(including employing or coincidentally attracting spammers with an unidentifiable stand), which largely deterred the negative trend, until around October 28, when spammers in favor of Ma reversed the trend. Although Wang's spammers still effected a resurge on November 4, the second more decisive period was from November 5. Spammers supportive of Ma exhibited sophisticated manipulating skills and successfully remained in control for more than 20 days, despite several minor efforts from the opposing side.

Figure 4: Sentiment, posting amount, and post type of spammers for the two main protagonists involved in the event.

Another noteworthy result drawn from our diachronic analysis is that opinion spammers on Sina Weibo displayed a deliberate micro-tactic of hiding. Given that it is not possible anymore to analyze the impact of spamming "liking" behavior (Sina Weibo no longer displays the users who "liked" a posting), this work focuses on reposting of existing articles, posting of original articles, and replies to posts. The result reveals a different finding from Allcott & Gentzkow [1], and further works, which posit that posting thematic articles serves a vital role in mobilizing endorsement to a specific political opinion. The opinion spammers on Sina Weibo, in contrast, deliberately avoid posting or reposting articles. Instead, they preferred to reply to existing posts to avoid mention (@) of their client's names, trying to alter the general attitude towards a post (tweet) with overwhelmingly sentimental replies. Since Sina Weibo typically displays replies to a posting one by one under that tweet, this practice can often create an exclusive "bubble filter" [13] that repels users on the opposite side. The replies may evoke the feeling in other readers that the opinion reflected in the article is false (if replies denounce it) or true (if replies support it). Original writing is also an option less frequently used. This specific combination of spamming tactics, while being effective in shifting the public sentiment towards an event as analyzed, makes the spamming activity more challenging to detect (Sina Weibo deletes tweets or posts that are deemed spam or for which they receive heavy complaints of it being such), and therefore, more subtle and effective.

7 CONCLUSIONS

This paper proposes a novel and principled method to exploit observed microblog posting behavior to detect spammers in the special setting of public opinion spamming on Sina Weibo, and examine the impact it exerts on the public opinion. The precision of model affirmed the estimated characterization of spamming behavior. Based on the precise detection of public opinion spamming, a diachronic analysis about the impact of opinion spammers on a widely noted case in China demonstrates that such spammers subtly manipulated the public sentiment on Sina Weibo, one of the top social media platforms in China. This work, therefore, sets the path towards new research on public opinion spamming, and calls for a more detailed and nuanced analysis of the spammers' impact on public opinion, and potentially, on the social justice and well-being of the society.

8 ACKNOWLEDGEMENTS

The authors wish to acknowledge the support provided by the National Natural Science Foundation of China (61503217, 91546203), the Key Research and Development Program of Shandong Province of China (2017CXGC0605) and China Scholarship Council (201606220187). Gerard de Melo's research is funded in part by ARO grant W911NF-17-C-0098 (DARPA SocialSim).

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. *Social Media and Fake News in the 2016 Election.* Working Paper 23089. National Bureau of Economic Research.
[2] Hao Chen, Jun Liu, Yanzhang Lv, Max Haifei Li, Mengyue Liu, and Qinghua Zheng. 2017. Semi-supervised Clue Fusion for Spammer Detection in Sina Weibo. *Information Fusion* (2017).
[3] Hao Chen, Jun Liu, and Jianhong Mi. 2016. SpamDia: Spammer Diagnosis in Sina Weibo Microblog. In *MobiMedia 2016.* 116–120.
[4] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. 2013. Exploiting burstiness in reviews for review spammer detection. In *ICWSM.*
[5] Song Feng, Ritwik Banerjee, and Yejin Choi. 2012. Syntactic stylometry for deception detection. In *ACL short.* 171–175.
[6] Song Feng, Longfei Xing, Anupam Gogar, and Yejin Choi. 2012. Distributional Footprints of Deceptive Product Reviews. In *ICWSM.*
[7] Kunal Goswami, Younghee Park, and Chungsik Song. 2017. Impact of reviewer social interaction on online consumer review fraud detection. *Journal of Big Data* 4, 1 (15 May 2017), 15.
[8] Ee-Peng Lim, Viet An Nguyen, Nitin Jindal, Bing Liu, and Hady Wirawan Lauw. 2010. Detecting product review spammers using rating behaviors. In *CIKM.*
[9] Yingcai Ma, Niu Yan, Ren Yan, and Yibo Xue. 2013. Detecting Spam on Sina Weibo. *CCIS-13* (2013).
[10] Arjun Mukherjee, Bing Liu, and Natalie Glance. 2012. Spotting fake reviewer groups in consumer reviews. In *WWW.* 191–200.
[11] Arjun Mukherjee, Bing Liu, Junhui Wang, Natalie Glance, and Nitin Jindal. 2011. Detecting group review spam. In *WWW Companion.* 93–94.
[12] Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T. Hancock. 2011. Finding Deceptive Opinion Spam by Any Stretch of the Imagination. (2011), 309–319.
[13] Eli Pariser. 2011. *The Filter Bubble: What the Internet Is Hiding from You.* The Penguin Group.
[14] Cornelius Puschmann and Jean Burgess. 2014. *The Politics of Twitter Data.* Peter Lang Publishing Inc.
[15] Yang Qiao, Huaping Zhang, Min Yu, and Yu Zhang. 2016. Sina-Weibo Spammer Detection with GBDT. In *Chinese National Conference on Social Media Processing.*
[16] Shebuti Rayana and Leman Akoglu. 2015. Collective Opinion Spam Detection: Bridging Review Networks and Metadata. In *SIGKDD.* 985–994.
[17] Laura Spinney. 2017. The Shared Past that Wasn't: Facebook, fake news and friends are warping your memory. 543 (2017), 168–170.
[18] Cass R Sunstein. 2009. *Going to extremes: How like minds unite and divide.* Oxford University Press.
[19] Zhuo Wang, Tingting Hou, Dawei Song, Zhun Li, and Tianqi Kong. 2016. Detecting Review Spammer Groups via Bipartite Graph Projection. *Comput. J.* 59, 6 (2016), 861–874.
[20] Sihong Xie, Guan Wang, Shuyang Lin, and Philip S. Yu. 2012. Review spam detection via temporal pattern discovery. In *SIGKDD.* 823–831.