



Catching Social Media Advertisers with Strategy Analysis

Meng Jiang

University of Illinois at Urbana-Champaign

www.meng-jiang.com



Advertisers: It's not a CyberSaftey issue?

□Tencent

Forbes The World's Biggest Public Companies

[f Share](#)

[Tweet](#)

[in Share](#)

[reddit](#)

[G+1](#)

[Submit](#)

Tencent 腾讯

#201 Tencent Holdings

Market Cap As of May 2016

\$197.4 Billion

Industry

Computer Services

Founded

1999

Country

China

Chief Executive Officer

Huateng Ma

Website

<http://www.tencent.com>

Employees

30,641

Sales

\$16.35 B

Tencent Holdings on Forbes Lists

#48 Innovative Companies

Asia's Fab 50 Companies

#201 Global 2000

#534 in Sales

#127 in Profit

#529 in Assets

#26 in Market value

Tencent Holdings Ltd
HKG: 0700 - Sep 20, 4:09 PM GMT+8

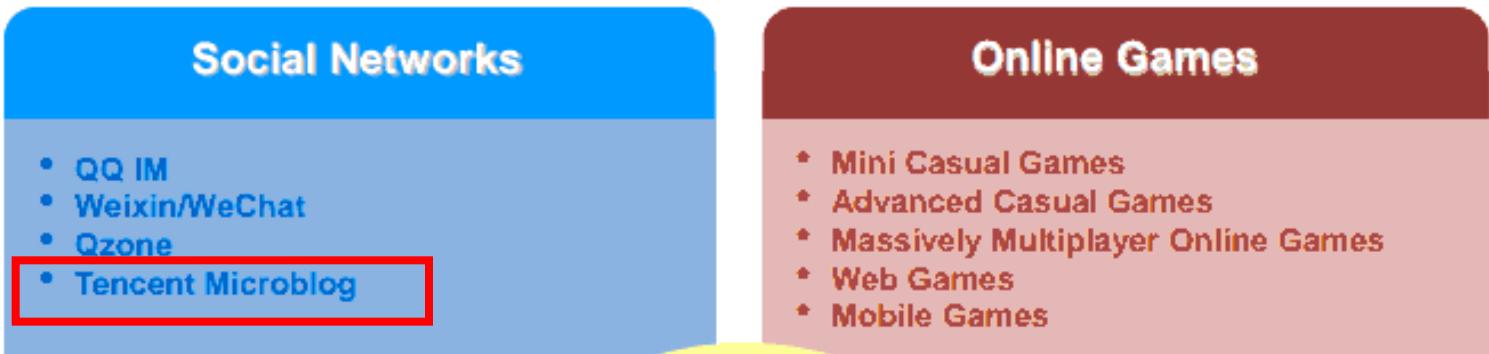
209.00 HKD -0.40 (0.19%)

1 day 5 day 1 month 3 month 1 year 5 year max



Diversified Product Portfolio

Fulfilling Online Lifestyle Needs



PC

Handset

Other Terminals



Advertisers: It's not a CyberSafety issue?

- ❑ NSFC Project (2011-) between Tencent and Tsinghua Univ.
- ❑ Tencent Weibo in 2011 told us
 - ❑ Low conversion rate (< 6%): #retweets per feed request
 - ❑ “I want a tweet recommender system!”
- ❑ M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu and S. Yang. “**Social Contextual Recommendation**” in **CIKM’12**. (cited by 149)
- ❑ **Deployed** in Weibo News Feed, 2012. Improved conversion rate from 5.78% to 8.27% (relatively **43%**).
- ❑ Tencent Weibo in 2013 told us
 - ❑ A more serious problem: zombie followers

Advertisers: It's not a CyberSafety issue?

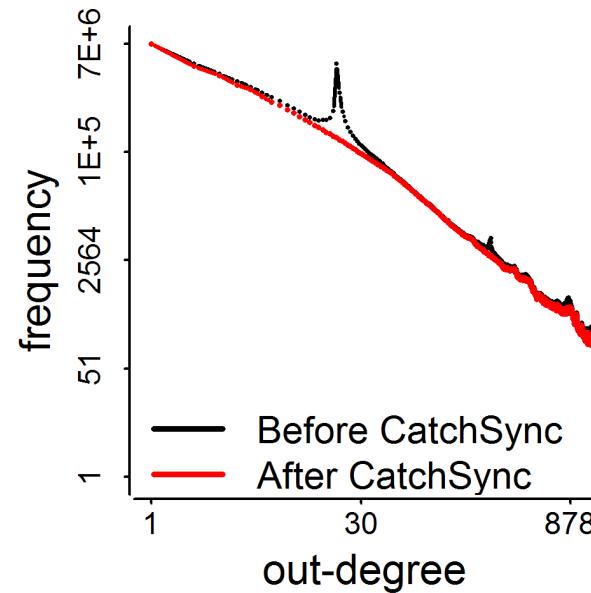
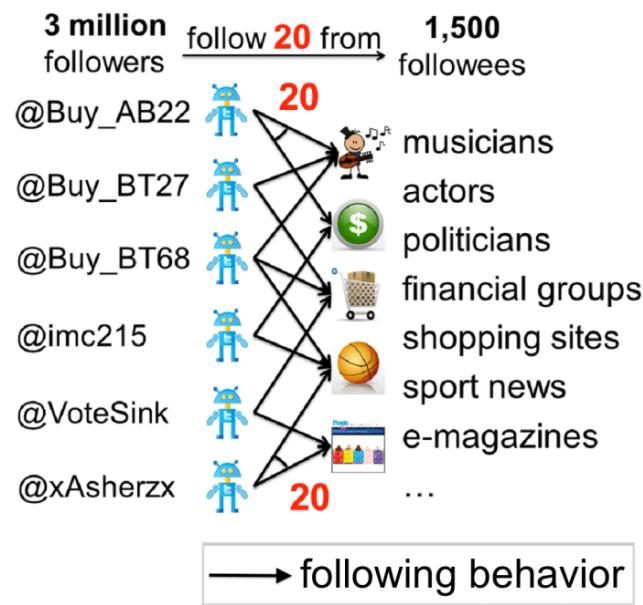


Experience-driven approaches: features of #followees, #hashtags, #URLs...



Advertisers: It's not a CyberSafety issue?

- ❑ M. Jiang, P. Cui, A. Beutel, C. Faloutsos and S. Yang. “**CatchSync**: Catching Synchronized Behavior in Large Directed Graphs” in **KDD’14 Best Paper Finalist**. (#citations = 36)



- ❑ Deployed in Weibo? Unfortunately, in July 2014...



Advertisers: It's not a CyberSafety issue?

- M. Jiang, A. Beutel, P. Cui, B. Hooi, S. Yang and C. Faloutsos. “A General **Suspiciousness Metric** for Dense Blocks in Multimodal Data” in ICDM’15.

User × hashtag × IP × minute	Mass c	Suspiciousness
$582 \times 3 \times 294 \times \mathbf{56,940}$	5,941,821	111,799,948
$188 \times 1 \times 313 \times \mathbf{56,943}$	2,344,614	47,013,868
$75 \times 1 \times 2 \times 2,061$	689,179	19,378,403

User ID	Time	IP address (city, province)	Tweet text with hashtag
USER-D	11-18 12:12:51	IP-1 (Deyang, Shandong)	#Snow# the Samsung GALAXY SII QQ Service customized version...
USER-E	11-18 12:12:53	IP-1 (Deyang, Shandong)	#Snow# the Samsung GALAXY SII QQ Service customized version...
USER-F	11-18 12:12:54	IP-2 (Zaozhuang, Shandong)	#Snow# the Samsung GALAXY SII QQ Service customized version...
USER-E	11-18 12:17:55	IP-1 (Deyang, Shandong)	#Li Ning - a weapon with a hero# good support activities!
USER-F	11-18 12:17:56	IP-2 (Zaozhuang, Shandong)	#Li Ning - a weapon with a hero# good support activities!
USER-D	11-18 12:18:40	IP-1 (Deyang, Shandong)	#Toshiba Bright Daren# color personality test to find out your sense...
USER-E	11-18 17:00:31	IP-2 (Zaozhuang, Shandong)	#Snow# the Samsung GALAXY SII QQ Service customized version...
USER-D	11-18 17:00:49	IP-2 (Zaozhuang, Shandong)	#Toshiba Bright Daren# color personality test to find out your sense...
USER-F	11-18 17:00:56	IP-2 (Zaozhuang, Shandong)	#Li Ning - a weapon with a hero# good support activities!

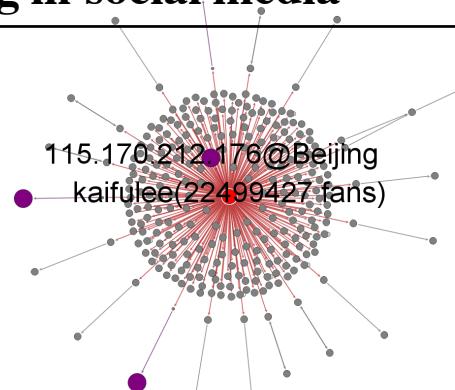
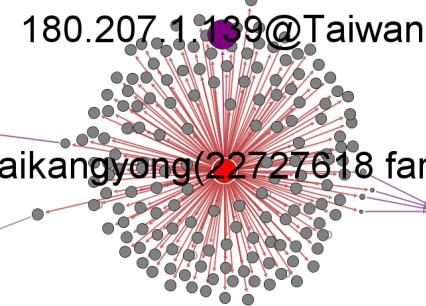


Advertisers: It's not a CyberSafety issue?

- It's really a CyberSafety issue! It can even dismiss your department/company.
- Q1: What are the advertising strategies? Are they new? Are they different?
- Q2: Can we detect the botnet advertisers (spammers) that were set up by the marketers?

Comparing Social Media Advertising and Traditional Advertising: S1

Celebrity branding/advertising in social media



Old-school celebrity branding



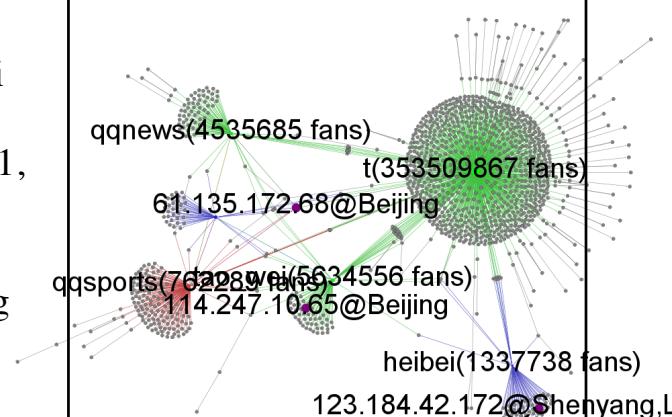
BILL &
MELINDA
GATES
foundation

Statistics: The 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**).

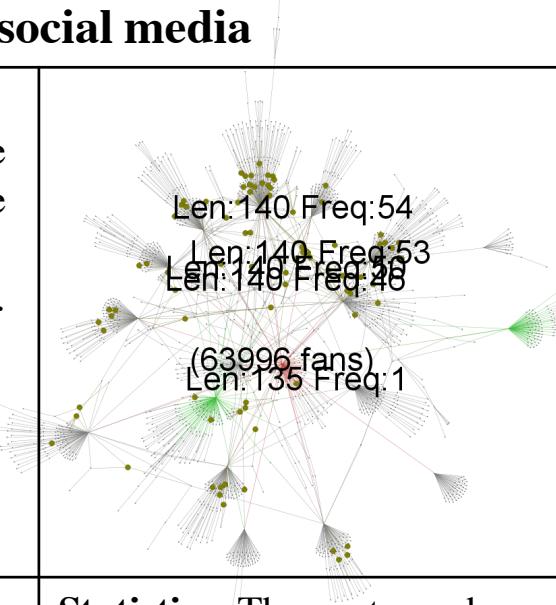
Statistics: The root user `@kaifulee` has 22.5M followers. The thread infected 6,346 users (**95.7% from followers**) and 5,357 devices (**97.3% from followers**). The number of retweets is 6,425 (**95.3% from followers**).

Strategy: 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. (Wikipedia)

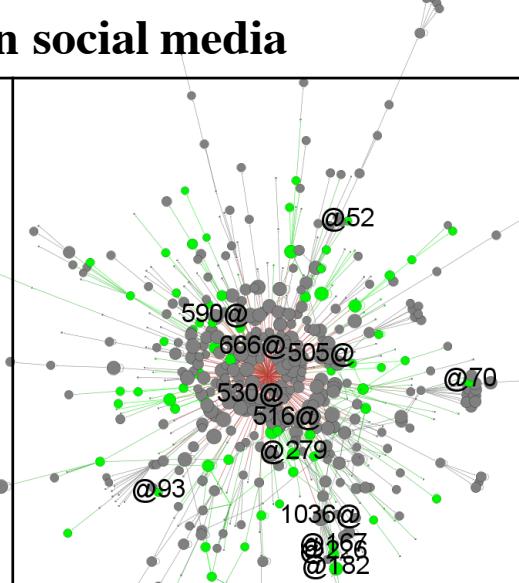
Comparing Social Media Advertising and Traditional Advertising: S2

Collaborative advertising in social media	Old-school collaborative advertising
<p>Retweet: by (1) @t with 353.5M followers, (2) @qqnews with 4.5M followers, (3) @hebei with 1.3M followers, (4) @taowei with 5.6M followers</p> <p>Root tweet: #China 0-1 Iraq# November 11, 2014 Brazil World Cup Asian Zone 20 the fourth round, the Chinese team 0-1 Iraq, Younis at stoppage time scored the winning goal. http://url.cn/1qkiWX</p> <p>Root user @qqsports has 762K followers.</p>	
<p>Strategy: in social media, no matter how 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.</p>	<p>Statistics: 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).</p> <p>Strategy: 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).</p>

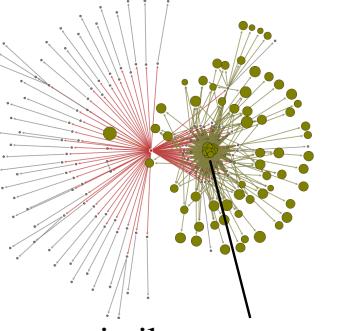
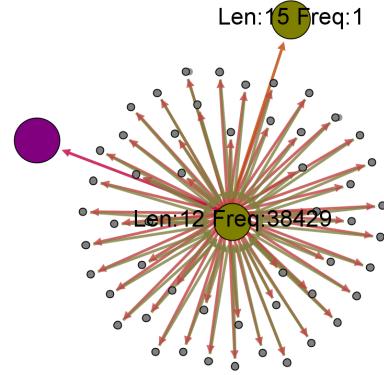
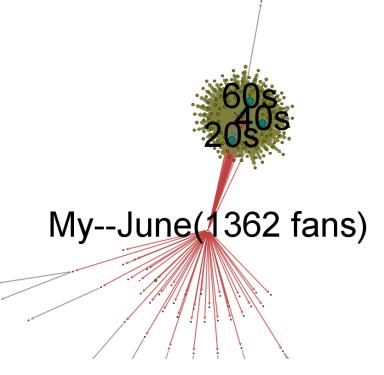
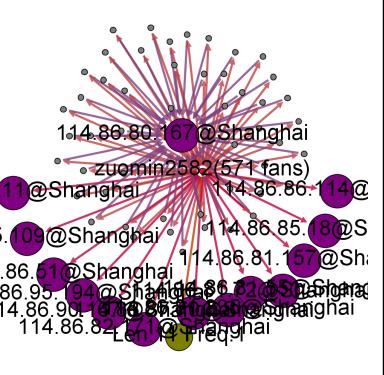
Comparing Social Media Advertising and Traditional Advertising: S3

Gift advertising in social media	Old-school gift advertising
<p>Rewtet: Our failures, one to the wrong time, the second is lost to the distance, the third is lost to himself. We always feel life is very long, very far away tomorrow, the dream will never be realized in waiting ...</p> <p>Root tweet: [6 tips to have thin legs] ... (2) jogging should be uniform, not when the fast slowly; ... (5) ... at a 90 degree angle standing; ... ❤ Want to know how to become a beauty Follow @zsyjkbk</p>	
<p>Strategy: 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 original text. Thus, the message can be widely diffused over the online network.</p>	<p>Statistics: 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).</p> <p>\$100 Amazon.com Gift Card</p> 

Comparing Social Media Advertising and Traditional Advertising: S4

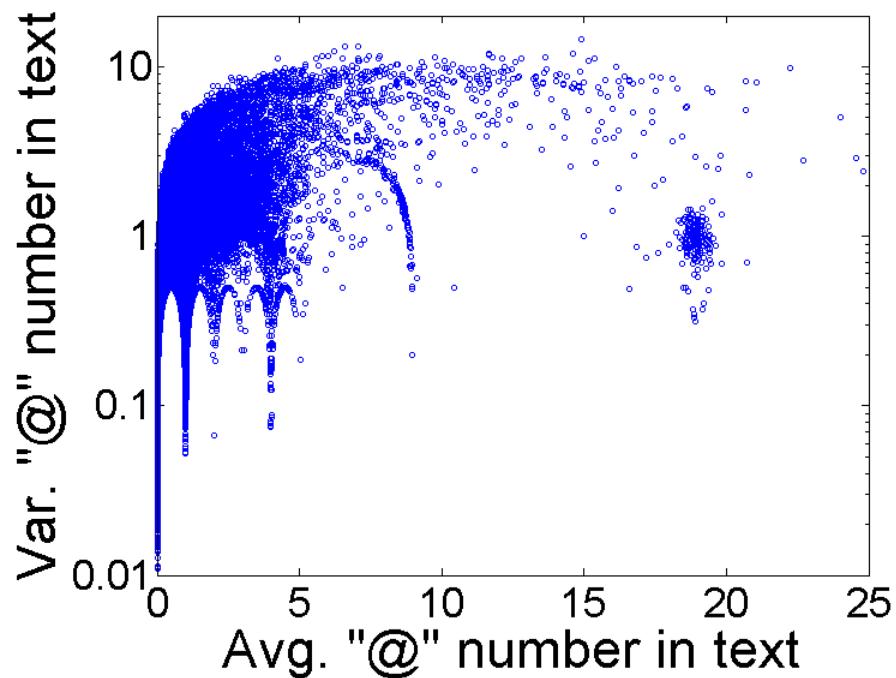
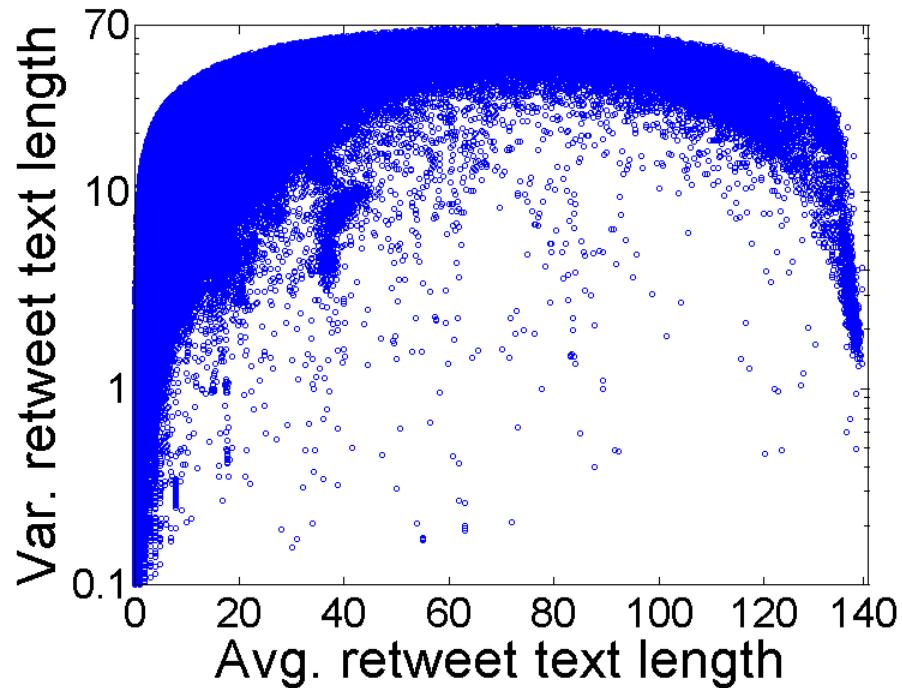
Multi-level marketing in social media	Old-school multi-level marketing
<p>Rewweet: #Qunaer# fair! good luck, Come together!@ lixianglin131407 @ snes986 @ cornett20096</p> <p>Root tweet: #Qunaer# activity Awards: #ipad2 # Participation: 1) become a fan @qunaer 2) forwarding this entry microblogging AT five friends; --- activities after the end of a winner drawn, presented an award ipad2! fair and just!</p> <p>Publication time: at 16:00 on Dec 30th</p>	
<p>Strategy: 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.</p>	<p>Statistics: 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).</p>
	<p>Strategy: 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.</p>

Solution 5: Synchrony Strategies in Social Media

Comment synchrony	URL synchrony	Time synchrony	Device synchrony
 <p>similar comments</p>		 <p>My--June(1362 fans)</p>	
<p>Strategy: (Galaxy Note ads) The botnets frequently retweet with similar phrases.</p>	<p>Strategy: (Porn URL) The botnets frequently retweet the same URL.</p>	<p>Strategy: (Galaxy Note ads) The botnets operate in lockstep with several fixed time intervals.</p>	<p>Strategy: (Galaxy Note ads) The botnets operate on the same group of devices in Shanghai.</p>
<p>Statistics: 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).</p>	<p>Statistics: 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).</p>	<p>Statistics: 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).</p>	<p>Statistics: 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).</p>

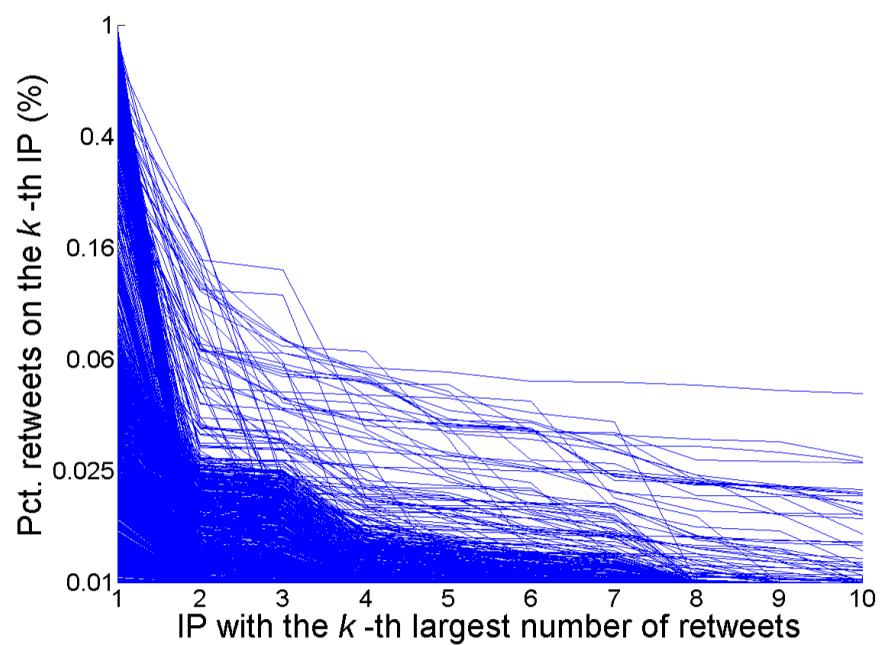
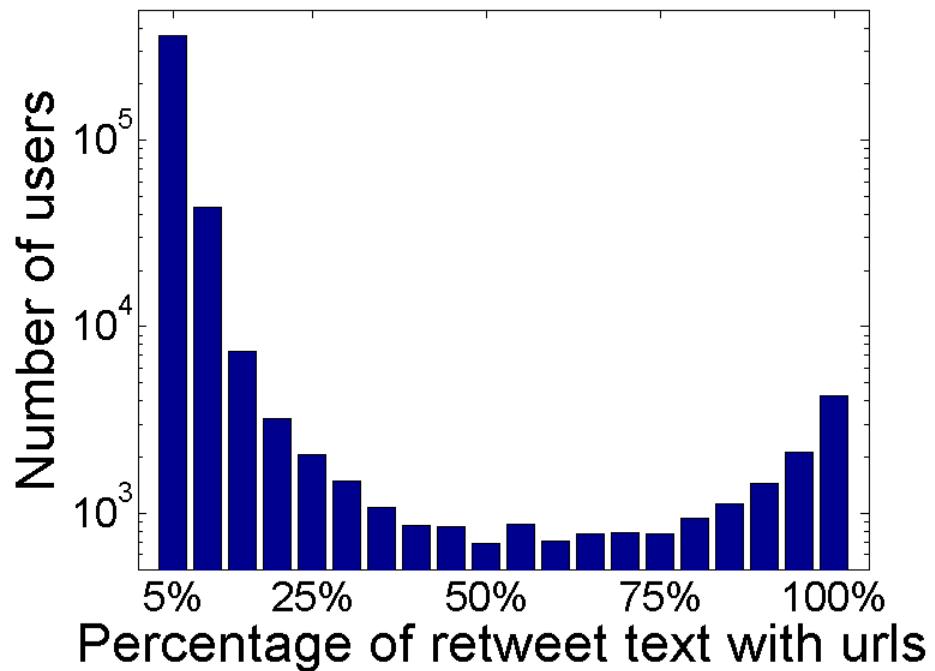
Observation: Features

- Comment features and mentioning features
 - Every dot is a user



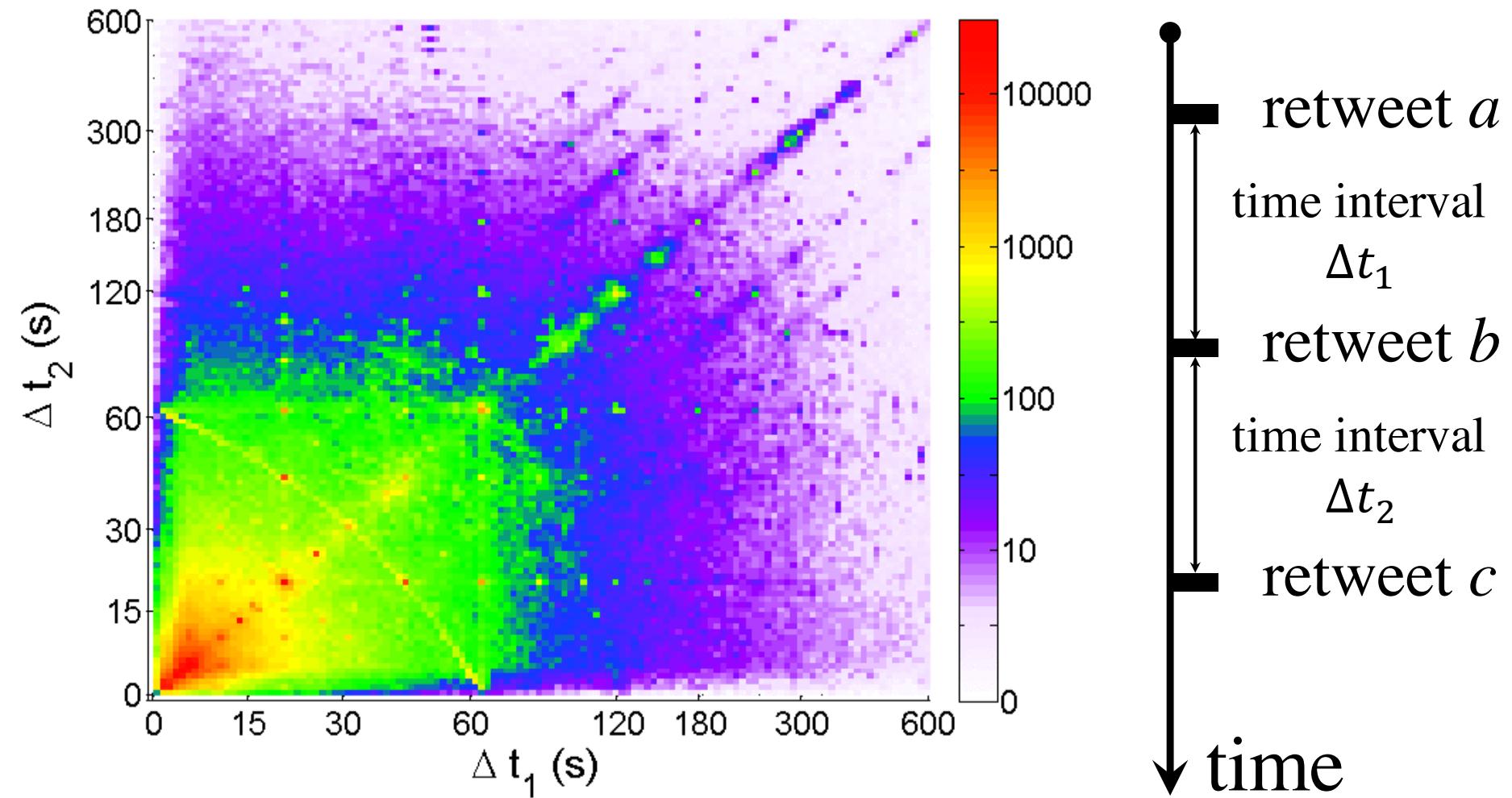
Observation: Features

- URL features and Device features



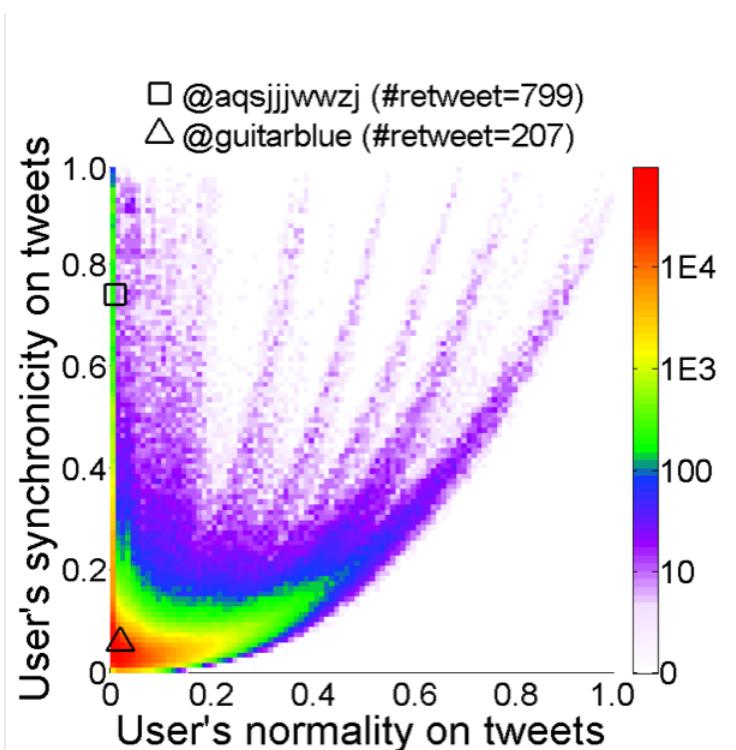
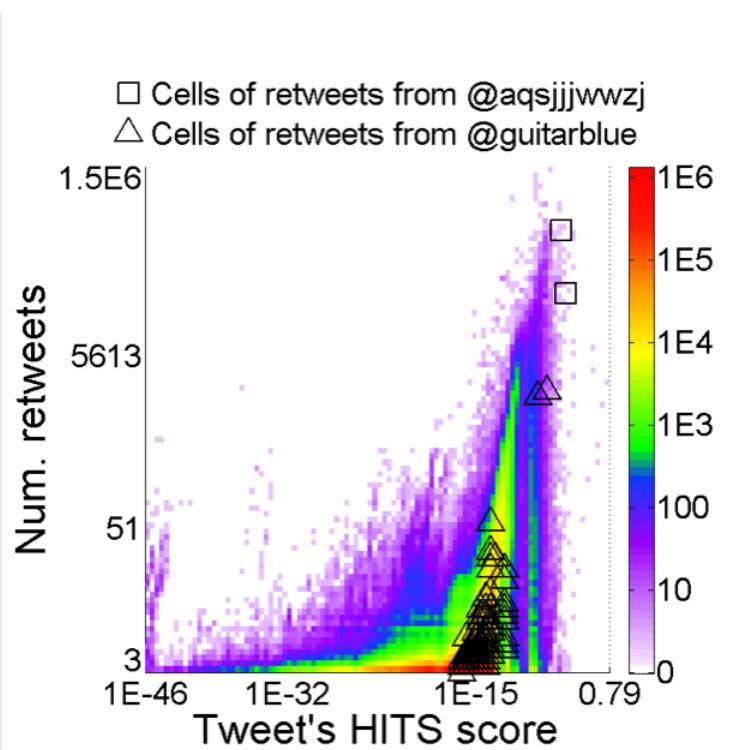
Observation: Features

- Lockstep behavioral features



Observation: Features

- Message synchrony [CatchSync, KDD'14]





Modeling: Features

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



Experimental Results

□ T1: Which strategy the tweet used? 5-class classification.

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

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



Experimental Results

- T2: Detecting botnet advertisers (spammers in synchrony): binary classification.

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

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



Summary

- ❑ Comparing social media advertising strategies and traditional (old-school) strategies
 - ❑ Celebrity branding
 - ❑ Collaborative advertising
 - ❑ Gift advertising
 - ❑ Multi-level marketing
 - ❑ Synchrony (social media botnet)
- ❑ Strategy classification: 0.889
- ❑ Spammer detection: 0.923
- ❑ Hope your company can avoid Weibo's misfortune.



Thank you!

Catching Social Media Advertisers with Strategy Analysis