Twitter Bots, Corporations, and Financial Markets*

Juhani Linnainmaa
Alessandro Previtero
Wei Wang

This version: April 5, 2019

Abstract

Social media play a significant role in shaping public opinions and beliefs. In this project we aim to provide novel evidence on the effects of tweets from automated accounts (bots) on stock prices. More specifically, we plan to test if bot-generated tweets proliferate at times when the shareholders and managers benefit the most from higher stock prices—that is, around M&As, IPOs, or stock option expirations. Our results could shed light on how social media affect stock prices and if—and to what extent—firms strategically use social media to boost their share prices.

*Juhani Linnainmaa is with the University of Southern California and NBER, Alessandro Previtero is with the Indiana University and NBER, and Wei Wang is with the Indiana University.

1 Research Objectives

Social media play an increasing role in shaping public opinions and beliefs. With more than 321 million active users worldwide as of the end of 2018, Twitter is one of the biggest social networks worldwide. Computer scientists have recently investigated the possibility that automated Twitter accounts—called bots—can manipulate public opinion. Shao, Ciampaglia, Varol, Flammini, and Menczer (2017), for example, find evidence that, during the 2016 U.S. presidential election, bots might have misled many human users into spreading misinformation. Preliminary evidence suggests that a sizable fraction of bot-generated tweets are related to business news (see Figure 1).

In this research project, we plan to investigate the effects of bot-generated tweets on stock prices. Specifically, we intend to study and measure the extent to which corporations use automated Twitter accounts (bots) to strategically spread positive information, increase investor attention and, ultimately, influence and inflate stock prices. We plan to use the Botometer, a machine learning-based tool developed by the Observatory on Social Media (OSoMe) at Indiana University. Armed with this tool, we intend to build a time series of bot-generated tweets for large listed corporations based on millions of tweets. Our objective is to test whether bot-generated tweets increase around important corporate events and, specifically, at times when corporations or managers can benefit from higher valuations.

The idea that firms might influence media coverage is not new in finance. Ahern and Sosyura (2014), for example, analyze press releases of bidders involved in fixed exchange-ratio stock mergers. In a fixed exchange-ratio stock merger, as opposed to a floating-ratio merger, the bidder derives benefits from higher valuations. Ahern and Sosyura find that, relative to floating exchange ratio acquirers, fixed-ratio bidders issue more press releases with a more positive tone. This result suggests that bidders engage in active media management precisely when it is in their best interest.

Compared to this and other studies on the influence of corporations on media, our project on bot-generated tweets has the potential to make three contributions to the existing literature. First, we can investigate the effect of social media news that—compared to traditional news—could be disseminated faster, more broadly (they could go "viral"), and, above all, in an unfiltered manner. Although firms can strategically issue press releases, it is ultimately the media that writes the news. More popular media have also stronger incentives to fact-check and to report accurate news. This filtering function is absent in social media. On the contrary, studies of social media report that bots facilitate the dissemination of less accurate news (Shao et al., 2017).

Second, with traditional media, firms can at best influence only their own news coverage. With social media, firms could also cheaply and effectively spread negative news about their competitors. In the Ahern et al. (2014) M&A study, both the bidder and the target have an incentive to issue press releases to alter their *relative* valuations. With social media, the bidder could potentially spread negative news about the target. Research has already provided evidence that social media could more easily disseminate false stories ("fake news"). Shao, Hui, Wang, Jiang, Flammini, Menczer, and Ciampaglia (2018), for example, analyze

political news and find that only 6 percent of Twitter accounts that the study identified as bots were sufficient to spread 31 percent of all the low-credibility information in the network.

Third, the richness of the social media data would allow us to investigate a more inclusive set of corporate events. Other than M&As, we plan to analyze other events around which firms have an incentive to manipulate their stock prices. These additional events include initial public offerings (IPOs), secondary offerings, earnings announcements, and the expiration of executive stock options.

Other studies have investigated the effect of Twitter on stock prices (e.g., Behrendt and Schmidt (2018) and Chen, De, Hu, and Hwang (2014)). By focusing on bot-generated tweets, our project has the potential to deepen our knowledge of the effects of social media on financial markets. More specifically, we can investigate the extent to which, if at all, firms use and manipulate bot-generated tweets with the goal of influencing stock prices when it is particularly favorable for their shareholders or managers.

2 Research Strategy

A research on bot-generated tweets and corporations faces two major empirical obstacles. First, collecting and analyzing data on millions of tweets is a significant task that requires the use of the most advanced techniques of big data analysis. Specifically, being able to identify bot-generated tweets represents a serious challenge. Second, establishing causality is not straightforward. For example, tweets might not cause stock prices to move; rather, tweets could reflect omitted (latent) news that cause both the price movements and the tweets.

To identify bot-generated tweets, we plan to use the Botometer, a machine learning-based tool developed by researchers in the computer science department at Indiana University. This tool uses an algorithm that extracts approximately 1,200 features related to a Twitter account activity and provides the likelihood of the account being automated (that is, a bot). The Observatory on Social Media (OSoMe) has stored a repository of tweets that includes a representative sample of 10% of all tweets starting from May 2016 up to recent months. The same Observatory has also designed Hoaxy, a search engine that shows how stories (from low-credibility sources) spread on Twitter. These two tools, Botometer and Hoaxy, are now integrated and a researcher can use them, first, to detect when information is spreading virally and, second, who is responsible for its spread. The Botometer Application Programming Interface (API) and the tweet repository are available to all researchers at the Indiana University.

We plan to expand on the existing database in two different ways. First, we aim to extend the time-series of tweets backwards to 2010, which is the time when Twitter reached 10 million active users.[1] In collecting these data, we plan to focus on tweets related to

[1]There are 66 million users in the U.S. in 2018.

specific listed company names—i.e., tweets that use the hashtag symbol to identify either a company name or ticker symbol. Then, we plan to analyze also the tone of the tweets to infer whether they contain a positive or negative message. For inferring tone, we plan to implement machine-learning techniques using Python, a leading programming language for textual analysis. Python provides easy-to-use libraries such as NLTK, Scikit-learn, and Gensim for classification, tokenization, and semantic reasoning. Last, we plan to use the Hoaxy tool to investigate how quickly and through what channels the tweets spread throughout the network.

To address the identification challenges in our project, we plan to use multiple sources of (potentially) exogenous variation in, first, the cost of using bots and, second, in investor attention. First, we intend to use a change in the Twitter corporate policy on how they handle automated accounts. Since July 2018, Twitter began to terminate suspected automated accounts. This change in the enforcement policy provides plausibly exogenous variation in the costs and effectiveness of spreading bot-generating tweets. After July 2018, we expect to see a reduction in the number of and the effects of bot-generated tweets. A similar approach could be to use time-series variation in electricity costs (or other costs associated with operating large servers) to instrument for the number of bot-generated tweets. In terms of variation in investor attention, research on peer-to-peer lending has successfully used weather (e.g., rainy days) as an instrument for capturing exogenous variation in the supply of funds. During bad weather periods, individuals are more likely to be online and therefore potentially more susceptible to bot-generated tweets.

Last, we acknowledge that it will not be possible to establish any direct link between bot-generated tweets and corporations (that is, there is no "smoking gun" that ties a company to specific acts of stock price manipulation). We will not know the identity of those who set up the automated accounts. We try to address this limitation by focusing on those corporate events in which managers have the highest incentives to manipulate stock prices such as, for example, expiration of stock options. The idea is to use the variation in incentives to understand the extent to which surge in automated tweets corresponds to period during which the corporations and managers would benefit the most from additional positive coverage.

3 Conclusion

There is mounting evidence that bot-generated tweets play a significant role in shaping public opinions and beliefs. In this project we aim to provide novel evidence on the effects of bot-generated tweets on stock prices. More specifically, we plan to test the extent to which bot-generated tweets are more common at times when shareholders and managers can benefit the most from higher prices—around, for example, M&As, IPOs, or stock option expirations. Our results can contribute to the public debate on the benefits and detrimental effects of social media.

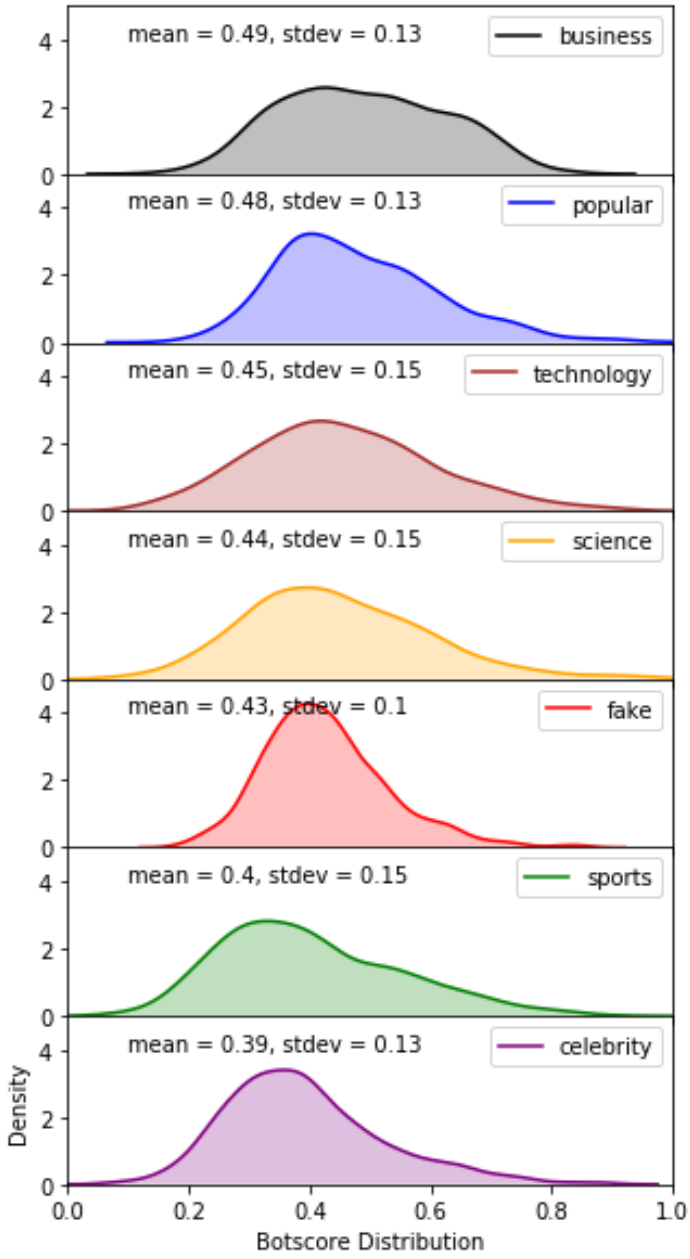


Figure 1: Botscore Distribution on Different News Categories (from Avram (2018))

References

Ahern, K. R. and D. Sosyura (2014). Who writes the news? corporate press releases during merger negotiations. *The Journal of Finance 69*(1), 241–291.

Avram, M. (2018). Hoaxy and fakey: Tools to analyze and mitigate the spread of misinformation in social media. Unpublished master thesis.

Behrendt, S. and A. Schmidt (2018). The twitter myth revisited: Intraday investor sentiment, twitter activity and individual-level stock return volatility. *Journal of Banking & Finance 96*, 355–367.

Chen, H., P. De, Y. J. Hu, and B.-H. Hwang (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies 27*(5), 1367–1403.

Shao, C., G. L. Ciampaglia, O. Varol, A. Flammini, and F. Menczer (2017). The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592*, 96–104.

Shao, C., P.-M. Hui, L. Wang, X. Jiang, A. Flammini, F. Menczer, and G. L. Ciampaglia (2018). Anatomy of an online misinformation network. *PloS one 13*(4), e0196087.