Efficiently Promoting Product Online Outcome: An Iterative Rating Attack Utilizing Product and Market Property

Yuhong Liu[1], Wenqi Zhou[2], and Hong Chen[2]

[1]Department of Computer Engineering, Santa Clara University, Santa Clara, CA, 95050
[2]Palumbo-Donahue School of Business, Duquesne University, Pittsburgh, PA, 15282
Emails:yhliu@scu.edu, zhouw@duq.edu, chenh4@duq.edu

***Abstract*—The prosperity of online rating system makes it a popular place for malicious vendors to mislead public's online decisions, whereas the security related studies are lagging behind. In this work, we develop a quantile regression model to investigate influential factors on online user choices and reveal that the promotion effect on products' market outcomes is determined by not only the attacker's manipulation power but also the specific property of the target product and the market self-exciting power. Inspired by these findings, we propose a novel iterative rating attack and validate its effectiveness through experiments.**

I. Introduction

With the rapid growth of e-commerce and social media, online rating systems that let users post ratings/reviews of products and services are playing an increasingly important role in influencing users' online purchasing/downloading decisions. On the one hand, users may directly rank products/services according to their rating scores. On the other hand, online recommender systems that help users identify their favorable items from vast amount of products/services also take such ratings/reviews as a critical input. According to a 2013 survey conducted by Dimensional Research [1], 88% online users have been influenced by an online user review when making a buying decision. A survey conducted by comScore Inc. and The Kelsey Group reveals that consumers are willing to pay at least 20% more for services receiving an "Excellent" or 5-star rating than for the same services receiving a "Good", or 4-star rating [2]. EBay sellers with established reputation can expect about 8% more revenue than new sellers marketing the same goods [3].

The huge profits provide great incentive for companies to manipulate online user ratings/reviews in practice. Book authors and eBay users are shown to write or buy favorable ratings for their own products [4], [5]. A recent study has identified that 10% of online products have had their user ratings manipulated [6]. Yelp has identified roughly 16% of its restaurant ratings [7] as dishonest ratings. The boom of rating companies, which provide sophisticated rating manipulation packages at affordable prices, reinforces the prevalence of such manipulations. For just $9.99, a company named "IncreaseYouTubeViews.com" can provide 30 "I like" ratings or 30 real user comments to boost video clips on YouTube. Taobao, which is the largest Internet retail platform in China, has identified these rating boosting services as a severe threat.

The protection of online rating systems in essence has roots in the thorough understanding of how attack strategy works. Hence, a number of studies have been conducted to investigate rating attack strategies. Generally speaking, rating attacks can be classified into two categories: self-boosting attacks, where malicious users aim to boost rating scores of their own products, and bad-mouthing attacks, where malicious users aim to downgrade rating scores of other competitors' products [8]–[10]. Specifically, a number of diverse rating manipulation strategies have been proposed, such as Sybil attack [11], Oscillation attack [12] and RepTrap attack [13].

Nevertheless, current studies on rating attacks are still immature due to several reasons. First, most of the current studies evaluate attack impact by measuring the distortion of a target product's rating score [14], or the number of unfair ratings bypassing the detection scheme [15], [16], while seldom considering the economic impact on **product market outcome**. The product market outcome can be measured by firm equity values, online sales, or downloads, if the product is in digital format. The lack of economic analysis often leads to impractical designs of attacks that are effective in changing products' rating scores while not necessarily attracting more real sales/downloads.

Second, the current design of attack strategies mainly focuses on malicious user behavior while ignoring product properties. Although comprehensive malicious user behavior has been extensively investigated, it is not the only impact factor determining the attack consequences. The same malicious user behavior may lead to completely different impacts on products with different properties, such as existing rating value and volume, existing sales/downloads, market ranks, etc.

Third, current attacks promote/downgrade products by considering only the "external energy" provided by unfair ratings while ignoring the "internal energy" generated by the market itself. For example, if a rating manipulation launched at time $t-1$ is able to increase the target product's sales at time t, we find that the greater sales and higher popularity can further bring in more sales at the next time point $t+1$ although rating manipulation has already stopped.

To fill the gap, we consider these three aspects in the design of the proposed attack and summarize our contributions as follows. **First**, we introduce economic analysis into the design of rating manipulations by modeling how manipulation related factors will influence products' online sales/downloads. **Second**, we further differentiate manipulation impact on products with different popularity by adopting a quantile regression model. **Third**, for the first time, we discover a "self-exciting" property in the online rating market which may provide extra energy beyond the manipulation power to push up target products to a higher rank than expected. Inspired by these findings, a novel iterative rating attack strategy is proposed and its effectiveness has been validated through experiment results. Note that we mainly focus on self-boosting attacks in this study. The same logic, however, may also help the design of bad-mouthing attacks.

The rest of the paper is organized as follows. Prior studies on rating manipulations and influential factors for product sales/downloads are reviewed in Section II. A quantile regression model is introduced in Section III-A and applied on data described in Section III-B. Observations of regression results are explained in Section IV. A novel iterative rating attack strategy is proposed in Section V, followed by experiment results in Section VI. Finally, Section VII concludes this paper.

II. Related Work

A. Influential Factors for Online User Choices

A thorough understanding of how different factors may affect online users' decision making serves as the foundation for the design of efficient rating manipulation strategies. Therefore, we first conduct a comprehensive literature review on influential factors for online user choices.

Rating value and volume are generally recognized as critical influential factors on a product's online market sales/downloads [17]–[22]. Rating values, which reflect prior users' preferences and perceptions of product quality, play an important role in influencing later users' choices [17], [19]. Specifically, the impact of rating value is found to be nonlinear [17], [23], meaning that a fixed increase in rating value can lead to disparate market sales/downloads for products with different existing ratings. On the other hand, rating volume, which indicates a product's visibility on the market, helps bringing the product to users' attention. Therefore, an increase in rating volume may help the product stand out from abundant competitors and lead to a larger chance of receiving greater market sales/downloads [20], [24].

More interestingly, various studies in recent years have found that the impact of rating value and volume differs over products with different popularity [25]–[28], which is often represented by market ranks. To accurately capture such impact, quantile regression models have been proposed by prior studies in marketing and information systems [26], [29] and achieved good results. Historically, this particular regression methodology has also been introduced in Econometrics for a long time and is shown to be robust and appropriate to estimate the differential impact of influential factors on the whole distribution of the outcome variable [30].

A product's sales/downloads may also be affected by the network effect and herding effect. First, the product diffusion theory indicates a network effect where the greater user base of a product, generally measured by past sales/downloads, will help expand its market share [25], [31], [32]. Second, the herding effect refers to that the empirical proof that online consumers follow others' adoption decisions [25], [33], [34]. In other words, if the products have become more popular (i.e. ranked more highly in the market), consumers may follow their predecessors' steps and also choose those more popular products.

We follow the above literature to adopt these influential factors in our quantile regression model, which is discussed in details in Section III-A.

B. Rating Manipulation Studies

To compare the proposed attack to existing ones, we further review state-of-the-art rating manipulation studies.

The design of rating attack strategies has been conducted by many security studies and is dynamically evolving. In simple attacks, unfair ratings are provided independently. For example, eBay users boost their own reputation often by buying and selling ratings from independent sources [4]. Such simple attacks, however, cannot cause severe damage to the system due to the limited power of individual user accounts.

Collusion attacks, where excessive number of online IDs coordinate to insert unfair ratings, are adopted by many rating manipulation strategies as a more powerful attack [15], [16]. The Sybil attack [11] is a typical example of collusion attacks. The colluding malicious users can (1) provide high ratings for self-promoting; (2) provide low ratings for bad-mouthing [8]–[10]; (3) restore their reputation by providing honest ratings to products that they do not care [35], [36]; or (4) whitewash their reputation by registering new user IDs [37].

Advanced collusion attacks, where malicious IDs perform more diverse yet coordinated tasks, are proposed to further strengthen manipulation impact and to avoid being detected. For example, in Oscillation attacks [12], multiple malicious user groups may perform different rating behavior to protect one another from being detected. The roles of these groups switch dynamically. Another example is RepTrap attack [13], where malicious user IDs coordinate to overturn the reputation of some products and turn them into traps one after another. This way, users who provide honest ratings on these trap products will be mistakenly identified as malicious by the rating defense scheme. As a consequence, malicious users are trusted by the systems and become more powerful to launch further attacks while honest users are marked as untrustworthy.

To defend against collusion attacks, many online rating systems increase the cost of acquiring multiple user IDs by binding identity with IP address [38], requiring entry fees [39], using network coordinates to detect Sybil attacks [40], and analyzing trust relationships in social networks to identify collusion groups [8]. In addition, a variety of advanced defenses are proposed to statistically analyze products' rating distributions [41], [42], to evaluate raters' feedback trust [43]–[45], and to adopt temporal and user similarity information in unfair rating detection [15], [16], etc.

Note that, in this study, we mainly focus on how to enhance manipulation impact when the same set of unfair ratings is applied on different target products. How these ratings can be inserted without being detected will be further studied in the future work and is beyond the scope of this paper. Therefore, defense solutions are not considered in this study.

III. Model and Data

A. Quantile Regression Model

In this study, we develop a quantile regression model to estimate the differential impact of various changes introduced by rating manipulations on online user choices. Specifically, a quantile is defined as the quantile of the outcome variable distribution. By examining a series of quantiles, the quantile regression model is designed to assess the different impacts of independent variables at various locations of the outcome variable distribution [30]. In particular, it models the conditional quantile of the outcome variable as a linear function of independent variables [30].

Compared to the widely adopted least-squares regression model, which assumes the impact of independent variables as uniform over the entire distribution of the outcome variable [26], [29], [46], the quantile regression model has its unique features and is particularly helpful in this study. Based on the literature review [25]–[28], in particular Section II-A, we agree with the prior work that the impact of user ratings, our key independent variables, is different on products with different popularities. And popularities on CNETD are evaluated by weekly downloads, our outcome variable. The adoption of quantile regression model allows us to capture the differential impact of user ratings on different quantile of weekly downloads, which could not be done through the least-squares regression model. In addition, quantile regression models have already been adopted by a number of prior studies to estimate the impact of online user ratings on product popularity [26], [29] and achieved convincing results.

The general form of the quantile regression model is expressed as:

$$Q\alpha(y|x) = x\beta(\alpha) \tag{1}$$

where $Q\alpha(y|x)$ denotes the αth quantile of the distribution of the outcome variable y, and x denotes the vector of independent variables.

In this study, we have followed the prior work [26] conducted in the same online software platform to develop our model. Specifically, we measure y by the weekly download number of product i at time t (d_t^i). A logarithm transformation is also applied on y to cope with the scale effect [17], [24], [25]. It allows us to estimate the effect of a change in the independent variables on the percentage change in the outcome variable. In other words, the change of independent variables is studied to affect a certain fraction of online user decisions.

The independent variable vector x includes those influential factors that will be affected by rating manipulations and other independent variables. Specifically, we **first** include the average rating value $\bar{r}_t^i$ of product i at week t, of which the impact is non-linear [20], [23]. The literature [20], [23] finds that the impact of increasing rating value depends on its original value. Accordingly, a square term on rating value is included. That is, $\beta_2(\alpha) * \bar{r}_t^i + \beta_3(\alpha) * \bar{r}_t^i * \bar{r}_t^i$. We also note that not all the products are reviewed. Actually, products are not randomly selected to be reviewed, and whether a product is reviewed also matters to its sales/downloads [47]. As a result, we include a binary variable Rev_t^i [46], of which the value is set as zero if product i has not received even a single rating by week t, or as one otherwise. **Second**, we include $\beta_5(\alpha) * log(\bar{v}_t^i)$ to capture the impact of rating volume (v_t^i) and set the value of $log(v_t^i)$ as one if product i does not receive any ratings at week t yet [17]. Online user decisions, which are captured by weekly downloads in this study, are identified by prior work to have a log-linear relationship with the number of user ratings. As a result, logrithm is applied to v_t^i. **Third**, the herding effect, cumulatively proxied by product rank (i.e. rank R_t^i), is represented by $\beta_4(\alpha) * R_t^i$ [25]. **Fourth**, we include $\beta_6(\alpha) * log(\widetilde{d}_t^i)$ to capture the network effect [25], [31], [32], where $\widetilde{d}_t^i$ is product i's total number of sales/downloads by week t. The logarithm transformation is applied to scale down the large variance of total product sales/downloads, otherwise the coefficient will not have enough degrees of freedom to be statistically estimated [46]. **In addition** to those key influential factors, there are some other factors that may also influence product sales/downloads but cannot be directly changed by rating manipulations, such as product age (Age_t), which indicates how long the product has been online, and its square term (Age_t^2) [25]. These factors are beyond the major focus of this study since we aim to design efficient rating attacks. To model product sales/downloads more robustly, we introduce control variables $Controls_{x,i,t}$ to represent such factors.

The proposed model has its own uniqueness as compared to the prior work [26]. First, a dummy variable Rev_t^i is introduced to differentiate the two scenarios as not being rated and having low rating value, because various prior work in the econometrics and business field has emphasized that missing value is not the same as the value of zero. Second, in addition to the rating value, the rating volume is also included in the proposed model, since recent literature has unanimously agreed on the power of rating volume on online user decisions. Third, a time lag of one week on key independent variables is considered to avoid reverse causality. A notorious confounding factor in econometrics model is the endogeneity caused by reverse causality [46]. In particular, if both the independent variables and the outcome variable are at the same time period, a significant coefficient on the independent variable may be partially caused by the causality of the outcome variable on the corresponding independent variable, instead of the other way around. To control for this issue, we follow the literature to adopt one time lag in all independent variables [19], [24], [46]. In other words, all influential factors at $t-1$ are included. This simple yet widely adopted statistic technique helps to exclude the possibility of reverse causality. Because all influential factors occur before the outcome variable, the outcome variable cannot affect any of those influential factors.

As a summary, we develop the following quantile regression model to estimate the impact of rating manipulations:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2017.2668992, IEEE Transactions on Information Forensics and Security

$$\begin{aligned} log(d_t^i)(\alpha) = & \beta_0(\alpha) + \beta_1(\alpha) * Rev\_u_{t-1}^i + \beta_2(\alpha) * \bar{r}_{t-1}^i \\ & + \beta_3(\alpha) * \bar{r}_{t-1}^i * \bar{r}_{t-1}^i + \beta_4(\alpha) * R_{t-1}^i \\ & + \beta_5(\alpha) * log(\widetilde{v}_{t-1}^i) + \beta_6(\alpha) * log(\widetilde{d}_{t-1}^i) \\ & + \beta_x(\alpha) Controls_{x,i,t} + \xi_{i,t}(\alpha) \end{aligned} \tag{2}$$

where $Controls_{x,i,t}$ is a $2 * 1$ matrix of control variables including $Age_{i,t}, Age_{i,t}^2$, α denotes αth quantile.

Quantile regression model, as similar to other frequency regression models, has an underlying asymptotic assumption that the sample data is sufficient to ensure unbiased estimations. According to the literature [46], our data set meets this assumption by having weekly data for around 300 products over 25 weeks, which will be elaborated in the next subsection. We also acknowledge that not all potential control variables may be included on the left side of the equation (2), which is a general concern of most econometrics model. In our work, those factors can be the marketing expenses of the product, product quality, and customer preferences. Fortunately, in equation (2), we have included the previous software ranking (R_{t-1}^i) and past total downloads ($\widetilde{d}_{t-1}^i$) to at least partially, if not fully, control for the impact of those omitted factors. Specifically, the software ranking and total downloads at previous week can reflect the effectiveness of advertising, the level of product quality and customer overall preferences.

The key variables used in the regression model is summarized in Table I.

Variable	Description
d_t^i	number of Weekly downloads of software i at week t
$\widetilde{d}_t^i$	total number of downloads of software i by week t
$\bar{r}_t^i$	average user rating of software i by week t
$\widetilde{v}_t^i$	total number of user rating of software i by week t
R_t^i	the rank of software i at week t measured by weekly sales/downloads
$Free_t^i$	a binary variable to measure if software i is free-to try at week t
$Rev\_u_t^i$	a binary variable to measure if software i is reviewed by users at week t
$Rev\_e_t^i$	a binary variable to measure if software i is reviewed by CNET editorial team at week t
Age_t^i	how long software i has been available on the market

TABLE I: Description of Key Variables

B. Context and Data

Our data is collected from CNET Download.com (CNETD), an online platform providing more than 30,000 free or free-to-try software programs for Windows, Mac, mobile devices, and Webware. It holds a leading user rating system with a large number of online user ratings/reviews. In addition, CNETD also shows download counts for each of its software programs. These software programs are also ranked according to their number of weekly downloads. The CNET editorial team also picks a small percentage of software programs to provide a rating in a five-star scale along with detailed comments.

There are several reasons for choosing this particular context. **First**, the experiential nature of a software product requires online users to generally rely on others' experiences to signal product quality and find matched products [18]. **Second**, product variety in online software markets has been increasing over years [26]. Online user ratings thus play a significant role in making a product visible out of abundant product choices. CNETD clearly displays the average rating star and the total number of user ratings for each listed product and updates that information daily. This enables us to compile a longitudinal dataset to quantify the impact of ratings. **Third**, online retail sales data, which captures the online market outcome well, is rarely available. Instead, weekly software downloads of free-trials on CNETD (i.e. d_t^i) are publicly available and have been used in prior studies to measure software programs' online market outcomes [18], [23], [25]. Those three features of CNETD ensure that it is an appropriate research context to estimate the impact of online user ratings and further investigate the effect of rating manipulations.

In particular, we collect weekly data of software downloads and online user ratings from CNETD over 26 weeks in four categories from August 2007 to February 2008. Those categories are Anti-virus, Download Managers, File Sharing and Web Browser, which are chosen to include both popular downloaded software programs as well as software programs with different application purposes. Specifically, we collect the number of weekly downloads (d_t^i), the cumulative number of downloads ($\widetilde{d}_t^i$), the average rating values ($\bar{r}_t^i$), the rating volume ($\widetilde{v}_t^i$), how long the software has been available on the market (Age_t^i), and product rank by weekly downloads (R_t^i), in addition to various software characteristics. Moreover, prior studies in the same context have found that being selected by CNET editors sends a positive signal to online users, leading to greater downloads [23]. To capture the impact of CNET editors' expert rating on users' choice [23], we add $Rev\_e_t^i$ as the third element to the control variable matrix $Controls_{x,i,t}$. As a result, the $Controls_{x,i,t}$ becomes a $3*1$ matrix of control variables including $Age_{i,t}, Age_{i,t}^2$ as well as $Rev\_e_t^i$.

The quantile regression model in equation (2) is then applied to this data set. Specifically, the first 25 weeks' data is used to estimate parameters of the quantile regression model and the last week's data is left alone for attack simulation later. We intentionally partition the data set in this particular way to assure that our quantile regression model estimation is independent of the subsequent attack simulation.

IV. Observations

By observing parameters estimated through the quantile regression model, in this section, we aim to understand how a product's market outcome can be promoted. We make an assumption that in a mature market, most products' market outcomes are relatively stable in short time periods, e.g. within a few weeks, since a product's reputation and customers' overall preferences do not change rapidly, unless some sudden changes occur [15]. In this study, we capture a product's market outcome through its absolute amount of **weekly downloads**.

A. An Analogy

We make an interesting observation that the quantile of product downloads in our proposed model actually captures the analogy between an electron's energy level transition and a product's market outcome level transition.

In atomic physics, energy level transition describes that an electron, after absorbing energy, may change its **energy level** to a higher energy excited state. Similarly, in the online market, a product, after absorbing some "promotion energy", may attract extra market downloads and jump to a higher **rank**. Note that in this paper, a higher rank denotes a rank with a smaller index value. For example, given two ranks as 40 and 20, we consider rank 20 is a higher rank. As an electron's energy level transition requires a sufficient amount of extra energy, in the online market, a product's rank transition also requires a sufficient number of extra downloads. In other words, the extra increased downloads have to exceed a certain threshold to make the rank transition occur.

Through this analogy, we ground our research questions to what sources can provide such "promotion energy". This question is addressed in the following sections.

B. External Energy - Rating Manipulation

The first potential source of "promotion energy" is self-boosting rating manipulations, which directly increase a target product's rating value and volume by inserting overly inflated ratings. We validate the impact of such manipulations by discussing how the change of rating value and volume affects products' future downloads. Specifically, we assume the original product rating value $\bar{r}_{t-1}$ and volume $\widetilde{v}_{t-1}$ have been changed by rating manipulations in week $t-1$, which leads to changes in weekly download as Δd_t. Note that unfair ratings which change $\bar{r}_{t-1}$ will inevitably change $\widetilde{v}_{t-1}$ in practice. Nevertheless, by introducing the statistic regression model, we are able to analyze the consequences of changing $\bar{r}_{t-1}$ and $\widetilde{v}_{t-1}$ independently.

1) **Change of Rating Value:** Based on values of $\beta_2(\alpha)$ and $\beta_3(\alpha)$, the impact of changing a product's rating value is shown in Figure 1. Specifically, four different markets including "Anti-virus", "Download Managers", "File Sharing", and "Web Browsers" are represented in the upper left, upper right, lower left and lower right subplots, respectively. For each subplot, the x-axis represents product original ranks R_{t-1} and the y-axis represents the increase in weekly downloads (Δd_t) in terms of percentage. The impact of increasing products' rating values by 0.5 star is illustrated. Specifically, five different curves are shown to represent if the products' original rating values (i.e. $\bar{r}_{t-1}$) were 1-star, 2-star, 3-star, 4-star and 4.5-star, respectively. Note that the fifth curve represents $\bar{r}_{t-1}$ as 4.5-star instead of 5-star since 5-star is the maximum rating value in most rating systems and cannot be increased any further. Several common trends are observed in all these four markets.

First, the future weekly downloads for top ranked products will not be influenced by the changes of their average rating values. The possible reason is that these products are so well-established in their own markets so that their reputation has been well recognized. Increasing their rating values can do little to change people's opinions and thus will not attract extra downloads from real users. For example, in the "Web Browser" market, the top ranked products include Mozilla Firefox 2.0.0.11, Internet Explorer 7, Mozilla Firefox 3 beta 2, Netscape Navigator 9.0.0.5 and Safari 3.0.4 public beta, which have been widely adopted by the public.

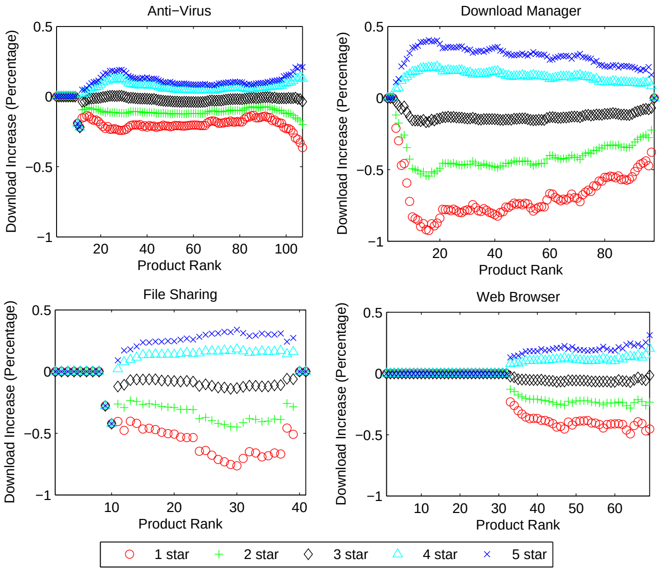


Fig. 1: Impact of Rating Value Change

Second, boosting a product's rating value will increase its weekly downloads if its original rating value $\bar{r}_t^i$ is high (e.g. above 3-star). Overly inflating the rating value of a product with low original ratings may even hurt its next week's downloads. For example, in the "Web Browser" market, for products ranked below 30, if their original rating values were 2-star (i.e. represented by green plus signs), increasing their rating values will cause drops of their future downloads.

The possible reason could be users' awareness of rating manipulations. In particular, users may be used to a product's original rating value $\bar{r}_t^i$, which exists for a certain time duration. A sudden boost of the rating score from a very low value may raise users' concerns instead of their trust, which will in turn lead to a drop of the future downloads. This is an interesting observation that has never been paid attention to by other rating manipulation studies, which often simply treat rating value increase and product download increase equally. However, we have shown that malicious attackers may shoot themselves in the foot by arbitrarily increasing a product's rating value while ignoring its original rating.

2) **Change of Rating Volume:** The impact of changing products' rating volumes is shown in Figure 2. Same as Figure 1, the four subplots represent four different markets. Specifically, for each subplot, the x-axis represents product ranks and the y-axis represents the increase in future weekly downloads in terms of percentage. The red circle at each particular rank represents the percentage of download increases caused by

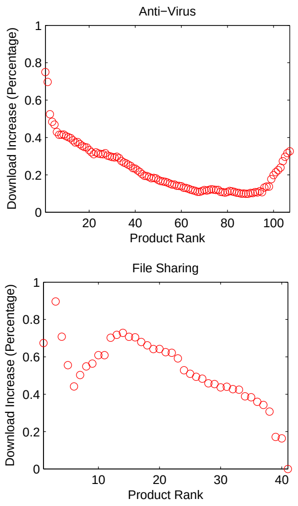


Fig. 2: Impact of Rating Volume Change

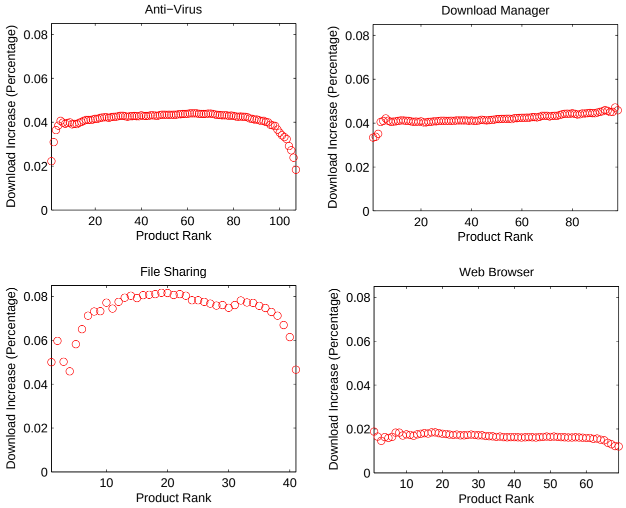


Fig. 3: Impact of Herding Effect

increasing the product's rating volume by 1% when all else being equal.

For all four subfigures, we make the following observations. First, the increase of rating volumes will always lead to positive impact on improving future downloads. Therefore, compared to boosting rating value, increasing rating volume at the existing rating level could be a more general strategy that is independent of the product's original rating status.

Second, the downloads of top ranked products are influenced most by rating volume changes, since more rating volume typically indicates higher visibility which plays an essential role in attracting more downloads.

Third, in spite of some occasional fluctuations, the impact of rating volume change dramatically drops for lower ranked products. It shows that manipulating higher ranked products often yield a better return. Nevertheless, we have to note that this figure shows the increase of rating volume in terms of percentage instead of absolute value. Generally speaking, higher ranked products often possess larger rating volume. Therefore, the same 1% rating volume may represent dramatically different absolute values for higher and lower ranked products. But this observation still makes sense because there often exist popular products that have not attracted too many ratings yet, such as newly released products. Manipulating such products could lead to maximum manipulation gain. An example is YouTube Grabber 4.2.7 which is ranked as $\#9$ in the "Download Managers" market while only possessing 27 ratings. Free Mobile Ringtones 1.1, another software in the same market, is ranked as 35 but has 100 ratings already.

Based on the above discussions on the impact of changes in rating value and volume, we would like to further investigate how these observations would affect the design of manipulation strategies in Section V.

C. Internal Energy - Market Self-exciting Power

Besides rating manipulations that influence products' market outcomes by directly changing rating value and volume, we also discover another source of "promotion energy" as the **market self-exiting power**, due to the aforementioned herding and network effect. Specifically, if the target product's downloads and even market rank are improved, some further pushing up power will be excited by the market, which may help the target product attract even more downloads in the future. We call such pushing up power as the market's "self-exciting power".

1) **Existence of Self-exciting Power:** We validate the existence of self-exciting power as follows. Recall that this power in essence is excited by the herding effect of rank improvement and network effect of download increase, which are captured by two quantile based parameters $\beta_4(\alpha)$ and $\beta_6(\alpha)$, respectively. By examining the impact of these two parameters on products' market outcomes in Figure 3 and Figure 4, we find them as always positive for all products in all four markets. It indicates that both a products' download increase and its rank improvement will cause positive increase in its future market outcome. In other words, the market's self-exciting power does exist.

2) **Consecutive Self-excitation:** More importantly, we observe that a consecutive self-excitation process may be initiated as a result of the market's self-exciting power. Specifically, the self-exciting power will directly lead to extra increase of the target product's downloads. Once the extra increased downloads exceed the rank transition requirement, the product rank will be further improved. We define the rank promotion caused by the self-exciting power as **self-excited rank promotion**. Such rank promotion and extra download increase will continue stimulating the market's self-exciting power, which may lead to another round of self-excited rank promotion and download increase. This creates consecutive self-excitation process over multiple rounds.

First, we validate the existence of self-excited rank promotion in Figure 5, where product rank and the offset of download increase are represented by the x-axis and y-axis respectively. Assume that product i's rank has been promoted by 1 from time $t-1$ to time t (i.e. $R_t^i = R_{t-1}^i - 1$),

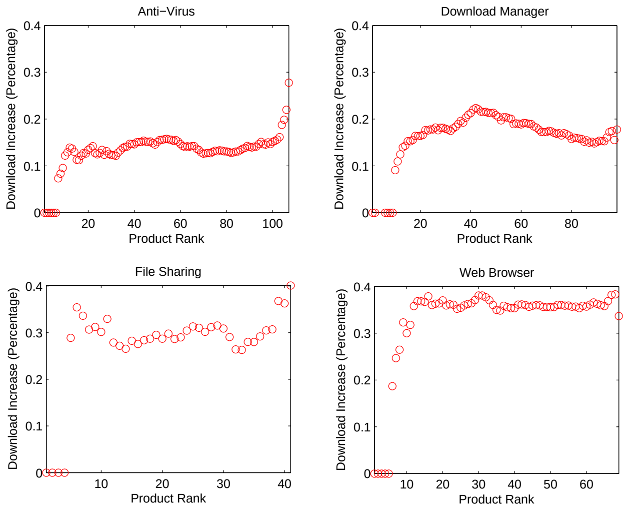


Fig. 4: Impact of Network Effect

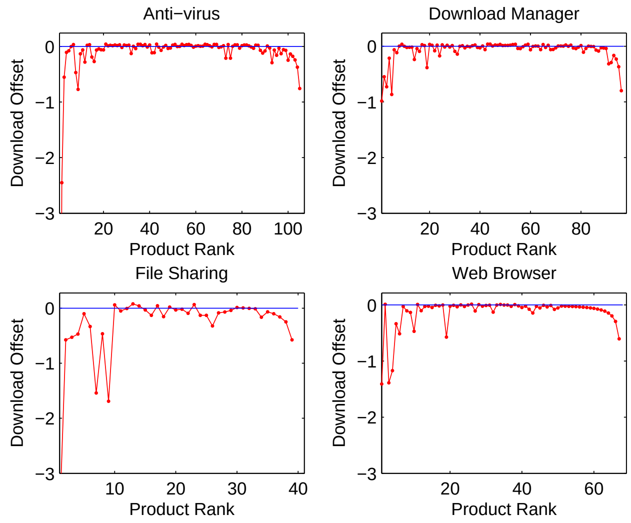


Fig. 5: Market Self-excitation

and such promotion comes from the extra increase of its total downloads from $\widetilde{d}^i_{t-1}$ to $\widetilde{d}^i_t$. Both of these two factors contribute to product i's market outcome improvement, which leads to an increase in the future weekly downloads by Δd_{achi}. Furthermore, we also assume that it requires Δd_{req} to continue climbing up by one rank (e.g. $R^i_{t+1} = R^i_t - 1$). In other words, Δd_{achi} represents the excitation power generated by products' market outcome improvement, while Δd_{req} represents the rank promotion requirement. The offset between these two values (i.e. $\Delta d_{achi} - \Delta d_{req}$) is illustrated for each product rank through the red curve in Figure 5. We also draw a blue horizontal line to indicate zeros for better illustration.

From Figure 5, we observe the same trend for all four markets. That is, $\Delta d_{achi} - \Delta d_{req}$ often yields non-negative values for medium ranks while negative values for top and low ranks. A non-negative offset at a given rank indicates that the self-exciting power is sufficient to cover the rank promotion requirement. In other words, if a product was pushed up to this rank from a lower one, it will be continuously and automatically excited to a higher rank with no need of further manipulations.

More important, in Figure 5, we only demonstrate the excitation power generated by one rank jump. If the product's rank has been improved by more than one position in the previous iteration, the self-excitation power will be stronger and may even turn the negative offset $\Delta d_{achi} - \Delta d_{req}$ to positive.

Second, we further illustrate the consecutive self-excitation through a specific example. The software in the File Sharing market named "BadBlue Personal Edition 2.7" is ranked 47 at week 26. Assume its rank is improved to 42 at week 27 by rating manipulations. As a result, we find that without introducing any further external energy (e.g. rating manipulations), its rank continues to climb up to 39, 38, 36, 35, 33, 31, 29 and 28 in the following 8 weeks and stops there at week 35. These rank transitions in week 28 $\sim$ 35 are all powered by the market's self-excitation.

To the best of our knowledge, we are the first to investigate such consecutive self-excitation of the online market. How to utilize this impact to facilitate rating manipulations will be further studied in Section V.

Through the above discussions, we discover that a product's rank can be promoted by either the "external energy" provided by rating manipulations or the "internal energy" generated by the market's self-exciting power. Inspired by this observation, we propose a novel iterative rating attack which enhances the manipulation impact on product sales/downloads by integrating these two types of energy.

V. Proposed Attack

Inspired by observations obtained in Section IV, in this section, we propose a novel iterative rating manipulation strategy by integrating both the external and internal promotion energy.

A. Attack Model

We make the following assumptions about the attack model.

- **Attack goal**: The attacker aims to boost the market outcome of the target product by inserting unfair ratings.
- **Attack power**: The attack power, measured by the total number of unfair ratings possessed by the attacker, is limited. We assume the number of unfair ratings is N. The attack scenario where attacker can overwhelm the market by inserting arbitrary number of unfair ratings is beyond the scope of this study.
- **Attack knowledge**: We assume that some product related information, product weekly downloads, market rank, existing rating value and volume, is made available to the public by the online market. Therefore, malicious attackers are also aware of such information for not only the target product but also all other products in the market.

In particular, we design the attack strategy by considering two aspects: (1) how to determine the unfair rating values and volume and (2) how to efficiently insert these unfair ratings, so that the target product's market outcome can be improved the most. These two aspects are analyzed in details in Section V-B and V-C.

B. Determine Unfair Rating Value and Volume

In Section IV, we observe that increasing a product's rating volume always leads to increase of its future downloads. However, overly boosting the rating value of a low-rating product may lead to drop of its future downloads. Based on such observations, we propose to determine unfair rating value and volume based on the target product's own property.

We first derive the equation to measure the impact of changes in rating value and volume on products' weekly downloads. Specifically, we take partial differentiation with respect to rating value and volume on both sides of equation (2) [46]. This leads to the following equation.

$$\Delta d = (\beta_2(\alpha) + 2\beta_3(\alpha)\bar{r}_o) * \Delta\bar{r} + \beta_5(\alpha)\Delta v \tag{3}$$

Specifically, Δd represents the increment of future weekly downloads; $\beta_2(\alpha), \beta_3(\alpha)$ and $\beta_5(\alpha)$ are parameters extracted from the market data by the regression model, where α represents product quantile; and $\bar{r}_o$ represents products' original average rating values. In addition, Δv and $\Delta\bar{r}$ are the increased rating volumes and values caused by rating manipulations. Hence, we have

$$\Delta v = v_{mal} \tag{4}$$

$$\begin{aligned}\Delta\bar{r} &= \frac{\bar{r}_{mal} * v_{mal} + \bar{r}_o * v_o}{(v_o + v_{mal})} - \bar{r}_o \\ &= \frac{\bar{r}_{mal} - \bar{r}_o}{v_{mal} + v_o} * v_{mal}\end{aligned} \tag{5}$$

where v_o denotes products' original rating volumes; $\bar{r}_{mal}$ and v_{mal} represent the value and volume of unfair ratings inserted by the attacker, respectively.

To boost the target product's market outcome, a malicious attacker has to ensure that the change in weekly downloads (Δd) is positive. By substituting Δv and $\Delta\bar{r}$ using equation (4) and (5), we obtain

$$\begin{aligned}&\Delta d = (\beta_2(\alpha) + 2\beta_3(\alpha)\bar{r}_o) * \frac{\bar{r}_{mal} - \bar{r}_o}{v_{mal} + v_o} v_{mal} + \beta_5(\alpha)v_{mal} > 0 \\ &\Rightarrow (\beta_2(\alpha) + 2\beta_3(\alpha)\bar{r}_o) * \frac{\bar{r}_{mal} - \bar{r}_o}{v_{mal} + v_o} + \beta_5(\alpha) > 0\end{aligned} \tag{6}$$

Note that we extract $\beta_2(\alpha) < 0$, $\beta_3(\alpha) > 0$ and $\beta_5(\alpha) > 0$ for all the products in the four markets from the regression model and therefore consider them true for a general market.

The inequality (6) can be analyzed in three cases according to the value of $(\beta_2(\alpha) + 2 * \beta_3(\alpha)\bar{r}_o)$.

$$(\beta_2(\alpha) + 2 * \beta_3(\alpha)\bar{r}_o) \begin{cases} > 0 & \text{if } \bar{r}_o > -\frac{\beta_2(\alpha)}{2\beta_3(\alpha)} \quad \text{case I} \\ = 0 & \text{if } \bar{r}_o = -\frac{\beta_2(\alpha)}{2\beta_3(\alpha)} \quad \text{case II} \\ < 0 & \text{if } \bar{r}_o < -\frac{\beta_2(\alpha)}{2\beta_3(\alpha)} \quad \text{case III} \end{cases}$$

Hence, we propose a rating value $r_c(\alpha)$, calculated as $r_c(\alpha) = -\frac{\beta_2(\alpha)}{2\beta_3(\alpha)}$, as the **critical rating value**, which changes across different product quantiles.

1) **Case I:** $(\beta_2(\alpha) + 2 * \beta_3(\alpha)\bar{r}_o) > 0$. Products with original rating values higher than the critical rating value fall in this case. To ensure positive impact on future downloads, manipulations on these products have to satisfy

$$\bar{r}_{mal} - \bar{r}_o > -\frac{\beta_5(\alpha) * (v_o + v_{mal})}{\beta_2(\alpha) + 2\beta_3(\alpha)\bar{r}_o} \tag{7}$$

Let

$$\lambda = \frac{\beta_5(\alpha) * (v_o + v_{mal})}{\beta_2(\alpha) + 2\beta_3(\alpha)\bar{r}_o}.$$

The value of λ is always positive, since $\beta_5(\alpha)$, $(\beta_2(\alpha) + 2 * \beta_3(\alpha)\bar{r}_o)$ and $v_o + v_{mal}$ are all positive. In addition, more powerful attacks (i.e. larger v_{mal}) on products with larger rating volumes (i.e. larger v_o) will yield larger λ values.

Feasible attacks: We consider attacks that have positive impact on products' weekly downloads (i.e. satisfying inequality (7)) as feasible attacks. Then, there are two types of feasible attacks. First, self-boosting attacks with $\bar{r}_{mal} - \bar{r}_o > 0$ can always increase products' weekly downloads. Second, an interesting observation is that even attacks with unfair rating values slightly smaller than the original ratings (i.e. $\bar{r}_{mal} \in (\bar{r}_o - \lambda,\ \bar{r}_o]$) may still cause positive impact on future downloads. And more powerful attacks on products with more existing rating volumes can tolerate an even larger offset between unfair ratings and original ratings. The potential reason is that high rating volume ensures the product's visibility which may then provide positive impact on the product's future downloads. Therefore, unfair ratings with their values in $(\bar{r}_o - \lambda,\ 5]$ can boost products' outcomes. The upper limit is set as 5 since it is assumed to be the maximum rating value.

Optimal attack: Among all the feasible attacks, the optimal one in this case is to insert all the N unfair ratings with the highest possible values, since Δd is monotonically increasing as $\bar{r}_{mal}$ and v_{mal} increase. Note that due to the employment of defense schemes, there is also a higher risk for the ratings with extremely high values to be identified as malicious ratings and removed from the system. The highest rating value that can avoid being detected is closely related with the specific defense scheme and is therefore beyond the scope of this study.

2) **Case II:** $(\beta_2(\alpha) + 2 * \beta_3(\alpha)\bar{r}_o) = 0$. Products with original rating values equal to the critical rating value fall in this case. In this case, inequality (6) is always satisfied since $\beta_5(\alpha)v_{mal}$ is always positive. The manipulation impact is only determined by malicious rating volume v_{mal}, and not related with $\bar{r}_{mal}$. Therefore, the **feasible attacks** include any attacks that insert ratings. The **optimal attack** in this case is to simply insert all the N unfair ratings with arbitrary values to avoid being detected.

3) **Case III:** $(\beta_2(\alpha) + 2 * \beta_3(\alpha)\bar{r}_o) < 0$. Products with original rating value lower than the critical rating value fall in this case. To achieve positive impact on future downloads, manipulations on these products have to satisfy:

$$\bar{r}_{mal} - \bar{r}_o < -\frac{\beta_5(\alpha) * (v_o + v_{mal})}{\beta_2(\alpha) + 2\beta_3(\alpha)\bar{r}_o} \tag{8}$$

Still let

$$\lambda = \frac{\beta_5(\alpha) * (v_o + v_{mal})}{\beta_2(\alpha) + 2\beta_3(\alpha)\bar{r}_o}.$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2017.2668992, IEEE Transactions on Information Forensics and Security

	Feasible Attacks	Optimal Attack
Type I $(\bar{r} > r_c(\alpha))$	$\bar{r}_{mal} \in (\bar{r}_o - \lambda,\ 5]$, $(\lambda > 0)$ $v_{mal} > 0$	$\bar{r}_{mal} = 5$ $v_{mal} = N$
Type II $(\bar{r} = r_c(\alpha))$	$\bar{r}_{mal} =$ any value $v_{mal} > 0$	$v_{mal} = N$
Type III $(\bar{r} < r_c(\alpha))$	$\bar{r}_{mal} \in [1,\ \bar{r}_o - \lambda)$, $(\lambda < 0)$ $v_{mal} > 0$	S1: $\bar{r}_{mal} = \bar{r}_o$ $v_{mal} = N$
		S2: $\bar{r}_{mal} = 1$ $v_{mal} = N$

TABLE II: Attacks for Different Product Types

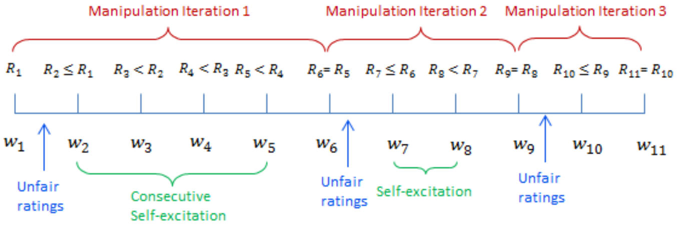


Fig. 6: Illustration of Proposed Attack

The value of λ is always negative, since the values of $\beta_5(\alpha)$, $v_o + v_{mal}$ and $(\beta_2(\alpha) + 2 * \beta_3(\alpha)\bar{r}_o)$ are positive, positive, and negative, respectively. Inequality (8) limits the maximum value of unfair ratings. In other words, a target product's market outcome may even drop if malicious attackers aggressively insert unfair ratings with arbitrary large values. Manipulations on these products are more tricky.

Feasible attacks: A straightforward self-boosting strategy is to carefully choose malicious rating values within the range $[\bar{r}_o,\ \bar{r}_o - \lambda)$. More powerful attacks on products with more existing rating volume can push the upper limit higher so that the range of feasible unfair rating values is larger. This strategy achieves the **best results** when $\bar{r}_{mal} = \bar{r}_o$, indicating that providing more ratings while not changing the original rating value could boost product downloads the most.

The second manipulation strategy is a little bit counter-intuitive because malicious attackers can insert unfair ratings with minimum rating values, such as 1-star, that are even lower than a product's original rating value $\bar{r}_o$, although they aim to boost the target product's market outcome. According to inequality (6), Δd is monotonically increasing as $\bar{r}_{mal}$ decreases. Therefore, this strategy achieves the **best results** when $\bar{r}_{mal} = 1$, where 1 is assumed to be the minimum rating value. Furthermore, this strategy can even outperform the straightforward strategy discussed in the above paragraph due to the low value of unfair rating. This matches the phenomenon that in reality, controversial items often obtain more visibility and attract more downloads. Despite its short term positive effect, this strategy will hurt the target product's market reputation in the long run. Therefore, businesses that aim to survive in the market for a long term may need to adopt it cautiously.

As a summary, we classify products into three types according to whether their original rating value is higher than, equal to, or lower than the corresponding critical rating value. The feasible and optimal manipulation strategies for each product type are investigated. Detailed information is listed in Table II. Note that the optimal attack is obtained by assuming the absence of defense scheme. When considering defense schemes, the attacker may need to find other feasible attacks to achieve the optimal boosting effect.

C. Insert Unfair Ratings in an Iterative Way

Once the unfair rating value and volume are determined, the next question is how to insert these unfair ratings in an efficient way. This question has been seldom investigated in existing studies. Through the following analysis, however, we have shown that different ways of inserting unfair ratings may yield significantly different attack results.

Inspired by the market's consecutive self-exciting property, we propose an iterative attack strategy (i.e. S_{iter}) that distributes the total N unfair ratings over multiple iterations. For each iteration, unfair ratings are only inserted in the first week, providing the initial power to stimulate the market's self-exciting power. If the self-excitation process occurs, the attacker will not insert unfair ratings until the target product's self-excited rank promotion stops.

Figure 6 illustrates three different manipulation iterations. Specifically, the time stamp w_i represents week i in this study, and R_i represents target product's rank at time w_i. A smaller R_i value indicates a higher/improved rank. In iteration 1, when unfair ratings are inserted after week 1, the product rank may get improved (i.e. $R_2 \leq R_1$). Consequently, a consecutive self-excitation process is triggered, leading to continuous product rank promotion in the following 3 weeks. Note that no further unfair ratings are inserted during these 3 weeks. Such process stops at week 5 and therefore, the product rank at week 6 is not further improved (i.e. $R_6 = R_5$). Being aware of the stop of self-excitation process, the attacker then launches the 2nd iteration after week 6. It's possible that the self-excitation is only stimulated once and does not cause continuous rank promotion, such as in manipulation iteration 2. Nevertheless, if the self-excitation is not stimulated by the insertion of unfair ratings, such as manipulation iteration 3, a new iteration will be launched (e.g. after week 11).

We further demonstrate such a process with a specific example. The software in the anti-virus market named "Defender Pro Anti Virus/Firewall 5.0.39" (DPAV), which has attracted 9088 total downloads and 24 ratings with an average value of 3, is ranked 71 at week 26 in the market. We manipulate this product by inserting 20 five-star ratings at week 26, which successfully boosts its rank to 70 at week 27. Without further inserting any unfair ratings, we continue tracking its rank transition and find that the rank climbs up to 69, 65, 62, 58, 52, 48, 47, and 46 in the following 8 weeks and stops at rank 43 at week 36. These rank transitions in week 28 $\sim$ 36 are all powered by the market's self-excitation since the manipulation only occurs at week 26. We consider the self-excitation stopped since the rank does not change any further, and then launch the second iteration by adding 30 more ratings at week 37. These 30 ratings further boost the product rank to 42 at week 38. Without any further manipulations, the product rank continues to be promoted to 41, 40, 38, 35, and stops at 33 at week 43.

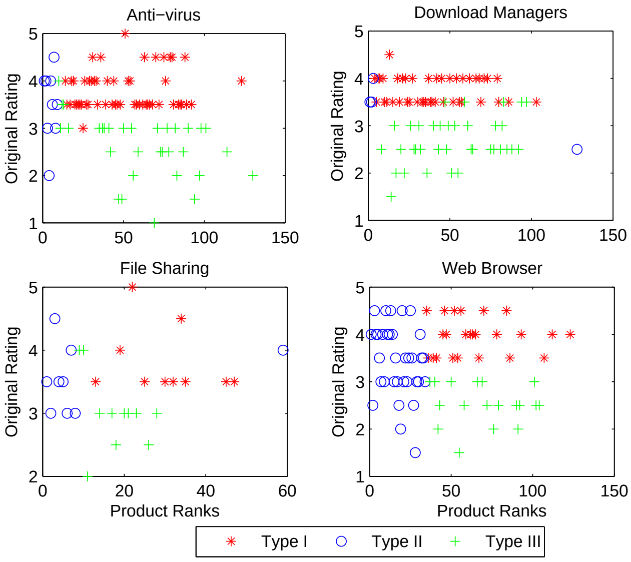


Fig. 7: Product Types

To further illustrate the efficiency of the iterative approach, we compare it to the commonly adopted approach, which inserts all the N unfair ratings for the target product at once. We call it the all-together strategy S_{all}. Still using the above example, we insert all the 50 unfair ratings at week 26, which boosts the product rank to 65 at week 27. Then the product rank is promoted by market self-excitation to 65, 60, 52, 47, 43 in the following 5 weeks and stops at 42 at week 32. The proposed scheme outperforms S_{all} due to the repeated utilization of the market self-exiting property.

VI. Experiment

In this section, we validate the impact of the proposed rating attack on products' market outcomes by using the 26th week data in our sample. As we mentioned earlier, the first 25-week data is used for estimating the regression model.

Generally speaking, a product's market outcome can be evaluated by either its download increase or its rank improvement. Download increase is adopted as the evaluation criteria in Section VI-A, since we focus on the impact of rating manipulations that directly target download increase. In Section VI-B, rank improvement is evaluated since it plays a significant role in the self-exciting process and we expect to push up the product rank within each iteration and over iterations. In such a scenario, rank improvement serves as a more direct and accurate way to assess attack impact.

A. Impact of Rating Changes

1) **Product Type Demonstration:** The analysis in Section V-B shows that the impact of rating manipulations is closely related to target product's own properties, including its market rank and existing rating. In this section, we would like to demonstrate product types based on the CNETD data set.

We illustrate the types of products in Figure 7. Specifically, four different market categories including "Anti-virus", "Download Managers", "File Sharing", and "Web Browsers" are represented in the upper left, upper right, lower left and lower right subplots, respectively. For each subplot, the x-axis represents product ranks and the y-axis represents products' original rating value.

Recall that the analysis in Section V-B identify product types based on the value of $\beta_2(\alpha)+2\beta_3(\alpha)\bar{r}_o$, which requires complex statistic studies on the entire market. Nevertheless, from Figure 7, we observe some much simpler rules to roughly identify product types based on their ranks and original rating values, which makes rating manipulations even easier.

- Top ranked or very low ranked products are often type II products, whose rating value changes does not affect their market outcomes much.
- Medium ranked products with higher original rating values are often type I products, whose rating value increase always yields positive impact on their market outcomes.
- Medium ranked products with low original rating values are often type III products, whose rating value increase may yield negative impact on their market outcomes.

The underlying rationale is that, facing a vast amount of products, online customers may first refer to extreme product ranks to make their decisions. On the one hand, products with extraordinary high ranks may still attract customers' attention even if their original rating values were very low. For example, in the "Web Browser" market, some type II products may even have their original rating values below 2-star. On the other hand, some extremely low ranked products may also belong to type II products because customers are scared by their ranking and are very unlikely to try them out, no matter how their rating values change.

For most products without extreme ranks in the market, customers are more sensitive to their rating value changes. In particular, for products with high original rating values (i.e. type I), customers tend to believe the rating value increase is caused by quality improvement, and are therefore more likely to download them. For products with low original rating values (i.e. type III), there is a higher chance for customers to relate the rating value increase, especially dramatic increase, to rating manipulations, leading to even negative impact on product downloads.

2) **Manipulations with Different Rating Values:** In this section, we examine the impact of manipulations with different unfair values on different types of products. The "Web Browser" market has been chosen as an example for demonstration. Specifically, 10 unfair ratings are inserted for each product and the increased downloads Δd are recorded. The results are shown in Figure 8, where the upper, medium and bottom subplots demonstrate manipulations on type I, type II and type III products, respectively. There are 5 curves in each subplot representing manipulations with 5 different unfair values from 1-star to 5-star.

From Figure 8, we observe different trends for different product types. (1) For each type I product, manipulations with higher rating value, such as 5-star, will attract more downloads than that with lower rating value do. (2) For each type III product, manipulations with higher rating values will attract fewer downloads. Moreover, overly inflated ratings (e.g. 5-star unfair ratings) can even cause negative impact on product

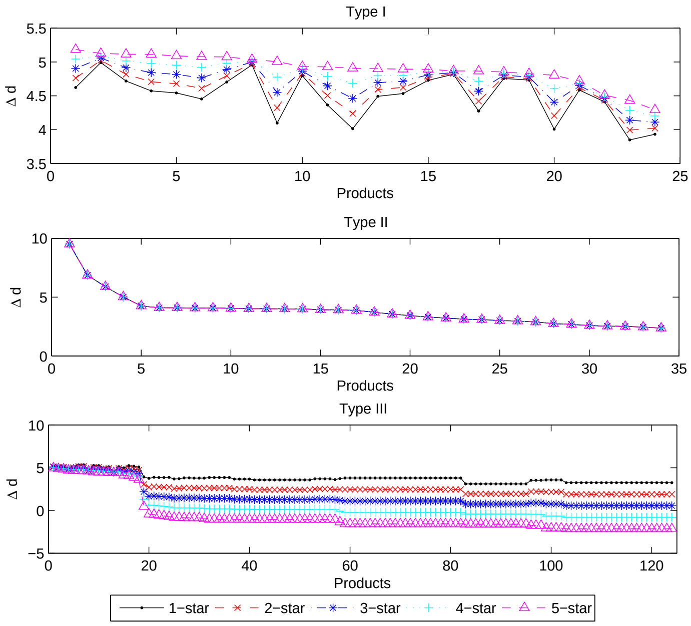


Fig. 8: Manipulations with 10 Unfair Ratings

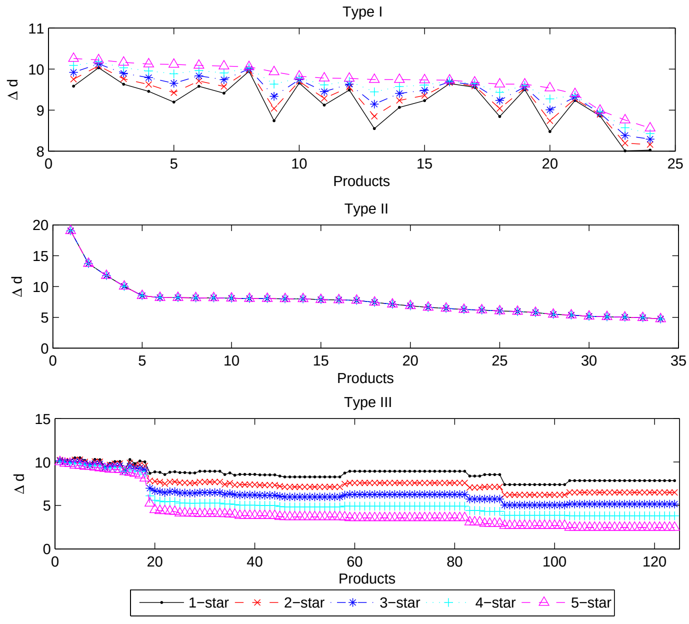


Fig. 9: Manipulations with 20 Unfair Ratings

downloads. (3) For each type II product, manipulations with same rating volume but different rating values always attract same amount of downloads. Note that, although the "Web Browser" market is chosen as an example, the same results are observed for all four categories.

3) **Manipulations with Different Rating Volumes:** In this section, we further investigate the impact of rating volumes on different types of products by repeating the above experiments with different unfair rating volumes. More unfair ratings, i.e. 20, are inserted for each product. Results from the " Web Browser" market are shown in Figure 9 for comparison purposes, while the same observations can still be made for the other three markets.

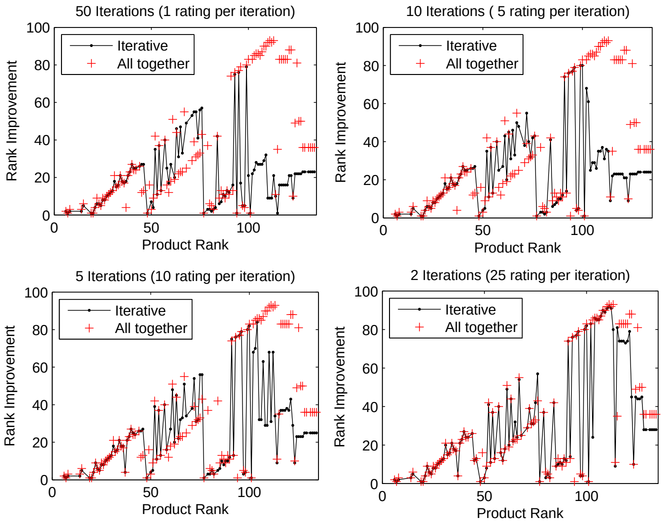


Fig. 10: Rank Improvement by Different Strategies

By comparing Figure 8 to Figure 9, we observe that for all three types of products, manipulations with larger volume lead to more future downloads. For example, with only 10 unfair ratings, the maximum Δd values are 5.2%, 10% and 5% for type I, II, and III products, respectively. With 20 unfair ratings, the maximum Δd values are increased to 10.4%, 20% and 10% for type I, II, and III products, respectively. It indicates that more powerful manipulations (i.e. larger volume) are able to boost products' downloads more.

The observations in this section and Section VI-A2 further validate the discussion in Section V-B.

B. Iterative Manipulations

To validate the feasibility of the proposed iterative manipulation strategy, we would like to compare its attack results to that of the conventional all-together strategy by fixing the manipulation power (i.e. unfair rating volume). In particular, our proposed iterative manipulation strategy starts with an initial rating attack followed by each iteration of continuously inserting fake reviews when the self-exciting power is exhausted.

In particular, we choose the Anti-virus market as an example to demonstrate the attack comparison, while similar observations are also found in the other three markets. The malicious attacker is supposed to have 50 unfair ratings in total. In the all-together strategy, all the 50 unfair ratings are inserted together at week 26. On the other hand, four iterative attacks are examined, where the 50 unfair ratings are evenly split into 50, 10, 5, and 2 iterations (i.e. weeks) to be inserted, respectively. In addition, each experiment chooses one product to be the target product. We repeatedly run this experiment for all the products in the Anti-virus market. The impact of these four iterative attacks on each product is compared to that of the all-together strategy and results are shown in Figure 10.

There are four subplots in Figure 10, comparing each of the four iterative strategies to the all-together strategy. In each plot, the x-axis represents product ranks, and the y-axis

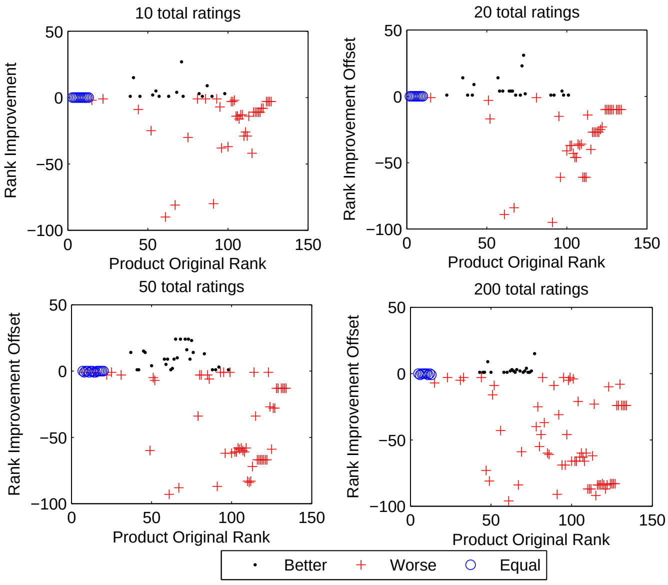


Fig. 11: Comparison between S_{iter} and S_{all}

represents the rank improvement caused by attacks. The rank improvement caused by the all-together strategy is marked as red plus sign, while that of the iterative strategies is marked as black points.

We can make two observations from Figure 10. First, the upper left (i.e. 50-iteration) strategy deviates the most from the all-together strategy, while the lower right (i.e. 2-iteration) strategy deviates the least. The reason is that as the number of iterations reduces, more ratings are inserted in each iteration, leading to more similar attack results as that of the all-together strategy. Second, for each specific subplot, the all-together strategy outperforms the iterative strategy on low ranked products and matches the iterative strategy on top ranked products. Nevertheless, on medium rank products, the iterative one often outperforms the all-together one. For example, among the 19 products ranking from 58 to 76, the 50-iteration strategy yields better rank improvement on 15 of them.

So far, we have fixed the manipulation power as 50 unfair ratings. Next, we would like to investigate if the same observations still exist for manipulations with different strength. Specifically, the manipulation power is adjusted as 10, 20, 50 and 200 unfair ratings in total, respectively. Since we observe in Figure 10 that the performance of the one-rating-per-iteration strategy deviates most from the all-together one, the following comparisons will focus on these two strategies only.

We will use S_{iter} to represent the one-rating-per-iteration strategy and S_{all} to represent the all-together strategy in the following discussions. Both S_{iter} and S_{all} are applied on each product. The rank improvement offset is calculated as the rank improvement caused by the iterative strategy subtracts that caused by the all-together one. This value, therefore, is positive when the iterative strategy yields better performance and is negative the other way round. The comparison results are shown in Figure 11.

Unfair Volume	Medium Rank Products		
	P_{range}	ΔR_{max}	P_{ratio}
10	$39 \sim 98$	43	81.67%
20	$25 \sim 101$	36	88.31%
50	$37 \sim 98$	26	77.42%
200	$43 \sim 77$	15	68.57%

TABLE III: Manipulation Impact on Medium Rank Products

In Figure 11, there are four subplots, representing four different manipulation power settings. For each subplot, the x-axis represents products' original rank while the y-axis represents the rank improvement offset. To assist the visualization of the results, we use black points, red plus and blue circles to represent products on which S_{iter} achieves better, worse or equal performance compared to S_{all}, respectively.

From Figure 11, we can still observe that regardless of the manipulation power, S_{all} always achieves the same manipulation impact as S_{iter} does on top rank products and beats S_{iter} on low rank products. On medium rank products, however, S_{iter} often outperforms S_{all}. We further conduct quantitative comparisons between S_{iter} and S_{all} on medium rank products in Table III.

Specifically, P_{range} represents the range of "medium" rank products; ΔR_{max} represents the maximum rank improvement offset, and P_{ratio} represents the percentage of "medium" rank products on which S_{iter} outperforms S_{all}. For example, for attacks with 10 unfair ratings, products ranking from 39 to 98 all belong to medium ranks, and the proposed attack S_{iter} achieves better performance on 81.67% of them. In the best case, the final rank of the target product manipulated by S_{iter} is 43 ranks higher than the case if it was manipulated by S_{all}.

We can observe high P_{ratio} values for all the four attack scenarios, indicating that the proposed S_{iter} outperforms S_{all} on most "medium" rank products. Moreover, such advantage varies over different manipulation power. The increase of manipulation power leads to range shrink of medium ranks as well as the value drop of P_{ratio} and ΔR_{max}, indicating less obvious advantages for iterative manipulations.

Through the above discussions, the attackers are able to derive their optimal manipulation strategy as follows. (1) If the target product is top ranked product, it does not really matter if the attacker chooses to insert unfair ratings all at once or through several iterations. (2) If the target product is low ranked product, the all-together strategy often yields better performance. (3) If the target product is "medium" ranked product, the attacker needs to choose two strategies alternatively according to his/her manipulation power. For "rich" attackers with sufficient manipulation power, it is generally better to insert the unfair ratings all at once. For attackers with constrained manipulation power, it is generally better to insert the unfair ratings through several iterations.

VII. Discussion and Future Work

The prosperity of online rating systems has significantly influenced the way people make their online purchasing/downloading decisions. Meanwhile, the simplicity of generating online ratings/reviews makes such systems vulnerable to diverse manipulations from malicious vendors in practice.

Being closely related to the root of this issue, the study of rating attack strategies, however, is still immature.

In this study, we first understand the economic impact of different influential factors on product sales/downloads by applying a quantile regression model on a real market data set that contains product download information. Based on such understandings, we then classify products into three types according to their own property and propose distinct feasible and optimal manipulation strategies for each product type. More important, we further disclose and validate the existence of the self-excited rank promotion. By integrating manipulation power with market's self-exciting power, we then propose a novel iterative rating attack and validate its advantage through experiments.

Beyond the specific CNETD context in this study, we would like to further discuss the feasibility of the proposed attack on other different online rating platforms. First, quantile regression models have been adopted by a number of prior work [25]–[28] to evaluate the impact of user ratings on products' market outcome on a variety of online rating platforms and have demonstrated convincing results. Therefore, we believe that the market self-exciting power observed based on such model also exists across different online rating platforms.

Second, we acknowledge that the proposed attack may be detected by existing defense schemes. One example is the user behavior pattern based unfair rating detection [8], [43]–[45], which identifies malicious reviewers by checking their review history and/or social networking connections. Furthermore, since the proposed attack aims to promote products' rank, another potential defense could be tracking products' ranking and identifying the products with rapid rank changes as suspicious products. In addition, many online rating platforms verify user ratings/reviews through manual checking, customer reporting, or transaction verification, which makes the success of the proposed attack more difficult.

Nevertheless, the proposed attack is still feasible on different online rating platforms due to two reasons. (1) It can be adjusted to partially avoid, if not completely, being detected by these defense schemes by slowly changing the target product's rating value/volume and spending more time/money to mimic normal behavior pattern. The tradeoff is that the manipulation cost is increasing and the process becomes longer, which may lead to a slower increase of the manipulation profits. (2) More important, in the proposed attack, malicious attackers can customize attack strategies based on their manipulation power, the specific property of the target product and the market self-exciting power. Most of current defense schemes, however, use same detection settings for all products and therefore cannot effectively detect each individual target product promoted by the proposed attack. Therefore, we believe that the proposed attack is feasible across different online rating platforms.

Through this study, we also identify several interesting directions for further investigation. First, we emphasize that the promotional effect on products' market outcome is determined by not only attacker's manipulation power but also the specific property of the target product and the market self-exciting power. To the best of our knowledge, this is the first work to make this statement. It also guides the future attack/defense studies in online rating systems towards the direction of customization, which is rarely investigated in current literature. Second, time factor may play an important role in influencing attack results. There have been many online articles/reports discussing the best time of a day to generate influential posts on different social networking sites, such as Facebook, Twitter, Instagram, and etc., whereas the reported best timing seems to vary across different platforms, regions, target audience and contents. Nevertheless, there is limited scientific work to study how different timing of fake rating injection will affect the manipulation impact. Our current data set only allows the analysis based on each week. We plan to collect data sets with finer-grained time stamps. Last but not least, we would like to study the tradeoff between bypassing defense schemes and making manipulation profits faster. An optimized solution to balance attack costs and profits will make the proposed attack even stronger.

References

[1] *The impact of customer service on customer lifetime value*, https://www.zendesk.com/resources/customer-service-and-lifetime-customer-value.

[2] L. Andeuw, *Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior*, http://www.prnewswire.com/news-releases/online-consumer-generated-reviews-have-significant-impact-on-offline-purchase-behavior-59899937.html, 2007.

[3] *Established eBay sellers get higher prices for good reputations*, http://www.ns.umich.edu/new/releases/329-established-ebay-sellers-get-higher-prices-for-good-reputations.

[4] J. Brown and J. Morgan, "Reputation in online auctions: The market for trust," *California Management Review*, vol. 49, no. 1, pp. 61–81, 2006.

[5] A. Harmon, "Amazon glitch unmasks war of reviewers," *The New York Times*, vol. 14, no. 8, 2004.

[6] N. Hu, I. Bose, N. S. Koh, and L. Liu, "Manipulation of online reviews: An analysis of ratings, readability, and sentiments," *Decision Support Systems*, vol. 52, no. 3, pp. 674–684, 2012.

[7] M. Luca and G. Zervas, "Fake it till you make it: Reputation, competition, and yelp review fraud," *Management Science*, vol. 62, no. 12, pp. 3412–3427, 2016.

[8] C. Dellarocas, "Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior," in *Proceedings of the 2nd ACM conference on Electronic commerce*. ACM, 2000, pp. 150–157.

[9] C. Dellarocas, "Mechanisms for coping with unfair ratings and discriminatory behavior in online reputation reporting systems," in *Proceedings of the twenty first international conference on Information systems*. Association for Information Systems, 2000, pp. 520–525.

[10] C. Dellarocas, "The digitization of word of mouth: Promise and challenges of online feedback mechanisms," *Management science*, vol. 49, no. 10, pp. 1407–1424, 2003.

[11] J. R. Douceur, "The sybil attack," in *International Workshop on Peer-to-Peer Systems*. Springer, 2002, pp. 251–260.

[12] M. Srivatsa, L. Xiong, and L. Liu, "Trustguard: countering vulnerabilities in reputation management for decentralized overlay networks," in *Proceedings of the 14th international conference on World Wide Web*. ACM, 2005, pp. 422–431.

[13] Y. Yang, Q. Feng, Y. L. Sun, and Y. Dai, "Reptrap: a novel attack on feedback-based reputation systems," in *Proceedings of the 4th international conference on Security and privacy in communication netowrks*. ACM, 2008, p. 8.

[14] P. Laureti, L. Moret, Y.-C. Zhang, and Y.-K. Yu, "Information filtering via iterative refinement," *EPL (Europhysics Letters)*, vol. 75, no. 6, p. 1006, 2006.

[15] Y. Liu and Y. Sun, "Anomaly detection in feedback-based reputation systems through temporal and correlation analysis," in *Social Computing (SocialCom), 2010 IEEE Second International Conference on*. IEEE, 2010, pp. 65–72.

[16] Y. Liu, Y. Sun, and T. Yu, "Defending multiple-user-multiple-target attacks in online reputation systems," in *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*. IEEE, 2011, pp. 425–434.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2017.2668992, IEEE Transactions on Information Forensics and Security

[17] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," *Journal of marketing research*, vol. 43, no. 3, pp. 345–354, 2006.

[18] W. Zhou and W. Duan, "An empirical study of how third-party websites influence the feedback mechanism between online word-of-mouth and retail sales," *Decision Support Systems*, vol. 76, pp. 14–23, 2015.

[19] D. Godes and D. Mayzlin, "Using online conversations to study word-of-mouth communication," *Marketing science*, vol. 23, no. 4, pp. 545–560, 2004.

[20] W. Duan, B. Gu, and A. B. Whinston, "The dynamics of online word-of-mouth and product salesan empirical investigation of the movie industry," *Journal of retailing*, vol. 84, no. 2, pp. 233–242, 2008.

[21] D. Mayzlin, "Promotional chat on the internet," *Marketing Science*, vol. 25, no. 2, pp. 155–163, 2006.

[22] X. Luo, J. Zhang, and W. Duan, "Social media and firm equity value," *Information Systems Research*, vol. 24, no. 1, pp. 146–163, 2013.

[23] W. Zhou and W. Duan, "Do professional reviews affect online user choices through user reviews?: An empirical study," *Journal of Management Information Systems*, vol. 33, no. 1, pp. 202–228, 2016.

[24] Y. Liu, "Word of mouth for movies: Its dynamics and impact on box office revenue," *Journal of marketing*, vol. 70, no. 3, pp. 74–89, 2006.

[25] W. Duan, B. Gu, and A. B. Whinston, "Informational cascades and software adoption on the internet: an empirical investigation," *Mis Quarterly*, pp. 23–48, 2009.

[26] W. Zhou and W. Duan, "Online user reviews, product variety, and the long tail: An empirical investigation on online software downloads," *Electronic Commerce Research and Applications*, vol. 11, no. 3, pp. 275–289, 2012.

[27] E. Brynjolfsson, Y. J. Hu, and M. D. Smith, "From niches to riches: Anatomy of the long tail," *Sloan Management Review*, vol. 47, no. 4, pp. 67–71, 2006.

[28] F. Zhu and X. Zhang, "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics," *Journal of marketing*, vol. 74, no. 2, pp. 133–148, 2010.

[29] A. Elberse and F. Oberholzer-Gee, *Superstars and underdogs: An examination of the long tail phenomenon in video sales*. Citeseer, 2006.

[30] R. Koenker and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.

[31] E. Brynjolfsson and C. F. Kemerer, "Network externalities in microcomputer software: An econometric analysis of the spreadsheet market," *Management Science*, vol. 42, no. 12, pp. 1627–1647, 1996.

[32] J. M. Gallaugher and Y.-M. Wang, "Understanding network effects in software markets: evidence from web server pricing," *Mis Quarterly*, pp. 303–327, 2002.

[33] A. V. Banerjee, "A simple model of herd behavior," *The Quarterly Journal of Economics*, pp. 797–817, 1992.

[34] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of political Economy*, pp. 992–1026, 1992.

[35] Y. L. Sun, Z. Han, W. Yu, and K. R. Liu, "A trust evaluation framework in distributed networks: Vulnerability analysis and defense against attacks." in *INFOCOM*, vol. 2006, 2006, pp. 1–13.

[36] K. Hoffman, D. Zage, and C. Nita-Rotaru, "A survey of attack and defense techniques for reputation systems," *ACM Computing Surveys (CSUR)*, vol. 42, no. 1, p. 1, 2009.

[37] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman, "Reputation systems," *Communications of the ACM*, vol. 43, no. 12, pp. 45–48, 2000.

[38] M. Abadi, M. Burrows, B. Lampson, and G. Plotkin, "A calculus for access control in distributed systems," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 15, no. 4, pp. 706–734, 1993.

[39] M. Feldman, C. Papadimitriou, J. Chuang, and I. Stoica, "Free-riding and whitewashing in peer-to-peer systems," in *Proceedings of the ACM SIGCOMM workshop on Practice and theory of incentives in networked systems*. ACM, 2004, pp. 228–236.

[40] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, "Sybilguard: defending against sybil attacks via social networks," in *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 4. ACM, 2006, pp. 267–278.

[41] A. Whitby, A. Jøsang, and J. Indulska, "Filtering out unfair ratings in bayesian reputation systems," in *Proc. 7th Int. Workshop on Trust in Agent Societies*, vol. 6, 2004, pp. 106–117.

[42] B. E. Commerce, A. Jøsang, and R. Ismail, "The beta reputation system," in *In Proceedings of the 15th Bled Electronic Commerce Conference*. Citeseer, 2002.

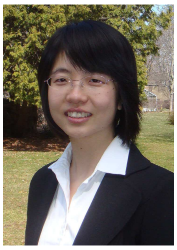
[43] M. Chen and J. P. Singh, "Computing and using reputations for internet ratings," in *Proceedings of the 3rd ACM conference on Electronic Commerce*. ACM, 2001, pp. 154–162.

[44] J. Zhang and R. Cohen, "Trusting advice from other buyers in e-marketplaces: the problem of unfair ratings," in *Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet*. ACM, 2006, pp. 225–234.

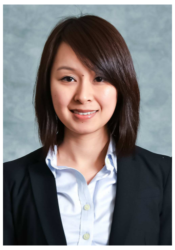
[45] W. Jianshu, M. Chunyan, and G. Angela, "An entropy-based approach to protecting rating systems from unfair testimonies," *IEICE TRANSACTIONS on Information and Systems*, vol. 89, no. 9, pp. 2502–2511, 2006.

[46] W. H. Greene, *Econometric analysis*. Pearson Education India, 2003.

[47] Y. Ying, F. Feinberg, and M. Wedel, "Leveraging missing ratings to improve online recommendation systems," *Journal of marketing research*, vol. 43, no. 3, pp. 355–365, 2006.



Yuhong Liu Assistant Professor at Department of Computer Engineering Santa Clara University, received her B.S. and M.S. degree from Beijing University of Posts and Telecommunications in 2004 and 2007 respectively, and the Ph.D. degree from University of Rhode Island in 2012. She is the recipient of the 2013 University of Rhode Island Graduate School Excellence in Doctoral Research Award. With expertise in trustworthy computing and cyber security, her research interests include developing trust models and applying them on emerging applications, such as online social media, cyber-physical systems and cloud computing. Her work on securing online reputation systems received the best paper award at the IEEE International Conference on Social Computing 2010 (acceptance rate = 13%). She also received best paper award at the 9th International Conference on Ubi-Media Computing (UMEDIA 2016).



Wenqi Zhou Assistant Professor of Information Systems Management at Palumbo-Donahue School of Business, Duquesne University. She received her Ph.D. in Information Systems from The George Washington University in 2013. Her research primarily focuses on understanding social, economic and managerial aspects of information technology and the Internet by analyzing large-scale online data. Specifically, her research interests include social media and online user-generated content, the economics of e-commerce, Internet marketing, and online communities and intermediaries. Her works have been published in Journal of Management Information Systems, Decision Support Systems, IEEE Computer, Electronic Commerce Research and Applications, and ICIS proceedings, among others. She also received Best Paper Award at the pre-ICIS workshop WeB 2012.

Hong Chen Visiting Assistant Professor of Information Systems Management in Palumbo-Donahue School of Business at Duquesne University in Pittsburgh. His research is interdisciplinary, and spans Information Systems and Analytics. His recent research examines the impact of online User Generated Content and cyber fraud on Internet market outcome. Prior to the currently role, he was VP for Research at Labonachip LLC, and Adjunct Faculty at University of Rhode Island. He received his Ph.D. degree from Department of Mechanical, Industrial and Systems Engineering at University of Rhode Island.