ABSTRACT
Advertisers worldwide spent $24 billion to reach consumers on social media in 2015. While such a new way of advertising has successfully turn the social media into generous profits, the strategies behind it is still mystery to users, advertisers and many businesses. In this paper, we uncover the underlying mechanisms of the social media advertising. Specifically, we compare them with the old-school advertising strategies that have been widely used since the early 1900s. The advertising on the high tech does not achieve beyond the wisdom of the elders but run faster at a unprecedented scale. We define a series of novel features from the strategies we discover. We further propose a classification method called SocAdDet based on the SVMs. Experiments on a real social dataset show that SocAdDet can accurately identify different advertising strategies and detect the social promoters. The high accuracy demonstrates that the social media advertising is stronger but not smarter.

Categories and Subject Descriptors
[Information systems]: Social networks; [Security and privacy]: Social aspects of security and privacy

Keywords
Advertising strategy; Social botnet; Synchronized behavior; Classification

1. INTRODUCTION
Advertisers worldwide spent $23.68 billion on paid media to reach consumers on social networks in 2015, according to new figures from eMarketer1, a 33.5% increase from 2014. Experts have put their marketing skills to turn the social media into generous profits. Their main goal in mind was to help small to medium or large sized businesses (i.e., threads in social media) succeed with the advertising strategies2,3. In this paper, we aim at uncovering the underlying mechanisms of the social media advertising. Specifically, we compare them with the old-school advertising strategies that have been widely used since the early 1900s. Furthermore, we propose novel features from our understanding of the strategies and develop an effective method to answer the following questions:

- Q1: Can we identify different marketing strategies that we will numerate in the comparison?
- Q2: Can we accurately detect the social botnet advertisers that are set up by the marketers in the network?

Now we give the terms and their defintions. Table 1 lists the comparision of terms for social media advertising and old-school advertising. An advertising thread is defined as follows.

Figure 1 illustrates the details of a social media advertising thread. The thread of diffusing a message in social media is like selling a product in real life, for example, the “product” is Xiaomi’s an-

---

1Digital Marketing Research & Insights: www.emarketer.com

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CyberSafety’16, October 28 2016, Indianapolis, IN, USA
© 2016 ACM. ISBN 978-1-4503-4650-4/16/10...$15.00
DOI: http://dx.doi.org/10.1145/3002137.3002143
Table 1: Comparing the terms of “resources” (users and contexts) in social media advertising and old-school advertising.

<table>
<thead>
<tr>
<th>Term in social media</th>
<th>Term in old-school advertising</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advertising thread</td>
<td>Promoting a product</td>
</tr>
<tr>
<td>Root user</td>
<td>Company</td>
</tr>
<tr>
<td>Root tweet</td>
<td>Product or service</td>
</tr>
<tr>
<td>Collaborated user</td>
<td>Advertising team</td>
</tr>
<tr>
<td>Infected user</td>
<td>Consumers who purchase the product</td>
</tr>
<tr>
<td>Device/IP address</td>
<td>Shop or street</td>
</tr>
<tr>
<td>Comment in retweet</td>
<td>Advertising words (slogan, etc.)</td>
</tr>
</tbody>
</table>

Figure 2: Celebrity branding/advertising. It has become the most popular advertising method in real life. In social media, a message from a celebrity account who has millions of followers can infect thousands of users. As the old-school branding, the influence of the celebrity is limited to the size of his/her fan community. Note that over 95% of the infected users are the followers.

In this section, we first provide case study and statistics to spot how the old-school strategies are used in social media. Dataset. It was crawled from Tencent Weibo for 43 days (Nov. 9 to Dec. 21, 2011), which has 19.3 million threads and 33.4 million users. The average number of infected users in the threads is 2.17.

2. ADVERTISING STRATEGIES AND FEATURES: COMPARING SOCIAL MEDIA AND TRADITIONAL ADVERTISING

In this section, we first provide case study and statistics to spot how traditional advertising strategies have been applied, as well as the “synchrony” strategy in social media. Finally, we define two problems, marketing strategy identification and botnet advertiser detection, and further define a series of novel features.

2.1 Spotting Traditional Strategies in Social Media Advertising

We introduce the case study and statistics of advertising threads to compare how the old-school strategies are used in social media.

Strategy S1 in old fashion: Celebrity branding is a type of advertising in which a celebrity becomes a brand ambassador and uses his or her status in society to promote or endorse a product, service or charity.

Strategy S1 in social media: If the root user is a celebrity who has numerous followers, the number of infected users is often larger than the threads in which the root users are ordinary accounts. Figure 2 shows two typical examples. Statistical results are as follows.

- Root user @caikangyong has 22.7M followers. The thread infected 5,138 users (97.6% from followers) and 3,246 devices (97.6% from followers). The number of retweets is 5,214 (97.4% from followers).
- Root user @kaifulee has 22.5M followers. The thread infected 6,346 users (95.7% from followers) and 5,138 devices (97.6% from followers). The number of retweets is 6,425 (95.3% from followers).

Both @kaifulee and @caikangyong had more than 22 million followers. The messages were “iPhone has exceeded Blackberry in smart phone sales...” and a poem to future lover. The numbers of infected users are very large, 6,345 and 5,138, however, over 95% of them are the followers of the root users. The ads are not able to diffuse over communities.

2.1.2 S2: Collaborative advertising

Strategy S2 in old fashion: Collaborative advertising is the process of sharing the same goal to increase brand and influence. For example, recruiting several famous basketball players can promote the product in different channels (i.e., the fan groups of the stars).

Strategy S2 in social media: In social media, no matter how
Figure 4: Gift advertising. The strategy of old-school gift advertising is to attract consumers to a new business with free gift cards. In social media, to increase the popularity of advertising content, the users add irrelevant but attractive content as gift cards when retweeting the message. We spot unexpectedly high frequency of the retweets’ length at the limit (140 characters). Only 4.5% infected users are the root user’s followers.

Figure 5: Multi-level marketing. The old-school multi-level marketing is to recruit sales force called “downline” to provide multiple levels of compensation. There are companies (e.g., Qunaer, Inc.) that manipulate botnets to mention several accounts in their retweets. We spot that the users are frequently mentioned in the diffusion network: “1036@” means the account mentions 1,036 users, and “@279” means the user is mentioned for 279 times.

2.1.3 S3: Gift advertising

Strategy S3 in old fashion: If you register an account or purchase a product, you will get a $100 gift card. Such a strategic behavior has been used to attract customers to a new business since the Mobil Oil Company introduced the first retail gift card in 1995. You definitely cannot purchase for gas from Mobil with an Amazon gift card but you can use the card to purchase for life goods which makes you satisfied. The card is irrelevant but attractive.

Strategy S3 in social media: When the root user’s followers retweet the message, they add irrelevant but attractive content (e.g., about “failure”, “life”, “dream”) as a “gift card” to replace the part of the original tweet text. Thus, the message can be widely diffused over the online network.

In Figure 4, the message can be widely diffused over the online network, infecting 13 thousand users, while only 4.5% follow the root user. Statistical results show that the root user has 63,996 followers. The thread infected 12,977 users (4.5% from followers) and 7,911 devices (6.2% from followers). The number of retweets is 14,904 (7.6% from followers).

2.1.4 S4: Multi-level marketing

Strategy S4 in old fashion: Multi-level marketing is a marketing strategy in which the sales force is compensated not only for sales they generate, but also for the sales of people that they recruit. This recruited sales force (referred to “downline”) provide multiple levels of compensation.

Strategy S4 in social media: the company manipulates their accounts to mention legitimate users with “@XXX”. The mentioned users become the “downline” to diffuse the message in multiple levels of the network. The message often refers to an activity of continue mentioning friends.

We spot that several accounts were manipulated to mention their friends with “@XXX” in the ads. The mentioned users become the “downline” to diffuse the message in multiple levels of the network. For example in Figure 5, the Qunaer Inc. (a TripAdvisor-like company in China) asked users to forward the microblogging and mention at least five friends, and thus they have a chance to win an ipad. This thread creates a shape of diffusion that looks quite similar as the traditional multi-level marketing network. Statistical analysis shows that the root user has 113,026 followers. The thread infected 1,060 users (69.2% from followers) and 1,013 devices (62.2% from followers). The number of retweets is 14,094 (7.6% from followers).

2.2 S5: Novel but Naïve Synchrony Strategy in Social Media Advertising

Besides the strategy transfer, the social media marketers generate multiple types of scripts to automatically post messages fast at a large scale [12, 10, 7, 5, 6]. However, scripts have to show some...
S5-1: Comment synchrony. A group of users retweet the same message with the same group of comments [4]. 26 different comments such as “share to my friends” and “support the good activity” were attached to the ad about Samsung Galaxy Note. We spot each comment was adopted by more than 100 users.

S5-2: URL synchrony. The accounts frequently retweet the same URLs (e.g., porn websites) [11]. We spot that a group of users retweet the porn message with the same link for 38,432 times.

S5-3: Time synchrony. One of the naïve settings for the botnets to inflate the popularity is to roll poll a group of them to retweet every $x$ seconds [1, 9, 4]. We can see the most common time interval values of two adjacent posts were $x \in \{20, 40, 60\}$ seconds.

S5-4: Device synchrony. @zuomin2582 manipulates 676 accounts among which all are his followers, 42 devices most of which are on 114.86.xx.xxx in Shanghai city. The root tweet was forwarded for 9,504 times, but this promotion has zero effect on legitimate users.

Note that over 98% of the infected users in most of the cases are the root user’s followers, which indicates that very few legitimate users outside the “community” consume the messages.

Important statistics of the threads in Figure 6 are as follows.

- Comment synchrony: (Galaxy Note ads) The botnets frequently retweet with similar phrases. The root user has 8,573 followers. The thread infected 3,059 users (98.4% from followers) and 42 devices (100% from followers). The number of retweets is 9,777 (99.5% from followers).

- URL synchrony: (Porn URL) The botnets frequently retweet the same URL. The thread infected 38,122 users (100% from followers) and 18,008 devices (99.989% from followers). The number of retweets is 38,432 (99.995% from followers).

- Time synchrony: (Galaxy Note ads) The botnets operate in lockstep with several fixed time intervals. The root user has 1,362 followers. The thread infected 598 users (98.0% from followers) and 12 devices (100% from followers). The number of retweets is 4,202 (99.6% from followers).

- Device synchrony: (Galaxy Note ads) The botnets operate on the same group of devices in Shanghai. The root user has 571 followers. The thread infected 676 users (100% from followers) and 42 devices (100% from followers). The number of retweets is 9,504 (100% from followers).

2.3 Problem Definition and Feature Selection

Now we define the two problems about social media advertising.

Problem 1 (Marketing Strategy Identification). Given a thread, identify the major strategy that was used in the thread from the set of strategies (S1 to S5).

Problem 2 (Botnet Advertiser Detection). Given users and threads in the social media, detect botnet advertisers from the users. A botnet advertiser is a root user who adopted the “synchrony” strategy to automatically post the retweets.

The first is a multiclass classification problem and the second is a binary classification. Table 2 lists the features for threads and users.

Comment, mentioning, URL and device features. The outliers of the big average value and small variance in Figure 7a-7b show the comment and mention synchrony. Figure 7c shows there are a group of retweets that have URLs, and in Figure 7d, we spot retweets in some threads were posted from few devices.

Lockstep behavioral features. The heatmap in Figure 8 has a logarithmic color scale that can show values are distributed in power law along the spectrum. Suppose there are three retweets $a$, $b$, $c$ from the same user in a thread. $\Delta t_1$ is the time interval between $a$ and $b$ and $\Delta t_2$ is the time interval between $b$ and $c$. We spot red clusters at the integer combinations such as $(30s, 30s), (30s, 60s)$ and $(45s, 45s)$, which indicates lockstep as the botnets behave.

Message synchronicity features. We apply CatchSync [2] on the user-message bipartite graph and spot the users’ synchronous be-
We propose a supervised learning method for identifying marketing strategies. For example, @aasijwczj made 799 retweets but only on two root tweets, so he has high synchronicity; @guitarblue posted 207 retweets of diverse content as triangles in the feature space, and thus his synchronicity is small. 

**SocAdDet method.** We propose a supervised learning method for the classification, *Social Advertising Detective*, which adopts the above features and uses SVMs as the training model. Next we demonstrate the effectiveness of our proposed novel features.

### 3. EXPERIMENTS

We visualize the threads whose popularity is over 1.000 and randomly label 2.000 of them with the codebook of advertising strategies. We also label 1.000 users as botnet advertisers or legitimate users by checking the collection of their tweets. The labeling was conducted by 5 student volunteers and we take the majority. The classification performances are evaluated with the accuracy metric.

#### 3.1 Marketing Strategy Identification

Table 3 shows that taking the full advantages of our feature collection can perform a 0.889 accuracy in identifying the marketing strategy from the 5 classes.

<table>
<thead>
<tr>
<th>Advertising strategy</th>
<th>Feature definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1: Celebrity branding</td>
<td>F1: the number of followers of the root users</td>
</tr>
<tr>
<td>S2: Collaborative advertising</td>
<td>F2((k)): the number of infected users who have more than (k) followers F3: the number of infected users who share the same device of the root user</td>
</tr>
<tr>
<td>S3: Gift advertising</td>
<td>F4: the largest frequency of the length of the comments F5: the most frequent length of the comments</td>
</tr>
<tr>
<td>S4: Multi-level marketing</td>
<td>F6((k)): the number of users who were mentioned more than (k) times F7((k)): the number of users who mentioned more than (k) users in total F8-F9: the average value (variance) of the number of mentions in the comments</td>
</tr>
<tr>
<td>S5-1: Comment synchrony</td>
<td>F10-F11: the message synchronicity (normality) of the user [2] F12-F13: the average value (variance) of the length of the retweet comments by the user</td>
</tr>
<tr>
<td>S5-2: URL synchrony</td>
<td>F14: the percentage of comments that have at least one URL by the user F15-F16: the average value (variance) of the number of URLs in the comments by the user</td>
</tr>
<tr>
<td>S5-3: Time synchrony</td>
<td>F17: the most frequent time interval (\Delta_t) between two retweets in a thread by the user F18-F19: the average value of the number of retweets (the time period) in a thread by the user</td>
</tr>
<tr>
<td>S5-4: Device synchrony</td>
<td>F20: if the user operates on the most frequent device in a thread</td>
</tr>
</tbody>
</table>

#### 3.2 Botnet Advertiser Detection

Table 4 shows that our proposed SocAdDet can catch the botnet advertisers more accurately than CatchSync \[2\]. The 0.923 accuracy demonstrates that the social marketers were not taking smarter strategies than the traditions. Social marketers should learn from how old-school advertisers work as their customers’ friends instead of explicitly being a promoter.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CatchSync [2] (F10-F11)</td>
<td>0.725</td>
</tr>
<tr>
<td>Comment synchrony (F10-F13)</td>
<td>0.796</td>
</tr>
<tr>
<td>URL synchrony (F14-F16)</td>
<td>0.725</td>
</tr>
<tr>
<td>Time synchrony (F17-F19)</td>
<td>0.831</td>
</tr>
<tr>
<td>Device synchrony (F20)</td>
<td>0.695</td>
</tr>
<tr>
<td>SocAdDet (F10-F20)</td>
<td>0.923</td>
</tr>
</tbody>
</table>

### 5. REFERENCES


