2018 IEEE Fourth International Conference on Big Data Computing Service and Applications

Analyzing Opinion Spammers' Network Behavior in Online Review Systems

Euijin Choo
Qatar Computing Research Institute
Email: echoo@qf.org.qa

***Abstract*—Review systems are indisputably important resource for users when making various decisions on products or services. Consequently, they become increasingly targeted by attackers who deliberately inject biased reviews, so called opinion spams, aiming to influence normal users' decisions for financial gain. In this paper, we perform an empirical analysis of opinion spammers on one of the famous online review systems, Yelp. Specifically, we analyze two different types of networks: implicit and explicit networks of opinion spammers and those of non-spammers. Through analyzing the network characteristics in different networks, we show similarities and differences between opinion spammers and non-spammers in terms of statistical characteristics and network properties. More specifically, through extensive analysis on Yelp dataset, we show that (i) the explicit network of non-spammers exhibits typical "small-world" properties of social networks. (ii) the implicit network of non-spammers is close to random networks. (iii) in both explicit and implicit networks, opinion spammers form near-isolated communities with dense inner connections among themselves, while exhibiting the lower level of "small-world" properties.**

I. Introduction

As public opinions shared via social media have started to play a key role in people's decision making process, it also opens possibilities for malicious parties to manipulate with fraudulent opinions, so called *opinion spams* [1–3, 11]. In fact, many reports suggested that nearly 20% of reviews in online review systems are written by opinion spammers [4, 5]. To help users get more credible information, researchers have proposed a few methods to detect opinion spams [6–11]. Although previous research has focused primarily on detecting opinion spams using pure content-based classifiers, it is often easy to manipulate review contents.

In this paper, we thus explore an alternative direction that does not attempt to detect opinion spams, but tries to study how spammers collaboratively interact with each other to promote themselves. To this end, we investigate similarities and differences between network behavior of opinion spammers and non-spammers, which will be helpful to design a new network-based approach to detect opinion spams without analyzing metadata.

Our work is done on Yelp dataset. Yelp provides a review filtering system to tag reviews as *not-recommended*. We assume that reviews tagged as not-recommended on Yelp are opinion spams and others as non-spam reviews. As normal users will not likely to involve in any opinion spams, we assume that if the reviewer posted at least one spam review, we consider them as opinion spammers; otherwise, non-spammers.

In this research, we analyze two different types of network behavior of reviewers: implicit and explicit. By an *explicit network*, we mean the actual social friend relationship on Yelp; by an *implicit network*, we mean users' relationships built through their interactions in the system. Specifically, users can compliment to/vote for each other/each other's review to show their positive feelings towards the reviewer/review on Yelp. We consider such actions (i.e., compliments) as interactions between users. Apparently, users will be likely to believe those reviewers with many compliments from others [1]. Indeed, Yelp encourages users to send compliments to each other and takes into account the number of compliments to recognize trustworthy users, called *Elite Squads* .

In our previous research, in online review systems, we have observed that non-spammers build natural implicit communities because of their common interests; whereas, opinion spammers tend to build artificial implicit networks to promote themselves, while having positive interactions (i.e., compliments) with their colluders [1]. In this paper, we further study whether there exists any relationship between such implicit networks and explicit networks.

Our main contributions are summarized as follows.

0-7695-6396-1/18/$31.00 ©2018 IEEE
DOI 10.1109/BigDataService.2018.00048

- We offer observations on users' implicit and explicit network behavior in online review systems by analyzing their social friend relationships and interactions in online review systems.
- To the best of our knowledge, this is the first attempt to study the differences between implicit and explicit networks in online review systems with regard to the opinion spam detection.
- We show that the significant differences between implicit and explicit networks of opinion spammers and non-spammers. More specifically, we show that both implicit and explicit networks of opinion spammers do not show typical "small-world" properties of social networks. We also show that whereas the implicit network of non-spammers is close to a random network, the implicit network of opinion spammers rather forms near-isolated communities with dense inner connections.

The remaining parts of this paper are organized as follows. In Section II, we review related work. Section III describes our datasets used in this study. In Section IV, we describe two different types of networks: implicit and explicit networks among users. Sections V presents our experimental results. Finally, in Section VI, we conclude the paper.

II. Related Work

To detect opinion spams, a number of approaches have been proposed [2, 3, 6, 13]. Arjun *et al*. defined 9 behavioral features of opinion spammers including bursty reviews, duplicate/near duplicate reviews, and the distribution of review ratings [6]. Ott*et al*. have integrated work from psychology and computational linguistics to develop and compare three approaches to detect deceptive opinion spam [3]. In *Cornell*, authors created and published large-scale and publicly available gold standard dataset for opinion spam research, containing 400 genuine reviews and 400 opinion spams and showed their classifier is nearly 90% accurate on their gold-standard opinion spam dataset. Kim *et al*. proposed a frame-based deep semantic analysis method [2].

A few graph-based approaches have also been proposed [7, 8, 14–16]. Wang *et al*. proposed a a heterogeneous graph model with three different types of nodes (i.e., reviewers, reviews, and businesses) to detect opinion spams through analyzing relationships among the three types of nodes [7]. Rayana *et al*. proposed a unified spam detection framework, SpEagle, to utilizes both the metadata such as texts and the relational data [16]. Akoglu *et al*. proposed a spam detection framework, FraudEagle, exploiting the network effect among reviewers and businesses [15].

III. Dataset

This section explains the dataset used in this study. Specifically, we describe a few characteristics of this dataset. We have implemented a data crawler with Java to collect data from Yelp, one of the most popular online review systems. Specifically, we collected information about businesses, reviews for businesses, reviewers, compliments for reviews, complimenters, and *Yelp Friends* information.

Yelp provides a review filtering system to tag reviews as *not-recommended* and/or to remove highly suspicious reviews. Yelp is not only a well-known system but also its filtering system has been shown to be effective over the past few years [12]. We thus assume that reviews tagged as not-recommended on Yelp are opinion spams and others as non-spam reviews. As normal users will not likely to involve in any opinion spams, we assume that if the reviewer posted at least one spam review, we consider them as opinion spammers; otherwise, non-spammers.

In details, our dataset includes 338,284 reviewers and 440,178 reviews. Specifically, we collected 249,356 recommended reviews from 182,122 reviewers (non-spammers) and 161,025 not recommended reviews from 133,740 reviewers (opinion spammers).

On Yelp, users can also send *compliments* to each other to be published in public. To study the difference between the implicit networks of opinion spammers and non-spammers, we collected *compliments* each user received. In our dataset, there are 462,074 compliments from 51,604 complimenters to 4,242 reviewers.

Yelp provides its own social network services, *Yelp Friends*. To study the explicit networks of opinion spammers and non-spammers, we collected *Yelp Friends* information of each reviewer.

IV. Network Type

In this section, we describe two different types of networks: implicit and explicit networks of users.

A. *Implicit Network*

To study the implicit network of users, we focus on two types of users' actions in a review system: reviewing and complimenting. A reviewer writes reviews about businesses and a complimenter posts compliments on other reviewers/reviews. We consider a user's complimenting action as an interaction between users. More specifically, an interaction from user u to user v is formed if u complimented on v's review. Also, we assume that there exists an *implicit* relationship from u to v, if there are interactions from u to v.

Unlike other review systems, compliments on Yelp are not necessarily related to specific reviews, but reviewers can interact with each other through compliments for typical message exchange. As we are interested in opinion spammers who want to make their reviews visible to others and to promote their businesses, we only consider complimenting on others' reviews as user interactions. Although the term, *compliment*, itself implies positive meaning, we notice that the sentiments of the compliments from reviewers are not always positive yet could be neutral. Because of the same reason discussed, we further analyze the sentiment of compliments, and focus on positive compliments.

User interactions often randomly occur depending on their interests on businesses and the sentiment of interactions will rely on the quality of the reviews [1]. However, if users have favoring connections with each other, those compliments can be intentionally posted by opinion spammers to promote themselves. In other words, the goal of opinion spammers would be getting more compliments to look like trustworthy reviewers.

We represent implicit relationships between opinion spammers and those between non-spammers as two separate graphs, called opinion spammers' *implicit network* (Fig.1) and non-spammers' *implicit network*, respectively. Each graph is a directed graph $G = (U, E)$ where U represents users (nodes) and E represents implicit relationships (edges). Each edge $\overrightarrow{e_{uv}}$ has direction from complimenter u to reviewer v. A complimenter has outgoing relationships (edges), and a reviewer has incoming relationships (edges) in a graph.

B. *Explicit Network*

To study the explicit network of users, we analyze *Yelp Friends* information. We assume that there exists an *explicit* relationship between reviewers, if they are Yelp Friends, as a counterpart to an implicit relationship. We represent explicit relationships between opinion spammers and those between non-spammers as two separate graphs, called opinion spammers' *explicit network* (Fig.2) and non-spammers' *explicit network*, respectively. Each graph is an undirected graph $G = (U, E)$ where U represents users (nodes) and E represents relationships (edges).

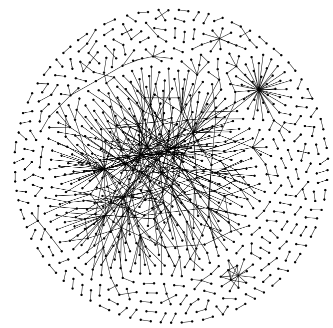


Fig. 1: An Implicit Network among Opinion Spammers

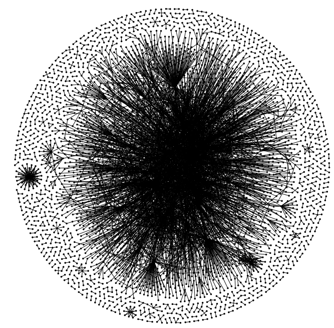
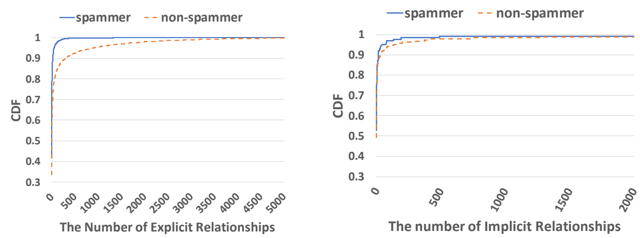


Fig. 2: An Explicit Network among Opinion Spammers

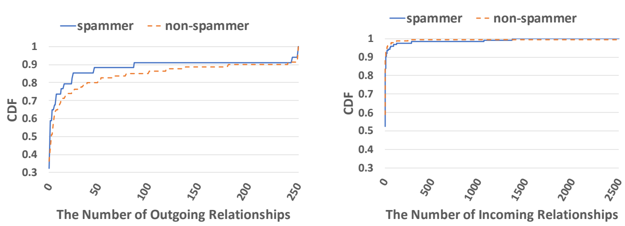
V. Analysis of Implicit and Explicit Networks of Opinion Spammers and Non-Spammers

In this section, we first discuss the statistical similarities/differences between implicit and explicit networks of opinion spammers and non-spammers (Section. V-A). Then, we analyze their characteristics in terms of their network properties (Section. V-B). We present the results in a series of plots where: the X-axis demonstrates the appropriate measure; the Y-axis presents Cumulative Distribution Function (i.e., corresponding portion of users); the dashed line represents non-spammers; the solid line represents opinion spammers.



(a) Explicit Network (b) Implicit Network

Fig. 3: The Distribution of the Number of Relationships in Explicit and Implicit Networks



(a) Outgoing Relationships (b) Incoming Relationships

Fig. 4: The Distribution of the Number of Outgoing and Incoming Relationships in Implicit Networks

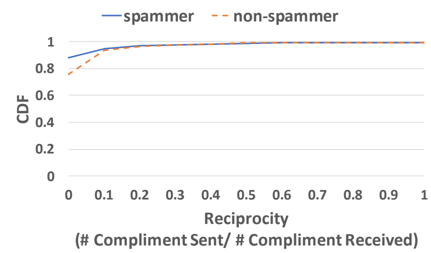


Fig. 5: The Distribution of Reciprocity

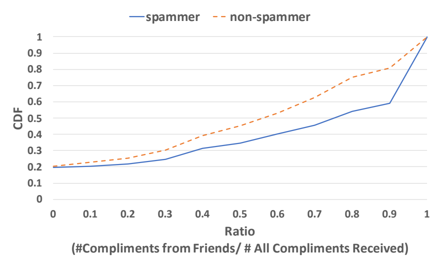


Fig. 6: The Distribution of the Number of Compliments from Explicit Relationships

A. *Statistical Characteristics*

Fig.3(a) and Fig.3(b) present the distribution of the number of explicit and implicit relationships, respectively. The X-axis represents the number of relationships, and the Y-axis represents CDF. As shown in Fig.3, opinion spammers have a relatively small number of implicit and explicit relationships compared to non-spammers.

We further plot the distribution of the number of outgoing and incoming relationships in implicit networks in Fig.4(a) and Fig.4(b), respectively. The X-axis represents the number of outgoing and incoming relationships, respectively; the Y-axis represents CDF. As shown in Fig.4, the number of outgoing relationships of non-spammers are relatively higher than opinion spammers; in contrast, the number of incoming relationships of opinion spammers are relatively higher than non-spammers. To study this difference, we further measure the reciprocity.

The *reciprocity* is defined by the ratio of the number of compliments a user sent to the number of compliments the user received. Fig.5 presents the distribution of the reciprocity of opinion spammers and non-spammers. The X-axis represents the reciprocity and the Y-axis represents CDF. As shown in Fig.5, opinion spammers have less reciprocity, compared to non-spammers. This is because, opinion spammers usually collaboratively compliment on a few targeted spammers, resulting in two types of spammers: a large group of spammers, each of whom has a few outgoing relationships (who compliment on other collaborators); and a few targeted spammers who have many incoming relationships from others and would benefit from spamming. We will further study how this phenomena will have influence on the network structure of opinion spammers in Section V-B.

To show that the compliments spammers received are indeed from colluding spammers, we measure the ratio of the number of compliments from users' explicit relationships to the number of all comments the users received. For spammers, we considered the number of compliments on their spam reviews. As it is often hard to judge the trustworthiness of reviews based on the texts, non-spammers could be fooled and might have complimented on the spam reviews. If spammers and their collaborating spammers are complimenting on their spam reviews, however, it would be suspicious.

Fig.6 presents the distribution of the ratio of the number of compliments from their explicit relationships. The X-axis represents the ratio of the number of compliments from users' explicit relationships, and the Y-axis represents CDF. Recall that non-spammers may naturally build implicit communities, as compliments are one means to typical message exchange on Yelp and the users sharing the same interest on business may compliment on each other's based on the quality of reviews. As shown in Fig.6, however, the ratio of compliments spammers received for their spam reviews from their explicit relationships are even higher than the ratio of compliments non-spammers received from their explicit relationships. On average, opinion spammers received 67.3 % compliments on their spam reviews from their explicit relationships.

B. Network Characteristics

In this Section, we measure the network properties of implicit and explicit networks of opinion spammers and non-spammers. Table.I represents the summary of network characteristics of opinion spammers' and non-spammers' implicit and explicit networks. In the following, we will describe each measure and the corresponding results in details.

Modularity is one means to measure the density of networks [18]. A network with high modularity has multiple communities having dense connections between the nodes within each community but sparse connections between nodes in different communities. As shown in Table.I, the modularity of opinion spammers' implicit and explicit networks are relatively high compared to modularity of non-spammers' implicit and explicit networks. On the other hand, the difference between the modularity of opinion spammers' explicit network and that of non-spammers' explicit network is only 0.022; whereas the difference between the modularity of opinion spammers' implicit network and that of non-spammers' implicit network is 0.332. This indicates that although opinion spammers have similar explicit network structure to non-spammers; their complimenting interactions are focused on specific targeted reviewers, resulting in high modularity.

The average shortest path length is one means to measure the efficiency of information flow on a graph, which is defined as the average number of steps along the shortest paths for all possible pairs of network nodes [19]. As shown in Table.I, the average shortest path lengths of opinion spammers' explicit network, non-spammers' explicit network, and non-spammers' implicit network are 4.272, 4.623, and 4.272, respectively; which is within the range of typical average shortest path length of social networks [20]. On the other hand, the average shortest path length of opinion spammers' implicit network is 2.173. Along with the result of modularity, this small value suggests that opinion spammers are building isolated implicit communities within which spammers are densely connected to each other.

The clustering coefficient of a node in a graph quantifies how close its neighbours are to being a complete graph [20]. Accordingly, a high clustering coefficient indicates that nodes form a complete graph with their immediate neighbors. We measure the average clustering coefficient to analyze the density of each network. As shown in Table.I, the average clustering coefficients of opinion spammers' and non-spammers' explicit networks are 0.116 and 0.164, respectively; which is within the range of typical average clustering coefficient of social networks [20]. On the other hand, the average clustering coefficients of opinion spammers' and non-spammers' implicit networks are much smaller than those of explicit networks; 0.011 and 0.045, respectively. Small average clustering coefficients along with long average shortest path often indicates that the network does not hold "small-world" properties yet it is similar to the random network [20]. Recall that, however, the average shortest path length of opinion spammers' implicit network is the shortest among those of four networks (i.e., the spammers' explicit and implicit networks, and non-spammers' explicit and implicit networks). This means that opinion spammers' implicit network is isolated and has dense inner connections.

VI. Conclusion

In this paper, we aim to empirically analyze two different types of network behaviour of reviewers: implicit and explicit in online review systems. To achieve visibility of reviewers in the system, reviewers are often advised to have more friends and interactions with each other. We thus analyze the Yelp Friend network (explicit) and compliment (interaction) network (implicit) of opinion spammers and non-spammers. We show that there exist significant differences between opinion spammers and non-spammers in terms of statistical network characteristics in implicit and explicit network behavior. More specifically, opinion spammers often build isolated dense inner network, which does not follow typical a "small world" property, in contrast to the explicit and implicit network of non-spammers.

Measure	Spammer Explicit	Spammer Implicit	Non-Spammer Explicit	Non-Spammer Implicit
Modularity	0.704	0.907	0.682	0.575
Average Shortest Path Length	4.272	2.173	4.623	4.272
Average Clustering Coefficient	0.116	0.011	0.164	0.045

TABLE I: Network Properties of Different Networks

REFERENCES

[1] E. Choo, T. Yu, and M. Chi, "Detecting opinion spammer groups and spam targets through community discovery and sentiment analysis," *Journal of Computer Security*, vol. 25, no. 3, pp. 283–318, 2017.

[2] S. Kim, H. Chang, S. Lee, M. Yu, and J. Kang, "Deep semantic frame-based deceptive opinion spam analysis," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 2015, pp. 1131–1140.

[3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 2011, pp. 309–319.

[4] M. Ott, C. Cardie, and J. Hancock, "Estimating the prevalence of deception in online review communities," in *Proc. of the 21st international conf. on World Wide Web*, 2012, pp. 201–210.

[5] BBC, "Yelp admits a quarter of submitted reviews could be fake," 2013, http://www.bbc.com/news/technology-24299742.

[6] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in *Proc. of the 19th ACM SIGKDD international conf. on Knowledge discovery and data mining*, 2013, pp. 632–640.

[7] G. Wang, S. Xie, B. Liu, and P. S. Yu, "Review graph based online store review spammer detection," in *Data Mining (ICDM), 2011 IEEE 11th International Conf. on*, 2011, pp. 1242–1247.

[8] J. Niu, L. Wang, Y. Chen, and W. He, "Detecting collusive cheating in online shopping systems through characteristics of social networks," in *Computer Communications Workshops, 2014 IEEE Conf. on*, 2014, pp. 311–316.

[9] Y.-R. Chen and H.-H. Chen, "Opinion spam detection in web forum: a real case study," in *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2015, pp. 173–183.

[10] Y. Yuan, S. Xie, C.-T. Lu, J. Tang, and S. Y. Philip, "Interpretable and effective opinion spam detection via temporal patterns mining across websites," in *Big Data (Big Data), 2016 IEEE International Conference on*. IEEE, 2016, pp. 96–105.

[11] M. Jiang, P. Cui, and C. Faloutsos, "Suspicious behavior detection: Current trends and future directions," *Intelligent Systems, IEEE*, vol. 31, no. 1, pp. 31–39, 2016.

[12] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing," in *Seventh International AAAI Conf. on Weblogs and Social Media*, 2013.

[13] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in *Proc. of the 21st international conf. on World Wide Web*, 2012, pp. 191–200.

[14] J. Ye and L. Akoglu, "Discovering opinion spammer groups by network footprints," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2015, pp. 267–282.

[15] L. Akoglu, R. Chandy, and C. Faloutsos, "Opinion fraud detection in online reviews by network effects." *ICWSM*, vol. 13, pp. 2–11, 2013.

[16] S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," in *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2015, pp. 985–994.

[17] D. Mayzlin, Y. Dover, and J. A. Chevalier, "Promotional reviews: An empirical investigation of online review manipulation," National Bureau of Economic Research, Tech. Rep., 2012.

[18] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[19] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu, "Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 71–80.

[20] C. Wilson, A. Sala, K. P. Puttaswamy, and B. Y. Zhao, "Beyond social graphs: User interactions in online social networks and their implications," *ACM Transactions on the Web (TWEB)*, vol. 6, no. 4, p. 17, 2012.