Detecting Crowdturfing in Social Media

Liang Wu, and Huan Liu

1 Synonyms

Astroturfing; Malicious crowdsourcing

2 Glossary

Astroturfing Astroturfing is the campaign that masks its supporters and sponsors to make it appear to be launched by grassroots participants.

Crowdsourcing Crowdsourcing is the process of obtaining needed services, ideas, or content by soliciting contributions from a group of people. Internet services facilitate the process by connecting customers and crowdsourcing workers.

Ground truth Ground truth is the accurate annotation of data examples, which is used in statistical models to prove or disprove research hypotheses.

Heterogeneous data Heterogeneous data are the data involving multiple modalities, such as a social media post containing texts and video clips.

Information diffusion Information diffusion happens between individuals when a flow of information travels from one individual to another.

Misinformation and Disinformation Misinformation and disinformation are the inaccurate or false information., while disinformation is intentionally spread to mislead other people, and misinformation is unintentionally spread.

L. Wu, H. Liu
Data Mining and Machine Learning Lab,
School of Computing, Informatics, and Decision Systems Engineering,
Arizona State University, Tempe, Arizona, USA
e-mail: {wuliang,huanliu}@asu.edu

Social media Social media can be defined as internet services that enable users to share content, and subscribe and receive information of interests. Social media sites allow for social interactions between users, where people can follow each other and make friends. Social media are immensely popular in the dissemination of breaking news and emerging stories.

Social media account A social media account is an entry of user in social media websites. A user may have multiple accounts, and an account can also be controlled by multiple users.

Social media post A social media post is an entry for content that may comprise text, images, videos, and links to other posts or external resources.

Social media user An individual who has a social media account, through which he/she posts information, receives and comments on other people, makes friends and follow other people.

Social network A social network is a network structure made up of social network users, and users are interconnected by the social ties.

Supervised learning Supervised learning is the machine learning task of inferring a decision function through learning from data with ground truth.

3 Definition

If participants of an astroturfing campaign are organized by crowdsourcing, the process is defined as **Crowdturfing**. Crowdturfing aims to gain or destroy reputation of people, products and other entities through spreading biased opinions and framed information.

Crowdturfing Detection is the process of detecting crowdturfing activities with software systems. The detection method can be unsupervised through discovering the anomalous traffic or on the basis of patterns of historical crowdturfing activities.

Crowdturfing in Social Media is the crowdturfing activities that regard social networking platforms as the main information channel of the campaign. Crowdturfing workers use social media accounts to spread information and may result in unfair popular popularity, such as a hijacked trending topic, in social networks.

4 Introduction

The pervasive use of social media services in recent years has revolutionized the way of information dissemination. Users of social media services have become not only information consumers, but also information producers. The openness and availabil-

ity of social media services also make it easier for malicious users to misuse social media services. Crowdturfing is such malicious activities, which are performed by real people and organized by crowdsourcing. We aim to discuss the phenomena in social media in this article.

With the increasing availability of social media, people listen to and trust opinions of online friends before they make decisions. People utilize information from social networks to find interesting movies, restaurants, etc. However, these positive opportunities also present a sinister counterpart: fake and inaccurate information intentionally spread to mislead people. Traditionally, people assumed that malicious activities were generated automatically by automated systems, so existing systems dependent on the assumption are easy to be bypassed by real users. Crowdsourcing facilitates the attacks through gathering crowdturfing workers and connecting them with potential customers. Sophisticated attacks of crowdturfing intimidate ordinary users with overwhelming unwanted information. Therefore, it is critical to detect crowdturfing on social media effectively.

Before we go further, it is essential to introduce the concept of crowdsourcing and how crowdsourcing organizes crowdturfing. Crowdsourcing websites, such as Amazon Mechanical Turks, Clickworker, and CrowdFlower, enable customers to post jobs and hire crowdsourcing workers to obtain needed services. The services can be various, ranging from answering surveys, annotating images, to translating texts. Since dedicated annotators are usually expensive to hire, it is hard to employ large-scale annotations or surveys. Crowdsourcing alleviates the burden of costs by segmenting a big task into small pieces, and connecting them with the spare time of crowdsourcing workers. In the meanwhile, however, crowdsourcing also paves the way for large-scale malicious campaigns.

Zhubajie.com (ZBJ) is one of the largest crowdsourcing websites in China, which started in 2006 and had been well established. On ZBJ, tasks are classified into different categories, such as programming (e.g., building websites) and graphic design. There is also a subsection dedicated for requests solely to crowdturfing. The crowdturfing tasks are mainly about advertising a particular product: customers usually ask crowdturfing workers to post a specific content and return corresponding screen shots for getting paid. The average budget of a crowdturfing task is around 12 to 15 dollars [28], and each a crowdturfing worker can earn 0.2 to 1 dollar on average for a single task.

The crowdturfing in social media is usually involved with spreading malicious URLs and form astroturf campaigns. The detection of crowdturfing in social media could be beneficial for improving experiences of users and maintaining the value of social media sites. In the following sections, we investigate into the existing literature about crowdturfing in social media. We discuss the challenges pertaining to the problem and explain the scientific fundamentals for studying crowdturfing. We then glean over the key research findings from related work. References to different tools, datasets, and commonly used evaluation metrics are provided for interested readers to further study the significance and challenges of crowdturfing detection in social media.

5 Key Points

The study of crowdturfing detection is subjected to various challenges. The availability of crowdsourcing sites and the openness of social media make the problem even more complicated. Below we present significant challenges noted by researchers and practitioners:

- **Absence of ground truth** Unlike traditional data mining problems, the ground truth of social media crowdturfing datasets can hardly be manually annotated. Though it is possible for human annotators to identify irregular patterns and adverse attacks, the way of organizing such campaigns, however, can hardly be told for sure. According to the definition of crowdturfing, being organized by crowdsourcing is a key element of crowdturfing. Therefore, there is a lack of availability of obtaining a benchmark dataset. The absence of ground truth exacerbates the difficulty of posing the problem as a supervised machine learning task.
- **Analyzing content** One of the biggest challenges is to analyze the massive amount of content information in social media. To detect organized crowdturfing activities, the larger size of samples may result in better and more ascertained results. However, take Twitter, for instance, the number of monthly active users exceeds 300 million and the number of tweets per day is over 500 million, it is almost impossible to process the massive data, which may be only a small portion, in a traditional setting.
- **Evolving strategies** Machine learning approaches focus on identifying patterns of adverse activities, which may be difficult to cope with novel strategies of emerging crowdturfing campaigns. For example, traditional crowdturfing detection methods [16] may only exploit the content, pictures, and videos have been increasingly used recently [11, 25]. Moreover, it is commonly assumed by existing methods that posts of crowdturfing workers are near duplicate since the content is usually pre-written by the supporters. However, sophisticated crowdturfing workers have also been found [10], who might compile more original content. The evolving crowdturfing strategy may easily bypass static filtering of traditional detection approaches.
- **Evolving participants** Crowdsourcing is popular since it provides a way for people to transform their spare time into monetization rewards. Many crowdturfing workers are working on the tasks only "part-time", so the corresponding social media accounts are usually not dedicated for crowdturfing, and contain content information about themselves instead of campaigns. The personal posts may serve as camouflage that disguises the campaigning information. So it is challenging for traditional approaches to detect such evolving crowdturfing participants with mixed content, and would require fine-grained analysis on user posts.
- **Heterogeneous data** Multimedia information is increasingly prevalent on social media sites. The multimedia information is brought by newly founded social media sites that are dedicated to images and videos such as Instagram and Pin-

terest. Existing social sites such as Facebook and Twitter are also widely used for disseminating pictures and short videos. Also, existing multimedia sites such as Youtube, are evolving with adding more social features. Therefore, images, video clips, and texts are simultaneously used for campaigns. The heterogeneity of data requires better adaptivity to cross-modal information, which is lacking in existing methods.

The different studies conducted in detecting crowdturfing have taken one or more of these challenges into consideration. We will present and discuss their major research findings in the section of Key Research Findings. Next, we will present the underlying scientific fundamentals for detecting crowdturfing in social media.

6 Historical Background

Astroturfing refers to campaigns that appear to be led by grassroots participants but is actually supported by intentionally masked sponsors. Crowdturfing can be regarded as a specific type of astroturfing, of which the campaign participants are organized by crowdsourcing. The study of astroturfing can be traced back to 1985, when insurance companies hired "astroturfs" to launch campaigns promoting their interests. Starting from 1998, researchers focused on measuring the difference between campaigns led by true grassroots and astroturfs [3]. Another example of astroturfing is the campaigns initiated by cigar industry in the late 1990s. With increasing taxes on tobacco and more regulations of smoking in the United States, cigar industry faces hard times from the early 1990s. In order to prevent the loss of income, tobacco companies, together with the *National Smokers Alliance*, initiated an aggressive campaign to protect their interests. The astroturfs sent cards and letters to advocate the rights of smokers [9].

With the development of information and communication technologies, astroturfing has evolved its way of reaching people. In addition to letters and cards, automated phone calls, websites, and emails are used to allow for making astroturfing prolific and economical. In 2001, Americans for Technology Leadership (ATL) and the Association for Competitive Technology began their advocates to make people send letters to their state attorneys general. The letters were pre-written and aimed to convince the attorneys general to drop a lawsuit against Microsoft. Later it was found out that ATL was heavily funded by Microsoft [20]. Crowdturfing becomes prevalent around 2010, benefiting from the availability of organizing a large number of participants economically. The lower cost of crowdturfing also enables small groups to launch their campaigns. On Internet platforms, such as forums, blogosphere, and social media, crowdturfing is found to be involved with political and commercial campaigns [1].

In order to detect astroturfing, traditional methods rely on checking the content manually. For example, letters from ATL and Association for Competitive Technology campaigns were found to be similar to each other; letters were even authored by people who died before they were signed [21]. However, with the availability

of crowdsourcing, manual checking is too expensive to cope with the large-scale crowdturfing. According to a study on crowdturfing websites citewang2012serf, a crowdturfing campaign may consist of 200 tasks. It would be extremely challenging to merely rely on manual checking to detect the massive amount of crowdturfing information, such as memes in social media [23]. In next section, we will present several challenges in identifying crowdturfing.

7 Detecting Crowdturfing in Social Media

Before introducing detecting crowdturfing activities in social media, we first introduce the following fundamental principles that have been used for investigating crowdturfing:

- **Mining and profiling social media users**
- **Modeling information diffusion in social media**

We will elaborate on these fundamentals in the next.

7.1 Mining and profiling social media users

In social media platforms, users post content, forward posts of friends, and receive information from their friends. People interact with each other through making friends and exchanging information. These activities induce an attributed network among social media users that has been well studied to understand patterns of crowdturfing and other adverse attacks [28, 30], which are explained below.

- **Social networks** Social media users could follow each other to subscribe other people's shared information. The behavior of following generates links, which result in a social network structure. The social network structure has been used to identify adverse behaviors in social networks. The information that has been used includes the number of links [18], contents and profiles of friends [19], and the community structures and hierarchies induced by the social network [8].
- **Conversations** In addition to sharing information, social network users can also comment each other to share and exchange opinions, which result in conversations between users. The ideas expressed in the comments can be regarded as an indicator of the quality of information. For example, if some emerging story is unlikely to be true, people usually try to verify and correct it by asking questions or commenting negatively [34].
- **Textual streams** In social networks, content is continuously generated that comprises a stream of texts. Since campaigns are time-sensitive and may emerge during specific time periods, specific patterns may exist, such as hibernation (low activity traffic volume in a long period) and abrupt bursts (high activity traffic

volume in a short period) [31]. These temporal and quantitative patterns can be a good measure for indicating abnormal activities.

- **Temporal and spatial patterns** Researchers have performed analysis on user behaviors to detect abnormal activities. A topic that is trending on social media is usually related to certain time and locations, so the time [31] and location of users [4], combined with the content they posted, can be taken as an indicator of the norm of information.
- **Information originality** The originality of information describes the uniqueness and novelty of information. Since crowdturfing is usually organized, the corresponding content should be less varied and has a more limited vocabulary. Though originality has been widely used for traditional approaches to identifying astroturfing, in the context of detecting crowdturfing, however, measuring originality of information is more challenging since the amount of information is massive, and many people may post other people's content without citation.

7.2 Modeling information diffusion in social media

Posting new content and being spread, forwarding information from other people to friends, and writing comments, these activities are imperative for social media users. The activities that portray these people as actors in social networks and help them to influence other people result in the diffusion of information between social media users over time. In this section, we will introduce fundamental principles of information diffusion, and how information diffusion models can be adapted to cope with information of crowdturfing.

- **General information diffusion models** The basic models of information diffusion include SIR model [13], tipping point model [6], independent cascade model [12] and linear threshold model [12]. Information diffusion models explain the cascade of information by studying the network topology and the mechanism of information exchange. Independent cascade model has been adopted to for social media information [7] since it takes the relations and influence of friends into consideration. Independent cascade model has also been used to model the information diffusion between social network users by assuming the relationships between people are independent [17].
- **Information diffusion for crowdturfing** In order to adapt general information diffusion models to the diffusion of information from crowdturfing, different kinds of information exchange and more roles can be introduced. For example, independent cascade model assumes the information exchange is informing, however, for disinformation from crowdturfing, the "exchange" is not only about informing but more about persuading other people to trust the information. Therefore, researchers introduce belief levels to the process of information exchange to study how the information spreads [2]. Another way is to introduce more roles. In addition to spreaders and receivers, Tian et al. introduced clarifiers

and persuaders for information diffusion, which adapts linear threshold model to the diffusion of disinformation [27].

Next, we discuss the key findings of crowdturfing detection. We segregate the key findings by the way they model social network users, i.e., content-based, behavior-based, and diffusion-based approaches.

7.3 Crowdturfing Detection

In this section, we discuss the key findings of crowdturfing detection. We segregate the key findings by the way they model social network users, i.e., content-based, behavior-based, and diffusion-based approaches. We also introduce available data resources that is useful for further investigation.

7.3.1 Content-based approaches

In order to detect crowdturfing, a possible one is to examine the social media content and identify the similar and duplicate posts. For example, Wang et al. generate behavioral signatures by exploiting user's contents [29]. Similarly, Yang et al. propose to discover the topics on which are focused by crowdturfing workers [32]. As discussed before, crowdturfing customers may ask crowdturfing workers to compile original content, so the corresponding posts may not share similar patterns that lead to distinguishable "behavioral signatures". The posts that appear to be different share the same goal of leading other users to a certain resource, such as a some user's account or an external web page. Therefore, the "targets" of posts can be used as an indicator of systematically organized content generation [26].

7.3.2 Behavior-based approaches

Since crowdturfing jobs are usually required to be done according to rules compiled by the customers, behaviors of social media users involved with crowdturfing are useful to reveal the abnormal activities. For example, the temporal pattern of uploading and clicking URLs can be used to differentiate crowdturfing [5]. In order to gain more influence on social media sites, crowdturfing workers also follow a certain account to farm links. Researchers have studied in using the behavioral patterns as well as the content information to identify crowdturfing workers on Twitter [15].

7.3.3 Diffusion-based approaches

Diffusion-based approaches aim to modify existing information diffusion models to understand and detect adverse activities. Through extending independent cascade

model and linear threshold model, information diffusion models can be used to represent the process of adverse attacks on social media sites [2, 27]. The diffusion models can not only facilitate detection of crowdturfing but also help intervene the process and minimize the aftereffects of disinformation. Ratkiewicz et al. extract the and nicely visualize the process of information diffusion [24], which helps understand the diffusion mechanism of crowdturfing campaign, and may provide a user-friendly interface to label crowdturfing content streams manually.

7.3.4 Datasets

Researchers and related practitioners have used different datasets for crowdturfing detection. An astroturfing dataset is available for Twitter textual streams [23], which is publicly available. The dataset consists of memes obtained from Twitter and contains ground truth data which are manually labeled. A malicious user dataset is available for studying behaviors of potential crowdturfing workers. The ground truth is available with the dataset, which is labeled automatically using social honeypots [14]. In order to study user behaviors from different social networks, more datasets are publicly available at ASU social media data repository [33]. It is also possible to obtain crowdturfing data through the crowdturfing websites, such as Zhubajie, Sandaha, and Fiverr.

8 Key Applications

- **TweetTracker** is an analysis tool that is designed to track, analyze, and monitor content information in social media. The tool incorporates the feature of location, and can assist researchers and practitioners with event-based data collection in media data. TweetTracker has a built-in module of detecting event-based malicious users [22], which can be used to identify propaganda participants. The tracking function can be used for crowdturfing detection by picking out duplicate contents. The tool is free to use for academic purposes upon approval.
- **Truthy** is a Twitter-based tool, which aims to detect the spread of astroturfs in social media by examining malicious accounts [23]. It is free to check an individual account from the tool.
- **Debot** is a freely available tool to check bot users on social media. Since the tool focuses on correlated and near-duplicate activities of different social media users, it could be used to detect crowdturfing workers who are organized for a campaign. The codes and data are freely available.

9 Future Directions

In this article, we have discussed major challenges, fundamental principles, and key concepts of crowdturfing detection. We focused on several noteworthy applications and datasets that afford noteworthy opportunities for further investigation. Although the topic has been extensively studied, we envision several directions for future work as our reliance on social media and its increases. With more social media sites being constructed and being popular, a crowdturfing campaign could be simultaneously launched on different social media sites. Therefore, it would be valuable to rigorously investigate cross-media crowdturfing detection. Also, with the increasing role of social media in real-world events, further research would entail not only different social media channels but also the effects of events happening in real world. For example, a company may launch campaigns during a public relation crisis; the incorporation of considering these real world events may further facilitate the detection of crowdturfing.

10 Cross References

Spam Detection on Social Networks

Anomaly Detection in Social Networks

Network Anomaly Detection Using Co-clustering

Fraud Detection Using Social Network Analysis, a Case Study

References

1. White house brushes off health-care protests. Wall Street Journal (August 4, 2009)
2. Acemoglu, D., Ozdaglar, A., ParandehGheibi, A.: Spread of (mis) information in social networks. Games and Economic Behavior **70**(2), 194–227 (2010)
3. Beder, S.: Public relations' role in manufacturing artificial grass roots coalitions. Public Relations Quarterly **43**(2), 20 (1998)
4. Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J., Gonçalves, M.: Detecting spammers and content promoters in online video social networks. In: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, pp. 620–627. ACM (2009)
5. Cao, C., Caverlee, J.: Detecting spam urls in social media via behavioral analysis. In: European Conference on Information Retrieval, pp. 703–714. Springer (2015)
6. Centola, D.: The spread of behavior in an online social network experiment. science **329**(5996), 1194–1197 (2010)
7. Chen, W., Yuan, Y., Zhang, L.: Scalable influence maximization in social networks under the linear threshold model. In: 2010 IEEE International Conference on Data Mining, pp. 88–97. IEEE (2010)
8. Gao, J., Liang, F., Fan, W., Wang, C., Sun, Y., Han, J.: On community outliers and their efficient detection in information networks. In: SIGKDD, pp. 813–822. ACM (2010)

9. Givel, M.: Consent and counter-mobilization: The case of the national smokers alliance. Journal of health communication **12**(4), 339–357 (2007)
10. Jagabathula, S., Subramanian, L., Venkataraman, A.: Reputation-based worker filtering in crowdsourcing. In: Advances in Neural Information Processing Systems, pp. 2492–2500 (2014)
11. Kaplan, A.M., Haenlein, M.: Users of the world, unite! the challenges and opportunities of social media. Business horizons **53**(1), 59–68 (2010)
12. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 137–146. ACM (2003)
13. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 115, pp. 700–721 (1927)
14. Lee, K., Eoff, B.D., Caverlee, J.: Seven months with the devils: A long-term study of content polluters on twitter. In: ICWSM (2011)
15. Lee, K., Tamilarasan, P., Caverlee, J.: Crowdturfers, campaigns, and social media: Tracking and revealing crowdsourced manipulation of social media. In: ICWSM (2013)
16. Lee, K., Webb, S., Ge, H.: The dark side of micro-task marketplaces: Characterizing fiverr and automatically detecting crowdturfing. arXiv preprint arXiv:1406.0574 (2014)
17. Lim, S.H., Kim, S.W., Kim, S., Park, S.: Construction of a blog network based on information diffusion. In: Proceedings of the 2011 ACM Symposium on Applied Computing, pp. 937–941. ACM (2011)
18. Mccord, M., Chuah, M.: Spam detection on twitter using traditional classifiers. In: Autonomic and trusted computing, pp. 175–186. Springer (2011)
19. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. Annual review of sociology pp. 415–444 (2001)
20. Menn, J., Sanders, E.: Report: Microsoft funded 'grass roots' campaign. Associated Press (August 21, 2001)
21. Menn, J., Sanders, E.: Lobbyists tied to microsoft wrote citizens' letters. The LA Times (August 23, 2001)
22. Morstatter, F., Wu, L., Nazer, T.H., Carley, K.M., Liu, H.: A new approach to bot detection: Striking the balance between precision and recall. The international conference on Advances in Social Network Analysis and Mining (2016)
23. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., Menczer, F.: Detecting and tracking the spread of astroturf memes in microblog streams. arXiv preprint arXiv:1011.3768 (2010)
24. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., Menczer, F.: Truthy: mapping the spread of astroturf in microblog streams. In: Proceedings of the 20th international conference companion on World wide web, pp. 249–252. ACM (2011)
25. Shen, J., Deng, R.H., Cheng, Z., Nie, L., Yan, S.: On robust image spam filtering via comprehensive visual modeling. Pattern Recognition **48**(10), 3227–3238 (2015)
26. Song, J., Lee, S., Kim, J.: Crowdtarget: Target-based detection of crowdturfing in online social networks. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pp. 793–804. ACM (2015)
27. Tian, R.Y., Zhang, X.F., Liu, Y.J.: Ssic model: A multi-layer model for intervention of online rumors spreading. Physica A: Statistical Mechanics and its Applications **427**, 181–191 (2015)
28. Wang, G., Wilson, C., Zhao, X., Zhu, Y., Mohanlal, M., Zheng, H., Zhao, B.Y.: Serf and turf: crowdturfing for fun and profit. In: Proceedings of the 21st international conference on World Wide Web, pp. 679–688. ACM (2012)
29. Wang, T., Wang, G., Li, X., Zheng, H., Zhao, B.Y.: Characterizing and detecting malicious crowdsourcing. ACM SIGCOMM Computer Communication Review **43**(4), 537–538 (2013)
30. Wu, L., Morstatter, F., Hu, X., Liu, H.: Mining misinformation in social media. Big Data in Complex and Social Networks (2016)

31. Xie, S., Wang, G., Lin, S., Yu, P.S.: Review spam detection via time series pattern discovery. In: Proceedings of the 21st International Conference on World Wide Web, pp. 635–636. ACM (2012)
32. Yang, X., Yang, Q., Wilson, C.: Penny for your thoughts: Searching for the 50 cent party on sina weibo. In: Ninth International AAAI Conference on Web and Social Media. Citeseer (2015)
33. Zafarani, R., Liu, H.: Social computing data repository at ASU (2009). URL http://socialcomputing.asu.edu
34. Zhao, Z., Resnick, P., Mei, Q.: Enquiring minds: Early detection of rumors in social media from enquiry posts. In: Proceedings of the 24th International Conference on World Wide Web, pp. 1395–1405. ACM (2015)

11 Recommended Reading

Mining misinformation in social media , Liang Wu, Fred Morstatter, Xia Hu, and Huan Liu. *To appear in Big Data in Complex and Social Networks, 2016.*

Social media mining: an introduction , Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. *Cambridge University Press, 2014.*