Adversarial Collusion on the Web: State-of-the-Art and Future Directions

Hridoy Sankar Dutta(✉) and Tanmoy Chakraborty

IIIT -Delhi, New Delhi, India
{hridoyd,tanmoy}@iiitd.ac.in

Abstract. The growth and popularity of online media has made it the most important platform for collaboration and communication among its users. Given its tremendous growth, social reputation of an entity in online media plays an important role. This has led to users choosing artificial ways to gain social reputation by means of blackmarket services as the natural way to boost social reputation is time-consuming. We refer to such artificial ways of boosting social reputation as *collusion*. In this tutorial, we will comprehensively review recent developments in analyzing and detecting collusive entities on online media. First, we give an overview of the problem and motivate the need to detect these entities. Second, we survey the state-of-the-art models that range from designing feature-based methods to more complex models, such as using deep learning architectures and advanced graph concepts. Third, we detail the annotation guidelines, provide a description of tools/applications and explain the publicly available datasets. The tutorial concludes with a discussion of future trends.

Keywords: Collusion · Blackmarkets · Online social networks · Social growth

1 Background

In recent years, we have seen an unprecedented popularity of online media, attracting a great number of people who want to share their thoughts and opinions. Gaining fame and reputation in online media platforms has become an important metric for several purposes – launching large-scale campaigns, promoting stocks, manipulating users' influence, and conducting political astroturfing, etc. The natural way of gaining appraisals and therefore attaining reputation can be a slow process. This has led to the creation of blackmarkets that help the online media entities to gain artificial appraisals such as followers, retweets, views, comments, etc. Gaining appraisals using blackmarket services violates the Terms and Service of online media platforms [2,3]. However, Google searches with keywords such as 'buy retweets', 'get YouTube views', etc. hit hundreds of websites offering 'easy', 'fast' and 'guaranteed' ways to increase appraisals [8,17]. This provides us with a clear picture on the popularity and impact of these services in providing appraisals across multiple online media platforms.

L. Bellatreche et al. (Eds.): BDA 2020, LNCS 12581, pp. 230–235, 2020.
https://doi.org/10.1007/978-3-030-66665-1_15

The problem is surprisingly prevalent; it is estimated that roughly 15% of Twitter accounts are operated by bots/collusive accounts [1], and more than 30% of the consumer reviews on online review/rating platforms are non-genuine [18]. Collusive entity detection is considered a challenging task as collusive users express a mix of organic and inorganic activities, and there is no synchronicity across their behaviors [4,8,10]. Thus, these users cannot be detected effectively by existing bot or fake user detection algorithms.

While the set of research for fake/fraud/spam entity detection in online media has expanded rapidly, there has been relatively little work that has studied the problem of collusive entity detection. The seminal work of [17] divided the blackmarket services into types based on their mode of operation: (i) **Premium** and (ii) **Freemium**. Premium blackmarkets provide appraisals when the customer pays money. On the other hand, in freemium services, customers themselves join the service and participate in providing collusive appraisals to other users of the service to gain (virtual) credits. In literature, it is seen that most of the existing studies on collusive entity detection are limited to online social networks and rating/review platforms. However, in reality, it is observed that collusive happens across multiple online media platforms such as video streaming platforms, recruitment platforms, discussion platforms, music sharing platforms and development platforms [9]. This further necessitates the development of advanced techniques to detect collusive entities across multiple online media platforms. Figure 1 shows the example of a blackmarket service which provides collusive appraisals to multiple online media platforms.

This tutorial presents a comprehensive overview of recent research and development on detecting the collusive entities using state-of-the-art social network analysis, data mining and machine/deep learning techniques. The topic is interdisciplinary, bridging scientific research and applied communities in data mining, human-computer interaction, and computational social science. This is a novel and fast-growing research area with significant applications and potential.

2 Objectives

The tutorial has three broad objectives:

1. To educate the community broadly about the problem of collusive activities on online media platforms.
2. To summarize advances made in the last few years in collusive entity detection with a description of annotation guidelines, publicly available collusive entity datasets, available tools and interfaces developed to detect collusive entities.
3. To encourage immediate adoption by researchers working in social network anomaly detection, as well as suggest future opportunities for short/long term investigations by researchers.

3 Organization and Outline

The tutorial will majorly cover the review progress in the area of collusive entity detection in different online media platforms. We will be presenting it in five

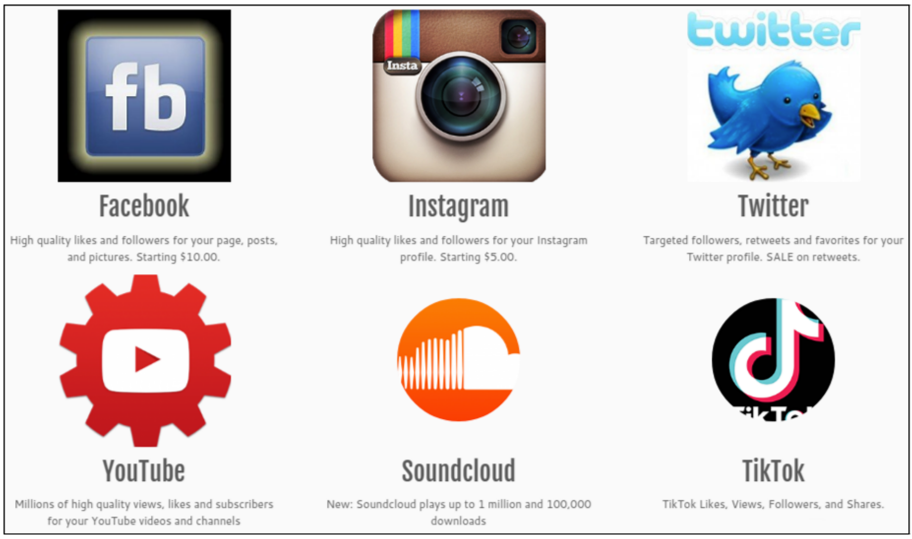

Fig. 1. Example of a blackmarket service which provides collusive appraisals to multiple online media platforms.

parts. Firstly, we start with an introduction to the problem with related definitions and concepts [9,10]. We briefly discuss the Terms and Service in online media platforms and the policy maintained by these platforms against fake and spam engagements. We also focus on how collusion is different from other relevant concepts such as fake, bot, sockpuppetry, malicious promotion, spam, content polluters, etc. We then show examples of how collusion happens across multiple online media platforms. We also highlight the challenges that we faced while designing models for collusive entity detection. We also shed some light on the limitations and restrictions of the APIs provided by the online media platforms and discuss the publicly available web scrapers and applications that helps to bypass the API rate limits. Secondly, we discuss the types of collusive activities: *individual collusion* [4–6,8,11,12,16] and *group collusion* [7,13,14,14,15,19]. We also highlight the differences between these types based on how they provide collusive appraisals. The third part is the core of our tutorial. Here, we address a number of models for collusive entity detection. Previous methods have employed different techniques ranging from feature-based methods, graph-based methods, topic modeling to complex structures using deep-learning based methods. We review the techniques that have been used for collusive entity detection and their different architectural variations. We also discuss the advantages and shortcomings of the models and their appropriateness to various types of problems. We then delve deeper into the characteristics of collusive entities by showing a variety of case studies. As case studies, we will show the effectiveness of the models on detecting collusive activities on Twitter and YouTube. In the fourth part, we focus on the annotation guidelines and description of publicly avail-

able datasets that are used for collusive entity detection. We first review the annotation guidelines for creating annotated datasets of collusive entities. Next, we provide pointers to existing studies with public links to the corresponding datasets, source codes and other related resources. A discussion of available tools and applications[1] for collusive entity detection will also be provided. In the final part, we spotlight some outstanding issues and discuss the future directions for collusive entity detection. These future research topics include: (i) collective collusion detection, (ii) understanding connectivity patterns in collusive network, (iii) event-specific studies, (iv) temporal modeling of collusive entities, (v) cross-lingual and multi-lingual studies, (vi) core collusive user detection, (vii) cross-platform spread of collusive content, (viii) multi-modal analysis of collusive entities, and (ix) how collusion promotes fake news. We believe our tutorial should interest the attending researchers and practitioners with interests in social network anomaly detection, user behavior modeling, graph mining, etc.

The outline of the tutorial is following:

1. **Introduction to collusion in online media.**
   - 1.1 Historical overview of online media platforms
   - 1.2 (Quick) Overview of collusion with definitions
   - 1.3 Example of collusive activities
   - 1.4 How collusion is different from other relevant concepts?
   - 1.5 Challenges in collusive entity detection
   - 1.6 How collusion happens across multiple online media platforms (social networks, rating/review platforms, video streaming platforms, recruitment platforms, discussion platforms, music sharing platforms, development platforms, other platforms)?
2. **Types of collusion**
   - 2.1 Individual collusion
   - 2.2 Group collusion
   - 2.3 How individual collusion differs from group collusion in providing collusive appraisals?
3. **Methods for collusive entity detection**
   - 3.1 Feature-based methods
   - 3.2 Graph-based methods
   - 3.3 Deep learning based methods
   - 3.4 Case studies
4. **Miscellaneous**
   - 4.1 How to annotate collusive entities?
   - 4.2 Description of publicly available datasets
   - 4.3 Tools, application and interfaces developed to detect collusion in online media
5. **Conclusion**
   - 5.1 Summary
   - 5.2 Open problems in collusive entity detection
   - 5.3 Future directions and discussion

[1] https://www.twitteraudit.com, https://followerwonk.com/analyze, https://botometer.iuni.iu.edu, https://www.modash.io/.

4 Target Audience

The intended audience for the tutorial are researchers of all levels seeking to understand the challenges, tasks and recent developments in collusive activities in online media. We will not assume any prerequisite knowledge and cover all the necessary concepts and definitions to ensure that the presentation is understandable to all tutorial attendees. Note that we will also cover some advanced topic materials as well.

We will also provide a detailed analysis of the existing works, a description of the publicly available datasets, tools/applications for collusive entity detection and an outline of future opportunities.

5 Presenters

Hridoy Sankar Dutta is currently pursuing his Ph.D. in Computer Science and Engineering from IIIT-Delhi, India. His current research interests include data-driven cybersecurity, social network analysis, natural language processing, and applied machine learning. He received his B.Tech degree in Computer Science and Engineering from Institute of Science and Technology, Gauhati University, India in 2013. From 2014 to 2015, he worked as an Assistant Project Engineer at the Indian Institute of Technology, Guwahati (IIT-G), India, for the project 'Development of Text to Speech System in Assamese and Manipuri Languages'. He completed his M.Tech in Computer Science and Engineering from NIT Durgapur, India in 2015. More details can be found at https://hridaydutta123.github.io/.

Tanmoy Chakraborty is an Assistant Professor and a Ramanujan Fellow at the Dept. of Computer Science and Engineering, IIIT-Delhi, India. He completed his Ph.D as a Google India PhD fellow at IIT Kharagpur, India in 2015. His primary research interests include Social Computing and Natural Language Processing. He has received several awards including Google Indian Faculty Award, Early Career Research Award, DAAD Faculty award. He leads a research group, LCS2 (http://lcs2.iiitd.edu.in), which primarily focuses on data-driven solutions for cyber-informatics. He is involved in mentoring several technology startups. For more details, please visit: http://faculty.iiitd.ac.in/~tanmoy/.

References

1. Politics and fake social media followers - lawsuit.org. https://lawsuit.org/politics-and-fake-social-media-followers/. Accessed 11 Aug 2020
2. Twitter: Platform manipulation and spam policy. https://help.twitter.com/en/rules-and-policies/platform-manipulation. Accessed 11 Aug 2020
3. Youtube: Fake engagement policy. https://tinyurl.com/yyvp68xh. Accessed 11 Aug 2020
4. Arora, U., Dutta, H.S., Joshi, B., Chetan, A., Chakraborty, T.: Analyzing and detecting collusive users involved in blackmarket retweeting activities. ACM Trans. Intell. Syst. Technol. **11**(3), 1–24 (2020)

5. Arora, U., Paka, W.S., Chakraborty, T.: Multitask learning for blackmarket Tweet detection. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 127–130 (2019)
6. Castellini, J., Poggioni, V., Sorbi, G.: Fake Twitter followers detection by denoising autoencoder. In: Proceedings of the International Conference on Web Intelligence, pp. 195–202 (2017)
7. Dhawan, S., Gangireddy, S.C.R., Kumar, S., Chakraborty, T.: Spotting collective behaviour of online fraud groups in customer reviews (2019)
8. Dutta, H.S., Chakraborty, T.: Blackmarket-driven collusion among retweeters–analysis, detection and characterization. IEEE Trans. Inf. Forensics Secur. **15**, 1935–1944 (2019)
9. Dutta, H.S., Chakraborty, T.: Blackmarket-driven collusion on online media: a survey. arXiv preprint arXiv:2008.13102 (2020)
10. Dutta, H.S., Chetan, A., Joshi, B., Chakraborty, T.: Retweet us, we will retweet you: spotting collusive retweeters involved in blackmarket services. In: ASONAM, pp. 242–249 (2018)
11. Dutta, H.S., Dutta, V.R., Adhikary, A., Chakraborty, T.: HawkesEye: detectingfake retweeters using Hawkes process and topic modeling. IEEE Trans. Inf. Forensics Secur. **15**, 2667–2678 (2020)
12. Dutta, H.S., Jobanputra, M., Negi, H., Chakraborty, T.: Detecting and analyzing collusive entities on YouTube. arXiv preprint arXiv:2005.06243 (2020)
13. Gupta, S., Kumaraguru, P., Chakraborty, T.: MalReG: detecting and analyzing malicious retweeter groups. In: CODS-COMAD, pp. 61–69. ACM (2019)
14. Kumar, S., Cheng, J., Leskovec, J., Subrahmanian, V.: An army of me: sockpuppets in online discussion communities. In: Proceedings of the 26th International Conference on World Wide Web, pp. 857–866 (2017)
15. Liu, S., Hooi, B., Faloutsos, C.: HoloScope: topology-and-spike aware fraud detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 1539–1548 (2017)
16. Shah, N.: FLOCK: combating astroturfing on livestreaming platforms. In: WWW, pp. 1083–1091 (2017)
17. Shah, N., Lamba, H., Beutel, A., Faloutsos, C.: OEC: open-ended classification for future-proof link-fraud detection. CoRR abs/1704.01420 (2017). http://arxiv.org/abs/1704.01420
18. Streitfeld, D.: The Best Book Reviews Money Can Buy, Tulsa, Okla (2012). http://www.todroberts.com/USF/BookReviews_for_Sale.pdf
19. Wang, Z., Gu, S., Zhao, X., Xu, X.: Graph-based review spammer group detection. In: KAIS, pp. 1–27 (2018)