Detecting Incentivized Review Groups With Co-Review Graph

Yubao Zhang
University of Delaware

Shuai Hao
Old Dominion University

Haining Wang
Virginia Tech, Arlington, Virginia

Original Publication Citation

Zhang, Y., Hao, S., & Wang, H. (2021). Detecting incentivized review groups with co-review graph. *High-Confidence Computing*, 1(1), 1-10, Article 100006. https://doi.org/10.1016/j.hcc.2021.100006

High-Confidence Computing 1 (2021) 100006

Contents lists available at ScienceDirect

High-Confidence Computing

journal homepage: www.elsevier.com/locate/hcc

Detecting incentivized review groups with co-review graph

Yubao Zhang[a], Shuai Hao[b], Haining Wang[c,*]

[a] Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA
[b] Department of Computer Science, Old Dominion University, Norfolk, VA 23529, USA
[c] Bradley Department of Electrical and Computer Engineering, Virginia Tech, Arlington, VA 22203 USA

ARTICLE INFO

Keywords:
Incentivized review groups
Co-review graph
Community detection

ABSTRACT

Online reviews play a crucial role in the ecosystem of nowadays business (especially e-commerce platforms), and have become the primary source of consumer opinions. To manipulate consumers' opinions, some sellers of e-commerce platforms outsource opinion spamming with incentives (*e.g.*, free products) in exchange for *incentivized reviews*. As incentives, by nature, are likely to drive more biased reviews or even fake reviews. Despite e-commerce platforms such as Amazon have taken initiatives to squash the incentivized review practice, sellers turn to various social networking platforms (*e.g.*, Facebook) to outsource the incentivized reviews. The aggregation of sellers who request incentivized reviews and reviewers who seek incentives forms *incentivized review groups*. In this paper, we focus on the incentivized review groups in e-commerce platforms. We perform the data collections from various social networking platforms, including Facebook, WeChat, and Douban. A measurement study of incentivized review groups is conducted with regards to group members, group activities, and products. To identify the incentivized review groups, we propose a new detection approach based on co-review graphs. Specifically, we employ the community detection method to find the suspicious communities from co-review graphs. We also build a "gold standard" dataset from the data we collected, which contains the information of reviewers who belong to incentivized review groups. We utilize the "gold standard" dataset to evaluate the effectiveness of our detection approach.

1. Introduction

Online reviews on commercial products and services extensively impact consumers' decision making. As reported, 90% of consumers read online reviews before purchasing a product or service, and 88% of consumers trust online reviews as much as personal recommendations [3]. Online reviews are today's "word of mouth" marketing; they could either make or break your business. About 80% of consumers reverse product purchase decisions after reviewing negative online reviews, and 87% affirm a purchase decision based on positive online reviews [9].

Therefore, today's merchants are strongly motivated to fabricate the online reviews in order to manipulate the custom opinions. One of the most popular way for fabricating positive reviews is called *incentivized reviews, i.e.*, merchants bribe reviewers by providing free products or even offer a compensation for favorable reviews (*e.g.*, five-star reviews on Amazon). With incentivized reviews, merchants could gain a competitive advantage over rival merchants, as customers prefer online products with larger number of favorable reviews.

To further affect people's thoughts and decisions, incentivized reviews are collected from a group of reviewers (*i.e.*, the *incentivized review groups*) so as to perform opinion spamming. In particular, incentivized review groups are online venues for trading reviews, where merchants can post the products that need favorable reviews and reviewers can write favorable reviews to obtain free products or even make extra compensation. Some of the merchants designate well-written reviews to reviewers such that they can guarantee the quality of incentivized reviews. As such, there emerges a shady business that acts as a go-between of merchants and consumers, such as review outsourcing websites.

Apparently, the underground industry of fabricating fake reviews mentioned above violates the rule of most e-commerce platforms, including Amazon. As Amazon consumer review policy [2] states, the violations include "a seller posts a review of their own product or their competitor's product" and "a seller offers a third party a financial reward, discount, free products, or other compensation in exchange for a review", etc. Despite the strict prohibition of Amazon (*i.e.*, banning accounts of both merchants and consumers), incentivized review groups are still thriving across different platforms, such as Facebook, WeChat, and websites. This shady industry produces a spate of fake reviews, which mislead the customers, damage the trust of reviews, and even endanger the healthiness of the e-commerce ecosystem.

In this paper, we focus on incentivized review groups on e-commerce platforms, *e.g.*, Amazon. To understand the breadth of the problem, we investigate incentivized review groups across several different platforms, including Facebook, Wechat, and Douban. With the data col-

* Corresponding author.
E-mail address: hnw@vt.edu (H. Wang).

https://doi.org/10.1016/j.hcc.2021.100006
Received 3 February 2021; Accepted 28 February 2021

lected from different platforms, we integrate the data from different data sources and examine incentivized review groups. We find that different platforms play different roles in the ecosystem of incentivized review groups. Specifically, incentivized review groups on Facebook act like blackboards, where a set of merchants post their products directly in these Facebook groups. Meanwhile, incentivized review groups on Douban are of service for merchants and brokers, which educate them how to effectively obtain incentivized reviews; and incentivized review groups on WeChat are most private and generally are owned by a single person, who recruits reviewers to join the group and posts review requests for a set of products.

To understand the incentivized review groups, we conduct a measurement study to characterize real review groups from different aspects. We investigate the number and the increment rate of review members, as well as the number of merchants in collected incentivized review groups. In terms of incentivized review requests, we inspect the incentivized review requests in different groups as well as from individual merchants. We also examine the categories, questions & answers, and the relationship between sellers[1] and manufacturers of products.

Based on the measurement study, we then propose a graph-based method to detect the incentivized review groups on Amazon. Our detection method is to leverage the co-review behavior among reviewers. *Co-review* means two reviewers post reviews on a same product. To this end, we first construct co-review graphs of reviewers and then employ the community detection method to find the suspicious communities. Specifically, we not only consider the frequency of co-reviews, but also use important features of the co-review behavior, such as co-reviews in a burst. A burst of favorable reviews of a product could imply the existence of incentivized reviews. Therefore, we leverage the burst of favorable reviews to improve the detection accuracy of incentivized review groups.

To evaluate our detection method, we construct a “gold standard” dataset from our data collection. The “gold standard” dataset is guaranteed by using the collection of real incentivized review groups, which enables us to validate the effectiveness of our method and even shed light on further fake review researches.[2] We also examine an extensive Amazon review dataset ranging from May 1996 to July 2014 and find that incentivized review groups posed nearly no threat on the ecosystem before 2014 [10].

The remainder of this paper is structured as follows. Section 2 presents the background of incentivized review groups. In Section 3, we demonstrate our data collection method. In Section 4, we conduct a measurement study on incentivized review groups in terms of members, review requests, and products. In Section 5, we present our detection method based on co-review graphs. Section 6 discusses the limitations of this work. Section 7 surveys the related work, and finally, Section 8 concludes the paper.

2. Background

Obtaining positive reviews is one major factor of being successful online sellers. Positive reviews, a vital form of social proof, indicate that the product is of a high quality and hence dissuade customers’ fears of purchasing the product online. When competing with similar products with similar price, the product with higher rate or better reviews is more likely to win out.

2.1. Incentivized reviews

To obtain positive reviews in a short term, sellers provide free products or even offer a compensation. These reviews are called “*incentivized reviews*”. With the incentive for reviewers, it is guaranteed that sellers can obtain positive reviews (such as five-star in Amazon) and enhance the rate of the products expeditiously. However, incentivized reviews violate the policy of Amazon since they are published in exchange for free products or compensation. Amazon changed the policy in 2016 to ban the incentivized reviews [1].

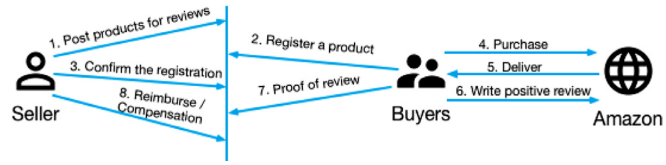


Fig. 1. Amazon incentivized review group.

2.2. Verified purchase

Around the same time when Amazon started the crackdown on incentivized reviews, Amazon introduced “verified purchase” tag. A “verified purchase” tag is placed on the review if Amazon can verify that the review was published by the account that made the purchase. Although “verified purchase” tag can highlight some authentic reviews and hinder the spam reviews to a certain degree, crooked sellers can bypass the hurdle or even exploit the “verified purchase” through review groups.

2.3. Incentivized review group

Incentivized review groups, also called incentivized review clubs, are communities created to connect consumers who want free products or compensation, and sellers who want positive product reviews. Fig. 1 shows how incentivized review groups work. First, a seller posts the products that need reviews and buyers register for particular products of their interest. After the registration is confirmed by the seller, buyers purchase the products in Amazon and write favorable reviews after the delivery. Up this point, they can show the proof of favorable reviews to the seller and obtain the reimbursement or compensation. The registration process enables the seller to follow up and ensure that the buyers have posted the reviews and the reviews are favorable (such as five stars in Amazon).

Since buyers make payments on Amazon at full price, they are eligible for posting “verified purchase” reviews. Once the reviews have been confirmed, sellers send the cost of their purchases back, sometimes plus a compensation. Despite Amazon’s strict policy against incentivized review groups (such as banning the accounts), a number of incentivized review groups are still operating on social networking platforms or websites. There are a great number of incentivized review groups on Facebook, which are set up specifically for Amazon sellers. Incentivized review groups usually set their groups as private or requiring sign-up to view the posts on Facebook to disguise themselves. Some of them claim the rules of incentivized review groups, including no scam, no hate speech, no cheating and encouraging users to report invalid posts (especially without stating refund term). Sellers also run the incentivized review groups in other websites (such as Reddit) or instant messaging applications (such as WeChat[3]).

3. Data collection

In this section, we describe the data collection mechanism of incentivized review groups and summarize the datasets. We collect incentivized review groups from various social networking platforms, including Facebook, WeChat, and Douban. WeChat is the most popular instant

[1] Throughout this paper, merchant is interchangable with seller.

[2] We plan to make our dataset publicly available with the publication of the article.

[3] A popular instant messaging application in China.

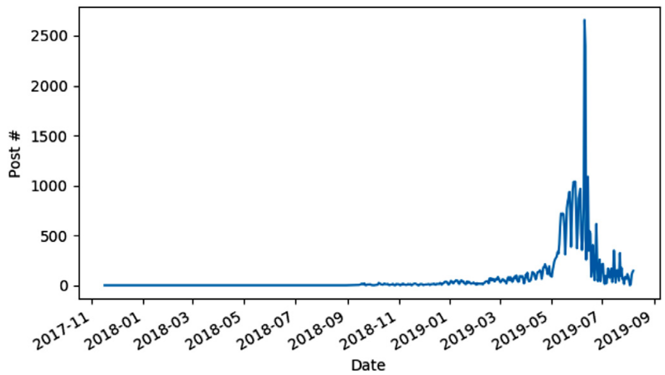


Fig. 2. Facebook review groups.

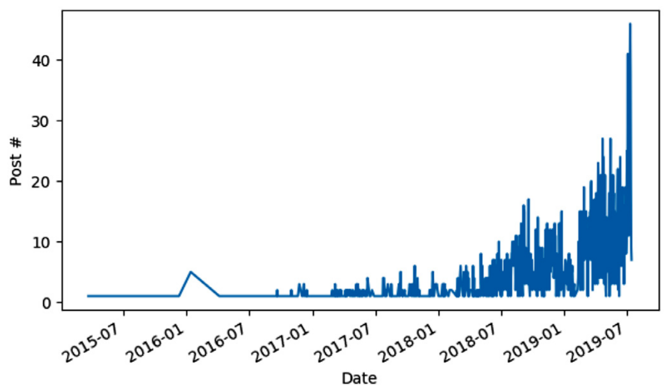


Fig. 3. Douban review groups.

messaging application in China, which allows users to create groups to broadcast events to group members. Douban is one of the most influential social networking service website in China, which allows users to create interest groups to share information.

Facebook:

There are a number of Facebook groups that are abused by incentivized review groups. Some of them are private and only allow group members to view the posts. We monitor 20 popular incentivized review groups in Facebook, including both private and public ones. Some of public groups turned into private during our collection and we need to send requests to join. Our collection of the groups ranges from November 1, 2017 to August 7, 2019. We collected a total of 47,148 posts created by 6260 Facebook accounts. Fig. 2 shows the number of posts over the collection period, which indicates the overall activity of these review groups over time.

Douban:

Sellers create interest groups in Douban to share review exchange information. We collect ten incentivized review groups in Douban ranging from May 1, 2015 to August 7, 2019. We collect a total of 3,762 posts from 1,226 authors. We find more than 1,000 WeChat accounts in these posts. Fig. 3 shows the number of posts against time. We find that the incentivized review groups have been becoming increasingly active over time.

WeChat:

WeChat group is an ideal place for sellers to broadcast their products since the WeChat group is private and it also offers convenience for further processing and making payment. We send requests to join one WeChat group found on Douban and collect the review requests and members' responses over a month ranging from July 7, 2019 to August 7, 2019. In this group, one broker is posting products for several sellers. Fig. 4 shows the number of products against time.

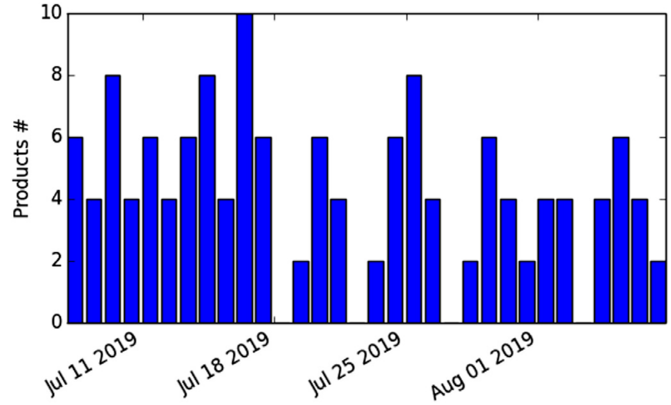


Fig. 4. A review group in WeChat.

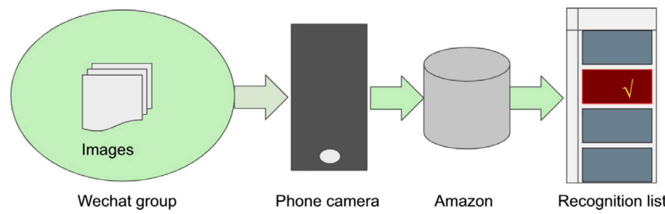
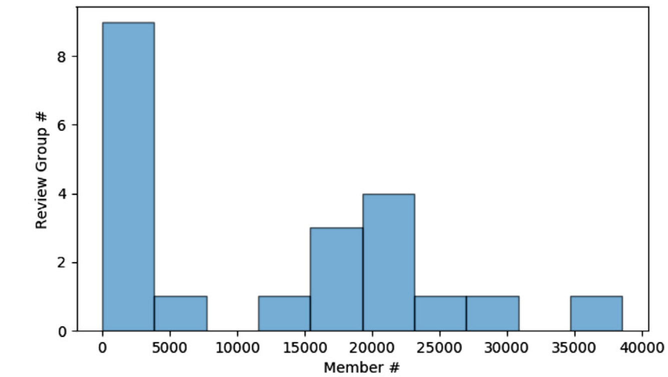


Fig. 5. Amazon product collection.

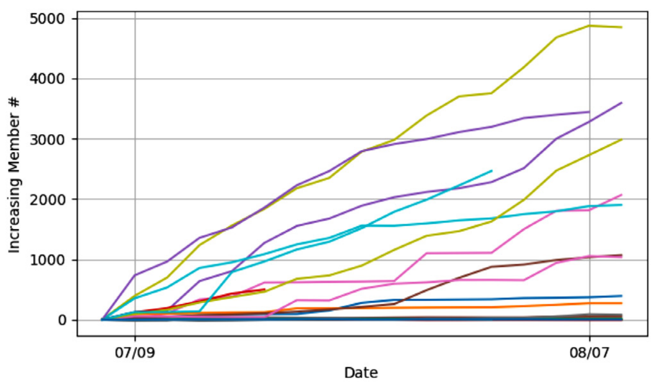
For the purpose of protecting them from detection, sellers are not publishing the URLs of products, but only images and a short introduction. It poses a challenge to the collection of products involved in the incentivized review groups. To this end, we employ the image recognition to collect the corresponding products with the images collected from the group, as shown in Fig. 5. Specifically, we utilize the *Camera Search* feature on Amazon mobile applications with mobile phones to search the products. If Amazon can identify the product, it will pop up its search results allowing you to view detailed information for the product your camera captured. We need to check multiple product images to guarantee that we find the right product in review groups. It is observed that sellers will copy some parts of product images from other sellers, but scarcely copy all of them. Therefore, we can distinguish the products with a collection of product images collected from incentivized review groups. We manually utilize the image recognition module of Amazon Apps on mobile phones to recognize one image at a time, identify the correct product from the search list, and collect the URLs of the product. We identify 93 products with image recognition in total from about 200 products posted in the incentivized review group. We then collect the reviews and product information of these products.

Summary: From above datasets, we find that different platforms play different roles in review groups. The review groups in Facebook are like blackboards, where a set of sellers can post their products directly. In our dataset, there are more than 6,000 sellers who posted products. In the review groups of Douban, most of the posts are to educate sellers how to obtain incentivized reviews and advertise the brokers who can help sellers. In the review groups of WeChat, there exists a single broker who owns the group. The broker acquires seller members and customer members in many different ways, such as advertisements in Douban. The broker posts the products for sellers and customers also request products that they want. Comparing with review groups in Facebook, the review groups in WeChat are private and hence make members feel sort of close to each other.

In view of difference of above platforms, we will characterize the review groups using the facebook dataset in Section 4, and develop the detection method with the WeChat dataset in Section 5. The Douban dataset sheds light on how brokers advertise their review groups and



(a) Member number.



(b) Member increment over time.

Fig. 6. Group members.

reach out more people, there are no review requests in the post and hence we just use the dataset to obtain the review groups of WeChat.

4. Measurement

In this section, we investigate the datasets and characterize incentivized review groups in terms of group members, review requests, and products.

4.1. Group members

We plot the histogram of member number for incentivized review groups we collected in Fig. 6(a) and also depict the increment of group members over time in Fig. 6(b). We observe that some groups attract a large number of group members. The largest group has more than 40,000 members. Over a month, there are seven groups that have more than 1,000 new members as Fig. 6(b) shows, indicating that these review groups are remarkably attractive and popular. It also implies that fake reviews from incentivized review groups are in a considerably large scale.

Sellers: Sellers play a key role in the review groups who post the review requests and attract members to join the groups. We plot the number of sellers for all groups in Fig. 7. We can see that there are a number of sellers in most of review groups, even more than 2,000 sellers in the largest group.

Sellers could join multiple review groups to reach more people and obtain more paid reviews. Fig. 8 shows the number of groups that sellers join. We can see that roughly 10% of sellers join more than one group and one of the most aggressive sellers even join nine review groups at the same time.

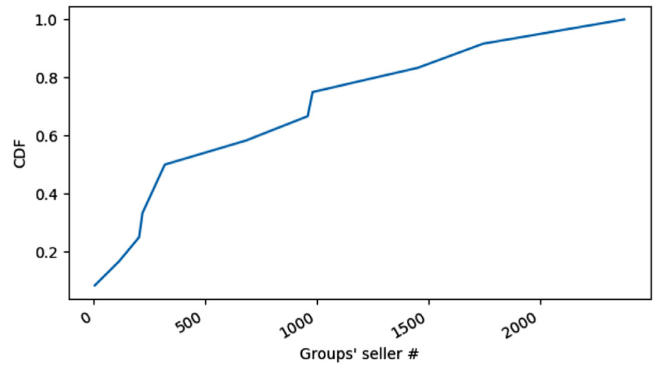


Fig. 7. The number of sellers.

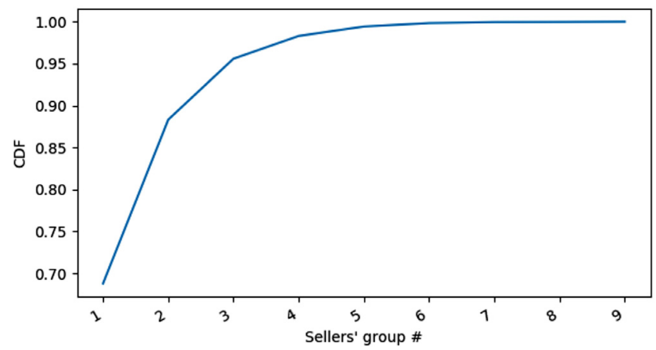


Fig. 8. The number of groups of sellers.

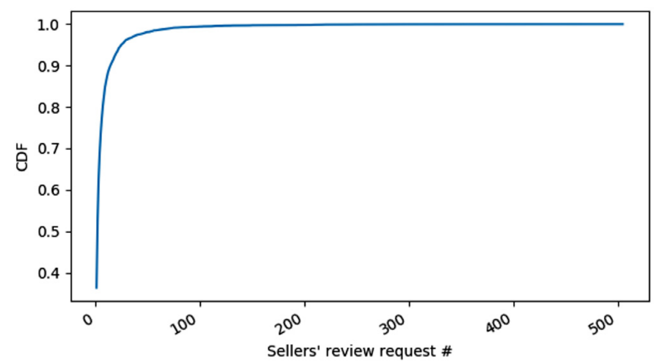


Fig. 9. Review requests of sellers.

4.2. Review requests

The number of review requests posted in a review group indicates how active the review group is. We plot the number of review requests of incentivized review groups against time in Fig. 13. We observe that some of review groups are notably active during our collection. The most active review group in our dataset has roughly 2,500 review requests each single day. Note that the remarkable drop of all groups on July 18, 2019 is due to the end of prime day of Amazon, which ranged from July 16, 2019 to July 17, 2019.

We also plot the number of review requests of sellers in Fig. 9. We can see that some of sellers are notably active who post more than 100 review requests. We further depict the number of review request against time for the seller who has the most review requests in Fig. 14. It is observed that the seller is quite active over the period. The review requests of this seller are across five different groups as time evolves, as shown in Fig. 10. He/she focuses on a certain group over a period and then switches to another group later on.

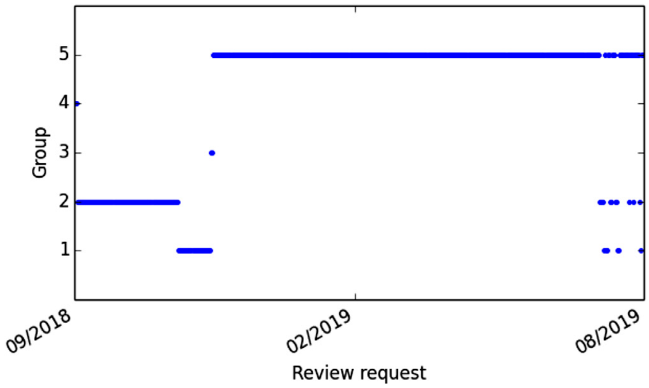


Fig. 10. Review requests across groups.

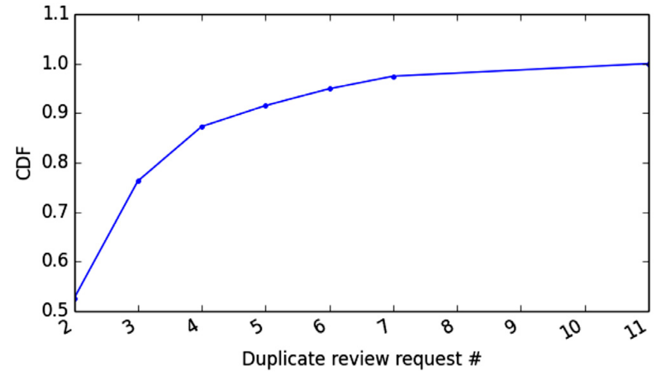


Fig. 11. Duplicate review requests.

Table 1
Product categories.

Category	Percentage(%)
Sports & Fitness	6.88
Accessories	5.94
Computers & Accessories	5.63
Cases, Holsters & Sleeves	5.31
Bedding	5.00
Men	4.69
Pest Control	4.69
Kitchen & Dining	3.12
Nursery	3.13
Women	3.13
Industrial Hardware	2.81
Motorcycle & Power sports	2.81
Outdoor Recreation	2.81
Hair Care	2.50
Replacement Parts	2.50
Building Toys	2.19
Heating, Cooling & Air Quality	2.19
Luggage & Travel Gear	1.88
Office & School Supplies	1.56
Tools & Accessories	1.25
Lighting & Ceiling Fans	1.25
Vacuums & Floor Care	1.25
Headphones	1.25
Material Handling Products	1.25
Home Audio	1.25
Gardening & Lawn Care	1.25
Car & Vehicle Electronics	1.25
Others	14.38

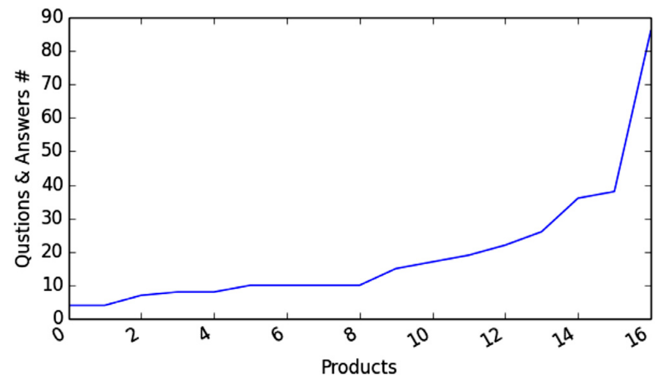


Fig. 12. Questions & Answers.

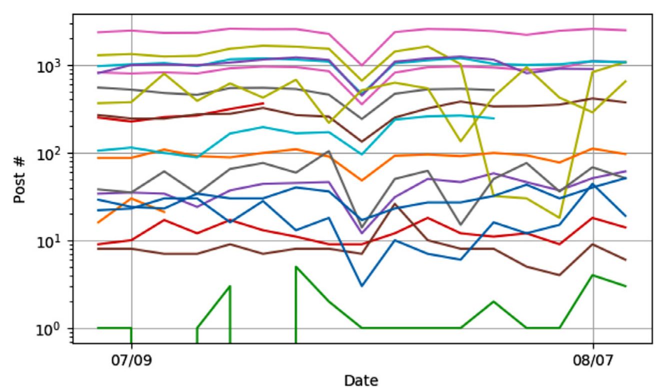


Fig. 13. Review requests posted in each group.

4.3. Products

4.3.1. Categories

We investigate the categories of products that stand in need of favorable reviews. Table 1 shows the percentage of product categories that have requested the favorable reviews in our collection. *Sports & Fitness* has the most review requests, accounting for 6.88%. It is followed by *Accessories* and *Computers & Accessories*, making up 5.94% and 5.63%, respectively. We find that 69.5% of the products we collect are fulfilled by Amazon, which means the inventory is kept in Amazon warehouse and will be packed by Amazon. Therefore, sellers can utilize the Amazon facility and platform to run their business. Another benefit of being fulfilled by Amazon could be concealing the place of origin.

Sellers send duplicate review requests for some products that are crying out for favorable reviews to boost sales. Fig. 11 depicts the number of duplicate requests, which could reach as high as eleven in our dataset, indicating the desperate need for positive reviews.

4.3.2. Questions & answers

Customers can ask questions in Amazon and the customers who bought the product are invited to answer the questions. Questions & Answers are helpful for resolving customers' doubts and hence improve the credibility of products. We investigate the Questions & Answers of products collected in the Wechat group. Fig. 12 plots the number of Questions & Answers. We observe that 16 out of 93 products have Questions & Answers. The largest number of Questions & Answers even reaches 87. It indicates that Questions & Answers could be utilized to promote reputation and credibility of products with favorable review requests.

4.3.3. Sellers and manufacturers

Sellers sell the products on Amazon and manufacturers provision the inventory. We investigate the relationship between sellers and manufacturers for the products with review requests. Fig. 15 shows three different types of relationship between sellers and manufacturers. The left means that the seller is the same as the manufacturer, the middle means

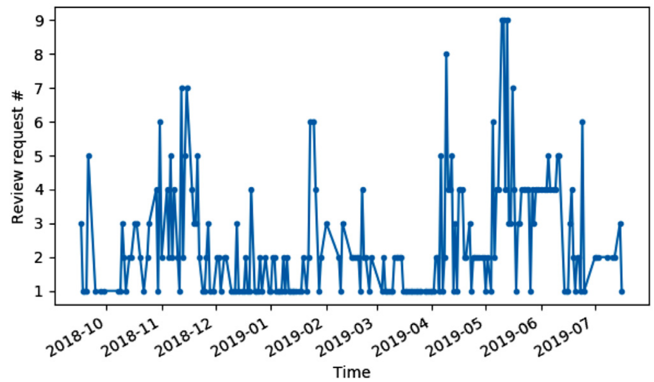


Fig. 14. Review requests of the most active seller.

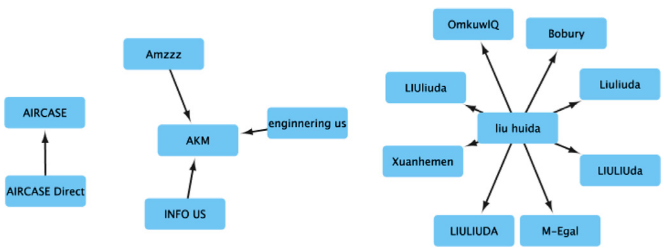


Fig. 15. Relationship types of sellers and manufacturers.

that multiple sellers work for one manufacturer, and the right represents that one seller works for multiple manufacturers. Understanding different types of relationship between sellers and manufacturers could be helpful for detecting the products that ask for incentivized reviews.

4.4. Strategies to evade detection

4.4.1. Private channels

Review groups opt for private channels, such as chat groups in Wechat and private groups in Facebook. Wechat groups are only visible to group members, and hence they perfectly fit the requirement of being private. The review group in Wechat we joined has only about 200 members when we joined but reached the maximum limit of 500 two months later. During the period, most of new members were invited by the members in the group. The private groups in Facebook are covert and also require the admission to join like Wechat groups. With the private groups, they can also find the members who are enthused about free product or compensation on the incentivized reviews, due to the effort members made to discover and join these groups. The detection of review groups in these private channels is difficult to reach a large scale, and sellers can easily transfer to other review groups.

4.4.2. Without sharing URLs

Even though sharing URLs of products could simplify the process of review requests and attract more customers, sellers always conceal the URLs of products in the review groups. Even in personal conversation, they are not willing to provide product URLs. The reason why not sharing URLs is because URLs from Amazon have referral information that can be utilized to track the source of sellers. If a number of customers visit a certain product with the same URL that refers to the seller, Amazon can detect the anomaly and probably ban the seller. Concealing URLs in review groups could bring a challenge to our study, which hurdles the collection of products with review requests as well as paid reviews. We utilize an Amazon image recognition procedure to overcome the barrier Section 3.

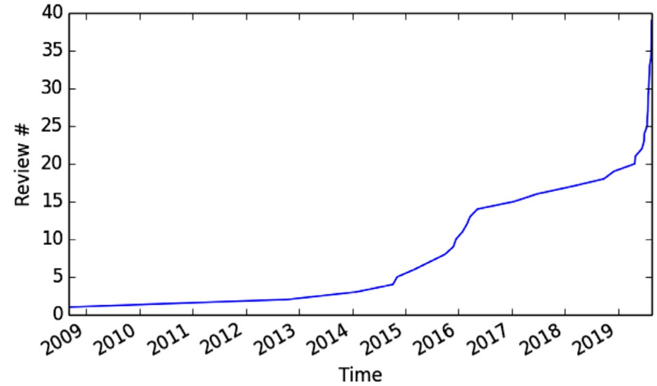


Fig. 16. An example of bursty reviews.

5. Detecting incentivized review groups with co-review graphs

In this section, we will model the reviewers as co-review graphs and refer the detection of incentivized review groups as a community detection problem. We then employ the graph analysis method to detect incentivized review groups. With a "gold standard" dataset, we evaluate different community detection algorithms. We also perform a retrospective study on an Amazon review dataset ranging from May 1996 to July 2014.

5.1. Model

We model the reviewers as an undirected graph $G = (V, E)$, where each node $v_i \in V$ represents a reviewer and each edge $\{v_i, v_j\} \in E$ represents a bilateral relationship between v_i and v_j. The bilateral relationship here means both v_i and v_j write reviews for at least one product. Therefore, we refer to the undirected graph as a *co-review graph*. In the graph, there are $n = |V|$ nodes and $m = |E|$ edges.

To detect the review groups, we employ the graph analysis to detect the communities in the graph and evaluate how accurately the identified communities reflect incentivized review groups. There are more edges inside the communities than the rest of the graph, and the nodes in the same community are considered to be similar to each other. Therefore, the communities of a co-review graph can reveal the cooperation pattern of reviewers in a review graph.

5.1.1. Features

To take various features into our detection, we construct multiple co-review graphs based on different features, such as frequency of co-review and co-review in bursts. Co-review graphs derived from different features can further improve our detection. Specifically, we consider the following features to construct co-review graphs to perform the community detection:

Frequency of co-review: The frequency of co-review between two reviewers is one of the most important features for indicating the probability of them belonging the same incentivized review group. There is no conclusion to draw if two reviewers only occur in one product together. If they occur in more than one products, it is likely that they belong to a same review group, especially when they occur many times together. Here, we construct the graph with reviewers occurring more than two times together.

Co-review in bursts: By checking the time series of reviews of the products that have incentivized reviews, we find that there exist evident bursts while the products requesting incentivized reviews. Fig. 16 shows an example of the burst. We can see the burst in July 2019 during our collection. We employ the Kleinberg's algorithm [12] to detect the burst in the time series. The algorithm models review number in a time series as an infinite hidden Markov model. With identification of bursts, we record the co-review of reviewers in the bursts. For the reviewers of review groups, they are required to post the most favorable reviews to obtain the free products or compensation. Therefore, we also check the rating of reviewers in the bursts, *e.g.*, five star in Amazon.

Posting nearness in time: The closer in time two reviewers post their reviews, the more possibly they belong to a same review group in some circumstances. Certainly, there exist some normal reviewers occur very close to each other. By collecting reviewers posting positive reviews (five star in Amazon) in the same product within two days, we construct the graph in terms of posting nearness in time.

5.1.2. Composing multiple graphs

We denote the graph from frequency of co-review as *FC graph*, the graph from co-review in bursts as *CB graph*, and the graph from posting nearness in time as *PN graph*. Multiple graphs derived from different features are complementary to each other. For example, CB graph have some important edges although two nodes of these edges co-occur only once and hence they do not exist in FC graph. Therefore, we compose multiple graphs according to the following equation:

$$\mathcal{G} = \mathcal{G}_{FC} + W_{CB}\mathcal{G}_{CB} + W_{PN}\mathcal{G}_{PN}. \tag{1}$$

First, we derive the FC graph by taking into account all pairs of nodes co-occurring more than once. Then, we compose CB graph into FC graph by adding edges that have at least one node in FC graph with weight W_{CB}, which measures the importance of co-review in burst feature. Similarly, we compose PN graph into FC graph with weight W_{PN}, which denotes the importance of posting nearness in time feature.

5.2. Community detection

5.2.1. Dataset

For further exploring the community of incentivized review groups, we collect the products posted in the review groups, including seller information, all reviews, and customer questions & answers. As mentioned in Section 3, sellers always conceal the products' URLs and are not willing to provide them even in personal conversation. We utilize an image recognition procedure to identify the products on Amazon. We identify 93 products posted in review groups by searching product images of more than 200 products. These identified products belong to 48 sellers. We further collect 531 products from these sellers. We find that sellers usually cooperate with more than one incentivized review groups and select different products for different time periods or different incentivized review groups. Therefore, some products from them are likely to be posted in the incentivized review groups that we do not have access or the periods out of our collection.

"Gold standard" dataset: Since we have knowledge of incentivized reviews posted by the incentivized review groups, we construct a "gold standard" dataset with these guaranteed incentivized reviews. The dataset consists of 764 incentivized reviews and 737 reviewers. With the dataset, we can extract the factual co-review connections of reviewers and evaluate the community detection algorithms. We obtain 5,950 co-review connections from the "gold standard" dataset.

5.2.2. Methods

We employ four different community detection methods to detect incentivized review groups. The following briefs these community detection methods:

CPM: The clique percolation method (CPM) [18] constructs the communities from k-cliques, which correspond to fully connected subgraphs of k nodes. Two k-cliques are considered adjacent if they share $(k-1)$ nodes and a union of adjacent k-cliques forms a community. Here, we consider $k = 4$.

Louvain: Louvain method [4] first finds small communities by optimizing modularity locally on all nodes and then group small communities into nodes. It repeats above two steps until achieving the optimal modularity. Modularity is a scale value between −1 and 1 that measures the density of edges inside communities to edges outside communities. Optimizing this value theoretically results in the best possible grouping of the nodes of a given network.

Table 2
AMI among algorithms.

	CPM	Louvain	LPA	Infomap
CPM	–	0.80	0.79	0.14
Louvain	–	–	0.83	0.12
LPA	–	–	–	0.14
Infomap	–	–	–	–

Table 3
Accuracy with only FC.

	CPM	Louvain	LPA
Accuracy	0.03	0.46	0.35

LPA: The Label Propagation algorithm (LPA) [19] works by propagating labels throughout the network and forming communities based on this process of label propagation. Intuitively, a single label can quickly become dominant in a densely connected group of nodes, but it is difficult to cross a sparsely connected region. The nodes that end up with the same label can be considered to be in the same community.

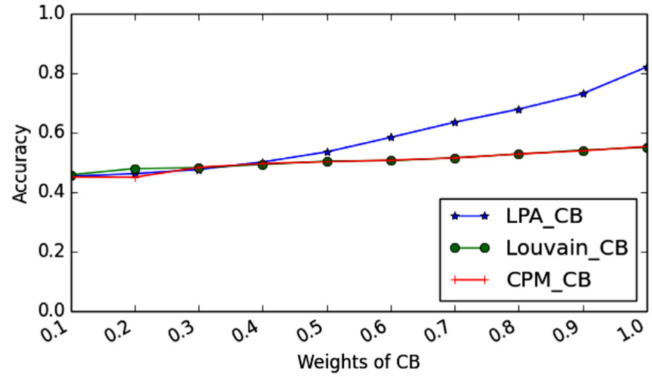
Infomap: Infomap [21] uses community partitions of the graph as a Huffman code that compresses the information about a random walker exploring the graph. A random walker exploring the network with the probability that the walker transits between two nodes given by its Markov transition matrix. Once the random walker enters the densely connected regions of the graph, it tends to stay there for a long time, and movements between the regions are relatively rare, which allows us to generate Huffman codes with modularity information. A modular description of a graph can be viewed as a compression of the topology of the graph.

5.2.3. Results

We employ above community detection algorithms to detect the communities corresponding to incentivized review groups. First, we compare the results from different algorithms by measuring the Adjusted Mutual Information (AMI) among different algorithms. AMI accounts for how similar two community detection results are to each other. Table 2 shows the results. We can see that most of algorithms are similar to each other, especially LPA and Louvain method. However, the result of Infomap algorithm is remarkably distinct with other algorithms. After careful inspection, we find that Infomap groups most of nodes to one huge community with various settings, such as number of levels. Therefore, we consider that Infomap is not suitable for this problem. Also, we consider $k = 3$ for CPM. We find that AMI between CPM and Louvain drops to 0.43 and AMI between CPM and LPA falls to 0.40. This inconsistency could indicate that $k = 3$ may underperform comparing with $k = 4$.

We then utilize the "gold standard" dataset to evaluate the accuracy of the above algorithms. The accuracy is measured by the proportion of the factual connections extracted from the "gold standard" dataset that a community detection algorithm can find out. Table 3 presents the accuracy of algorithms with only FC graph. We observe that the accuracy is quite low for Louvain and LPA methods, especially CPM method which is 0.03.

Varying weights of composing different graphs: To improve the accuracy, we compose PN and CB graphs into FC graph, respectively. We measure the importance of PN graph and CB graph by varying the composing weights. Fig. 17 shows the results. We can find that composing CB graph can constantly improve the accuracy of LPA method. When fully composing CB graph, the accuracy achieves 81%. In the meanwhile, CPM method achieves a higher accuracy comparing with 3% by composing CB graph, while Louvain method gains trivial improvement. When composing PN graph into FC graph, CPM method gains constantly improving



(a) Varying CB weights.

Fig. 17. Varying weights.

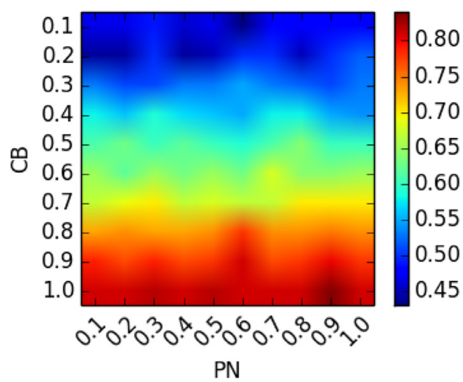


Fig. 18. Varying CB and PN weights for LPA method.

accuracy and Louvain method obtains trivial improvement, while the accuracy of LPA method first drops and then rises up. Overall, composing CB graph achieves a higher accuracy than PN graph. Moreover, PN graph is roughly 10 times bigger than FC graph and CB graph. Although it could improve the accuracy in some cases, it also carry a bunch of unwanted nodes and edges. This is probably the reason why LPA method is not gaining constant improvement as increasing the weight of PN graph.

We vary the weights of CB graph and PN graph at the same time, such that we can find out the optimal weight combination to achieve the best performance. Fig. 18 depicts the accuracy of LPA method. The side bar represents the scale of accuracy. We can see that ($W_{CB} = 1.0$, $W_{PN} = 0.9$) achieves the best accuracy in our experiment, although it is just a bit higher than ($W_{CB} = 1.0$, $W_{PN} = 0.1$). It indicates that CB graph remarkably improves the community partition comparing with PN graph.

We then investigate the distribution of communities, which could measure the performance of community detection method to a certain degree. If the biggest community of a community detection method includes most of nodes and it covers nearly all of test edges from our "gold-standard" dataset, this detection method would achieve a extremely high accuracy but useless. Therefore, we prefer a balanced community detection method. We select LPA method that achieves the best performance above as an example and plot the distribution of communities in Fig. 19. The area of the circle means the number of nodes in the community and the coverage of communities represents how many edges from our "gold-standard" dataset they cover. We can see that LPA method partitions notably balanced communities. The left two communities which cover most of edges from our "gold-standard" dataset are apparently the communities engaged in incentivized review groups. They are both of moderate size, 1015 and 654 respectively. We will further inspect the nodes of these two communities in Section 5.3. We also select a subset

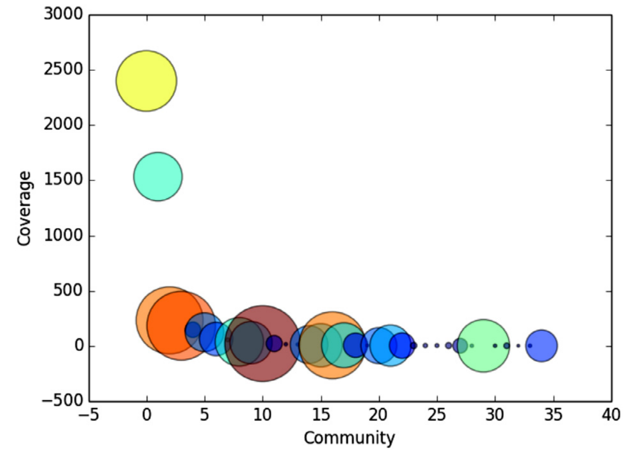


Fig. 19. An example of communities.

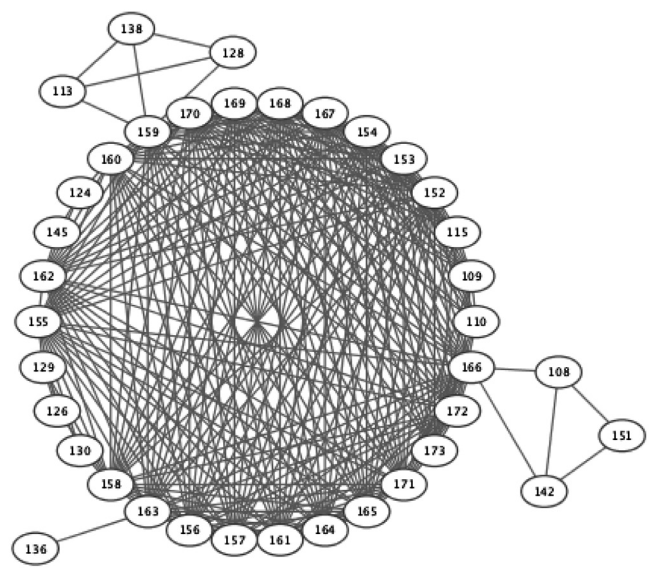


Fig. 20. An example of community graph.

of the community with 1015 nodes and display the community graph in Fig. 20. We can see that the subset in the community is a connected component with graph density 0.4, which means that these reviewers are tightly connected to each other.

5.3. *Reviewer profiles*

For the reviewers in two communities mentioned above, we collect their profiles from Amazon. We investigate their ranking, number of reviews, and number of helpful votes. Amazon ranks reviewers with a private algorithm. Smaller ranking represents higher reputation. Reviewers with higher reputations are preferred for sellers who ask for incentivized reviews, since their reviews seems more authentic. Number of reviews of a reviewer demonstrates how active the reviewer is, while number of helpful vote of a reviewer is to indicate how helpful the reviews of the reviewer are. In other words, it suggests to which extent the reviewer helps other customers. Fig. 21 depicts ranking (top), number of reviews (middle), and number of helpful votes (bottom) of reviewers.

We can observe that in the left, reviewers with higher reputation also have more reviews and helpful votes. In the middle, there are some spikes which represents that a few reviewers have outstanding amount of reviews or helpful votes. While in the right, reviewers with relatively lower reputation have extraordinary amount of both reviews and helpful

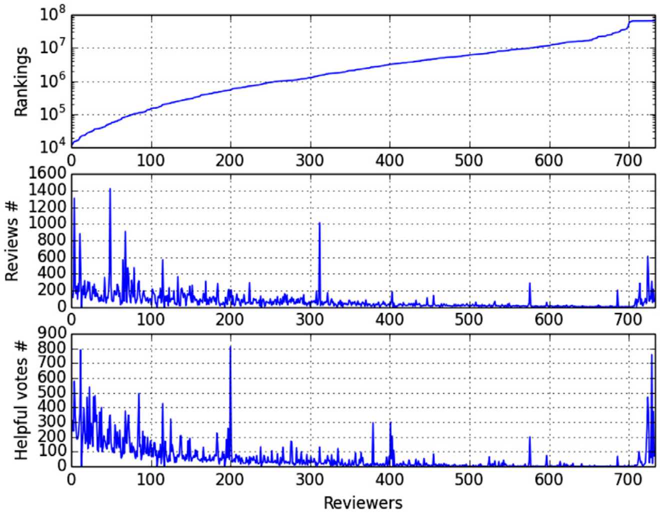


Fig. 21. Ranking, review #, and helpful vote # of reviewers.

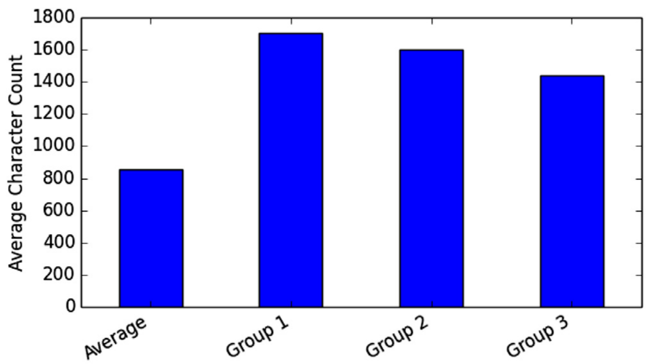


Fig. 22. Average character count.

votes. They can write a bunch of reviews as they want, however helpful votes must come from the others. It implies the possibility that these reviewers could obtain helpful votes reciprocally from other customers or fake accounts. We inspect these reviewers and find that they post a number of reviews within a short period.

5.4. *A retrospect of Amazon dataset*

We conduct a retrospective study of Amazon review groups with a public dataset. The dataset [10] contains product reviews and metadata from Amazon, including 142.8 million reviews spanning from May 1996 to July 2014. We construct the co-review graph with frequency of co-review and find that there are only 1,022 reviewers in the co-review graph. It indicates that incentivized review groups were not in an extensive scale before 2014.

We employ LPA community detection method on the co-review graph with frequency of co-review and find 31 groups. We inspect three largest groups, which contains 115, 109, and 71 nodes respectively. We name them as “Group 1”, “Group 2”, and “Group 3”. Fig. 22 plots the average character count of reviews across different groups. “Average Group” means the average character count over all reviews in the dataset. We can see that these three groups have remarkably more characters than the average, which implies that the reviewers from these groups are possibly professional critics who are invited to write professional reviews.

We examine the distribution of rating and helpful index of “Group 1”. Fig. 23 shows the result. It is observed that there exist a number of average reviews less than 4 and also a spate of reviews’ helpful index

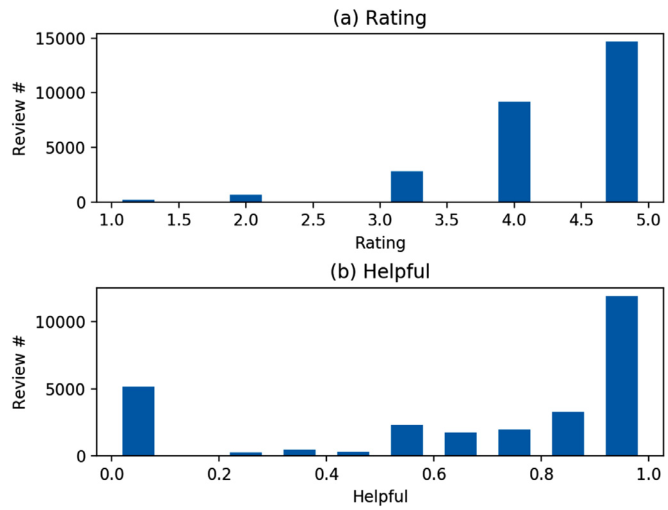


Fig. 23. Rating and helpful index of Group 1.

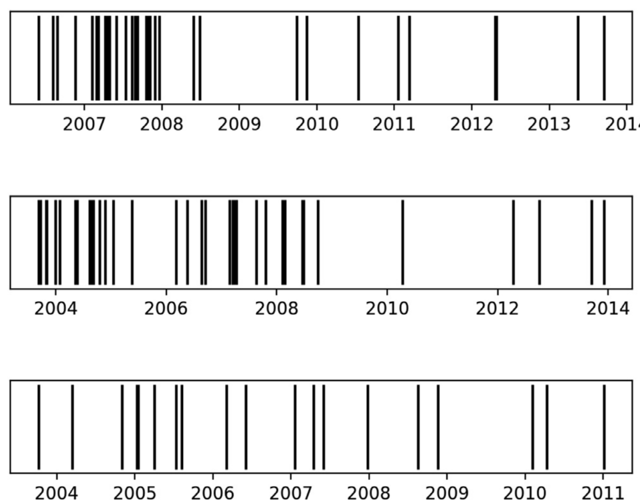


Fig. 24. Review timestamps of three products in Group 1.

less than 0.5, which implies that the reviews from “Group 1” are not considerably biased. We also inspect review timestamps across different products and find that the reviews of products quite evenly distribute in a long range, as Fig. 24 shown.

Overall, we can conclude that incentivized review group was not thriving yet before 2014. Sellers tended to invite professional critics to write long review to promote their products, which completely conformed to Amazon’s policy. The reviews from professional critics were not remarkably biased, since there existed a number of low rating reviews from them.

6. Limitations

We collect only one incentivized review group from WeChat for our “gold-standard” dataset due to the massive labor of collection and low accuracy of image recognition as mentioned in Section 3. Therefore, our dataset is limited to a single certain incentivized review group. Even though our “gold-standard” dataset sheds light on the detection of incentivized review groups to a certain degree, an extensive collection of incentivized review groups can definitely lead to a better detection effectiveness.

We perform community detection method on co-review graphs to partition reviewers to different groups. We also extensively investigate identified groups and evaluate the community detection methods with our "gold-standard" dataset. The detected groups can be further labeled as benign or malicious groups, *e.g.*, via spam behavior indicators [14]. Identifying malicious groups is out of scope of this paper and we will consider the labeling in our future work.

7. Related works

7.1. Spam review detection

Yao et al. [26] presented a potential attack against online review systems by employing deep learning to automatically generate fake reviews. They also proposed countermeasures against these fake reviews. Xie et al. [23] utilized temporal patterns to detect singleton review spam. Wang et al. [22] built review graphs to capture the relationships among reviewers, reviews, and stores, and then quantified the trustiness of reviewers. Zheng et al. [27] attempted to detect elite Sybil fake reviews in Sybil campaigns. Rayana and Akoglu [20] exploited behavioral data, text data, and relational data to detect spam reviews and reviewers. Ott et al. [16,17] detected deceptive reviews from both positive and negative sentiment review datasets. Feng et al. [7] investigated syntactic stylometry for deception detection. Li et al. [13] detected deceptive opinion spam across different domains. Mukherjee et al. [15] examined filtered reviews of Yelp and inferred their filtering algorithms. Fusilier *et al.* [8] employed character n-gram features to detect deceptive opinion spam. Harris [9] examined a variety of human-based, machine-based, and hybrid assessment methods to detect deceptive opinion spam in product reviews. In [11], Jamshidi et al. examined the explicitly incentivized reviews which state their incentives explicitly in the reviews. Different from Jamshidi et al. [11], we investigate the underground economy of incentivized reviews across different social networking platform and propose a detection method for the incentivized review groups. Also, Mukherjee et al. [14] identified opinion spam groups based on a set of spam behavior indicators. These spam behavior indicators could also be applicable to improve our detection of incentivized review groups.

7.2. Reputation manipulation

In online markets, sellers' reputation is closely related to profitability. Dishonest sellers have been reported to maneuver the reputation system by manipulating the transaction history. Xu et al. [25] investigated the underground market by which sellers could easily harness human labor to conduct fake transactions for improving their stores' reputation. They referred to this underground market as Seller-Reputation-Escalation (SRE) markets. Cai et al. [5] employed reinforcement learning methods to detect reputation manipulation in online markets. In addition, Xie and Zhu [24] inspected the underground market where mobile app developers could misuse positive reviews illegally. They also analyzed the promotion incentives and characteristics of promoted apps and suspicious reviews. In [6], the authors exploited the unusual ranking change patterns of apps to identify promoted apps and detected the collusive groups who posted high app ratings or inflated apps' downloads.

8. Conclusions

The focus of this work on the incentivized review groups on Amazon. We first investigate incentivized review groups across different platforms to understand the breadth of the problem and conduct a measurement study in terms of group members, review requests, and products. After the measurement study, we propose a detection method based on co-review graphs. We leverage community detection method to find the suspicious communities from the co-review graphs. To evaluate our detection method, we construct a "gold standard" incentivized review group dataset, which could shed light on further study on incentivized reviews. We also examine an extensive Amazon review dataset ranging from May 1996 to July 2014 and find that incentivized review groups posed nearly no threat on the ecosystem before 2014.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Amazon bans incentivized reviews, 2016a.
[2] Amazon Consumer Review Policy, 2016b.
[3] Consumer Review Survey, 2017.
[4] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) (2008) P10008.
[5] Q. Cai, A. Filos-Ratsikas, P. Tang, Y. Zhang, Reinforcement mechanism design for fraudulent behaviour in e-commerce, in: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018.
[6] H. Chen, D. He, S. Zhu, J. Yang, Toward detecting collusive ranking manipulation attackers in mobile app markets, in: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, 2017, pp. 58–70.
[7] S. Feng, R. Banerjee, Y. Choi, Syntactic stylometry for deception detection, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, Association for Computational Linguistics, 2012, pp. 171–175.
[8] D.H. Fusilier, M. Montes-y Gómez, P. Rosso, R.G. Cabrera, Detection of opinion spam with character n-grams, in: Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics, Springer, 2015, pp. 285–294.
[9] C.G. Harris, Detecting deceptive opinion spam using human computation, in: Proceedings of the Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence, 2012.
[10] R. He, J. McAuley, Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering, in: Proceedings of the 25th International Conference on World Wide Web (WWW), 2016, pp. 507–517.
[11] S. Jamshidi, R. Rejaie, J. Li, Trojan horses in Amazon's castle: understanding the incentivized online reviews, in: Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2018, pp. 335–342.
[12] J. Kleinberg, Bursty and hierarchical structure in streams, Data Min. Knowl. Discov. 7 (4) (2003) 373–397.
[13] J. Li, M. Ott, C. Cardie, E. Hovy, Towards a general rule for identifying deceptive opinion spam, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1), 2014, pp. 1566–1576.
[14] A. Mukherjee, B. Liu, N. Glance, Spotting fake reviewer groups in consumer reviews, in: Proceedings of the 21st international conference on World Wide Web (WWW), ACM, 2012, pp. 191–200.
[15] A. Mukherjee, V. Venkataraman, B. Liu, N.S. Glance, What yelp fake review filter might be doing? in: Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), 2013, pp. 409–418.
[16] M. Ott, C. Cardie, J.T. Hancock, Negative deceptive opinion spam, in: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2013, pp. 497–501.
[17] M. Ott, Y. Choi, C. Cardie, J.T. Hancock, Finding deceptive opinion spam by any stretch of the imagination, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, Association for Computational Linguistics, 2011, pp. 309–319.
[18] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (7043) (2005) 814.
[19] U.N. Raghavan, R. Albert, S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, Phys. Rev. E 76 (3) (2007) 036106.
[20] S. Rayana, 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.
[21] M. Rosvall, C.T. Bergstrom, Maps of random walks on complex networks reveal community structure, Proc. Natl. Acad. Sci. 105 (4) (2008) 1118–1123.
[22] G. Wang, S. Xie, B. Liu, S.Y. Philip, Review graph based online store review spammer detection, in: Proceedings of the IEEE 11th International Conference on Data Mining, IEEE, 2011, pp. 1242–1247.
[23] S. Xie, G. Wang, S. Lin, P.S. Yu, Review spam detection via temporal pattern discovery, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2012, pp. 823–831.
[24] Z. Xie, S. Zhu, Appwatcher: unveiling the underground market of trading mobile app reviews, in: Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks, 2015, p. 10.
[25] H. Xu, D. Liu, H. Wang, A. Stavrou, E-commerce reputation manipulation: The emergence of reputation-escalation-as-a-service, in: Proceedings of the International Conference on World Wide Web (WWW), 2015, pp. 1296–1306.
[26] Y. Yao, B. Viswanath, J. Cryan, H. Zheng, B.Y. Zhao, Automated crowdturfing attacks and defenses in online review systems, in: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, ACM, 2017, pp. 1143–1158.
[27] H. Zheng, M. Xue, H. Lu, S. Hao, H. Zhu, X. Liang, K. Ross, Smoke screener or straight shooter: Detecting elite sybil attacks in user-review social networks, in: Proceedings of the NDSS, 2018.