# A Novel Iterative Online Rating Attack based on Market Self-exciting Property

Yuhong Liu<sup>1</sup> and Wenqi Zhou<sup>2</sup>

<sup>1</sup>Department of Computer Engineering, Santa Clara University, Santa Clara, CA, 95050 <sup>2</sup>Palumbo Donahue School of Business, Duquesne University, Pittsburgh, PA, 15282 Emails:yhliu@scu.edu, zhouw@duq.edu

Abstract—The prosperity of online rating system makes it an important place for malicious vendors to mislead public's online decisions, whereas the security related studies are lagging behind. In this work, we adopt a quantile regression model to investigate influential factors on online user choices and reveal the "self-exciting" property of online market. Inspired by these findings, we propose a novel iterative rating attack and validate its advantage through experiments.

Index Terms-rating manipulation, social media, security attack

# I. Introduction

With the rapid growth of e-commerce and social media, online rating systems that let users post ratings/reviews for products and services are playing an increasingly important role in influencing users' online purchasing/downloading decisions. The significant impact of such systems provides 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 [1]–[3]. A recent study has identified that 10% online products have manipulated user ratings [4].

The protection of online rating systems in essence grounds in the thorough understanding of how attack strategy works. Hence, a number of researches have been conducted to investigate rating attack strategies. Generally speaking, rating attacks can be classified into two categories as self-boosting attacks, where malicious users aim to boost rating scores of their own products and badmouthing attacks, where malicious users aim to downgrade rating scores of other competitors' products [5]–[7]. Specifically, diverse rating manipulation strategies have been proposed, such as Sybil attack [8], Oscillation attack [9] and RepTrap attack [10].

Nevertheless, existing rating attack studies are still immature in three aspects. First, the attack impact is only evaluated by how much distortion has been caused on target product's rating score, or how many unfair ratings have bypassed the detection scheme, while the economic impact on target product sales/downloads is seldom considered. 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, current attack studies often treat products homogeneously while not differentiating the impact of the same attack on products with different properties, such as existing ratings and market ranks.

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 product's sales get increased by a rating manipulation at time t - 1, the greater sales and higher popularity at time t 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 based 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. 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 selfboosting 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 based regression model is introduced in Section III-A and applied on data described in Section III-B. Observations 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. Rating Manipulation Studies

The design of rating attack strategies has been conducted by many security studies and is evolving dynamically. 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 [1].

Collusion attacks, where excessive number of online IDs coordinately insert unfair ratings, are adopted by many rating manipulation strategies as a more powerful attack [11], [12]. The Sybil attack [8] is a typical example of collusion attacks. The colluded malicious users can (1) provide high ratings for self-promoting; (2) provide low ratings for bad-mouthing [5]–[7]; (3) restore their reputation by providing honest ratings to products that they do not care [13], [14]; or (4) whitewash their reputation by registering new user IDs [15].

Advanced collusion attacks, where malicious IDs perform more diverse yet coordinate tasks, are proposed to further strengthen manipulation impact and to avoid being deected. For instance, in Oscillation attacks [9], multiple malicious user groups may perform different rating behavior to protect one another from being detected. In RepTrap attacks [10], attackers overturn the reputation of products to undermine honest users' reputation.

Note that, in this study, we mainly focus on how to enhance manipulation impact when a certain number of unfair ratings are 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.

# **B.** Influential Factors for Product Sales/Downloads

Many researches have already been conducted to investigate how different factors may influence product sales/downloads. Rating value and volume are generally recognized as critical influential factors [16]–[21]. While rating volume is often believed to have linear impact [19], [22], the impact of rating value is found to be nonlinear [16], [23], indicating that a fixed increase in rating value can lead to disparate market sales/downloads for products with different existing ratings.

More interestingly, researches in recent years have found that the impact of rating value and volume differs over products with different popularity [24]–[27], which is often represented by market ranks. To accurately capture such impact, quantile regression models have been proposed by prior studies [25], [28] and are shown to be robust and appropriate to estimate the differential impact of influential factors on the whole distribution of the product sales/downloads variable [29].

A product's sales/downloads may also be affected the network effect and herding effect. First, the network effect indicates that the greater user base of a product will help expand its market share [24], [30], [31]. An extreme example is that if a user's friends all use a particular chatting software program, very likely this user will adopt the same software in order to stay in touch. Second, the herding effect [32], [33] describes that online consumers are empirically proved to follow others' adoption decisions [24]. In other words, if a product gets popular and ranked higher in the market, consumers may follow their predecessors' steps and also choose this more popular product.

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

# III. Model and Data

#### A. Quantile Regression Model

As quantile is defined as the quantile of the outcome variable distribution, the quantile regression model as a superior estimation method is designed to evaluate the impact of influential factors on the entire distribution of the outcome variable [29]. The general form of a quantile regression model is expressed as:

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

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

In this study, we adopt the quantile regression model to estimate the impact of any changes introduced by rating manipulations on product sales/downloads. Therefore, we use y to measure product sales/downloads. A logarithm transformation is applied on y to cope with the scale effect [16], [22], [24].

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 [19], [23]. That is,  $\beta_2(\alpha) * \bar{r}_t^i + \beta_3(\alpha) * \bar{r}_t^i * \bar{r}_t^i$ . Moreover, we include a binary variable  $Rev_u_t^i$  [34] to indicate if product i has received any user rating by week t. 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 [16]. Third, the herding effect, cumulatively proxied by product rank (i.e. rank  $R_t^i$ ), is represented by  $\beta_4(\alpha) * R_t^i$  [24]. Fourth, we include  $\beta_6(\alpha) * log(d_t^i)$  to capture the network effect [24], [30], [31], where  $d_t^i$  is product i's total number of sales/downloads by week t. The logarithm transformation is applied to deal with the large variance of total product sales/downloads, otherwise the coefficient will not have enough degrees of freedom to be statistically estimated [34]. In addition to those key influential factors, we also adopt two control variables: product age  $(Aqe_t)$  and its square term  $(Aqe_t^2)$  proposed in [24] to model product sales/downloads more robustly.

A notorious confounding factor is the endogeneity caused by reverse causality [34]. To control for this issue, we adopt one time lag in all independent variables [18], [22], [34], so that all independent variables at time t - 1 are included to eliminate the possibility of reverse causality.

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

$$log(d_{t}^{i})(\alpha) = \beta_{0}(\alpha) + \beta_{1}(\alpha) * Rev_{u_{t-1}^{i}} + \beta_{2}(\alpha) * \tilde{r}_{t-1}^{i} + \beta_{3}(\alpha) * \tilde{r}_{t-1}^{i} * \tilde{r}_{t-1}^{i} + \beta_{4}(\alpha) * R_{t-1}^{i} + \beta_{5}(\alpha) * log(\tilde{v}_{t-1}^{i}) + \beta_{6}(\alpha) * log(\tilde{d}_{t-1}^{i}) + \beta_{x}(\alpha)Controls_{x,i,t} + \xi_{i,t}(\alpha)$$
(2)

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

#### B. Context and Data

Our data is collected from CNET Download.com (CNETD), an online platform providing more than 30,000 free or free-totry 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, which well captures products' online market share that is rarely available on other platforms.

In particular, we collect a weekly panel data of software downloading 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  $(\tilde{d}_t^i)$ , the average rating values  $(\tilde{r}_t^i)$ , the rating volume  $(\tilde{v}_t^i)$ , 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, 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  $Controls_{x,i,t}$ .

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**

Based on parameters estimated through the quantile regression model, we aim to answer the question as how a product's popularity, measured by its market downloads, can be promoted. Specifically, we understand this problem by using the analogy of energy level transition to product popularity level transition (or promotion). 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 popularity level, which is measured by its market rank. Then the problem becomes what sources can provide such "promotion energy".

The most obvious one is the rating manipulation which inserts overly inflated ratings, as we observe positive impact on product downloads from rating value and volume based on the values of  $\beta_2(\alpha), \beta_3(\alpha)$ , and  $\beta_5(\alpha)$ .

In addition, we also discover an interesting phenomenon as the **self-excited rank promotion**. Specifically, if a product gets its popularity promoted, which includes either rank improvement or downloads increase, it will receive some pushing up power from the market which helps it attract even more downloads. We call such pushing up power as the market's "self-exciting power". The extra increased downloads, once exceed the rank transition requirement, may cause further rank promotion. The rank promotion caused by the self-exciting power is so called self-excited rank promotion.

We validate the existence of the self-excited rank promotion through two steps: (1) whether a product's popularity improvement can boost its future downloads and (2) whether the increased downloads can be large enough so that the product can jump to a higher rank.

As the first step, we check the impact of a product's popularity improvement on its future downloads. Recall that such impact is captured by two quantile based parameters  $\beta_4(\alpha)$  and  $\beta_6(\alpha)$ . By examine the values of these two parameters, we find them as always positive for all products in all four markets, indicating a positive impact of product popularity improvement on its future downloads.

We validate the second step in Figure 1, where product rank and download increase are represented by the x-axis and y-axis respectively. Assume that product *i*'s rank has been improved by 1 from time t - 1 to time t (i.e.  $R_t^i = R_{t-1}^i - 1$ ), and such improvement also comes together with the extra increase of its total downloads from  $\tilde{d}_{t-1}^i$  to  $\tilde{d}_t^i$ . Both of these two factors contribute to product *i*'s popularity improvement, which leads to an increase in the future downloads by  $\Delta d_{achi}$ . Furthermore, we also assume that it requires  $\Delta d_{req}$  to continue climbing up by one rank (e.g.  $R_{t+1}^i = R_t^i - 1$ ). In other words,  $\Delta d_{achi}$  represents the excitation power generated by popularity improvement, while  $\Delta d_{req}$  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 1. We also draw a blue horizontal line to indicate zeros for better illustration.

A non-negative offset at a given rank indicates that the selfexciting power is sufficient to cover the rank promotion requirement. In other words, if a product is pushed up to this rank from



Fig. 1: Market Self-excitation

a lower one, it will be automatically excited to a higher rank with no need of manipulations.

From Figure 1, we observe the same trend for all four categories. That is,  $\Delta d_{achi} - \Delta d_{req}$  often yields non-negative values for medium ranks while negative values for top and low ranks, indicating a higher possibility of self-excitation effect occurring at medium ranks. More important, in Figure 1, we only demonstrate the excitation power generated by one rank jump and its corresponding total download increase. If the product's rank has been improved by more than one in the previous iteration, the self-excitation power will be larger and may even turn the offset  $\Delta d_{achi} - \Delta d_{req}$  from negative to positive.

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 energy.

# V. Proposed Attack

We assume the attack model as that an attacker with a fixed number of unfair ratings (i.e. N) aim to promote target product's rank as much as possible. Inspired by the market's 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 beginning, providing the initial power to enable target product's rank promotion. Then the attacker just wait and see if the market self-exciting power can provide further energy to continue pushing up target product's rank. A new iteration will not be launched until target product's rank transition stops.

We demonstrate such process through a specific example. For instance, 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 as 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



Fig. 2: Comparison between  $S_{iter}$  and  $S_{all}$ 

self-excitation since the manipulation only occurs at week 26. We consider the self-excitation stops 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.

Through this example, we see that by inserting unfair ratings through multiple iterations, the target product absorbs not only the manipulation power, but also the market self-exciting power at each iteration.

## VI. Experiment

To validate the feasibility of the proposed iterative attack strategy  $S_{iter}$ , we would like to compare its attack results to that of the conventional all-together strategy  $S_{all}$ , which inserts unfair ratings all at once, by fixing the manipulation power (i.e. unfair rating volume). In particular, we choose the Anti-virus market as an example to demonstrate the attack comparison, while similar observations can also be extracted from the other three markets. We simulate four attack scenarios where the total number of unfair ratings is adjusted as 10, 20, 50 and 200, respectively. In  $S_{all}$ , all unfair ratings are inserted together at week 26. In  $S_{iter}$ , to maximally utilize the market's self-excitation property, only one unfair rating is inserted in each iteration. Furthermore, each experiment chooses one product to be the target product. We repeatedly run this experiment for all the products in the Antivirus market. The impact of these two attacks on each product is compared and results are shown in Figure 2.

In Figure 2, 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, which is calculated as the rank improvement caused by  $S_{iter}$  subtracts that caused by  $S_{all}$ . This value, therefore, is positive when  $S_{iter}$  yields better performance and is negative the other way round. To assist the visualization of the results, black points are used to represent the products on which  $S_{iter}$  outperforms  $S_{all}$ , while red plus signs are used to represent the rest products.

From Figure 2, we can observe that  $S_{all}$  achieves the same manipulation impact as  $S_{iter}$  does on top rank products, while beating

Unfair	Medium Rank Products		
Volume	$P_{range}$	$\Delta R_{max}$	$P_{ratio}$
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 I: Manipulation Impact on Medium Rank Products

 $S_{iter}$  on low rank products, regardless of the manipulation power. 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 I.

Specifically,  $P_{range}$  represents the range of "medium" rank products;  $\Delta R_{max}$  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 attack 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  $\Delta R_{max}$ , indicating less obvious advantages for iterative manipulations.

Through the above discussions, we argue that malicious attackers should dynamically adjust their attack strategy according to the target product's property as well as their own manipulation power. (1) If the target product is a top rank 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 a low rank product, the all-together strategy often yields better performance. (3) If the target product is a medium rank product, inserting unfair ratings in an iterative way yields better performance for most of the time. Nevertheless, the range of medium ranks is related to the attacker's manipulation power. For attackers with constrained manipulation power,  $S_{iter}$  yields better performance on a larger range of products, while for "rich" attackers with overwhelming manipulation power, such advantage becomes less obvious.

#### VII. Conclusion

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 very vulnerable to diverse manipulations from malicious vendors in practice. The study of rating attack strategies, however, is still very simple and immature.

In this study, we first understand the impact of different influential factors on product sales/downloads by applying a quantile based regression model on a real market data set that contains product download information. We further disclose and validate the existence of the self-excited rank promotion. By integrating market's self-exciting power, we then propose a novel iterative rating attack and validate its advantage through experiments. We conclude this paper by arguing that the attack impact is determined by not only attacker's manipulation power but also the market and target product's own property. Such observation will benefit the future attack/defense studies in online rating systems.

# REFERENCES

- J. Brown and J. Morgan, "Reputation in online auctions: The market for trust," *California Management Review*, vol. 49, no. 1, pp. 61–81, 2006.
- [2] S. Charmen-Anderson, "Fake reviews: Amazons rotten core," 2012.
- [3] A. Harmon, "Amazon glitch unmasks war of reviewers," *The New York Times*, vol. 14, no. 8, 2004.
- [4] 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.
- [5] 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.
- [6] ——, "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.
- [7] ——, "The digitization of word of mouth: Promise and challenges of online feedback mechanisms," *Management science*, vol. 49, no. 10, pp. 1407–1424, 2003.
- [8] J. R. Douceur, "The sybil attack," in International Workshop on Peer-to-Peer Systems. Springer, 2002, pp. 251–260.
- [9] 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.
- [10] 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.
- [11] 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 International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE, 2011, pp. 425–434.
- [12] Y. Liu and Y. Sun, "Anomaly detection in feedback-based reputation systems through temporal and correlation analysis," in *Social Computing (Social-Com), 2010 IEEE Second International Conference on*. IEEE, 2010, pp. 65–72.
- [13] 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.
  [14] K. Hoffman, D. Zage, and C. Nita-Rotaru, "A survey of attack and de-
- [14] 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.
- [15] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman, "Reputation systems," *Communications of the ACM*, vol. 43, no. 12, pp. 45–48, 2000.
  [16] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales:
- [16] 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.
- [17] 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.
- [18] D. Godes and D. Mayzlin, "Using online conversations to study word-ofmouth communication," *Marketing science*, vol. 23, no. 4, pp. 545–560, 2004.
- [19] W. Duan, B. Gu, and A. B. Whinston, "The dynamics of online word-ofmouth and product salesan empirical investigation of the movie industry," *Journal of retailing*, vol. 84, no. 2, pp. 233–242, 2008.
- [20] D. Mayzlin, "Promotional chat on the internet," *Marketing Science*, vol. 25, no. 2, pp. 155–163, 2006.
- [21] X. Luo, J. Zhang, and W. Duan, "Social media and firm equity value," *Information Systems Research*, vol. 24, no. 1, pp. 146–163, 2013.
- [22] 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.
- [23] W. Zhou and W. Duan, "Do professional reviews affect online user choices through user reviews?: An empirical study," *Wenqi Zhou and Wenjing Duan*, 2016.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] A. Elberse and F. Oberholzer-Gee, Superstars and underdogs: An examination of the long tail phenomenon in video sales. Citeseer, 2006.
- [29] R. Koenker and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.

- [30] 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.
- [31] J. M. Gallaugher and Y.-M. Wang, "Understanding network effects in software markets: evidence from web server pricing," *Mis Quarterly*, pp. 303–327, 2002.
- [32] A. V. Banerjee, "A simple model of herd behavior," *The Quarterly Journal of Economics*, pp. 797–817, 1992.
- [33] 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.
- [34] W. H. Greene, Econometric analysis. Pearson Education India, 2003.