the social advertising factors that trigger the attention of the target audience.

Social advertising is unique in that users are engaged in not only publishing the advertisement, but also recreating the advertisement by writing slogans as illustrated in Figure 1(b). Thus, in order to understand the factors influencing social advertising performance, we find it necessary to study both the advertised contents and advertising user. Past studies on online advertisement performance are primarily interested in content-related features that influence the success of advertisements. [14, 19, 25, 17, 21, 22, 15, 13] However, successful advertisements in this new social advertising paradigm are the results of not only effective "contents" but also influential "publishers", because even the same contents can yield different results depending on the influence of the publishers in the social context. Thereby, we look into content-related features as well as user-related features of the advertisements. Note that the terms "users" and "publishers" are used interchangeably for the rest of the paper. The contributions of this paper are as follows:

- We discover through content analysis that in social advertising paradigm, the widely-used promotional techniques do not necessarily bring improvements in advertising performance. This implies that the traditional advertisement content analysis method alone cannot fully explain the advertisement performance in social advertising.
- We verify that the users with high indegree who are presumed to be influential on OSNS do indeed show prominent performance in a social advertising setting. However, users who spawn many retweets are not necessary influential in terms of advertisement performance.
- To the best of our knowledge, our research is one of the first studies to examine the impact of both the content-related and user-related factors on social advertising performance. We perform our analysis on real-world data from a popular social media advertising platform, which realistically represents the real-life social advertising phenomenon.

2. RESEARCH PROBLEMS

In this study, we focus on studying social advertising performance through the analysis of clicks on advertisements. The research problems we address in this work are as follows:

1. What is the overall click distribution of the advertisements?
   We first want to observe how many clicks the advertisements receive through social advertising, in general. We expect the number of clicks to be unevenly distributed, with a few advertisements with high clicks, because skewness in popularity distribution is often observed in many other phenomena on the internet, such as web visits [1, 8] and video viewing activities [7]. By exploring the frequency distribution of clicks of the advertisements that are published through AdbyMe, we want to understand what portion of the population achieve high clicks, and quantify what we mean by "high" number of clicks. By exploring the frequency distribution of clicks on the advertisements that are published through AdbyMe, we also want to understand whether the advertisements on social media advertisement platforms receive similar number of clicks or not.
2. What are the characteristics of the advertisements that drive high number of clicks?
   We want to find out the features associated with the advertisements that attract users to click on the advertisements. The content-related features that we want to focus on are sweepstakes and prize giveaways, celebrity endorsement, sexual appeals, and curiosity components embedded within the advertised slogans, which were proven to be effective promotional strategies by previous studies. In addition to these content-related features, we want to examine the user-related features. In particular, we study whether the influence of a social publisher has an impact on the clicks of an advertisement. We take three different measures of user influence on Twitter: number of followers, retweet likelihood, and post count.

3. RELATED WORK

Our research is related to two bodies of research: researches on analyzing factors on advertising performance, and researches on social advertising.

3.1 Factors on Advertising Performance

Analyzing the factors that influence advertising performance has been an area of interest for many researchers and practitioners. In the past, print-based advertisements and TV commercials were major targets of interests [18, 23, 25]. Starting from the 1990s, when the commercialization of the World Wide Web was actively taking place, researches expanded their reach to online advertisements [12, 20, 11, 26].

The common research objective of these studies is to examine the different types of advertisement appeals. A large number of studies aim to analyze the verbal and visual features associated with the advertisement contents that draw the attention of users [14, 19, 25, 17, 21, 22, 15, 13]. These features include, but are not limited to, sweepstakes and prize giveaways, celebrity endorsement, sexual appeals, and curiosity components. Sweepstakes, prize giveaways, and contests are very common promotional strategies, intended to increase brand awareness by generating enthusiasm among viewers. Studies in the past show that consumer valuation of the advertised products and their response rates can be increased through the use of prizes and contests in advertisements [15, 13]. Many researchers study the effect of having well-known individuals such as pop stars or athletes using their fame to promote brands or products [14, 22, 19]. Such advertisement techniques are referred to as celebrity endorsement, and are found to be effective in arousing interests of the public. Some [25, 21] find that that more than 20% of online advertisements incorporate sexually provocative messages or images, and it increases the initial click response rate of the audience. [17] claims that advertisements that generate curiosity from the audience are shown to be more effective than the advertisements that only provides product information.

Although these promotional strategies were found to be effective by many of previous researches, it has not yet been verified if it holds true in a social advertising context. In a social advertising setting, an advertisement is delivered through a personal connection unlike in other advertising settings, and it is possible for the audience to perceive social advertising contents differently from other types of advertisements. Thus the traditional promotional strategies may result in an unexpected outcome, which makes it worthwhile for us to analyze contents in a social advertising context.

3.2 Social Advertising and Social Influence

It has only been in recent years that OSNS have introduced the new social advertising paradigm. Recent studies explore the social factors associated with the user relationships and how they impact the users' responses to advertisements in social advertising settings. [4] examine the effect of social signals on Facebook users' tendency to further spread information, and find that those who are exposed to social signals are significantly more likely to spread information. Through a large-scale observational study, [3] suggest that probability of adopting a behavior increases with the adopting peers. These studies were performed in the context of information diffusion, where the advertisement performance was measured using the likelihood of activating further information cascades. Although it may well represent the audience's perceived value of advertised content, it does not necessarily reflect how well it catches the attention of users. We use the number of clicks as a measure of advertising performance, which better represents how successful an advertisement is in grabbing the audience's attention.

[2] measures the strength of interpersonal ties between the users and its impact on consumer response to advertisements in terms of clicks, and find that effects of advertisements are greatest for strong ties. The main interests of these previous studies are in studying the egocentric network properties such as relationships among users and tie strengths. Little attention has been paid to understand how the user influence at a macro level, such as indegree, is related to the overall performance of social advertisements [6]. Do users with many friends and followers who are presumed to be influential on OSNS actually perform well in social advertising? Do users who generally spawn many cascades tend to be effective social advertisers? The purpose of this study is to answer these questions that have not yet been addresses by existing researches.

4. DATA DESCRIPTION AND COLLECTION METHODOLOGY

AdbyMe is a social media advertisement platform that serves as a bridge between advertisers and advertisement publishers. AdbyMe was founded in October 2010 in Korea, and rapidly gained popularity over the past three years. As of 2013, AdbyMe consists of 17,260 registered users who participated in publishing in total of 80,612 slogans. The vast majority of AdbyMe users are Koreans, while AdbyMe has recently expanded its business to other countries including Japan and US. Similar social media advertisement platforms around the world include, but are not limited to, Mylikes[2], Ad.ly[3], and SponsoredTweets[4].

We asked AdbyMe administrators to allow us to access their database to gather data from their service. They kindly provided an access account to their database, which allowed us to gather large amounts of data. Additionally, we collected data from Twitter using the Twitter API[5]. We focused on the registered users on AdbyMe who use Twitter as a main channel for publishing advertisements, and collected the tweets posted on their timelines.

We collected the entire set of data generated through AdbyMe in 2012. The data we collected consists of three main parts: advertisement request data, user data, and publication data. Advertisement request data contains information about the advertiser, title and description of the requested advertisement, and the address of the webpage displaying the full advertised content. User data contains information about the social publishers and the OSNS they mainly use for advertising. If a user mainly uses Twitter for example, his/her Twitter screenname is recorded as a part of the user information. Lastly, publication data consists of the advertisement id and the user id when he/she selects an advertisement to post on their OSNS. In addition, the publication data contains the slogan created by the user when publishing it on their OSNS as well as the unique URL assigned to the user for publishing the specified advertisement. Most importantly, publication data also contains the total number of clicks a publication received. Note that click counts are based "unique" click counts, which means that repeated clicks by a single user or duplications from a single IP are only counted as one. In total, we collected data from 3,468 users who contributed in total of 606,707 publications using 79,765 different slogans on 844 advertisement requests.

In addition, we crawled the Twitter profile and timeline of the AdbyMe users who labeled Twitter as their main advertising medium. Data from Twitter was used to measure the influence of a user on OSNS, which will be elaborated with more detail in the next section. For 98% of these users, we collected all the tweets they wrote in 2012. The remaining 2% of the population represent those who generated more than 3,200 tweets in one year. Because Twitter API only allows one to collect up to 3,200 tweets per user, we were only able to partially collect the tweets written in 2012 for these heavy-users. For the heavy users, we make estimations based on these partial data, which will be explained in further detail in the following section. Analysis was performed on anonymized and aggregated data.

5. FEATURE DESCRIPTION AND EXTRACTION METHODOLOGY

In order to understand the factors affecting the performance of social advertisements, we study the content-related features and user-related features of social advertisements. Choice of features is motivated by the findings from previous studies as well as the following questions: Are OSNS users attracted to prize giveaways displayed on advertisements? Are they sensitive to celebrity endorsement? Do curiosity and sex-appealing elements in advertised contents instigate users to click on the advertisement? Does it make a difference who publishes the advertisement? In summary, the content-related features examined in this study include *sweepstakes and prize giveaways*, *celebrity endorsement*, *sexual appeals*, and *curiosity components*.

The user-related features we are interested in are social influence measures of a user within the network which include *indegree*, *retweet likelihood*, and *post count*. A previous study [6] suggests that indegree represents a user's popularity and directly indicates the size of audience. Retweet indicates the ability of a user to engage others in propagating the contents. Post counts indicate the extent to which a user is actively engaged in information sharing activities on OSNS. We expect these user influence measures to have an impact on the success of an advertisement, along with the content-related features. In the following subsections, we describe in detail the methods we used to extract these features from advertised contents and publishers.

[2]http://mylikes.com/
[3]http://ad.ly/
[4]http://sponsoredtweets.com/
[5]https://dev.twitter.com/

SP: Sweepstakes & Prize Giveaways CE: Celebrity Endorsement
CC: Curiosity Components SA: Sexual Appeal

	SP	CE	CC	SA
SP	1**			
CE	-0.043*	1**		
CC	-0.124**	0.162**	1**	
SA	-0.155**	0.051**	0.368**	1**

Table 1: Phi Correlation Coefficient among Content-Related Features (*p<0.1, **p<0.01)

F: Number of Followers R: Retweet Likelihood
P: Tweet Post Count

	F	R	P
F	1**		
R	-0.150**	1**	
P	-0.102**	0.145**	1**

Table 2: Pearson Correlation Coefficient am ong User-Related Features (p<0.01)**

5.1 Content-Related Feature Extraction and Classification

We take a semi-automatic approach to extract content-related features from the advertised contents, or slogans. We first build keyword lists pertaining to each feature for automatic extraction and classification. For example, our list of keywords related to sweepstakes and prize giveaways feature includes the following keywords: "prize", "win", "contest", "sweepstakes", and many others. The keyword lists for sweepstakes and prize giveaways, and sexual appeals were constructed collectively by the members of this research team. For the celebrity endorsement feature, we crawl the names of popular individuals in 2012 from the people search ranking list offered by Nate[6]. Nate is one of the most widely used search engines/portals in Korea, and they provide the ranked list of most searched people names on a daily basis. People on this list are mostly well-known pop stars, athletes, or politicians, and the list well represents those who are at the center of attention during the specific time period.

Based on the keywords on these lists, each advertisement is automatically assigned a binary score for each feature through keyword matching. For instance, if an advertised slogan contains messages on prizes, but does not mention any names of famous figures, it will receive a score of 1 for sweepstakes and prize giveaways feature and 0 for celebrity endorsement feature.

Keyword matching can yield false positives and false negatives due to the limitation in keyword lists as well as ambiguity of natural language. We repeatedly selected a sample of slogans and manually double-checked the results to check the false hit rate, and made sure to improve the keyword lists so that false hit rate is less than 0.05 across all features.

Extracting the curiosity components from the slogans is more complicated. We classified a slogan as the one that contains a curiosity-component if 1) the slogan does not explicitly mention the name of the advertised product, or/and 2) the slogan triggers a viewer's need to obtain further information on the advertised product. Thus, instead of building a keyword list in the aforementioned fashion, we first automatically checked whether the slogans contain the name of the advertised product through keyword matching. Then, for the slogans that do not explicitly mention the product names, we manually checked if the slogan triggers a viewer's desire to know more about the product. Manual classification was based on the majority rule, in order to maintain objectivity in our classification decision.

61% of the entire slogans contains at least one of the four content-related features, and 16% contains more than two content-related features. In order to examine the uniqueness of these features, we perform correlation analysis among the variables. We compute Phi coefficients [10] to measure the pair-wise associations among the binary variables, where 0 indicates no relationship. We observe weak correlations among the features, as shown in table 1. Curiosity component and sexual appeal exhibit a relatively stronger correlation, mainly due to the fact that many of the sexually appealing contents also do tend to trigger curiosity. However, we find it necessary to study both factors, since not every slogan with curiosity components contain sexually appealing contents, and each factor can hold different implications. Note that the correlations are statistically significant with the indicated p-values.

[6]http://www.nate.com/

5.2 User-Related Feature Extraction

The user-related features observed in this work are indegree, retweet likelihood, and post counts. On Twitter, indegree of a user is simply denoted by the number of his/her followers, which can be collected using Twitter API. To measure retweet likelihood and post counts, we have collected the Twitter timeline of AdbyMe users. We counted the number of tweets each user has written in 2012, which denotes the post counts of a user. Instead of simply counting all the tweets a user has written since the creation of his/her Twitter account, we focus on those written in 2012, because it correctly represents how active a user was engaged in using OSNS during the time he/she participated in social advertising. For the heavy users who wrote more than 3,200 tweets in a year, we estimated the post count measure based on the partial data that was collected. In particular, we determined how many tweets they generated on average in a month, or in a week depending on the total quantity, then made an estimated count of the tweets they would have generated in 2012. Retweet likelihood is measured by the portion of total tweets originally written by a user that have spawned further retweets by other users. If a user has written 10 tweets and 7 of them were retweeted by the others, the user's retweet likelihood is 0.7.

Correlation analysis is an important step in validating if it is appropriate to study the effect of each feature as an independent variable. One may, for instance, assume that the number of followers of a user and his/her post count are strongly correlated, questioning the validity of independent variables. We perform correlation analysis among the user-related features to examine the uniqueness of the features.

Table 2 shows the Pearson coefficients for every pair of user-related features. Pearson coefficient ranges from -1 to +1, where ± 1 indicates perfect agreement or disagreement, and 0 indicates no relationship. The features are shown to be weakly correlated. Correlations are statistically significant with the p-value less than 0.01.

In the following section, we examine the overall distribution of advertised slogans and advertising users, in terms of the number of clicks they received. We further perform in-depth analyses of the slogans and users with respect to the extracted features.

6. CLICK DISTRIBUTION OF SOCIAL ADVERTISEMENTS

We begin our study by describing the click distribution of advertisements that were published through AdbyMe in 2012. In total we have data for 79,765 slogans and 3,468 users. We want to understand how many clicks these slogans and users received. Note that through AdbyMe platform, a user can create multiple slogans and same slogans can be used by many different users. Thus, a click distribution of slogans is different from that of users; the former is an outcome of the content-related features regarding the advertisement while the latter is an outcome of the user-related features. We now describe both distributions in detail.

6.1 Click Distribution of Slogans

Figure 2 shows the distribution of clicks on the set of slogans written in 2012. The histogram on the left show frequency, and on the right is a cumulative frequency graph. In these graphical representations, we only consider the slogans that received at least 1 click, which leaves us with 72,738 slogans. Note that the y-axis is in log scale, and as expected, we can observe the unevenly distributed clicks with a heavy-tail. The distribution shows that there are many slogans that receive only a few clicks, and a few slogans that received many clicks.

A slogan with the best advertisement performance received 13,318 clicks. We did not plot those that received more than 750 clicks on the histogram for a better visualization of the long tail effect; the rest not shown on the graph accounts for top 1% of the entire sample. Only 726 out of the 72,738 slogans received more than 750 clicks. Slogans with the top 10% performance rate receive more than 81 clicks, and top 20% receive more than 32 clicks. Approximately half of the slogans receive more than 5 clicks and half gets less than or equal to 5 clicks.

6.2 Click Distribution of Users

We take the same approach to study the click distribution of users. Once again, we only consider the users who received at least 1 click from their publications of advertisements, which are 2,443 users. As illustrated in Figure 3, users also exhibit a long-tailed distribution pattern; many users receive only a few clicks, while a few users receive many clicks. A user with the best advertisement performance received in total of 204,324 clicks. The users with the top 1% advertisement performance receive more than 27,000 clicks. The top 10% receive more than 1,250 clicks, and the top 20% receive more than 350 clicks. Approximately half of the users received more than 7 clicks.

The number of slogans that a user published on Twitter within a year varied from 1 to 297. It also follows a long-tail distribution with the median value of 10 publications. Note that we only take into consideration the users who received at least 1 click on their advertisement.

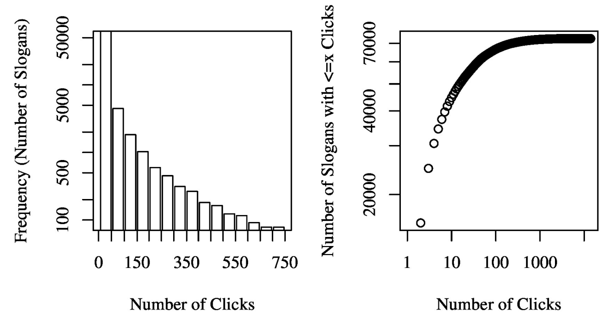

Figure 2: Overall Click Distribution of Slogans (left: frequency, right: cumulative frequency)

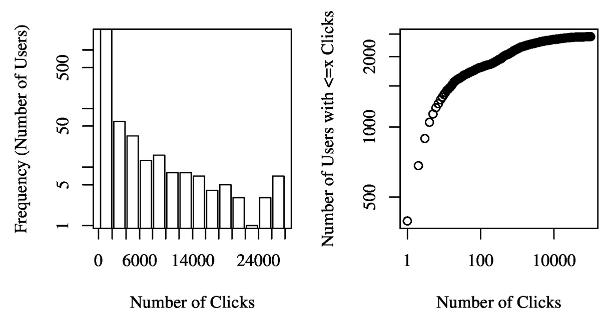

Figure 3: Overall Click Distribution of Users (left: frequency, right: cumulative frequency)

7. WHICH FACTORS IMPACT SOCIAL ADVERTISING PERFORMANCE?

We now pay attention to the content-related features and user-related features that we extracted and study how these features affect the number of clicks on advertisements.

Figure 4 shows the advertisement performance of slogans in the presence and absence of each of the four content-related features: sweepstakes and prize giveaways, celebrity endorsement, sexual appeals, and curiosity components. Because the clicks are non-normally distributed, we used Mann-Whitney U Test [16], a non-parametric statistical hypothesis test, to evaluate the differences in clicks. Note that the y-axis is in log-scale and the outliers are eliminated. The test results showed that there are statistically significant differences in the advertisement performance of slogans. Surprisingly, mentioning of sweepstakes and prize giveaways in the advertisement slogan resulted in a poorer advertising performance ($p<0.01$), on the contrary to a common notion. Celebrity endorsement feature was found to be an effective way of improving the performance ($p<0.05$), although the effect was not dramatic. The features that strongly affected the clicks were sexual appeal and curiosity component; the results showed that the advertising performance of slogans with and without these features differed by an order of magnitude ($p<0.01$). Our results imply that SNS users are interested in acquiring new information, which is consistent with the previous finding that the main purpose of using microblogs is to communicate and to share information. The reason why prizes and giveaway messages resulted in poorer

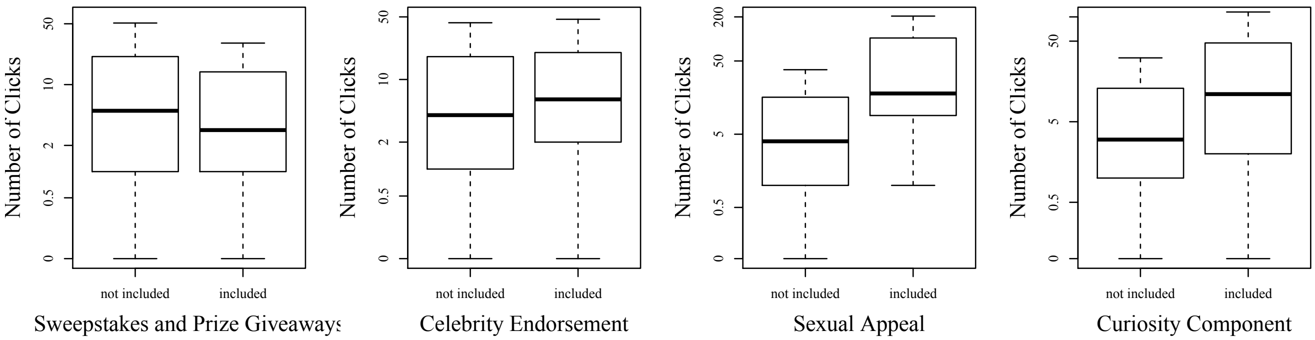

Figure 4: Click Distribution of Slogans with respect to Content-related Features

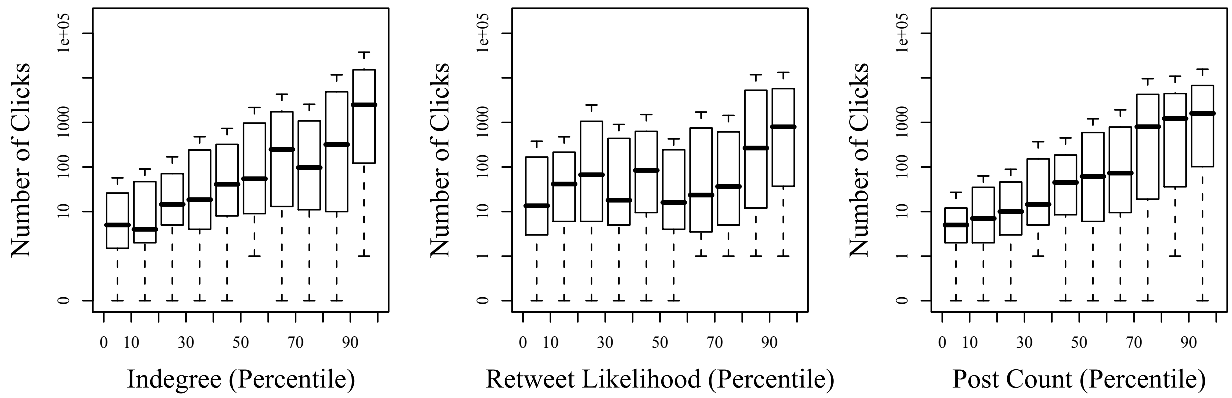

Figure 5: Click Distribution of Users with respect to User-related Features

advertisement performance may be attributed to overexposure to online phishing scams [5], while further analysis is required to understand the exact cause.

Figure 5 shows the advertisement performance of users with respect to their social influence measure on the network. We first sort the users by each measure and determine their percentile ranks, since percentile ranks are a good way of showing relative standing of an individual in a population. We then determine at which percentile range they fall into, and group the users accordingly; those at the higher percentile range indicate the ones with relatively higher indegree, retweet likelihood, and post counts.

We can immediately notice the growth in the number of clicks as the indegree and post counts go up, although slight fluctuations can be observed along the way. Overall, we see the ascending trend in clicks with respect to indegree, which indicates that users with the large target audience are likely to yield many clicks when they post advertisements. This serves as actual evidence that indegree can measure the influence of a user on a social network; we have verified that having a larger audience does indeed lead to a larger response rate in the social advertising setting. We also find that the total post counts of a user, which can represent how actively they are engaged on OSNS, is a strong indicator of their success as social publishers. We see a large gap between the number of clicks received by the top 30% of the users and the rest, in terms of post counts. At this transition, the number of clicks differed by more than an order of magnitude. One may question whether the more number of tweets a user posts on Twitter, the more number of advertisements he/she publishes on Twitter, which can eventually affect the total number of clicks. However, the two are found to be rather weakly correlated ($r^2 = 0.19$). Furthermore, we do not find a strong correlation between the number of total advertisement publications and the total number of clicks received by a user ($r^2 = 0.21$), which indicates that it does not necessarily mean that the more advertisements a user posts, the more likely for them to received many clicks.

For retweet likelihood, we cannot find a noticeable ascending pattern up until the 80th percentile is reached. We detect a great leap in the average, or median, for those with percentile rank greater than 80, but we do not find the evidence showing that the higher the likelihood of diffusion, the higher the advertising performance. It tells us that a user's ability to prompt audience to share his/her contents is not strongly correlated with his/her ability to arouse the interests of the audience to click on the links on their contents. This also implies that the audience's interests in further sharing the content are not consistent with their interests in viewing the content.

8. CHARACTERIZING THE TOP ADVERTISEMENTS

The click distributions of advertisements showed that there are few slogans and users that receive many clicks from the

Rank	No. of Clicks	Slogan (Translated)	SP	CE	SA	CC
1	9,800	Everyone's talking about him on the messenger these days!	N	N	N	Y
2	5,804	A hot girl on the street told me to try "Clinical Pro-plex" - I asked her what it is, and this is what she showed me! lol	N	N	Y	Y
3	5,572	A secret that only Korea wasn't aware of... Gives me chills!	N	N	N	Y
4	5,286	Wow! Vega Racer2 is insane! hope iPhone5 is as good as this...	N	N	N	N
5	5,101	Shocking! A European secret that only South Korea didn't know for 30 years..	N	N	N	Y
6	4,981	never imagined this would be a true story.. brilliant!	N	N	N	Y
7	4,947	Recommended by my friends - "Sometimes Sane"... Really enjoyed this book :)	N	N	N	N
8	4,941	Sora Kang is gorgeous even when she's eating Tacos! :)	N	N	N	N
9	4,900	Vega was waiting for the right moment to compete against iPad3(New iPad)! Amazing spec!!	N	N	Y	Y
10	4,661	How can this sexy dancing queen be the wife of a Seoul city mayor?! I envy her style!	N	N	N	Y

SP: Sweepstakes & Prize Giveaways CE: Celebrity Endorsement CC: Curiosity Components SA: Sexual Appeal

Table 3: Top 10 Slogans with High Advertising Performance

Rank	No. of Clicks	Indegree	Post Count	Retweet Likelihood
1	204,324	284,484 (99.68%)	3,530 (77.67%)	4.84 (63.59%)
2	182,246	331,711 (99.79%)	8,268 (91.22%)	53.60 (97.38%)
3	102,498	243,122 (99.25%)	3,687 (78.83%)	60.19 (98.12%)
4	83,847	145,485 (97.98%)	83,328 (98.94%)	21.10 (87.53)
5	76,447	268,508 (99.47%)	2,880 (71.53%)	31.05 (92.14%)
6	68,317	215,309 (99.04%)	4,404 (83.60%)	41.65 (96.00%)
7	65,896	16,264 (82.85%)	5,566 (87.20%)	4.69 (61.97%)
8	65,652	35,970 (91.16%)	13,128 (94.39%)	62.20 (98.37%)
9	62,052	37,788 (91.91%)	9,252 (92.17%)	79.59 (99.00%)
10	41,395	18,247 (83.49%)	3,500 (77.46%)	1.82 (36.28%)

Table 4: Top 10 Users with High Advertising Performance

audience. These slogans and users can be denoted as "effective" advertisement contents and publishers. In this section, we focus on these effective advertisements at the upper right end of the distribution, and explore the characteristics associated with the effective advertisements. Slogans and users who have been ranked among the top 10 for achieving the highest advertising performance have been selected for a closer look.

Table 3 shows the list of top 10 slogans with the highest overall number of clicks. Surprisingly, none of the top 10 slogans has explicit mentions of prizes or celebrities, which contradicts a common belief that adding prize promotions and celebrities to an advertisement motivates the users to click on the advertisement. The most common content-related feature shared by the top slogans was the curiosity component feature, followed by the sexual-appeal feature. Furthermore, it is interesting to note that more than half of the top 10 slogans never actually mention the name of the product, and leaves audience with no clue of what they were advertising.

Table 4 shows the list of top 10 users who received the most clicks. Users on the list ranked high in terms of indegree; half out of the top 10 was placed within the top 1% of the indegree percentile rank, and 8 of them on the list were ranked within top 10%. The finding indicates that the publishers with high advertising performance are composed of users with a large target audience. We also find that the post counts of the top 10 publishers are relatively high. Nearly half of them were within top 10% post count percentile rank, and all but one user on the top 10 list falls in the upper quartile. Retweet likelihood seems high at a first glance, for many of them falls within the top 10% percentile. However, a closer look at the list reveals that there is a high variability within the top 10 list; 2 of them have 60% percentile rank and one of them falls within the 30% percentile rank.

Slogan Click Percentile	Slogan Count	Slogan by Top 1% Users	Slogan by Top 10% Users
0-10%	67,851	5,111 (7.53%)	22,351(32.94%)
10-20%	2,364	694 (29.35%)	1,580 (66.83%)
20-30%	888	270 (30.40%)	569 (64.07%)
30-40%	470	162 (34.46%)	299 (63.61%)
40-50%	305	124 (40.65%)	202 (66.22%)
50-60%	233	102 (43.77%)	162 (69.52%)
60-70%	196	84 (42.85%)	127(64.79%)
70-80%	158	73 (46.20%)	107(67.72%)
80-90%	139	69 (49.64%)	96(69.06%)
90-100%	133	87 (65.41%)	102(76.69%)

Table 5: Overlap between Top Slogans & Top Users

Observations on the top slogans and users are in accordance with the findings from the previous section, where high indegree and post counts were found to be good indicators of a successful social advertisement publisher. These observations lead to another question: are any relationships between the slogans and users? In other words, how much of top slogans are composed by top users? To measure the degree of top user involvement, we first group the slogans by the click percentile rank then count how many slogans in each group were written by the top users. Table 5 shows what portion of the slogans at each percentile range involves users with top 1% and top 10% advertising performance. The result clearly shows that the higher the percentile rank of slogan clicks, the higher the degree of top user involvement. The growing pattern is even more distinct for the top 1% users. We also find that the top users write a fairly large amount of slogans overall, which relates to the idea that effective social advertisers tend to have high post counts as shown in Figure 5. This indicates that successful advertisers are active in publishing slogans through their social networks. Our findings imply that user influence on the network plays a significant role in social advertising, suggesting that analyzing the content-related features need to be supplemented by user influence analysis.

9. DISCUSSION

The key observations of this paper are twofold: first, social advertising performance is strongly influenced by the characteristic of its publisher. We show that social advertising performance increases as the users' indegree and post counts increase, indicating that those with larger target audiences are more likely to yield successful advertising results than the others, and those who are actively engaged in social activities on OSNS generally perform better in social advertising. This reveals the truly "social" nature of social advertising; the users' social standings and activity level on the network are significant indicators of the success of the advertisement.

Second, content analysis results demonstrate that curiosity component is the main factor associated with the successful social advertisements. A majority of the top slogans did not explicitly state what it was trying to sell, and only contained slight hints or descriptions of the product. Mentions of prizes or celebrity endorsement, which are traditionally believed to be effective promotional techniques, were not shown to produce significant improvements in social advertising performance. Our findings illustrate that social advertisements can effectively capture the attention of users, even without specific references to the advertised products. Being able to capture the attention of an audience is an important preliminary step in advertising; the AIDA marketing model explains that four stages in advertising include Attention, Interest, Desire, and Action [24]. Our study on clicks on advertisements is a direct measure of "Attention" of the audience, and it tells us that social advertising can help businesses in taking the first important step in achieving their marketing goal.

We have studied the social advertising performance in terms of "attention", and another interesting issue worth exploring is on understanding the audience's "attitude" towards social advertisements. Marketing reports [5] show that too much advertisements can overwhelm the audience and reduce marketing effectiveness. This leads to the idea that users who post too many advertisements on their social network may be perceived as spammers. Interestingly, our observation from Table 5 showed that the top users were overall actively engaged in publishing top slogans. A possible explanation behind why these active users show high advertising performance may be due to the fact that they tend to post many non-advertisements as well. An in-depth analysis of the relationship between tweet post counts and advertisement post counts and how they affect the advertising performance would yield useful insights in understanding the perceived attitude towards the advertisements.

Our study is based on a single dataset from AdbyMe, in which majority of the users are Koreans who created slogans in Korean language. Thereby, one must note that, to a certain extent, there can exist cultural bias within our analyses results. For example, spam advertisements in Korea are oftentimes in the form of flashy animated banners with sweepstakes messages, which may have affected the advertisement performance. Although it is true that cultural factors cannot be ignored, we must also take into account that our study is reproducible in different cultural contexts, since existing social advertising platforms worldwide are comparable in their functions and structures. Performing cross-cultural comparisons and studying the cultural differences in perception of social advertisements would be an interesting research direction for the future.

We would also like to further examine other features not considered in this work. As a part of our study, we observed that half of the top 10 slogans were about the newly released movies. We want to study the effectiveness of social advertising across different product categories, which can lead to a better understanding of the nature of social adverting and OSNS usage. It would also be interesting to observe if there is any relationship between a user's area of interests or expertise and the contents they publish through social media advertising platform. Furthermore, our analysis on the content-related and user-related features opens up the possibility of generating models that can predict the success of an advertisement or building recommendation algorithms in the context of social advertising.

10. CONCLUSION

In this research, we study the factors that may affect the number of clicks on advertisements being published on online social networking services. We collect real-world data from a popular social media advertising platform, and perform content analysis as well as user influence analysis on the advertisement data. Surprisingly, some of the promotional techniques widely used in traditional advertising media were found to be not as effective in a social advertising setting. We also find that social advertising performance increases as the user's indegree and level of activity increase. This implies that user influence on the network plays a significant role in social advertising, suggesting that both the advertised contents and the advertising publisher need to be considered to understand the social advertising phenomenon.

11. ACKNOWLEDGEMENTS

This research was supported in part by the National Research Foundation of Korea grant funded by the Korean government (MSIP) (No. NRF-2009-0081365), and in part by WCU (World Class University) program under the National Research Foundation of Korea and funded byÂăthe Ministry of Eduation, Science and Technology of Korea (No. R31-30007).

12. REFERENCES

[1] K. Ali and M. Scarr. Robust methodologies for modeling web click distributions. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 511–520, New York, NY, USA, 2007. ACM.

[2] E. Bakshy, D. Eckles, R. Yan, and I. Rosenn. Social influence in social advertising: evidence from field experiments. In *ACM Conference on Electronic Commerce*, 2012.

[3] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 65–74, New York, NY, USA, 2011. ACM.

[4] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 519–528, New York, NY, USA, 2012. ACM.

[5] BurstMedia. The perils of ad clutter, December 2008.

[6] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*, 2010.

[7] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, IMC '07, pages 1–14, New York, NY, USA, 2007. ACM.

[8] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Reviews*, June 2007.

[9] ComScore. It's a social world. top 10 need-to-knows about social networking and where it's headed. Technical report, ComScore, 2011.

[10] H. Cramer. *Mathematical Methods of Statistics*. Princeton landmarks in mathematics and physics. Princeton University Press, 1999.

[11] D. R. Fortin and R. R. Dholakia. Interactivity and vividness effects on social presence and involvement with a web-based advertisement. *Journal of Business Research*, 58(3):387–396, 2005.

[12] C. F. Hofacker and J. Murphy. World wide web banner advertisement copy testing. *European Journal of Marketing*, 32(7-8), 1998.

[13] A. Kalra and M. Shi. Consumer Value-Maximizing Sweepstakes and Contests. *Journal of Marketing Research*, 47:287–300, 2010.

[14] M. A. Kamins, M. J. Brand, S. A. Hoeke, and J. C. Moe. Two-sided versus one-sided celebrity endorsements: The impact on advertising effectiveness and credibility. *Journal of Advertising*, 18(2):pp. 4–10, 1989.

[15] D. Kesmodel. Brand marketers return to the web, driving new growth in display ads. *The Wall Street Journal*, May 2006. May 10.

[16] H. B. Mann and W. D. R. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1):50–60, 1947.

[17] S. Menon and D. Soman. Managing the power of curiosity for effective Web advertising strategies. *Journal of Advertising*, 31(3):1–14, 2002.

[18] A. A. Mitchell. The effect of verbal and visual components of advertisements on brand attitudes and attitude toward the advertisement. *Journal of Consumer Research*, 13(1):12, 1986.

[19] R. Ohanian. The impact of celebrity spokespersonsÊij perceived image on consumers' intention to purchase. *Journal of Advertising Research*, 31(1):46–54, 1991.

[20] J. Parsons, K. Gallagher, and K. D. Foster. Messages in the medium: An experimental investigation of web advertising effectiveness and attitudes toward web content. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*.

[21] T. Reichert. Sex in advertising research: a review of content, effects, and functions of sexual information in consumer advertising. *Annual Review of Sex Research*, 13:241–273, 2002.

[22] D. Silvera and B. Austad. Factors predicting the effectiveness of celebrity endorsement advertisements. *European Journal of Marketing*, 38(11/12):1509–1526, 2004.

[23] S. N. Singh and C. A. Cole. Advertising copy testing in print media. *Current Issues and Research in Advertising*, 11(1-2):215–284, 1988.

[24] E. K. Strong Jr. Theories of selling. *Journal of applied psychology*, 9(1):75, 1925.

[25] R. Vezina and O. Paul. Provocation in advertising: A conceptualization and an empirical assessment. *International Journal of Research in Marketing*, 14(2):177 – 192, 1997.

[26] C. Y. Yoo and K. Kim. Processing of animation in online banner advertising: The roles of cognitive and emotional responses. *Journal of Interactive Marketing*, 19(4):18 – 34, 2005.