Penny for Your Thoughts: Searching for the 50 Cent Party on Sina Weibo

Xiaofeng Yang and **Qian Yang** and **Christo Wilson**
Northeastern University
xiaofeng@ccs.neu.edu, *yang.qia@husky.neu.edu*, *cbw@ccs.neu.edu*

Abstract

Evidence suggests that the Chinese government employs "Internet Commentators" to post propaganda on social media. This group is pejoratively nicknamed the "50 cent party" or *Wumao*. In this study, we make the first attempt to quantify the size and behavior of the Wumao. Our study leverages a large corpus of data from Sina Weibo (Twitter in China) that includes 26M tweets and comments from 2.7M users over the span of one year. Unfortunately, detecting the Wumao is difficult because there is no ground truth information about them. To overcome this challenge, we apply a series of unsupervised techniques to filter our dataset and isolate *suspicious* users who exhibit characteristics indicative of being Wumao.

1 Introduction

Social media has become a global platform for political discussion. Twitter alone played a key role during the Arab Spring, Occupy Wall Street, and Ferguson, just to name a few events. However, because of social media's open nature, several governments around the world have taken measures to stifle or block these services. For example, Twitter has been blocked in Turkey, Iran, and Pakistan (Liebelson 2014), while crowdturfers suspected of working for the Russian government have flooded pro-democracy #hashtags with spam (News 2011).

The situation in China exemplifies this tension between social media and government. Anecdotal evidence suggests that the Chinese government employs "Internet Commentators" to post propaganda on social media (Henochowicz 2014). This group has been given the pejorative nickname "五毛党" ("50 Cent Party") or *Wumao*, after the amount of money each worker supposedly earns per post. Although the Chinese government runs a public program to train Commentators (Kaiman 2014), almost nothing is known about the Wumao. Estimates of the Wumao's size range wildly from hundreds to hundreds of thousands (Fareed 2008; Henochowicz 2014).

In this study, we make the first attempt to identify the Wumao and measure their behavior. We focus on Sina Weibo, since it is the largest microblogging service in China. The goal of our work is to determine if there are large-scale, coordinated efforts to sway political discussion on Weibo, and if so, how many users are involved in this effort?

To answer these questions, we downloaded a large corpus of data from Sina Weibo that includes 26M tweets and comments from 2.7M users over the span of one year. However, detecting the Wumao is difficult because nobody has ground-truth information about the Wumao's activities. To overcome this challenge, we apply a series of three unsupervised techniques to filter our dataset and isolate *suspicious* users who exhibit characteristics indicative of being Wumao.

- *First*, we cluster the users in our dataset based on the similarity of their messages. The intuition behind this step is that the Wumao are given orders by a central organization, and prior work has shown that crowdturfers produce content that is very similar (Motoyama et al. 2011; Wang et al. 2012).
- *Second*, we analyze the messages generated by users in each cluster to filter out spammers, since they are not of interest in this study.
- *Third*, we use Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003) to analyze the topics that are discussed by each cluster of users. The intuition behind this step is that we expect Wumao to discuss political topics more frequently than normal users.

After applying these steps, we identified 12 clusters containing 290 users that discuss political topics, and are not spammers. We manually analyzed the content produced by these users, but found no evidence of Wumao activity. Overall, there were 75 pro-government users, but they were spread across clusters, and did not exhibit any signs of coordination. Thus, our results suggest that either the Wumao did not operate on Sina Weibo during our measurement period, or estimates of the size of the group are vastly inflated.

2 Background

We begin by discussing background information related to Sina Weibo and paid commenters on Chinese social media.

2.1 Sina Weibo

Sina Weibo is one of the most popular OSNs in China, boasting 500M users who post 100M tweets per day (Wenlin 2013). Weibo provides many similar features to Twitter: users have personal profiles, follow each other, and

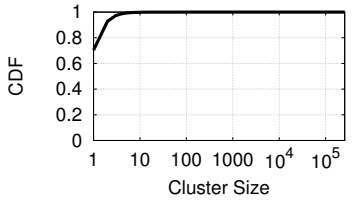

Figure 1: Sizes of clusters of users that generate similar content.

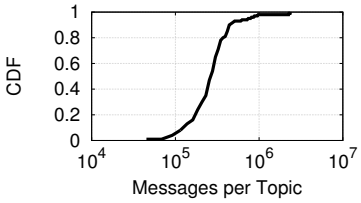

Figure 2: Messages per topic for the 100 topics located by LDA.

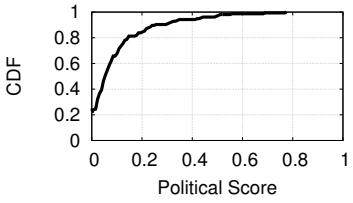

Figure 3: Political scores for the 151 large, non-spammer clusters.

post 140-character tweets (or "weibos") that may include retweets, @mentions, and #hashtags. Unlike Twitter, Weibo users may also comment on tweets. Prior work has shown that there is an order of magnitude more comments on Weibo than tweets (Chen, Zhang, and Wilson 2013).

Studies have shown that Weibo is a major target of *crowdturfing* due to its popularity (Wang et al. 2012). Crowdturfing is the practice of paying real users to spread ads, spam, and malicious rumors on social media (typically using fake accounts). Crowdturfers are referred to as the "Internet Water Army" ("网络水军") in China.

2.2 The Wumao

The Wumao, or "50 Cent Party" ("五毛党"), are Internet users who are paid to post content on social media in support of the Chinese government. The Wumao use many of the same tactics as the Water Army, but they are a distinct organization focusing only on politics. One of the earliest records of the Wumao is a leaked government report from 2006 that describes the responsibilities of "Internet Commentators" (the sanctioned name for the Wumao) (Liao 2013). The report states that Commentators are paid based on how many social media accounts they control and how often they tweet. A more recent official leak in 2014 divulged the identity and work details of 300 Commentators (Henochowicz 2014).

Although the Wumao are known to exist, very little is known about the size of the group or the websites/topics they target. To date, nobody has been able to obtain ground truth information about the Wumao or their command hierarchy (unlike crowdturfing marketplaces, which have been infiltrated and studied in detail (Wang et al. 2012)).

3 Collecting and Filtering Data

In this study, we make the first attempt to locate the Wumao and analyze their behavior. We focus on Weibo, since it is extremely popular in China and often used to discuss politics (Zhu et al. 2013; Chen, Zhang, and Wilson 2013). Since we do not have ground truth information on the Wumao, we apply a series of unsupervised filters to locate *suspicious* users. In §4, we manually examine these suspicious users to determine if they are Wumao.

3.1 Data Collection

We collected data by crawling all tweets generated by 2,066 politically active Chinese celebrities between August 2012 and August 2013. We also gathered all comments on these tweets. We chose this dataset because it includes political discussions. Additionally, prior work has shown that there are an order of magnitude more comments on Weibo than tweets, and that comments are most frequent on celebrity tweets (Chen, Zhang, and Wilson 2013). Thus, if the Wumao want to reach a large audience, the best way to do that is to comment on celebrity tweets.

In total, we collected 26M *messages* (*i.e.,* tweets and comments) from 2.7M users. However, 88% of users message <10 times, giving us too little data to analyze. After filtering these users out, 20M messages from 470K users remain.

3.2 Clustering Similar Users

The first step in our analysis is clustering users based on the similarity of their messages. The intuition behind this step is that the Wumao are given orders by a central organization, and prior work has shown that content produced by crowdturfers ends up being very similar (Motoyama et al. 2011; Wang et al. 2012). Thus, if Wumao are present in the dataset they may cluster together, unlike normal users who generate unique content.

To calculate the similarity between messages, we leverage the MinHash algorithm. MinHash was developed to quickly locate similar (but not necessarily duplicate) strings in spam email. It works by dividing message T_i into all possible substrings of length n, hashing the substrings, and placing the smallest k hashes in set S_i. Finally, MinHash computes the Jaccard Index between S_i and S_j derived from T_j. A MinHash score of 0 means the messages are very dissimilar, while 1 means the messages are extremely similar.

As done in prior work (Gao et al. 2010; Thomas et al. 2014), we cleaned the messages by removing URLs and @mentions before computing MinHash. We also ignore messages with $< k$ hashes. We set $n = 4$ and $k = 7$ based on guidance from prior work (Thomas et al. 2014) and our own experimental validation. Finally, we compute the similarity between users i and j as the average MinHash score of all pairs of messages generated by i and j. Overall, <1% of user pairs have similarity >0.2. This is to be expected, since there are 220B possible user pairings in our dataset and most honest users generate unique content.

After computing the similarity scores between all pairs of 470K users, we constructed and clustered a *similarity graph*. The similarity graph is a complete graph where each user is a node, and each edge is weighted using the average MinHash score. We used the Louvain algorithm (Blondel et al. 2008) to cluster the similarity graph because it does not require

Description	Key Words	% of Msgs.
Discussion of Senkaku Islands	钓鱼岛 (Senkaku Islands), 日本 (Japan), 中国 (China), 美国 (U.S.)	1.5
A famous lawyer discusses democracy	律师 (lawyer), 民主 (democracy), 政府 (government), 中国 (China)	1.0
Discussion of current affairs and history	李承鹏 (journalist's name), 真相 (fact), 历史 (history), 杨锦麟 (historian's name)	0.5
Discussion of violent law enforcement	城管 (urban management officers), 破坏 (demolish), 广州 (Guangzhou)	0.4

Table 1: Political topics identified by LDA, along with the percentage of all messages that fall into each topic.

selecting the number of clusters ahead-of-time. To speed up Louvain, we filtered out all edges with scores $\leq$0.2.

In total, Louvain located 84K clusters. Figure 1 plots the size of these clusters, revaling that 99.8% contain $<$10 users. Since our goal is to identify large-scale suspicious behavior, we filter out all users in clusters of size $\leq$10. This leaves the 155 largest clusters which contain 386K users.

3.3 Removing Spammers

At this point, we have identified clusters of users that produce similar messages. However, some of these clusters may contain spammers, since prior work has shown that spammers on OSNs generate similar content (Gao et al. 2010; Thomas et al. 2014). We are not interested in spammers, so we must filter them out.

To identify spammers, we analyze the content of messages. Prior work has shown that spammers on OSNs tweet links to shady websites (Yardi et al. 2010; Gao et al. 2010; Grier et al. 2010; Thomas et al. 2011; 2014). In contrast, the Wumao's goal is to influence political discussion, which can be done without tweeting URLs. Thus, we anticipate that clusters of spammers will produce more messages with URLs than clusters of normal users or Wumao.

To quantify this intuition, we calculate a *spam score* for the 155 remaining clusters in our dataset. We define the spam score for a cluster c as the fraction of messages generated by users in c that contain a URL. Intuitively, a cluster with a spam score close to 1 means that a group of users are sending similar messages that include URLs, which is indicative of a spam campaign. Using this methodology, we identified four clusters containing 29K users that have spam scores $>$0.5. Manual analysis confirms that the users in these clusters were spamming. We filter these clusters out, leaving us with 151 clusters and 357K users.

3.4 Identifying Political Topics and Clusters

The last step in our filtering process is calculating a *political score* for each cluster. The intuition behind this step is that we expect the Wumao to discuss political topics more frequently than normal users. To identify topics in our dataset, we leverage Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003). LDA takes a corpus of documents as input, and outputs K topics, each of which is a list words sorted by the strength of their association with that topic.

In this work we use the same topic extraction methodology that has been successfully applied to Weibo data by prior work (Chen, Zhang, and Wilson 2013). *First*, we use OpenCLAS (jadesoul 2013) coupled with a crowdsourced dictionary of Chinese words from Sogou Pinyin (Sogou 2013) to segment each message into words. *Second*, we combine each tweet with its comments to form a single document, which improves the accuracy of LDA (Ramage, Dumais, and Liebling 2010; Hong and Davison 2010; Quercia, Askham, and Crowcroft 2012). *Third*, we filter out the top 10% most frequent words and words that appear $<$5 times from the corpus. This eliminates stop words and speeds up LDA's runtime.

Fourth and finally, we executed LDA on a randomly sampled 10% subset of documents from our corpus. LDA is CPU and memory intensive, and thus we were unable to run it on our whole corpus. We experimented with several values of K, but eventually settled on $K = 100$ as larger values did not produce a greater number of meaningful topics. Figure 2 plots the number of messages per topic, and reveals that the popularity of topics varies by an order of magnitude. This is not surprising: discussions about celebrities and entertainment engender much more engagement than political topics.

After executing LDA, we had three native Chinese speakers manually examine the top 20 words in all 100 topics, and independently pick topics that were political. To be conservative, we labeled a topic as political if $\geq$2 raters identified it as political. In total, eight out of 100 topics were identified as political. Four example political topics are shown in Table 1. Two of the eight political topics are within the top 10 most discussed topics in our dataset.

Calculating Political Scores. After manually identifying the political topics, we calculate a *political score* for the remaining clusters in the dataset. We define the political score of a cluster c as the number of words in messages generated by users in c that overlap with the top 10 words in our eight political topics, divided by the number of words in c's messages that overlap with the top 10 words in all 100 topics. Thus, political scores are between 0 and 1.

Figure 3 shows the political score distribution of the 151 remaining clusters. Only 12 clusters have scores $>$0.3, which is not surprising given that only 8% of topics in dataset are political. As shown in Table 2, these highly political clusters are all small, although clusters #17, #27, and #107 generate many messages. There are two large clusters (13K and 19K users) that have relatively high political scores (0.25 and 0.22), but manual analysis of the messages from these users reveals that they have been clustered because they all retweet similar content. Thus, after three rounds of filtering, we are left with 12 clusters containing 290 highly political, suspicious users.

4 Analysis and Conclusions

In §3, we use a series of three unsupervised filters to identify 290 suspicious users in our dataset. In this section, we con-

Cluster #	#Users	#Msgs.	Political Score
43	18	35	0.7757
123	10	292	0.6865
17	77	248905	0.5533
65	14	30	0.5062
150	10	26	0.4892
97	11	30	0.4267
28	28	120	0.4180
30	27	337	0.3482
27	28	2626	0.3346
107	11	1137	0.3215
147	10	13	0.3131
29	27	74	0.3100

Table 2: Statistics on the 12 political clusters in our dataset.

clude our study by manually analyzing these users to determine if they are Wumao, *i.e.,* do they engage in large-scale, coordinated, political propaganda campaigns?

Results. After manually analyzing *all* content produced by these 290 users, we located no evidence that any of these users are Wumao. Five of the 12 clusters received high political scores overall due to a single prolific user. The remaining seven clusters do include 194 political users. Of these users, 112 are neutral towards the government, 7 are negative, and 75 are pro-government. Even if we skeptically assume that neutral and pro-government users are suspicious, these users are spread across several clusters, and there is no evidence of coordination between them.

In conclusion, we find no evidence of large-scale Wumao activity on Weibo. Although it is impossible to prove that the Wumao are not on Weibo without ground truth, our methodology did not identify any large-scale political crowdturfing.

Limitations. There are several potential reasons why we may not observe Wumao. First, although the Wumao have existed since 2006 (Liao 2013), it is possible that they were not active on Weibo during 2012–2013. Second, it is possible that our Weibo dataset did not capture the correct users. However, we specifically crawled a politically engaged subgraph of Weibo, which seems like the region Wumao would be active in. Third, sensitive posts may be censored, although most censored posts are against the government, while Wumao posts are in favor of the government.

Finally, our methodology relies on the assumption that Wumao behave like typical crowdturfers. However, it is possible that Wumao are stealthy and do not follow these patterns. We attempted to have manual labelers identify stealthy Wumao in our dataset, but this only succeeded in identifying pro-government users. Clearly, our goal is not to demonize political speech; the only way to separate propaganda from individual expression is to look for large-scale coordination, which is exactly what our methodology does.

Acknowledgements

This research was supported in part by NSF grants CNS-1054233 and CHS-1408345.

References

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *the Journal of Statistical Mechanics: Theory and Experiment* 10.

Chen, L.; Zhang, C.; and Wilson, C. 2013. Tweeting under pressure: Analyzing trending topics and evolving word choice on sina weibo. In *Proc. of COSN*.

Fareed, M. 2008. China joins a turf war. The Guardian.

Gao, H.; Hu, J.; Wilson, C.; Li, Z.; Chen, Y.; and Zhao, B. Y. 2010. Detecting and characterizing social spam campaigns. In *Proc. of IMC*.

Grier, C.; Thomas, K.; Paxson, V.; and Zhang, M. 2010. @spam: the underground on 140 characters or less. In *Proc. of CCS*.

Henochowicz, A. 2014. Thousands of local internet propaganda emails leaked. China Digital Times.

Hong, L., and Davison, B. D. 2010. Empirical study of topic modeling in twitter. In *Proc. of SOMA*.

jadesoul. 2013. Open Chinese Lexical Analysis System. Github.

Kaiman, J. 2014. China to train leaders to manage online public opinion. The Guardian.

Liao, R. 2013. 探析微博在政府反腐倡廉过程中现实意义 (The significance of Weibo in fighting against corruption in the government process). 论文网.

Liebelson, D. 2014. Map: Here are the countries that block facebook, twitter, and youtube. Mother Jones.

Motoyama, M.; McCoy, D.; Levchenko, K.; Savage, S.; and Voelker, G. M. 2011. Dirty jobs: The role of freelance labor in web service abuse. In *Proc. of Usenix Security*.

News, B. 2011. Russian twitter political protests 'swamped by spam'. BBC News. `http://www.bbc.co.uk/news/technology-16108876`.

Quercia, D.; Askham, H.; and Crowcroft, J. 2012. Tweetlda: Supervised topic classification and link prediction on twitter. In *Proc. of WebSci*.

Ramage, D.; Dumais, S. T.; and Liebling, D. J. 2010. Characterizing microblogs with topic models. In *Proc. of ICWSM*.

Sogou, I. 2013. Sogou pinyin. Sogou. `http://pinyin.sogou.com/`.

Thomas, K.; Grier, C.; Paxson, V.; and Song, D. 2011. Suspended accounts in retrospect: An analysis of twitter spam. In *Proc. of IMC*.

Thomas, K.; Li, F.; Grier, C.; and Paxson, V. 2014. Consequences of connectivity: Characterizing account hijacking on twitter. In *Proc. of CCS*.

Wang, G.; Wilson, C.; Zhao, X.; Zhu, Y.; Mohanlal, M.; Zheng, H.; and Zhao, B. Y. 2012. Serf and turf: crowdturfing for fun and profit. In *Proc. of WWW*.

Wenlin, Z. 2013. Weibo has over 500 million users, and 4.6 million active users daily. Xinhua News.

Yardi, S.; Romero, D.; Schoenebeck, G.; and Boyd, D. 2010. Detecting spam in a twitter network. *First Monday* 15(1).

Zhu, T.; Phipps, D.; Pridgen, A.; Crandall, J. R.; and Wallach, D. S. 2013. The velocity of censorship: High-fidelity detection of microblog post deletions. In *Proc. of USENIX Security*.