Catching Social Media Advertisers with Strategy Analysis

Meng Jiang
University of Illinois at Urbana-Champaign
201 N Goodwin Ave, Urbana, IL 61801, US
mjiang89@illinois.edu

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 eMarketer[1], 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 strategies* [8, 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.

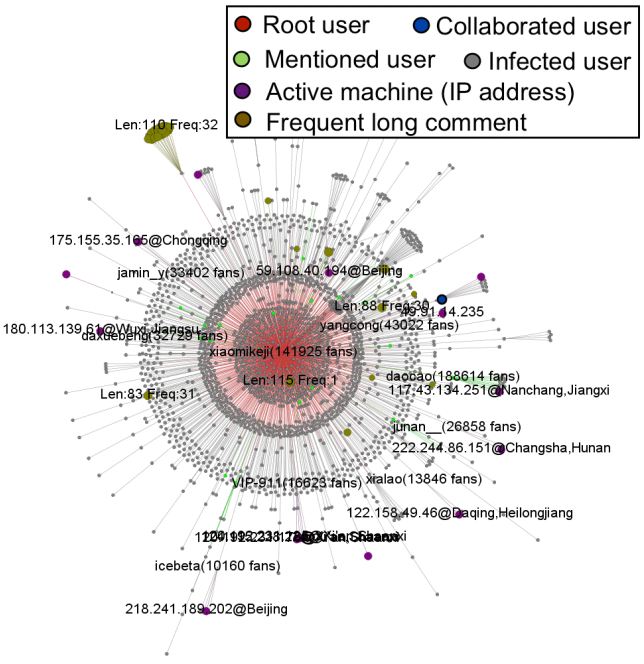


Figure 1: *A social media advertising thread and its "resources".* We spot different contexts and roles of users in the social media advertising. Good strategies lead to high popularity on the network.

DEFINITION 1 (SOCIAL MEDIA ADVERTISING THREAD). *A social media advertising thread is the process that starts from a user ("root user") posting a message ("root tweet"), controls one or several accounts to retweet/broadcast the message with some strategies, and generates a high popularity in the network.*

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-

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

CyberSafety'16, October 28 2016, Indianapolis, IN, USA

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.

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

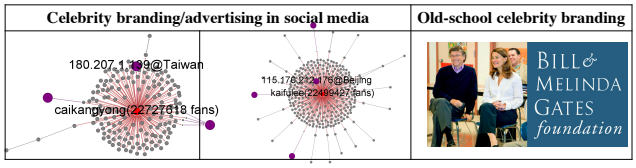


Figure 2: *Celebrity branding/advertising.* It has become the most popular advertising method in real life. In social media, a message from a celebrity account who has millions of followers can infect thousands of users. As the old-school branding, the influence of the celebrity is limited to the size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

nouncement to launch a new device. The root user *Xiaomi Inc.* (red) is like the big seller, a "celebrity" with 142K followers. The collaborated user (blue) is like the friends of the seller who are also celebrities. The infected user (grey) who retweets the message is like a customer who purchases the product. Mentioning a friend (green) using "@" by an infected user is like building a multi-level marketing relationship. The device or IP address (purple) is like the shop or street where the company sells the product. The comment (brown) includes the advertising words of the promotion. Only devices and comments that frequently appear are displayed.

Contribution 1. We compare the social media advertising and traditional advertising: we spot four major old-school strategies (celebrity branding, collaborative advertising, gift advertising and multi-level marketing) in the social media environment, and highlight the main difference of the high tech. Botnets can be easily set up for promoting the messages fast at a unprecedented scale. However, they do not achieve beyond the wisdom of the elders.

Contribution 2. With the deep understanding of the advertising strategies, we propose a number of novel features and further propose a classification method called SocAdDet based on the SVMs. Experiments on a real 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.

The rest of this paper is organized as follows. Section 2 compares the strategies by social media advertisers and traditional advertisers. Section 3 introduces our experimental design and results. Section 4 concludes the paper.

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.

Figure 3: *Collaborative advertising.* The old-school advertising recruits celebrities to promote a brand, product or service. The celebrity endorsements can cover multiple groups of their fans. In social media, if several big nodes of millions of followers share the same content or URL, the message can be adopted by multiple communities. In the above example, only 9.0% of the infected population are the root user's followers.

2.1 Spotting Traditional Strategies in Social Media Advertising

We introduce the case study[2] and statistics of advertising threads to compare 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.

2.1.1 S1: Celebrity branding/advertising

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. Statisital 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 5,357 devices (97.3% from followers). The number of retweets is 6,425 (95.3% from followers).

Both @*kaifulee*[3] and @*caikangyong*[4] 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

[2]The tweets were translated with *Google Translator*.
[3]A Taiwanese IT venture capitalist and computer scientist.
[4]A writer and television host in Taiwan.

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.

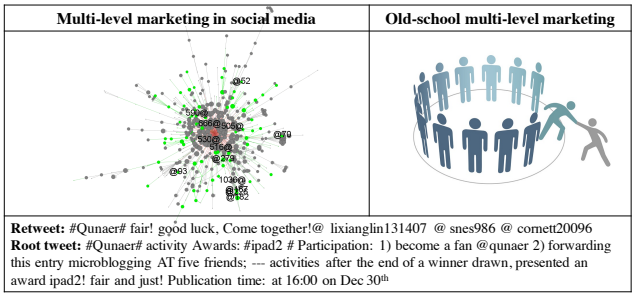


Figure 5: *Multi-level marketing.* The old-school multi-level marketing is to recruit sales force called "downline" to provide multiple levels of compensation. There are companies (e.g., *Qunaer, Inc.*) that manipulate botnets to mention several accounts in their retweets. We spot that the users are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

popular an account is, only his followers can directly receive his message. The more big nodes in the network retweet/share the content and URL, the larger number of the infected followers the message has.

Collaborative advertising is the process of sharing the same goal to increase brand and influence. For example in Figure 3, the more big nodes who have many followers in the network share the content and URL, the larger number of the infected users the message has. Therefore, only 9% of the infected users are the root user's followers. Statistical analysis shows that the root user has 762,289 followers. The thread infected 21,807 users (9.0% from followers) and 15,446 devices (7.8% from followers). The number of retweets is 23,625 (9.1% from followers).

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 inrelevant 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

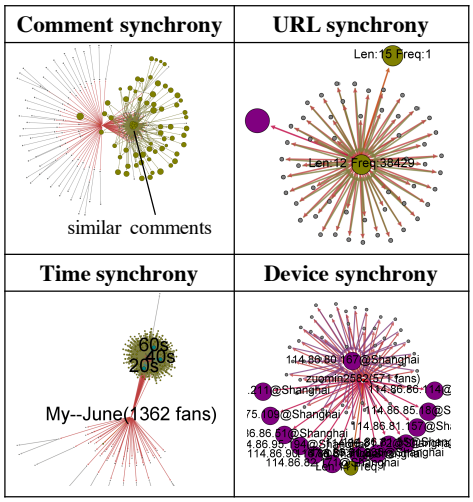


Figure 6: *Synchrony for "easy" social media advertising.* Botnet accounts were set up to post messages (1) of the same content of the comments, (2) of the same URL in the comments, (3) in lockstep with the same time intervals, (4) on the same devices. We spot the comment, URL, time and device synchrony and generate corresponding features to capture the advertising behavior.

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. Statistial resutls 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

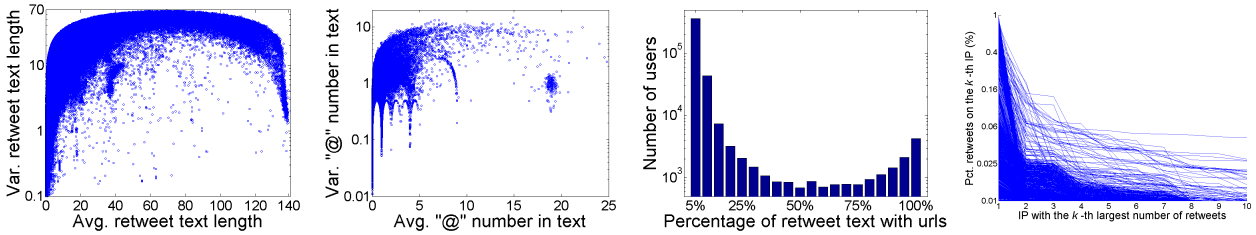
Strategy S4 in old fashion: Multi-level marketing is a marketing strategy in which the sales force is compensated not only for sales they generate, but also for the sales of people that they recruit. This recruited sales force (referred to "downline") provide multiple levels of compensation.

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 (45.2% from followers). The number of retweets is 14,282 (91.5% 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



(a) Comment features (b) Mentioning features (c) URL features (d) Device features

Figure 7: *Distributions of text-based and device features.* (a) Each dot is a thread. A thread with long but similar comments may be an ad. (b) Each dot is a thread. A thread with too many mentions in the comment may be an ad. (c) If all the retweets of a user have at least one URL, the user may be an advertiser. (d) Each curve is a thread. If the thread has many retweets from few devices, it may be an advertisement.

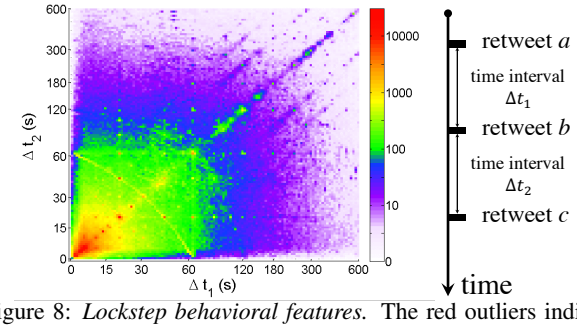
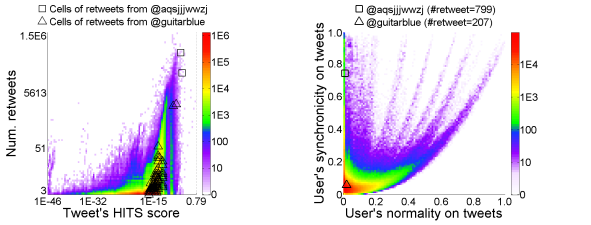


Figure 8: *Lockstep behavioral features.* The red outliers indicate the lockstep patterns in the time intervals of retweeting.



(a) Root tweets in the degree-HITS feature space (b) Synchronicity and normality of users in the left feature space

Figure 9: Message synchronicity features.

pattern. Specifically, the botnet accounts show synchronous behaviors. The major types of synchrony are as follows (see Figure 6).

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 18 devices (61.1% 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, Δt_1 is the time interval between a and b and Δt_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-

Table 2: Advertising features that we extract from the understanding of advertising strategies.

Advertising strategy	Feature definition
S1: Celebrity branding	F1: the number of followers of the root users
S2: Collaborative advertising	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
S3: Gift advertising	F4: the largest frequency of the length of the comments F5: the most frequent length of the comments
S4: Multi-level marketing	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
S5-1: Comment synchrony	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
S5-2: URL synchrony	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
S5-3: Time synchrony	F17: the most frequent time interval Δ_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
S5-4: Device synchrony	F20: if the user operates on the most frequent device in a thread

Table 3: Integrating the feature collection performs well in identifying the marketing strategy from the 5 classes.

Method	Parameters	Accuracy
Random	-	0.200
S1 (F1)	-	0.447
S2 (F2-F3)	F2(1000)	0.326
S3 (F4-F5)	-	0.352
S4 (F6-F9)	F6(5), F7(10)	0.257
S5 (F10-F20)	-	0.525
SocAdDet (F1-F20)	F2(100), F6(5), F7(10)	0.852
	F2(1000), F6(2), F7(10)	0.855
	F2(1000), F6(5), F7(5)	0.867
	F2(10000), F6(10), F7(20)	0.776
	F2(1000), F6(5), F7(10)	**0.889**

havioral pattern. For example, *@aqsjjjwwzj* 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 for advertising threads from the five classes.

From the results we have observations of the performance comparisions as follows.

- Taking the series of features from S5 is more accurate than only taking the traditional strategy into account. It demonstrates that there has been a large amount of botnets using the "novel" strategy of synchronity to promote their message.
- For the traditional strategies, the best settings of featurs are (1) the number of infected users who have more than $k = 1,000$ followers, (2) the number of users who were mentioned more than $k = 5$ times, and (3) the number of users who mentioned more than $k = 20$ users in total.

Table 4: SocAdDec outperforms CatchSync in detecting botnets.

Method	Accuracy
CatchSync [2] (F10-F11)	0.725
Comment synchrony (F10-F13)	0.796
URL synchrony (F14-F16)	0.725
Time synchrony (F17-F19)	0.831
Device synchrony (F20)	0.645
SocAdDet (F10-F20)	**0.923**

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.

4. CONCLUSION

In this paper, we compared the strategies of the social media advertising and traditional advertising. We defined a series of features and proposed SocAdDet to identify the strategy and advertisers in social media. The experimental results show that the social advertising strategies are faster, stronger but not smarter than the human intelligence of a long history.

5. REFERENCES

[1] O. Dabeer, P. Mehendale, A. Karnik, and A. Saroop. Timing tweets to increase effectiveness of information campaigns. In *ICWSM*. Citeseer, 2011.

[2] M. Jiang, P. Cui, A. Beutel, C. Faloutsos, and S. Yang. Catching synchronized behaviors in large networks: A graph mining approach. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2016.

[3] M. Jiang, P. Cui, and F. Christos. Suspicious behavior detection: Current trends and future directions. *IEEE Intelligent Systems*, 31(1):31–39, 2016.

[4] M. Jiang, C. Faloutsos, and J. Han. Catchtartan: Representing and summarizing dynamic multicontextual behaviors. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2016.

[5] S. Kumar, F. Spezzano, and V. Subrahmanian. Vews: A wikipedia vandal early warning system. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.

[6] S. Kumar, R. West, and J. Leskovec. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *Proceedings of the 25th International World Wide Web Conference*, 2016.

[7] H. Li, A. Mukherjee, B. Liu, R. Kornfield, and S. Emery. Detecting campaign promoters on twitter using markov random fields. In *2014 IEEE International Conference on Data Mining*, pages 290–299. IEEE, 2014.

[8] Y.-M. Li and Y.-L. Shiu. A diffusion mechanism for social advertising over microblogs. *Decision Support Systems*, 54(1):9–22, 2012.

[9] G. Liu, Y. Fu, T. Xu, H. Xiong, and G. Chen. Discovering temporal retweeting patterns for social media marketing campaigns. In *2014 IEEE International Conference on Data Mining*, pages 905–910. IEEE, 2014.

[10] C.-T. Lu, H.-H. Shuai, and P. S. Yu. Identifying your customers in social networks. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 391–400. ACM, 2014.

[11] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599. Association for Computational Linguistics, 2011.

[12] X. Zhang, S. Zhu, and W. Liang. Detecting spam and promoting campaigns in the twitter social network. In *2012 IEEE 12th International Conference on Data Mining*, pages 1194–1199. IEEE, 2012.